Title
The Cream of the Crop? Inequality and Migrant Selectivity in Ireland during the Age of Mass Migration

Author
Connor, Dylan S

Publication Date
2018-03-01

The Cream of the Crop?

Inequality and Migrant Selectivity in Ireland during the Age of Mass Migration

Version: January 2018

Dylan Shane Connor
dylan.connor@colorado.edu

University of Colorado Boulder

Abstract

As over 30 million people moved to North America during the Age of Mass Migration (1850-1913), governments feared that Europe was being depleted of its most talented workers. I use new longitudinal data from the early twentieth century to study the geography and selectivity of migration from Ireland, the European country with the highest emigration rate. I find that Irish-speakers and the sons of farmers in poorer communities, where emigrant networks were strongest, were more likely to move to the United States. These results indicate that highly skilled workers were, in fact, underrepresented in the flow to America.

1. Introduction

During the Age of Mass Migration (1850-1913), over four million people moved from Ireland to the United States. Sluggish economic growth in Ireland combined with high overseas demand for labor to spur over half a century of mass migration. From 1845 to 1911, this migration flow contributed to a decline in the Irish population from over eight million to less than four and a half. With as many as 13 per 1,000 people emigrating annually, Ireland's emigration rate was more than double any other major European source-country (Hatton and Williamson, 1993). The contraction of rural Irish opportunities is thought to be a major driver of this emigration and was at the time believed to have exacerbated the emigration of Ireland's "best men" (Royal Commission on Labour, 1893, p. 49).

This paper first asks: was it, in fact, the best and brightest who left Ireland for the United States? To answer this question, I construct a new longitudinal dataset linking non-migrants, internal migrants and emigrants to the United States, back to their childhood homes in Ireland at the turn of twentieth century. To examine how conditions in origin communities differentially affected the migration decisions of people from different socioeconomic backgrounds ("migrant selection"), I integrate these data with new measurements of landholding and emigrant networks in rural communities. This is, to my knowledge, the first micro-analysis of migrant selection from Ireland in the early twentieth century, and more broadly, the first to examine how childhood

communities condition selection into internal and international migration.

Successive Irish reports speculated that the loss of workers with greater economic potential was impeding development in Ireland. The Royal Commission on Labor, for example, suspected that Irish agricultural growth was being stymied by the emigration of Ireland's best agricultural workers – the sons of farmers (Royal Commission on Labour, 1893, p. 49). Protracted emigration throughout rural Ireland partly had its roots in the precipitous decline of laboring, landownership and marriage opportunities (Fitzpatrick, 1980a; Guinnane, 1997). Emigration from agricultural communities lacking in economic opportunity also resulted in streams of cash remittances, which financed further emigration, and provided income for the young and old who remained at home (Arensberg & Kimball, 1940).

These issues are of consequence to the development of the American economy. Evidence continues to show that immigration from Europe generally had positive long-run effects on economic growth in the United States (Ager & Brückner, 2013; Glaeser, La Porta, Lopez-de-Silanes, & Shleifer, 2004; Rodríguez-Pose & von Berlepsch, 2014). Specifically, immigration tends to benefit the receiving economy when the skills of immigrants compare favorably to natives, or when immigrants change receiving country institutions for the better (Borjas, 2014). Thus, migration from Ireland in the early twentieth century would likely have benefited the American economy if the most highly skilled Irish people did, in fact, emigrate to the USA.

My findings provide quite strong evidence that Ireland's higher skilled and wealthier sons were *underrepresented* in the flow to America. I estimate that the sons of farmers were 1.3 percentage points more likely to emigrate than the sons of urban skilled workers and laborers (from a base probability of 3%). I show that the sons of farmers who emigrated tended to come from the poorest landholding families. In addition, Irish speakers – who tended to be of low socioeconomic status – were significantly more likely to emigrate. Thus, I conclude that while emigrants were more likely to come from landholding families, they tended to be drawn from poorer backgrounds ("negatively selected").

This broader finding of *negative selection* from Ireland is consistent with standard economic models of migrant selection (Borjas, 1987, 1991). These models assume that migrants have two choices – stay in the home country or move abroad – and that they then compare the relative returns to skill in each place, generating predictions of either negative or positive selection. Unskilled labor was rewarded relatively better in the United States than in Ireland in the early twentieth century, and this is consistent with my findings of greater emigration among poorer farmers' sons, who were generally lower skilled.

The profiles of people migrating within Ireland looked quite different from those moving abroad. Compared to urban laborers, I find white-collar workers to be up to three times as likely to move to a town or city in Ireland, and farmers and Irish-speakers to be significantly less likely to do so. Thus, while emigration attracted sons from poorer farming and Irish-speaking families, migration within Ireland drew from the upper echelons of the Irish population in terms of skill and status. Given the relatively high demand for clerical and skilled labor in Irish towns and cities in the early twentieth century, these findings may also be consistent within Borjas' framework.

My findings indicate that differences in family and community conditions could account for why, compared to their similarly skilled counterparts (e.g. laborers' sons), the sons of farmers

tended to emigrate rather than move within Ireland. After controlling for the strength of emigrant networks, per acreage land value and landholding inequality in childhood communities, I find no significant difference in the probability of emigration for farmers' and laborers' sons. Thus, the emigration of farmers' sons appears to have been precipitated by their residence in poorer western communities with longer histories of emigration. These emigrants also tended to come from larger families and were less likely to inherit family land (as indicated by birth order).

The proclivity for emigration over internal migration in poorer farming and Irish-speaking communities could reflect the economic climate of time. During the Age of Mass Migration, immigrants in the New World sent hundreds of millions of dollars back to Europe (Magee & Thompson, 2006). These remittance payments often eased the cost for friends and family-members to travel to the United States (Wegge, 1998, 2002). Beyond financing migration, however, many Irish families used remittances to finance expenditures, with many of the poorest communities developing a "dependence on American money" (Gaeltacht Commission, 1926, p. 45). The reliance of large families on income streams from America - where high wages provided opportunities to save and remit – to supplement household expenses and travel, may help explain why lower skilled workers were particularly likely to emigrate from these communities.

Differences in decisions to leave Ireland may also reflect social and political preferences. Given that one would expect Irish-speakers to face higher cultural penalties in American labor markets, it is surprising that they were more likely to emigrate than move within Ireland. One explanation may be related to the situation of this study of emigration at a tense political juncture following the agrarian unrest of Ireland's late nineteenth century Land War and just prior to the uprising against British rule in 1916. Evidence from emigrant letters suggest that many Irish-speakers viewed their emigration – to a land perceived to offer social and political freedoms that were unavailable in Ireland – as a form of political exile (Miller, 1988). This political interpretation of emigration is speculative, but may be worthy of further investigation.

I draw these conclusions from new longitudinal data on over 66,000 Irish males who decided to either stay in Ireland or move to America in the early twentieth century. Using methods pioneered by Abramitzky, Boustan and Eriksson (2012) and Ferrie (1996), I built this dataset by linking individuals between the 1901 and 1911 censuses of Ireland, or to the 1910 American decennial census. I linked this new sample to high quality data on the value and distribution of land wealth across 158 local areas (Poor Law Unions), which I digitized from the Irish land census. I further complement these rich data by using information on surname similarity to devise a new measure of the geographical prevalence of migrant networks. I show that my results are largely robust to bias from individuals erroneously linked between censuses, generally known as "false positives" (Bailey, Cole, Henderson, & Massey, 2017; Massey, 2017).

2. Migrant Selection in the Age of Mass Migration

The Age of Mass Migration continues to provide insight on the motivations of migrants. Recent studies of historical migration from Mexico, Britain, Norway, Italy and Spain have produced results consistent with the predictions of canonical economic models of migrant selection (Abramitzky et al., 2012; Kosack & Ward, 2014; Long & Ferrie, 2013). Further investigation from this period, however, suggests that migrant selection may also be scale-dependent, in that it can vary regionally, and may be conditioned by regional, community and family influences

(Abramitzky, Boustan, & Eriksson, 2013; Juif, 2015; Spitzer & Zimran, 2015). Thus, understanding the forces generating migration selection requires measurements of both national and local determinant of migration.

2.1. Cross-country wage differences

The canonical Borjas model can be used to generate predictions of migrant selection by comparing earnings inequality across locations. If, for example, higher skilled workers earned relatively more in the United States than in Ireland, the Borjas model would predict emigrants to be above average in skill relative to the Irish population (positively selected). If the returns to skill were lower in the United States than in Ireland, the Borjas model would predict migrants to be drawn from the lower end of the Irish skill distribution or negatively selected.

From what we know of how labor markets have historically rewarded skill, the Borjas model predicts negative selection from Ireland to the United States. The basis for these predictions is evident in the earnings differences of skilled workers and laborers in Figure 1. Although the wages of Irish carpenters and laborers in 1901 were roughly similar in being 60 percent of their American levels, building laborers earned relatively more in the United States than skilled fitters. From the relatively higher returns for laborers, the Borjas model would predict that lower skilled workers had greater incentives to emigrate (negative selection). It may also be the case, however, that the migration of the poorest people may have been constrained by poverty, as has been noted for historical Ireland and across many other contexts (Belot & Hatton, 2012; Ó Gráda, 1994).

Note: The bars in Figure 1 show the real wage gaps between occupations in Ireland and the United States. A value of 1 implies real wages are identical in Ireland and the US while 0.6 implies a given occupational wage in Ireland is 60 percent of its US level.

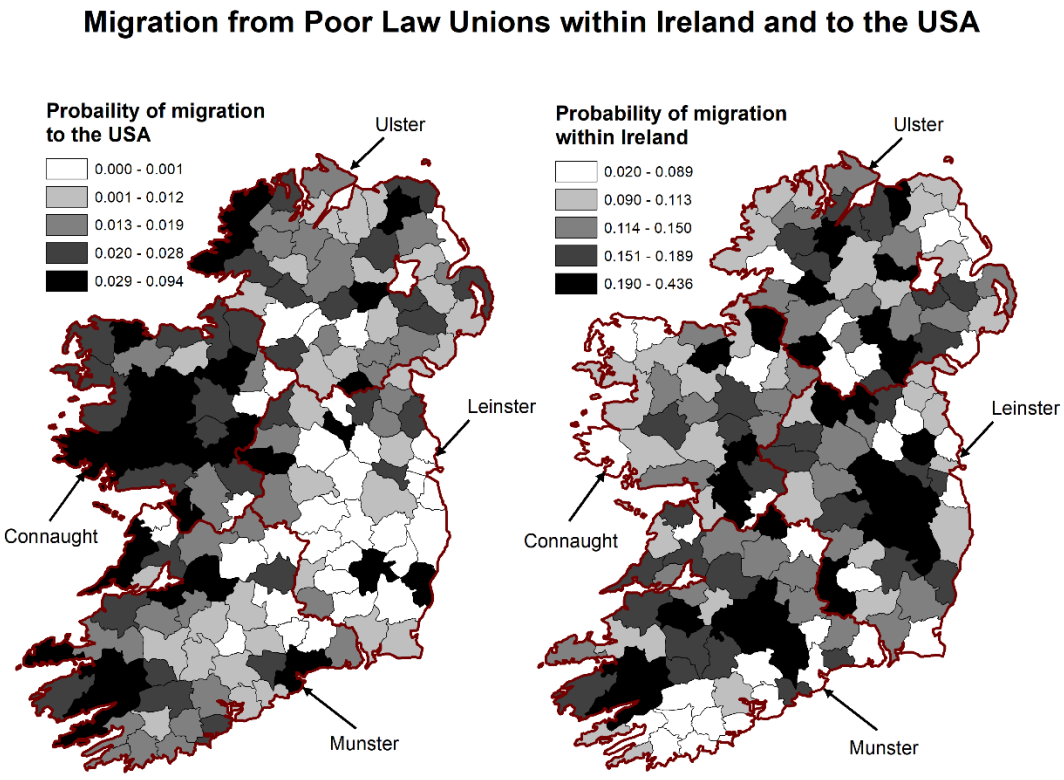
Figure 1. Average Irish occupation wages relative to wages in the USA

Earlier studies of emigrant selection from Ireland support this country-scale prediction of

negative selection. While the very poorest people in Ireland faced poverty constraints on migration during the Great Famine (1845-1852), emigrants still tended to be poorer than the Irish population as a whole (Anbinder & McCaffrey, 2015; Mokyr & Ó Gráda, 1982). Although we know relatively little about migrant selection in the half century following the Famine, we *do* know that up to 1900, most emigrants were unskilled and aged between 15 and 24 (Guinnane, 1997, p. 106). This study tests this negative selection hypothesis by constructing the data necessary to reliably compare the profiles of emigrants to non-migrants.

2.2. Family and community influences on migration

While differences in how countries reward skill surely influenced migration decisions, historical accounts suggest that the configuration of inequality across regions, communities and families mattered too. The modernization of the Irish economy occurred through the transformation of rural Ireland (Miller, 2008). Efforts aimed at increasing agricultural competitiveness and productivity such as farm mechanization, the consolidation of holdings and declines in the subdivision of land among heirs ("impartible inheritance"), reduced opportunities for agricultural labor, farm ownership and even marriage (Fitzpatrick, 1980b; Guinnane, 1997). Thus, while these transformations may have led to improvements in average Irish wages (Begley, Geary, & Stark, 2014), they did little to circumscribe emigration.



Note: Darker colors are associated with higher probabilities of migration.

Figure 2. Map of migration within Ireland and to the USA by Poor Law Union of childhood

Changes in the rural economy exacerbated regional inequality across the island of Ireland and emigration from poorer rural communities. Combinations of constrained life chances, family strategies to secure remittances, and chain migration ensured a persistent flow of emigrants across the Atlantic throughout the late nineteenth and early twentieth century. Estimates from my preferred sample in Figure 2 show that the probability of emigration was consistently higher across the west coast of Ireland, particularly in the provinces of Connaught, Munster and Ulster.[1] These Poor Law Unions (PLUs) were occupied by some of the poorest Irish communities and were often characterized by subsistence farming, longer distance from ports and large cities, and greater family sizes (Fernihough, 2017).

The lack of proximate economic opportunities in these more western regions is suggested by patterns of migration across counties within Ireland. Panel B of Figure 2 shows higher rates of internal migration for males growing up near the economic cores of Belfast and Dublin, and the more prosperous agricultural regions of the southeast. The demand for labor in these regions appears to have drawn heavily from surrounding counties. From comparing the maps shown in Figure 2, it appears that people were less likely to emigrate from these eastern regions, which may suggest that proximity to larger labor markets reduced emigration to the USA. Earlier evidence suggests that even in poorer western regions, land acquisition opportunities may have discouraged emigration (Hatton and Williamson, 1993).

Additionally, family obligations and personal connections were key to sustaining emigration. Conservative estimates place the remittance flow from North America to Britain over the 1848-1900 period at $260 million, and an unknown but sizeable share of this sum flowed to Ireland (Schrier, 1958). Although the costs of travelling to the United States had declined considerably up to the late nineteenth and early twentieth century, they were still sizeable. Although travel costs fluctuated, the price of passage could drop to a level equivalent to the weekly wage of a lower skilled worker in the United States or the annual rent of a typical farm in the poorer regions of the Irish West (Fitzpatrick, 1984; Guinnane, 1997). Given these costs, it is unsurprising that 35 percent of Irish steerage passengers reportedly crossed the Atlantic between 1904-1910 on tickets paid for, likely by somebody known to them, in the United States (Harkness, 1931).

As observed by Harkness and others, a key motivation for migration and familial migration chains was to secure remittances to support siblings and elder family members facing economic hardship in Ireland. While the Gaeltacht Commission (1926) suggested that migration effectively become a livelihood strategy for many struggling communities, ethnographic accounts such as Arensberg & Kimball (1940) note that "the young men must travel … a great many farms, especially in West and North Clare, are partially supported by Christmas gifts sent from children living abroad" (p. 143). The ability for Irish emigrants to save sizable sums of money - to be remitted or otherwise - has been noted back to the Great Famine (Wegge, Anbinder, & Ó Gráda, 2017).

Thus, the current historical evidence suggests that economic differences within and across regions, and the prevalence of migrant networks may have influenced decisions around

[1] The shape files for Poor Law Unions produced by Gregory (2008) were obtained online from the British Data Archive website (www.britishdataarchive.com).

emigration. I measure rural economic opportunities using a Gini index of landownership and the per acreage value of land. Hypothetically, wealthier communities and those with more equal distributions of landholding may have discouraged emigration or had differential effects on how landholding status affected who decided to leave. Additionally, I use birth order and family size to study the effect of family opportunity (inheritance) and constraints on emigration. Finally, I measure the prevalence of emigrant networks to capture possible economic assistance with travel through migration chains and possible remittance motives.

3. Data and Estimation

To study migrant selection and its determinants, I constructed a sample of Irish-born males whom I could observe in their homes in Ireland as children ("Child Sample") or as young men ("Adult Sample") before deciding whether to move within Ireland or to the United States. I created this sample using a modified version of the record linkage techniques developed by Abramitzky, Boustan and Eriksson (2012) and Ferrie (1996). In general, my preferred "Birthplace Match" sample yields comparable results to other common linkage techniques and in Section 5 I show my results to be largely robust to the issue of false positives.

I use the Adult Sample to assess how migrants were selected from the Irish labor force as of 1901. The Adult Sample is comprised of men aged 16-25 who reported holding an occupation of their own in Ireland in 1901. It is important to measure occupational status prior to migration, as immigrants often hold occupations below their true level of skill in the new county (Abramitzky, Boustan, & Eriksson, 2014). As emigration is often a response to transitory income shocks, measures of occupational status should also pre-date the immediate migration period. Otherwise, data sources such as departure records may lead to underestimates of the permanent income or occupational status of migrants (Fernández-Huertas Moraga, 2011). These issues lead me to favor the linking of immigrants back to the Irish census to study migrant selection.

The Child Sample provides the cleanest estimates from this analysis, as differences from the Adult Sample may be biased by prior migration patterns from source communities. Moreover, occupational attainment and future migration decisions may be jointly determined: decisions to acquire land and become a farmer, for example, would likely have been made alongside expectations around future migration. I overcome these issues by focusing on the occupations of co-resident fathers while their sons, who may later have decided to emigrate, were in childhood. Further, focusing on younger sons aged 6-18 helps avoid reverse causality in fathers' wealth through the migration of their adult children.

3.1. Record Linkage and sample construction

This linkage procedure relies on uniquely matching a refined population of males between census years. I searched for individuals that were enumerated in the complete-count 1901 census of Ireland to either the 1911 census of Ireland or the 1910 American census.[2] These individuals were aged from 6 to 25 in 1901 and were linked using their full names and ages. I undertake basic text standardization so that full and shortened first names (e.g. Joe and Joseph) are considered

[2] The Irish data has been digitized by the National Archives of Ireland and prepared by Connor (2017) and Connor, Mills, & Moore-Cherry (2011). The data from the United States has been made available by the Minnesota Population Center and its collaborator Ancestry.com.

equivalent. To simultaneously match individuals from 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.

I account for age misreporting and differences in the enumeration dates of the Irish and US census by using an iterative age matching procedure.[3] I first search for individuals from the 1901 census of Ireland with an identical name and who are exactly ten years older in 1910-1911 censuses. Where no match exists, I iteratively broaden the search criteria to people who are up to +/- 2 years older or younger. Individuals are included in the sample when a person with uniquely matching characteristics is found in the later census. Cases matched to multiple individuals by name and age are deemed ineligible for inclusion and dropped from the sample.

Inconsistently reported information on location of birth in the American and Irish censuses posed a challenge to this analysis. While county of birth is reported in the Irish censuses, only country of birth ("Ireland") is reported for Irish immigrants in the American census. Linking samples without detailed information on place of birth inflates the rate of false positives and introduces error into the relationship between the observed locations of origin and destination. This is because the quality of successful links tends to be lower when matching on fewer characteristics, and when false positives do occur, they are more likely to be between individuals born in different counties.

I mitigate this problem while constructing my preferred "Birthplace Match" sample by focusing on individuals with less common names, and last names containing geographical information. Due to its history of low inter-county migration, Ireland was peculiar in having last names which are strongly tied to counties (Matheson, 1901). I exploit this information and use the 1901 census to calculate how concentrated last names are to specific counties of birth. Using these measures, I limit the original census records to the 15 percent of men whose last names provide a strong indication as to their county of birth. The threshold for determining inclusion in this sample was men whose last names were 60 percent confined to a single county. I chose this cut-off as it is the point at which four out of five men have an identical observed county of birth to that imputed from their last name.[4]

This restricted population forms the basis for my preferred high quality linked sample. There are 125,317 men with a name above the 60 percent threshold in the 1901 census. I successfully link 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. The overall average annual emigration rate from the official statistics is 7.9 per thousand over the 1901-1909 period and the correlation of the county level emigration rate from the linked sample and the official statistics is +0.90. I further benchmark my results from my preferred sample against estimates derived from more common

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

[4] 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.

linkage approaches.

I constructed my main sample for comparison using a similar technique to that employed by Abramitzky, Boustan and Eriksson (2012) in their study of migration from Norway ("Standard Match"). In this sample, no birthplace restriction was made on last names and individuals were linked using the iterative age procedure describe above. To help account for name changing or misreporting between censuses, I used the New York State Identification and Intelligence Phonetic Code ("NYSIIS") to standardize common names based on how they sound.[5] To ensure that my restriction to geographically concentrated names does not generate radically different results to more flexible matching procedures, I provide key results for the Birthplace and Standard Match samples.

3. 2. Summary statistics

Table 1 presents the basic summary statistics for the Birthplace and Standard Match, and for the baseline census data from which they were produced. Overall, for the Birthplace Match, the forward linkage rate from 1901 census males aged 6-25 in 1901 is 29 percent, and the specific linkage rates for the Adult and Child Samples resemble this. The overall backward linkage rate for Irish-born males aged 15-34 in the 1910 census is 11 percent. Such discrepancies in linkage rates have also been reported in previous studies linking populations within and across countries (see Abramitzky, Boustan and Eriksson, 2012).[6]

The match rate for the Birthplace Match is twice as high as in the Standard Match. While 32 percent of sons who were co-resident with their fathers and aged 6-18 in 1901 were successfully linked, the linkage rate was only 16 percent for their counterparts with less geographically concentrated names. For the Adult Sample, the match rates were 26 and 13 percent for the Birthplace and Standard Match, respectively. It is not surprising that the linkage rates in the Birthplace Match is higher than in the Standard Match or to other studies, as these names are, by definition, more uncommon.

Although successfully linked people tend to be of higher socioeconomic status across both samples, the Birthplace Match disproportionately captures males from agricultural backgrounds. While the share of sons in the Birthplace Match whose father is a farmer is as high as 54 percent, it is only 47 percent in the Standard Match. As shown by comparison to their respective base populations, this difference is mainly driven by the higher share of farmers who have last names concentrated in specific counties. In fact, the record linkage process leads to a reduction in the share of sons with farmer fathers by 2-3 percentage points across both the Standard and Birthplace Match. Relatedly, people from skilled and white-collar backgrounds and from the wealthier provinces of Leinster and Ulster are more likely to be linked between censuses. Although the sample is not fully representative of the Irish population of 1901, it is of sufficient quality to study migrant selection within the population.

[5] The appendix also contains results from a linkage approach which only links individuals who are unique on name and age within a five-year age band. Estimates from this sample are consistent with those presented here.

[6] In the Standard Match, the backward linkage rate is 13% (10,579/79,376) and the forward linkage rate is 16% (132,528/828,944). The calculations for the backward and forward linkage rates in the Birthplace Match are 1,956/19,060 (11%) and 36,060/125,317 (29%).

	Birthplace last names only		Full 1901 census	
	Base Population	Birthplace Match	Base Population	Standard Match
Child Sample				
Age in 1901	12	11	12	11
	(3.7)	(3.7)	(3.7)	(3.7)
Share Catholic	0.79	0.71	0.76	0.66
Share with literate fathers	0.73	0.77	0.75	0.79
HISCAM index of father	53.5	53.5	52.67	52.8
	(7.9)	(8)	(8.4)	(8.5)
Total Children	4.35	5.5	4.17	3.81
	(2.64)	(2.13)	(2.67)	(2.59)
Father's occupation				
White-collar	0.08	0.08	0.08	0.1
Skilled worker	0.17	0.20	0.20	0.22
Urban laborer	0.09	0.09	0.11	0.1
Farmer	0.56	0.54	0.50	0.47
Farm laborer	0.10	0.09	0.11	0.10
Birth order				
First son	0.28	0.28	0.29	0.27
Last son	0.25	0.25	0.25	0.27
Middle son	0.37	0.37	0.36	0.36
Only son	0.10	0.10	0.11	0.11
Province				
Connaught	0.23	0.19	0.18	0.15
Leinster	0.12	0.15	0.22	0.24
Munster	0.34	0.28	0.26	0.21
Ulster	0.31	0.37	0.34	0.41
Origin characteristics				
Share Urban	0.18	0.21	0.23	0.26
Network strength	1.002	1.002	1.002	1.002
	(0.002)	(0.001)	(0.001)	(0.001)
Land value per acre	0.6	0.66	0.69	0.75
	(0.38)	(0.40)	(0.41)	(0.42)
Gini	0.41	0.40	0.41	0.40
	(0.10)	(0.11)	(0.10)	(0.10)
Observations	55,589	17,669	378,928	60,015
Match rate from base population	-	32%	-	16%
Adult Sample				
Age in 1901	20	20	20	20
	(3)	(3)	(3)	(3)
Share literate	0.91	0.93	0.92	0.93
HISCAM index	50.08	50.4	49.44	50
		(6.82)	(7.4)	(7.2)
Occupation				
White-collar	0.08	0.09	0.09	0.11
Skilled worker	0.20	0.21	0.23	0.24
Urban laborer	0.11	0.10	0.14	0.11
Farmer	0.41	0.43	0.35	0.36
Farm laborer	0.20	0.17	0.19	0.18
Observations	53,565	13,785	386,559	51,601
Match rate from base population	-	26%	-	13%

Note: 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.

Table 1. Summary statistics

3.3. Estimating Selection by Occupation and Skill

I test for selection by studying the influence of economic characteristics on the decision to migrate across counties or to the United States. The selection variables of interest are measures of occupation, literacy, religion and Irish-speaking ability. Although the literacy, religion and linguistic measures could be coded directly from the census, the measures of occupation required further work, as the stated occupations in the Irish census have not been ranked with respect to occupational status or earnings.

I converted the occupational strings from the census to measures of occupational status using recent work Fernihough, Ó Gráda, and Walsh (2015) linking Irish census returns to corresponding codes in the Historical International Classification of Occupations (HISCO) (van Leeuwen & Maas, 2011). By implementing these codes, I could assign each occupational string to an occupational class from a reduced version of the Historical International Social Class Scheme (HISCLASS). The HISCLASS categories correspond to identifiable socioeconomic classes and provide valuable information with respect to skill, class and land ownership.

I estimate selection into internal or international migration for Adult and Child Samples within a probit framework. These models predict whether migrants and non-migrants differed from each other with respect to the variables of interest described above. The regression equation is specified as:

$$p(Migrate = 1) = \phi(\beta_0 + \Sigma_{k=1\,..K}\beta_k X_k) \quad (1)$$

where Y refers to the probability of an Irish male in 1901 moving county to an urban or rural area within Ireland or to the United States by 1910/1911.[7] All coefficients are presented as marginal effects. Thus, β_k can be interpreted as the percentage point change in the probability of an individual migrating to a destination associated with a one-unit change in the k^{th} independent variable. I infer selection from the differences in the probability of migrating to a given location by occupation and literacy, where the reference group are individuals in all other destinations in 1910/1911.[8] The baseline estimates of selection into emigration are only conditional on age, so that I can account for potential cohort differences in education and emigration.

3.4. Estimating Selection by Source Region Conditions

My analysis of how selection varies with conditions in source communities is a main contribution of this study. Specifically, I measure the local value and inequality in landholdings, and the strength of emigrant networks in source communities or PLUs. To derive interpretable estimates, I standardize these measures so that a one-unit increase in any of these variables is associated with a standard deviation change. The standard deviations of quantitative variables are included

[7] I categorized enumeration districts as urban based on whether 20 percent or less of the heads of household are farmers or farm laborers.

[8] Binary logistic regressions are preferred over a multinomial logit approach because the base category for measuring selection into emigration needs to include both internal migrants and non-migrants. Otherwise, selection into international migration will be overestimated for groups with higher rates of internal migration.

in Table 1. The models including PLU characteristics focus on migration within Ireland or to the United States between 1901 and 1910-1911 from rural areas only. These analyses aim to assess the effect of mainly rural community conditions on migration decisions, and to use these as control variables to better understand occupational differences in migration.

3.4.1. Migrant network

I use a new surname-based approach to measure the strength of emigrant social networks in Irish Poor Law Unions. I extend on previous work measuring migrant networks using migration rates based on last names (Wegge, 1998, 2008) by measuring the broader similarity of last names between Irish Poor Law Unions and the Irish-born population in the US. This approach, which is described in detail in the Appendix, accounts for differences in name reporting and enumeration between Ireland and the US, and exploits similarities in the roots of last names across Ireland.

I measure the strength of emigrant social networks by calculating how similar the last name distributions of PLUs are to last names of Irish-born people in the United States. I begin by calculating the string distance between every pair of names observed in Ireland and the United States in the 1900-1901 period using the *jarowinkler* package in Stata (Feigenbaum, 2016). Using this matrix, I assigned a population-weighted average distance between every last name in Ireland in 1901, every last name in the United States in 1900, and every last name in the combined Irish-US population for 1900-1901. I used these statistics to calculate a ratio of the relative similarity of an Irish last name to all Irish-born American last names and to the last names of the Irish-born population as a whole:

$$network\ strength_i = \frac{JW\left(Name_{i_{Ire1901}} Name_{i\ldots k_{USA1900}}\right)}{JW\left(Name_{i_{Ire1901}} Name_{i\ldots k_{USA1900-Ire1901}}\right)} \quad (2)$$

where the numerator is the average Jaro-Winkler (*JW*) distance between the last name of individual *i* and the *k* last names of Irish-born people in the United States in 1900. The denominator standardizes for baseline differences in name similarity by taking the average distance between last name *i* and *k* last names in the combined Irish-US population. I assign each individual their respective ratio based on their last name. I average these distances across PLUs and counties to measure the geographical prevalence of migrant *network strength*.[9]

The suitability of using name similarity to measure migrant networks is supported by its correlation with independently collected historical emigration statistics. Figure 3 plots the official county-level emigration rates against the county aggregation of the network variable. Although the numerical values of the network variable only range from around 1.000-1.004, deviations on this scale are qualitatively meaningful. This is evident in that the network measure holds a strong positive correlation (+0.64) with the emigration rate of counties from 1851 and 1902. The strong correspondence between the network measure and the county emigration rates provides compelling evidence for using last names to measure local migrant networks.

[9] 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.

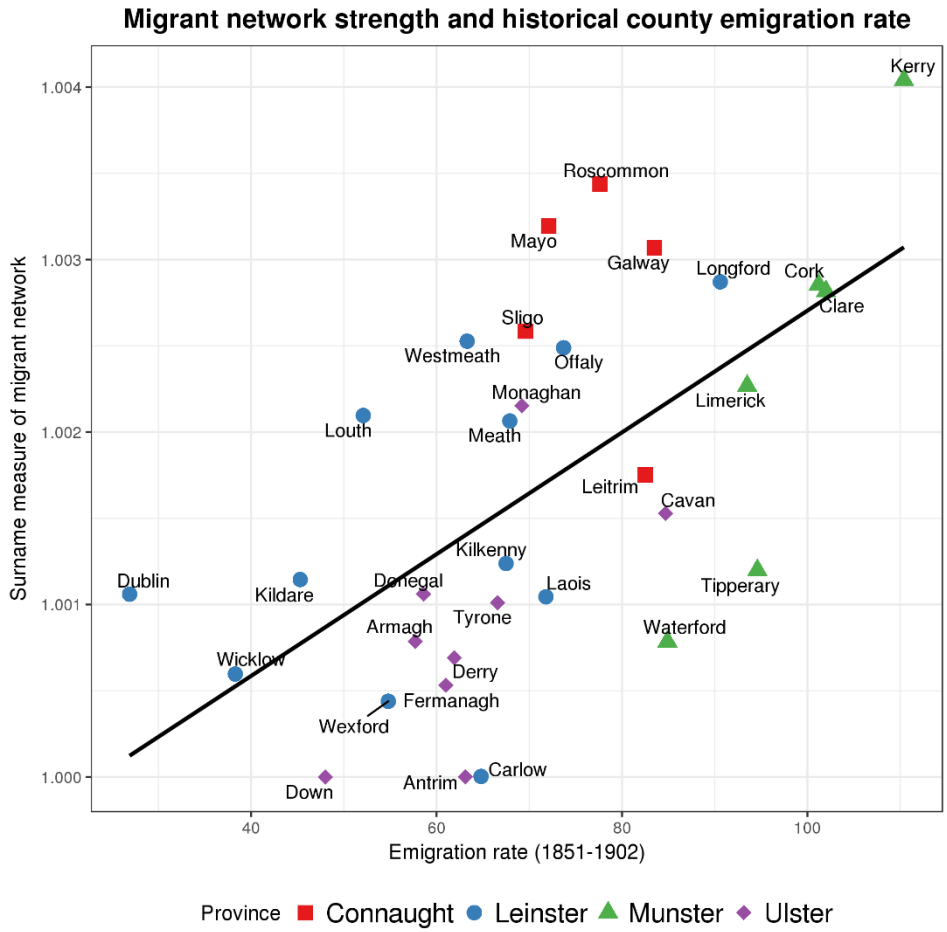


Figure 3. Scatter plot of surname migrant network measure and county emigration rate

There are two main reasons for why I measured migrant networks with a surname-based approach instead of using county-level emigration rates. County emigration rates may be subject to considerable measurement error due to underreporting (e.g. Ó Gráda, 1975), and the fact that beyond a certain point, mass emigration may actually weaken migrant networks, as entire families and communities move. Thus, a local measure of migrant networks based on last names can plausibly capture still existing social and familial transnational ties. Further, from a practical concern, the last name based measure can be flexibly derived at the scales available in the data (i.e. PLUs) and are not confined to the county-level, as is the case for official emigration statistics.

3.4.2. Land structure

The measures of the distribution and value of PLU 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.[10] 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 acreage value of landholdings and to construct a Gini index of inequality for each PLU. These data

[10] 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.

from the land census apply only to families living on agricultural holdings. Thus, the Gini index reflects inequality among agricultural landholders and is mainly applicable to rural areas.

This land register data needed to be manually linked to the census records. Although the indexed 1901 census contains information on street address, electoral division and county of residence, it does not report the Poor Law Union. Thus, I needed to use the Irish Topographical Index to compile and link the 3,000 enumeration districts from the 1901 census to one of the 158 PLUs listed in the land census. I successfully matched all electoral divisions to a PLU, and assigned everyone their corresponding PLU land value and Gini coefficient.

The value of holdings may be a better indicator of income or consumption than wealth. The history of tenancy in Ireland makes it difficult to draw a sharp distinction between owners and occupiers (Turner, 2002). Thus, while land value may be a limited measure of wealth, the value of agricultural land may reflect agricultural rents (Burt, 1986; Featherstone & Baker, 1987) and the capitalization of future land use opportunities (Borchers, Ifft, & Kuethe, 2014; Plantinga, Lubowski, & Stavins, 2002). Therefore, the value of land may be a reasonable proxy for economic opportunity in agricultural areas.

3.4.3. Birth order and family size

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. These data were also used to count the number of children in the household ("Family size"). Unfortunately, these variables will be mismeasured for families where older siblings have already left home. I mitigate this problem by confining the birth order and family size analyses to sons whose mothers were aged under 45 in 1901.[11]

4. Results

I present the main results over four sections. I begin by measuring occupational selection to the United States from the Birthplace and Standard Match while only controlling for age. I then compare the profiles of international migrants to men who moved within Ireland and incorporate measure of religion and Irish-speaking ability with further control variables. The final sections of the analysis focus on community and family influences on migrant selection.

4.1. Selection to the US

4.1.1. Probit estimation of emigration for all occupations

Table 2 estimates the baseline occupational selection of migration to the United States, and suggests that emigrants were more likely to come from farming backgrounds. This is expressed in Columns 1 and 2, which show occupational differences in the probability of emigration relative to urban laborers. The occupational classes refer to fathers' occupation in the Child Sample and own occupation in the Adult Sample. As these models only control for age, the estimated effects

[11] This restriction only affects the magnitude of the results but not differences in significance or interpretation.

	Birthplace Match		Standard Match	
	(1)	(2)	(3)	(4)
	Moved to USA	Moved to USA	Moved to USA	Moved to USA
Age	0.035***	0.021***	0.052***	0.025***
	(0.011)	(0.002)	(0.006)	(0.002)
Age squared	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Occupation (ref= Urban laborer):				
White-collar	-0.008	0.000	-0.009*	0.001
	(0.010)	(0.005)	(0.005)	(0.003)
Skilled worker	-0.003	-0.001	-0.000	-0.001
	(0.008)	(0.004)	(0.005)	(0.002)
Farmer	0.015*	0.013***	0.019***	0.014***
	(0.008)	(0.004)	(0.004)	(0.002)
Farm laborer	0.021**	0.004	0.025***	0.001
	(0.009)	(0.005)	(0.005)	(0.003)
Observations	13785	17669	51093	59701
Sample	Adult	Child	Adult	Child
Province FE	No	No	No	No

* p < 0.10, ** p < 0.05, *** p < 0.01

Note: Marginal effects from probit estimation.

Table 2. Estimates of migrant selection to the United States

Column 1 indicates that men employed in agriculture were more likely to emigrate than men employed in urban occupations, particularly in higher skilled urban occupations. Compared to urban laborers, the probability of emigration is 1.5 percentage points higher for farmers and 2.1 percent point higher for farm laborers (from a base probability of 7%), and these margins widen when the comparison is made to white-collar workers. It is notable that the probability of emigration is not significantly different between farmers and farm laborers, or urban laborers and white-collar workers. These results suggest that the dominant form of selection in the Adult Sample is across sectors, and there appears to be relatively weak selection within agricultural and non-agricultural occupations.

The strongest patterns of selection based on fathers' occupation is the higher rate of emigration among the sons of farmers. The estimates in Column 2 show the sons of farmers to be 1.3 percentage points more likely to move to the United States than the sons of urban laborers (from a base probability of 2%). Although the base probability of emigration is lower in the Child Sample than in the Adult Sample, the relative occupational selection is larger: while farmers in the Adult sample are around 15 percent more likely to emigrate than urban laborers, farmers' sons are more than 60 percent more likely to emigrate than urban laborers' sons, and 37 percent more likely to do so than the sons of farm laborers.

The Birthplace and Standard Match yield very similar conclusions with respect to selection to the United States. Results from the Adult Sample in Column 3 suggest that emigration was relatively high from the agricultural sector, and if anything, slightly higher among men employed as farm laborers. This similarity in outcomes between farmers and laborers may be accounted for by the young age of farmers in the Adult sample (mean age of 20 in 1901), whom may not yet have accumulated sizable enough holdings or stock to significantly deter migration. The results from the Child Sample in Column 4 are consistent with the Birthplace Match in showing the sons

of farmers to be more likely to emigrate.

4.1.2. Probit estimation of emigration for agricultural occupations

Due to the heterogeneity within Irish agricultural occupations, it is challenging to draw direct conclusions of selection from the higher emigration rate of farmers and their sons. The sons of smallholding farmers from the poorer western counties of Ireland, for example, likely had access to far fewer resources and opportunities than the sons of wealthier grazers in the more fertile and productive agricultural regions in the Irish south and east. Unfortunately, the indexed 1901 census does not provide direct measures of wealth and status for farmers. Instead, I use the per acreage land values of farmers' home areas to indirectly categorize agricultural families as poor, medium or wealthy.[12] I use these data to predict emigration from rural areas in Table 3. Columns 1-2 provide unconditional results, while Columns 3-4 control for cultural and geographical characteristics with fixed effects for province of origin, religion and Irish-speaking ability.

	Birthplace Match			
	(1) Moved to USA	(2) Moved to USA	(3) Moved to USA	(4) Moved to USA
Occupation (ref= Farmer (poor area)):				
Laborer (poor area)	0.035**	0.012	0.030*	0.013
	(0.018)	(0.016)	(0.017)	(0.015)
Laborer (medium area)	-0.035***	-0.032***	-0.023*	-0.023***
	(0.012)	(0.005)	(0.012)	(0.005)
Farmer (medium area)	-0.026***	-0.008*	-0.018*	-0.003
	(0.010)	(0.004)	(0.009)	(0.004)
Laborer (wealthy area)	-0.037***	-0.017***	-0.017	-0.005
	(0.011)	(0.006)	(0.012)	(0.006)
Farmer (wealthy area)	-0.057***	-0.024***	-0.037***	-0.015***
	(0.009)	(0.004)	(0.010)	(0.004)
Observations	8270	11196	8270	11196
Sample	Adult	Child	Adult	Child
Province FE	No	No	Yes	Yes
Irish-speaking FE	No	No	Yes	Yes
Religion FE	No	No	Yes	Yes

* p < 0.10, ** p < 0.05, *** p < 0.01

Note: Marginal effects from probit estimation. Controls include: age and age squared.

Table 3. Estimates of migrant selection from agriculture to the United States

These estimates suggest that the high rate of migration to the USA for farmers and laborers is driven by emigration from the poorest agricultural regions. Column 1 shows that in poorer areas, laborers are 3.5 percentage points more likely to emigrate than farmers, and both are more likely to emigrate than their counterparts in wealthier areas. These results suggest that while emigration from the agricultural sector generally varied across areas with land value, landless laborers tended to be more likely to emigrate than their counterparts who had obtained land by a relatively young age. Despite some attenuation, these effects generally persist with the addition of geographical and cultural controls in Column 3.

These results and the higher emigration rate among farmers' sons can be interpreted as

[12] The classifications of poor, medium and wealthy are made by cutting the land value variable into terciles based on the farmer and farm laborer sample.

negative selection into emigration to the United States. Column 2 suggests that when comparisons are made within areas, farmers' and laborers' sons are similarly likely to emigrate, and both are more likely to leave poorer areas. Thus, the earlier results showing farmers' sons to be more likely to emigrate is driven by the disproportionate residence of the farmer population in poorer areas. Moreover, landless men in Column 1 appear to be more likely to emigrate than landholders. These estimates suggest that emigrants generally came from poorer agricultural background and were negatively selected from the population as a whole.

4.2. Selection into Internal and International Migration

4.2.1. Descriptive Analysis

While people leaving Ireland tended to come from poorer families, individuals moving within Ireland were more bifurcated in their backgrounds. This is evident in Table 4, which decomposes migration decisions in the Adult and Child sample by occupation and choice of destination. Columns 1 and 2 provide qualitative and quantitative rankings of these occupations, which are further divided by whether they were within the agricultural or non-agricultural sector. Although it is difficult to interpret differences across sectors (e.g. a poor farmers' migration to an urban laborers'), it is more sensible to examine differences within sectors. We can, for example, be more confident that white-collar and skilled workers earned more than urban laborers, and landholding farmers were typically better off than landless farm laborers.

While farmers and their sons were the least likely to leave their home counties, males from white-collar backgrounds were the most likely to migrate. This is evident from the share of the Adult Sample who stayed in their origin counties in Column 3. Although 77 percent of farmers were still in the same county by 1911, only 62 percent of white-collar workers appear to have been in the same situation. With 71-74 percent staying in the same county, farm laborers, urban laborers and skilled workers exhibit quite similar inter-county migration patterns to one another. The ordering of these occupational differences is similar to outcomes observed in the Child Sample, but they are less pronounced. This is not surprising as the Child Sample is less mobile, and their 1901 locations are less likely to reflect previous occupational decisions.

Table 4 also provides evidence that sectoral differences in migration may partially reflect the sorting of workers across locations based on skill. Column 4, for example, shows that the sons of farm laborers in 1901 are the most likely to have moved to a rural area by 1911. In the urban case, men in non-agricultural occupations were more likely to move county to a different urban area by 1911. There also appears to be a consistent skill gradient in migration to urban areas: white-collar males were more likely to move than skilled workers, who were slightly more likely to move than urban laborers. This pattern may reflect the relatively high demand for skilled and clerical work in Irish towns and cities, and the lower demand for unskilled labor as compared to other countries in the early twentieth century.

	1	2	3	4	5	6	7
	Occ. status	HISCAM Index (mean)	Did not move	Moved to rural area	Moved to urban area	Moved to USA	Total
Occupation of son (Adult sample)							
Non-agricultural							
White-collar	High	54	795 / 62%	195 / 15%	207 / 16%	81 / 6%	1,278 / 100%
Skilled	Medium	49	2,145 / 74%	317 / 11%	249 / 9%	205 / 7%	2,916 / 100%
Urban laborer	Low	45	964 / 73%	168 / 13%	91 / 7%	98 / 7%	1,321 / 100%
Total			3,904 / 71%	680 / 12%	547 / 10%	384 / 7%	5,515 / 100%
Agricultural							
Farmer	High	58	4,607 / 77%	572 / 10%	270 / 5%	535 / 9%	5,984 / 100%
Farm laborer	Low	52	1,630 / 71%	295 / 13%	139 / 6%	222 / 10%	2,286 / 100%
Total			6,237 / 75%	867 / 10%	409 / 5%	757 / 9%	8,270 / 100%
Occupation of father (Child sample)							
Non-agricultural							
White-collar	High	58	1,133 / 76%	136 / 9%	186 / 12%	45 / 3%	1,500 / 100%
Skilled	Medium	48	2,850 / 83%	234 / 7%	262 / 8%	104 / 3%	3,450 / 100%
Urban laborer	Low	37	1,257 / 83%	121 / 8%	98 / 6%	47 / 3%	1,523 / 100%
Total			5,240 / 81%	491 / 8%	546 / 8%	196 / 3%	6,473 / 100%
Agricultural							
Farmer	High	59	8,006 / 83%	732 / 8%	370 / 4%	503 / 5%	9,611 / 100%
Farm laborer	Low	51	1,303 / 82%	152 / 10%	72 / 5%	58 / 4%	1,585 / 100%
Total			9,309 / 83%	884 / 8%	442 / 4%	561 / 5%	11,196 / 100%

Table 4. Migration decisions by destination and occupational ranking

Column 6 reaffirms that the agricultural classes were overrepresented in the American emigration flow, and the apparent reluctance of farmers to move within Ireland. In the Adult and Child Samples, males from farming and farm laboring backgrounds are around two percentage points more likely to move to the United States than their non-agricultural counterparts. This difference is even more striking when made conditional on having left one's county of origin. Focusing on movers only in the Child Sample: 29 percent of farmers' sons moved to the United States, while only 22 percent of farm laborers' sons, and 13 percent of white-collar sons did the same. Thus, it appears that when farmers' sons did decide to move, they were unusual in being more likely to choose the United States as their new destination.

4.2.2. Probit estimation of internal and international migration

This section further compares decisions to move within Ireland and abroad by controlling for differences in cultural and geographical origins. By incorporating these control variables, I can better assess the source of occupational differences in migration decisions. These probit models take the following form:

$$p(Migrate = 1) = \phi(\beta_0 + \beta_1 Occupation_1 + \beta_2 Literacy_2 + \Sigma_{k=1\,..K}\beta_k X_k)$$

(3)

where the outcome variable *Migrate* refers to whether an individual decided to move county to one of three discrete destinations: a rural area in Ireland ("Moved to rural area"), an urban area in Ireland ("Moved to urban area"), or to the United States ("Moved to USA"). The probability of migrating to a specific destination for the Adult and Child Sample are reported in Tables 5 and 6.

	(1) Moved to rural area	(2) Moved to urban area	(3) Moved to USA
Age	0.005	0.010	0.029***
	(0.013)	(0.010)	(0.011)
Age squared	-0.000	-0.000	-0.001***
	(0.000)	(0.000)	(0.000)
Occupation (ref = Urban laborer):			
White-collar	0.025*	0.079***	-0.003
	(0.014)	(0.012)	(0.010)
Skilled worker	-0.013	0.012	0.002
	(0.011)	(0.008)	(0.009)
Farmer	-0.033***	-0.016**	0.006
	(0.010)	(0.007)	(0.008)
Farm laborer	-0.006	-0.008	0.013
	(0.011)	(0.008)	(0.009)
Can Read and Write	0.004	-0.009	-0.004
	(0.010)	(0.009)	(0.009)
Catholic	0.045***	-0.015***	0.016***
	(0.006)	(0.006)	(0.006)
Speaks Irish	-0.017**	0.001	0.014**
	(0.007)	(0.006)	(0.006)
Parents present	-0.043***	-0.014***	-0.011**
	(0.006)	(0.005)	(0.005)
Married	-0.015	-0.020**	-0.026**
	(0.012)	(0.008)	(0.012)
Province (ref = Leinster)			
Connaught	0.045***	-0.023***	0.049***
	(0.010)	(0.007)	(0.008)
Munster	-0.004	-0.005	0.055***
	(0.008)	(0.006)	(0.007)
Ulster	-0.001	0.014**	0.012**
	(0.008)	(0.007)	(0.006)
Observations	13785	13785	13785
Sample	Adult	Adult	Adult
Province FE	Yes	Yes	Yes

* p < 0.10, ** p < 0.05, *** p < 0.01

Note: Marginal effects from probit estimation. Sample is confined to males aged 16-25 in 1901.

Table 5. Migrant selection regression for Adult Sample

Table 5 presents the estimates from the Adult Sample and confirms that migration within Ireland appears to have been particularly attractive to higher skilled workers, and less so to farmers. Across Columns 1 and 2, farmers tend to be 1.6-3.3 percentage points less likely to move to a rural or urban area in Ireland than urban laborers. With control variables in the model, there also appears to be no significant differences in the probability of migration for urban and farm laborers. White-collar workers, however, are by far the most likely occupational class to move within Ireland: this is particularly so for migration to urban areas, where white-collar workers are

almost 8 percentage points more likely to move than urban laborers. As the greater mobility of highly skilled workers and the low mobility of farmers within Ireland does not appear to be driven by differences in geographical origin, cultural characteristics or religion, this suggests that occupational differences in destination choice may reflect sorting across locations based on skill.

Although selection to the US on occupation and literacy tends to be weak, Catholics and Irish-speakers are significantly more likely to leave Ireland. Column 3 shows weak effects for literacy and occupation on emigration once controls are added for province of origin and cultural characteristics. Catholics and Irish speakers, however, appear to be 1.4-1.6 percentage points more likely to emigrate, and men in the poorer provinces of Connaught and Munster are around 5 percentage points more likely to leave the country. Thus, the results from the Adult Sample suggest that selection to the United States in the Adult Sample are likely being driven by geographical differences in agricultural conditions and specific forms of cultural selection.

	(1) Moved to rural area	(2) Moved to urban area	(3) Moved to USA
Age	0.007*	0.003	0.020***
	(0.004)	(0.003)	(0.002)
Age squared	-0.000	-0.000	-0.000***
	(0.000)	(0.000)	(0.000)
Occupation (ref = Urban laborer):			
White-collar	0.017	0.059***	0.001
	(0.011)	(0.010)	(0.005)
Skilled worker	-0.006	0.008	0.000
	(0.009)	(0.007)	(0.004)
Farmer	-0.013*	-0.021***	0.006
	(0.008)	(0.006)	(0.004)
Farm laborer	0.010	-0.014*	0.003
	(0.010)	(0.008)	(0.005)
Can Read and Write	-0.008	0.009**	-0.003
	(0.005)	(0.004)	(0.003)
Speaks Irish	-0.015***	-0.006	0.006*
	(0.005)	(0.005)	(0.003)
Catholic	0.037***	-0.006	0.001
	(0.005)	(0.004)	(0.003)
Province (ref = Leinster)			
Connaught	0.024***	0.000	0.027***
	(0.007)	(0.005)	(0.004)
Munster	0.009	0.011**	0.019***
	(0.006)	(0.005)	(0.003)
Ulster	-0.004	0.030***	0.011***
	(0.006)	(0.005)	(0.003)
Observations	17672	17672	17672
Sample	Child	Child	Child
Province FE	Yes	Yes	Yes

* p < 0.10, ** p < 0.05, *** p < 0.01

Note: Marginal effects from probit estimation. Sample is confined to sons aged 6-18 in 1901.

Table 6. Migrant selection regression for Child Sample

The results based on fathers' occupation tend to mirror those from the Adult Sample. Table 6 presents an identical set of models to Table 5, but where own occupation and literacy are substituted with the characteristics of fathers, and the models are estimated using the Child Sample. Like the results generated by the Adult Sample, internal migration rates are higher for the sons of white-collar workers and lower for the sons of farmers. Likewise, Column 3 shows slightly higher rates of emigration among the sons of farmers, but the province controls render the effect non-significant. The only persisting effects on emigration are the higher emigration rates for people living outside of Leinster and Ulster, and that Irish-speakers are half a percentage point more likely to leave Ireland than their typically monolingual counterparts. Further, the Catholic effect tends toward zero in these models, suggesting that differences in parental occupation may account for the religious selection shown in Table 5.

Taken together, these results imply that the emigration flow was disproportionately comprised of Irish-speakers and the sons of farmers in poorer agricultural areas. Although univariate regressions showed being Catholic or illiterate to be positively associated with emigration (not shown), these effects disappear once I control for province of origin and whether a son spoke Irish. These control variables appear to be capturing higher rates of emigration in western emigration-prone communities, where people tended to be poorer and more likely to be Catholic, illiterate and Irish-speaking.

4.3. Community and family influences on selection

While regional migration patterns can account for much of the occupational- and skill-based selection into emigration, these models do not directly examine *why* regions matter. The New Economics of Labor Migration framework provides a series of hypotheses for how regional and community influences may affect migration decisions. Specifically, sons and daughters from economically deprived families are hypothesized to leave poorer and more unequal communities with stronger migrant networks, often as a means of securing remittances (Stark & Bloom, 1985; Stark & Taylor, 1989). In this vein, the following model attempts to explain occupational variation in migration decisions with family and community characteristics:

$$\begin{aligned} p(Migrate = 1) = \phi(\beta_0 + \ & \beta_1 Fathers'occupation_1 + \ \beta_2 Catholic_2 \\ & + \beta_3 Spoke\ Irish_3 + \beta_4 Gini_4 + \beta_5 Land\ value_5 \\ & + \beta_6 Network\ strength_6 \ + \beta_7 Family\ size_7 \\ & + \beta_8 Birth\ Order_8) \end{aligned} \tag{4}$$

where the outcome variable is whether a son in the Child Sample moved from their rural community of origin to a different county in Ireland or to the United States. The primary selection variables of interest are *Fathers' occupation* and whether a son was *Catholic* or *Spoke Irish.* I collapse the occupational categories to the sons of landholding farmers, landless laborers, or skilled and white-collar workers ("Skilled/W.C."). The community variables of interest are the local *Gini*, *Land value* and emigrant *Network strength* and the family characteristics pertain to *Birth order* and *Family size*. Columns 1-3 of Table 7 present the results for emigration, and Columns 4-6 refer to migration across counties within Ireland.

	(1) Moved to the USA	(2) Moved to the USA	(3) Moved to the USA	(4) Moved within Ireland	(5) Moved within Ireland	(6) Moved within Ireland
Age	0.026***	0.025***	0.025***	0.016***	0.017***	0.017***
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)
Age squared	-0.001***	-0.001***	-0.001***	-0.001**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Occupation (ref = Farmer):						
Laborer	-0.008**	-0.003	-0.003	0.032***	0.038***	0.038***
	(0.003)	(0.004)	(0.004)	(0.009)	(0.010)	(0.010)
Skilled/W.C.	-0.005	-0.002	-0.002	0.067***	0.072***	0.073***
	(0.004)	(0.003)	(0.003)	(0.010)	(0.010)	(0.010)
Speaks Irish	0.013***	0.008*	0.008*	-0.013	-0.018**	-0.019**
	(0.005)	(0.005)	(0.005)	(0.008)	(0.009)	(0.008)
Catholic	0.003	-0.007	-0.007	0.027***	0.006	0.006
	(0.004)	(0.005)	(0.005)	(0.009)	(0.011)	(0.011)
PLU characteristics:						
Gini		0.002	0.002		0.018***	0.018***
		(0.002)	(0.002)		(0.005)	(0.005)
Land value (per acre)		-0.005	-0.005		-0.004	-0.004
		(0.003)	(0.003)		(0.007)	(0.007)
Network strength		0.004***	0.004**		0.002	0.002
		(0.002)	(0.002)		(0.006)	(0.006)
Family characteristics:						
Family Size		0.002**	0.001		-0.001	-0.002
		(0.001)	(0.001)		(0.001)	(0.001)
Birth order (ref = First born)						
Last son			0.005			0.008
			(0.003)			(0.008)
Middle son			0.008**			0.004
			(0.003)			(0.007)
Only son			-0.000			-0.013
			(0.005)			(0.011)
Observations	13974	13974	13974	13974	13974	13974
Sample	Child	Child	Child	Child	Child	Child
Province FE	No	No	No	No	Yes	No
Rural Only	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	PLU	PLU	PLU	PLU	PLU	PLU

* p < 0.10, ** p < 0.05, *** p < 0.01

Note: Marginal effects from probit estimation. Sample is confined to sons aged 6-18 in 1901 and living in a rural area. I adjust the standard errors to account for the clustering of observations across PLUs (Moulton, 1990). The family size control does not change the effects related to Poor Law Unions.

Table 7. Poor Law Union characteristics and rural outmigration for Child Sample

As expected, sons were more likely to emigrate from poorer, more unequal communities with stronger migrant networks. The estimates in Column 2 show that controlling for fathers' occupation, sons are more likely to emigrate from PLUs with higher Gini index values (greater inequality) and lower land values. These effects are statistically significant in models without the control for migrant *network strength* (model not shown). In Column 2, a standard deviation increase in family size and network strength are associated with a 0.02 and 0.04 percentage point increase in the probability of emigration, respectively. Thus, family size and the strength of migrant networks are stronger proximate predictors of emigration than land value and structure.

As one might expect, the family size effect attenuates with the inclusion of birth order controls in Column 3, which shows sons in the middle of the birth order to be almost a full percentage point more likely to emigrate than their first-born counterparts.

These community and family characteristics explain much of the occupational selection into migration to the United States. Column 1 shows laborers' sons to be 0.8 of a percentage point less likely to emigrate than farmers' sons. With the addition of controls for PLU characteristics and family size, however, this estimate more than halves and is no longer significant. Thus, it appears that the higher emigration rate among farmers' son can be accounted for by their residence in poorer communities with stronger emigrant networks, and their larger than average family sizes. Although the Irish-speaking effect also attenuates with these controls, it remains marginally significant.

The significance of these findings for emigration are brought into sharper relief by the analyses of occupational selection into internal migration. Column 4 shows the sons of laborers, skilled and white-collar workers to be more likely to move within Ireland than the sons of farmers. In contrast to the case of emigration, however, these effects persist with the addition of family and community controls in Columns 5 and 6. Further, while family size, birth order and network strength have significant effects on the probability of emigration, only the Gini measure of land inequality significantly affects internal migration outcomes. The stronger effect of land inequality on internal migration than emigration could reflect preferences of individuals to acquire land and thus, stay in Ireland. These findings extend on earlier community studies of emigration (Hatton & Williamson, 1993) by showing that the factors driving internal and overseas migration differed, and should not be ubiquitously reduced to being "push factors".

These results provide quite compelling evidence that emigration was an option pursued by those experiencing some of the harshest economic conditions. Emigration appears to have been concentrated in the poorest communities, which had stronger migrant networks and larger families. These communities had large numbers of poor and smallholding farming households, and it is emigration from these places which accounts for much of the national-scale differences in migration by farming and laboring background. Moreover, the positive association between family size and emigration (but not migration within Ireland) indicates that emigration may have been motivated by efforts to support, or relieve pressure on, larger families living on uneconomical farm holdings in Ireland.

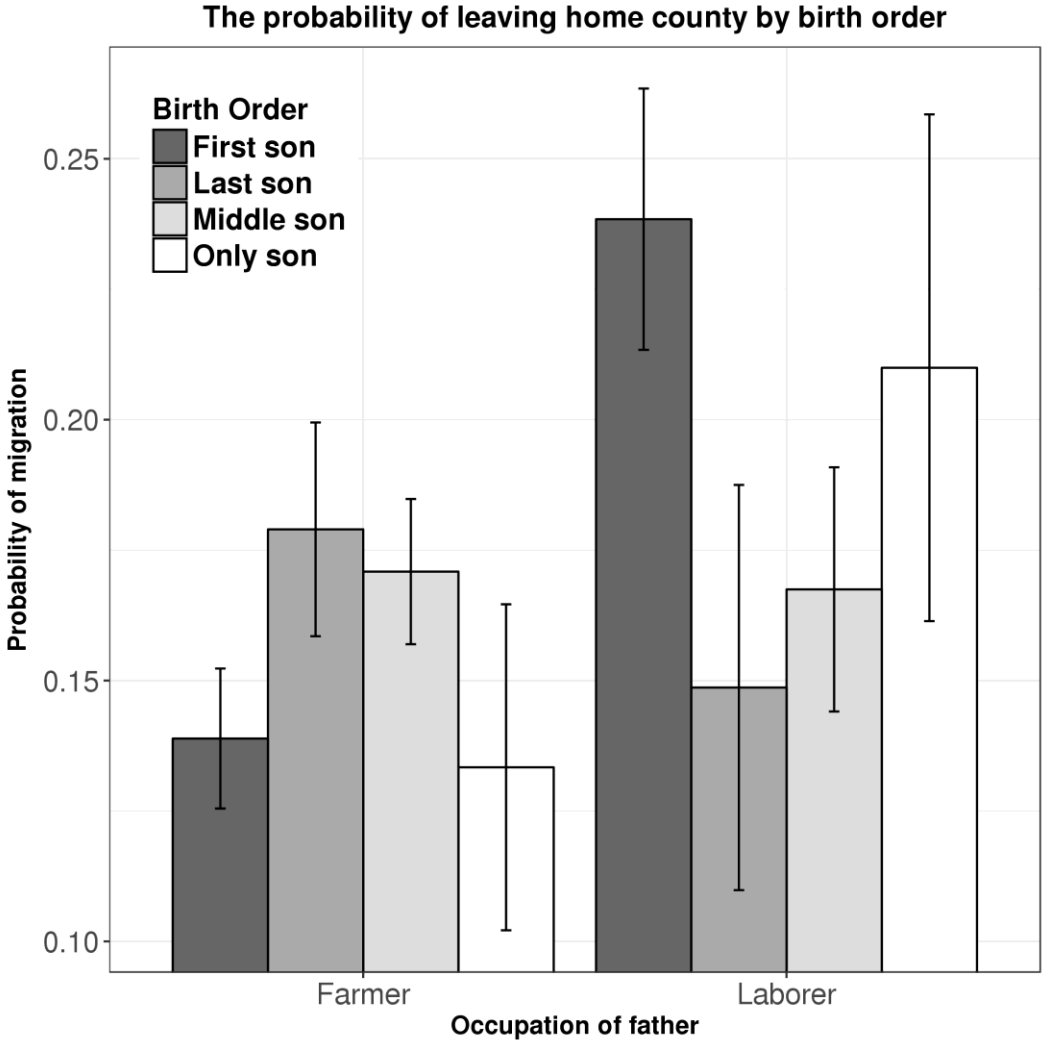
Although Irish-speakers tended to live in relatively impoverished conditions in Ireland, the positive association between Irish-speaking ability and emigration - and its negative relationship with internal migration - persists even with controls for community and family characteristics. This is puzzling as, due to the cultural and linguistic distance between Irish-speaking communities and American cities, one might expect Irish-speakers to face greater economic handicaps in US labor markets and be less incentivized to move.[13] One explanation may be that these patterns do not reflect economic motivations but, perhaps, social and political preferences. Although historians of Ireland have argued that higher emigration among Irish-speakers was driven by a

[13] Although many younger males growing up in Irish-speaking families would have spoken Irish and English, there would, at the least, have been detectable accent differences, which would as recent studies suggest, have likely been penalized in American labor markets in the 1900s (Abramitzky, Boustan, & Eriksson, 2016)

greater discontentment with British rule and the modernization of Irish society, this hypothesis is challenging to directly test here.

4.4. Birth order and family size effects on selection

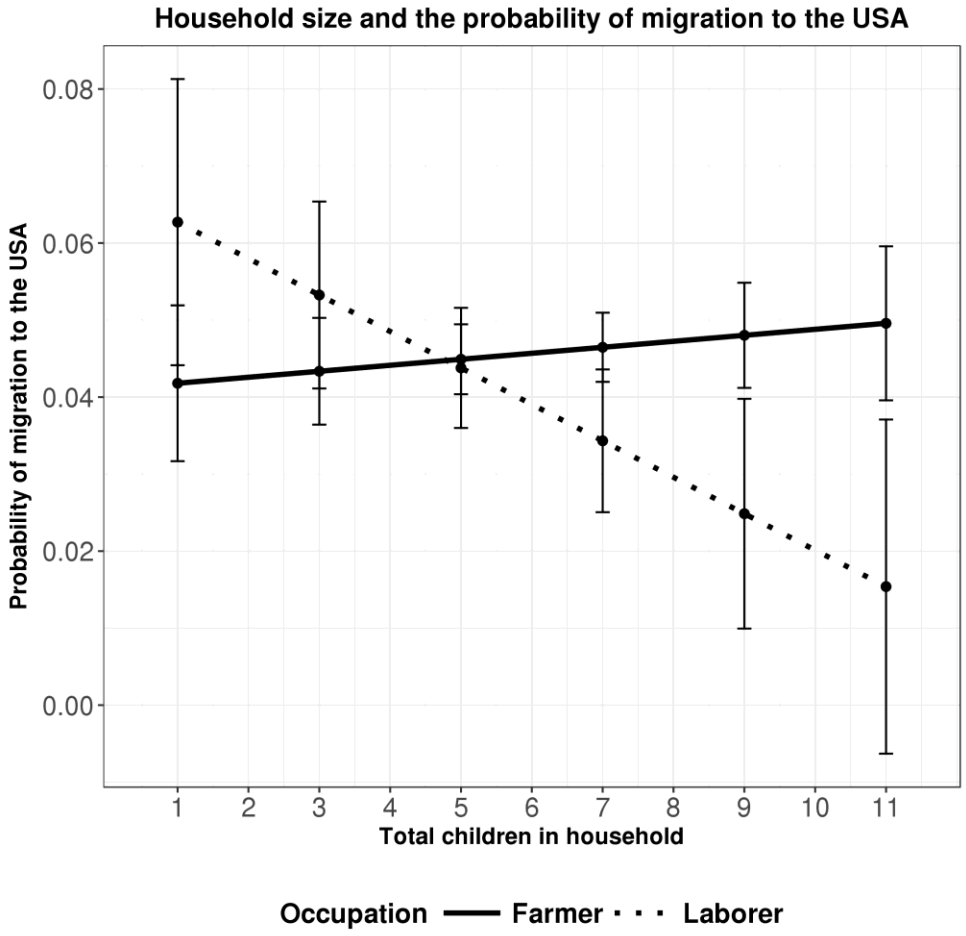
Despite these community influences on migration, it is striking that the sons of landholding farmers were more likely to emigrate than their landless counterparts, as one would expect their greater chances of inheriting land to discourage migration (Abramitzky et al., 2013). At the same time, however, farmers' sons may have also been better positioned than laborers' sons to finance migration to the United States. I test these hypotheses by interacting fathers' occupation with birth order and family size, as evidence suggests that first and last born sons of farmers were more likely to inherit land than middle-born sons of farmers (Ó Gráda, 1980) and, other things equal, larger families would likely be more constrained in financing emigration for their children.



Notes: These estimates were derived from a mode controlling for family size and PLU characteristics. The p-values are lower when family size is not included in the model. Confidence intervals are the 10% level.

Figure 4. Rural emigration and birth order in farming and laboring families

It appears that the sons of farmers who were less likely to inherit land were more likely to leave their home areas. Figure 4 plots the predicted probability of a son migrating away from their county of origin to a location within Ireland or the United States by birth order and fathers' occupation. Migration is significantly higher for the middle and last sons of farmers than their first-born counterparts. While the probability of migration for middle born sons – those least likely to inherit – is around 0.17, it is only 0.14 for first born sons. These birth order differences within farming families are generally consistent for migration within Ireland and abroad. Thus, it appears that while primogeniture (first born inheritance) may not necessarily have been the rule in historical Ireland (Guinnane, 1997), a stronger tendency to bequest land to first born sons did appear to have a knock-on effect in increasing migration among the middle sons of farmers.[14]



Notes: There is a significant difference ($p < 0.05$) in the slope of the total number of children in the household by whether the father was a farmer or a laborer. This model includes controls for PLU characteristics, and the effect holds when controlling for birth order. Confidence intervals are at the 10% level.

Figure 5. Household size and the probability of migration to the USA

[14] In the appendix, I show that while these effects are weaker, they also hold for emigration.

The decomposition of family size by fathers' occupation also provides evidence that farming families may have been better positioned to assist with the cost of moving to the United States. Figure 5 plots the predicted probability of emigration for farmers' and laborers' sons by family size and shows that at smaller family sizes, there appears is no significant difference in the probability of emigration by fathers' occupation. However, as family sizes increases – and likely budget constraints on migration and economic need with them – laborers' sons become significantly less likely to emigrate than farmers' sons. The probability of emigration for farmers' sons appears unchanged or even slightly increases with family size. This divergence is such that the sons of farmers in large families are significantly more likely to emigrate than their laboring counterparts.

These birth order and family size differences suggest that families may have affected emigration decisions in providing opportunity and possibly assisting with the costs of migration. Farming families provided greater opportunities to acquire land and the birth order analysis suggests that such opportunities discouraged migration. Although I cannot directly assess differences in the ability to finance the journey to America, the family size analysis provides evidence that laborers' sons may have been more cost constrained in terms of migration. Even poorer farming families had access to dowries and other revenue sources that could be gifted to non-inheriting children for migration (Guinnane, 1997). These cost differences may help account for the higher emigration rate of farmers' sons. As described in historical reports, such assistance may have also come with expectations that migrants would later support their families through cash transfers or assisting siblings with subsequent migration.

5. Robustness

5. 1. Quality of Birthplace Match

My relatively conservative approach to constructing the Birthplace Match aimed to improve the quality of successful linkages. This is important as recent evidence suggests that incorrect linkages can bias estimates from linked samples (Eriksson, 2016). Although no automated linkage procedure produces fully representative samples, name standardization and iterative age matching techniques can increase the rate of false positives (Bailey et al., 2017; Massey, 2017). These recent findings motivated my decision to take the more conservative linkage approach of not standardizing first names and using geographically concentrated last names.

The quality of the Birthplace Match is evident in its comparison to other linkage approaches and to previous historical estimates. Although 34 percent of the Standard Match moved county between 1901 and 1911, this rate halves to 17 percent in the Birthplace Match. Compared to the Standard Match, the internal migration patterns observed in the Birthplace Match more closely resemble Guinnane's (1997, p. 122-123) earlier maps and calculations showing that 13 percent of people in Ireland in 1911 lived in a county different to where they were born. Thus, the Birthplace Match appears to produce credible historical estimates of migration and is suggestive of the overall quality of the data.

Although the focus of the Birthplace Match on people with geographically concentrated last names produced a disproportionately rural sample, the compositional change between the linked and base samples are remarkably similar between the Standard and Birthplace Match. As reported for both samples in Table 1: successful linkage was less likely among laboring and

farming families and for males living in the poorer provinces of Connaught and Munster. Under both linkage approaches, the share of sons with parents who are farmers is 2-3 percentage points lower in the linked samples relative to their baseline samples. Further, while the Birthplace Match has a higher linkage rate than the Standard Match, the sample inclusion bias is quite similar.

5. 2. False positives in the linked sample

I use a series of samples linked with different criteria as evidence that false positives are not seriously distorting my results. The possibility that false positives are leading to erroneous conclusions are significantly reduced by my conservative linkage approach. It is possible, however, that the changing or misreporting of names between censuses and the iterative age matching procedure could still affect my results (see Bailey et al., 2017). To allay these concerns, Table 8 presents two further matching approaches.

	(1) Moved to the USA	(2) Moved to the USA	(3) Moved to the USA	(4) Moved to the USA
Age	0.020***	0.032***	0.013***	0.026***
	(0.003)	(0.003)	(0.003)	(0.003)
Age squared	-0.001***	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Occupation (ref = Farmer):				
Laborer	-0.009***	-0.014***	-0.006*	-0.013***
	(0.003)	(0.003)	(0.003)	(0.002)
Skilled/W.C.	-0.011***	-0.015***	-0.011***	-0.012***
	(0.003)	(0.003)	(0.003)	(0.003)
Speaks Irish	0.012***	0.014***	0.011**	0.013***
	(0.004)	(0.004)	(0.005)	(0.004)
Catholic	0.005	0.015***	0.005	0.009***
	(0.003)	(0.003)	(0.004)	(0.003)
Linkage approach	Birthplace Match	Standard Match	Birthplace-Standard Match	Strict-Age Match
Observations	17671	60017	10615	33529
Sample	Child	Child	Child	Child
Province FE	No	No	No	No
Rural Only	No	No	No	No

Standard errors in parentheses * p < 0.10, ** p < 0.05, *** p < 0.01

Table 8. Comparisons of estimates of emigration to the USA across linked samples

I investigate whether name changing is biasing my results by creating a third sample comprised of the 60 percent of individuals matched to the same person in the Birthplace Match *and* the Standard Match with NYSIIS standardized names ("Birthplace-Standard Match"). This sample of individuals have been successfully matched on their full name and their standardized names, and are highly likely to be linked to the correct person across censuses. Results from the Birthplace-Standard Match in Column 3 are very similar to those produced by the separate analyses of the Birthplace (Col. 1) and Standard Match (Col. 2). The sons of farmers are more likely to emigrate than the sons of laborers, skilled and white-collar workers, and the higher emigration rate of Irish speakers and Catholic is largely consistent across each sample.

To ensure that privileging individuals with more closely matching ages is not distorting my results, I created a highly constrained sample by linking individuals using their full first and last names but restricting the sample to individuals who have unique name and age characteristics within a five-year age window ("Strict-Age Match"). This approach uses the full population, so the sample size is larger than in the Birthplace Match. The results from the Strict-Age Match are, however, highly consistent with those generated by the other three samples, suggesting that false positives due to iterative age matching are not likely to be seriously affecting my conclusions.[15]

5. 3. Alternative migration flows

Perhaps the two most important migratory flows this study does not directly address is migration across the Irish Sea to Britain and female migration to North America. More than half of the 365,024 Irish-born people in the 18-37 range in the 1910 American census were female. There were also around 159,534 Irish-born people in the same age range living in Britain.[16] I make a cursory assessment of the characteristics of these flows, and how their omission from this analysis could affect my conclusions by describing the scale of these flows and, in lieu of linked data, using age-heaping techniques to gauge migrant selection with respect to numeracy.

Age-heaping methods can be used to assess numeracy in a population by comparing the share of persons erroneously rounding their ages to a number ending in a zero or a five (e.g. 30, 35, 40) (Mokyr, 1983). Age-heaping can be used to assess numeracy differences between challenging to link populations such as Irish-born females in Britain and the United States. I calculate the commonly used Whipple Index for Irish-born males and females aged 18-37 in Ireland in 1911, Britain in 1911, the USA in 1910 and Canada in 1911. The Whipple Index is computed as a ratio from the observed frequency of ages ending in a 0 or 5 in a population to the frequency predicted from a uniform distribution (A'Hearn, Baten, & Crayen, 2009). This is measured as:

$$W = \frac{\sum(n_{20} + n_{25} + n_{30} + n_{35})}{\frac{1}{5}\sum_{i=18}^{37} n_i} x\ 100 \qquad (5)$$

where a Whipple Index (W) value of 500 would indicate perfect heaping on multiples of five (lowest numeracy) and a value of 100 would indicate no heaping at all (highest numeracy).

Table 9 presents the Whipple Index for each population and confirms the finding that Irish migrants to the United States tended to be negatively selected, but migrants to Britain may have been more positively selected. Compared to their counterparts in Ireland, the Whipple Index is around 10 points higher for Irish-born males and females in the United States, suggesting that numeracy levels were lower in the Irish American population and not substantially different between males and females. In Britain, in contrast, the Whipple Index is 11 points lower than for Irish-based males, and 15 points lower for females. This suggests that younger Irish-born adults

[15] The appendix provides more in-depth analyses from these samples.

[16] Data from the North Atlantic Population Project suggest that share of young Irish males in Canada in 1911 was relatively smaller at around 14,000.

in Britain, particularly women, may have been more numerate on average than their counterparts in Ireland, and particularly so compared to Irish-born people in the USA. Compared to the British and American cases, the Whipple Index is quite similar for the Irish-born in Canada and their counterparts in Ireland.

	Male		Female	
	Whipple	Obs.	Whipple	Obs.
Ireland 1911	120	606,791	119	619,848
USA 1910	130	152,989	130	212,035
Britain 1911	109	87,822	105	71,712
Canada 1911	118	1,019	124	624

Note: There are no arrival year restrictions placed on the Irish-American population as the same restrictions cannot be placed on the British population. The Canadian data are a 5% sample of the population.

Table 9. Age-heaping in Ireland, Britain and North America

These findings support my overall conclusion of negative selection into emigration. This is because if migrants from Ireland to Britain were positively selected, my comparison of Irish-born males in Ireland and the USA may underestimate the degree of negative selection into emigration. However, a more detailed comparative study of the Irish emigration flow to these alternate destinations could be a valuable future project. While sex and gender differences in Irish emigration are also of interest, the similarity in the Whipple Index of males and females in Ireland and the USA suggests that my omission of females from the analysis is not likely to lead me to seriously mischaracterize the occupational and class backgrounds of Irish emigrants in the United States.[17]

6. Conclusion

A century ago, industrial growth in Ireland was sluggish and hundreds of thousands of Irish people moved to American cities. Government leaders and officials were concerned that Ireland was losing its most talented and able individuals to the United States. Yet by showing that emigrants were typically drawn from farming households in the poorest Irish communities, my findings provide only limited support for these concerns. Although these migrants were more likely to be the sons of landholders, and may have been wealthier - perhaps even healthier - than their laboring counterparts (Royal Commission on Labour, 1893), they came from poorer and lower status families, suggesting that they were negatively selected from the population as a whole. This finding of negative selection adds to historical evidence providing confirmatory support for predictions from canonical economic models of migration (Boustan & Abramitzky, 2016).

While these patterns of migrant selection may reflect a pursuit of economic opportunity, they could also be rooted in attempts to improve conditions in Ireland. Economists and

[17] Another omitted migration flow which may be of consequence are return migrants: recent estimates that for each of the 644,574 Irish immigrants arriving in the United States from 1900-1910, 477,324 or 74 percent returned to Ireland (Bandiera, Rasul, & Viarengo, 2013). Studying whether the characteristics of return migrants (from abroad or within Ireland) differed from those who stayed in their destinations would be a challenging but valuable topic for further study.

sociologists of migration have widely noted that families in low-income countries often encourage the migration of sons and daughters to secure remittances, and the diffusion of such behaviors can lead entire communities to specialize in migration (Rapoport & Docquier, 2006; Taylor, 1999). My findings of higher emigration rates among Irish-speakers, smallholders, larger families, and from areas with denser migrant networks, may suggest that Irish communities were engaging in similar strategies – a point frequently made in reports from the early twentieth century. The fact that these families appear to have favored emigration to high wage American cities over locations within Ireland provides further credence to this interpretation.

More broadly, my findings suggest that taking a broader perspective on source-country conditions could be valuable for understanding migrant selection. In addition to the emphasis of recent studies on cultural inheritance systems, discrimination and migrant networks in shaping selection patterns (Abramitzky et al., 2013; Spitzer, 2013; Spitzer & Zimran, 2015), political and cultural preferences may have mattered too. Reports of the social and political disconnect between Irish-speaking communities and modernizing early twentieth century Irish society (Miller, 1988), are supported by my findings of higher rates of emigration among Irish-speakers.[18] Thus, there may be considerable scope for further quantitative assessment of this issue, and other political and cultural motivations for emigration during the Age of Mass Migration.

This study broadly illustrates the richness of historical data which can be brought to bear on questions of migrant selectivity. Concerns around the 'quantity and quality' of immigrants have continued to gain salience in the contemporary political context. Yet, we still know surprisingly little of how migrants are selected from their populations or of the interacting conditions which generate patterns of migrant selection. In lieu of high-quality data from the contemporary context, rapid innovations in historical data analysis provide a valuable alternative to better understand these important economic and social science questions.

7. References

Abramitzky, R., Boustan, L., & Eriksson, K. (2012). Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, *102*(5), 1832–1856.

Abramitzky, R., Boustan, L., & Eriksson, K. (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, R., Boustan, L., & Eriksson, K. (2014). A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy*, *122*(4).

Abramitzky, R., Boustan, L. P., & Eriksson, K. (2016). *Cultural assimilation during the age of mass migration*. National Bureau of Economic Research.

Ager, P., & Brückner, M. (2013). Cultural diversity and economic growth: Evidence from the US during the age of mass migration. *European Economic Review*, *64*, 76–97.

A'Hearn, B., Baten, J., & Crayen, D. (2009). Quantifying quantitative literacy: age heaping and the history of human capital. *The Journal of Economic History*, *69*(3), 783–808.

Anbinder, T., & McCaffrey, H. (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, C. M., & Kimball, S. T. (1940). *Family and community in Ireland*. Harvard University Press.

Bailey, M., Cole, C., Henderson, M., & Massey, C. (2017). *How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth*. National Bureau of Economic Research.

Bandiera, O., Rasul, I., & Viarengo, M. (2013). The making of modern America: Migratory flows in the age of mass migration. *Journal of Development Economics*, *102*, 23–47.

[18] See Waldinger (2013) for discussion of emigrants and political engagement with the homeland

Begley, J., Geary, F., & Stark, T. (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, M. V., & Hatton, T. J. (2012). Immigrant Selection in the OECD*. *The Scandinavian Journal of Economics*, *114*(4), 1105–1128.
Borchers, A., Ifft, J., & Kuethe, T. (2014). Linking the price of agricultural land to use values and amenities. *American Journal of Agricultural Economics*, *96*(5), 1307–1320.
Borjas, G. J. (2014). *Immigration economics*. Harvard University Press.
Boustan, L., & Abramitzky, R. (2016). Immigration in American Economic History. *Journal of Economic Literature*.
Burt, O. R. (1986). Econometric modeling of the capitalization formula for farmland prices. *American Journal of Agricultural Economics*, *68*(1), 10–26.
Connor, D. S. (2017). Poverty, Religious Differences and Child Mortality in the Early 20th Century: The Case of Dublin. *Annals of the American Association of Geographers*.
Connor, D. S., Mills, G., & Moore-Cherry, N. (2011). The 1911 Census and Dublin city: A spatial analysis. *Irish Geography*, *44*(2–3), 245–263.
Featherstone, A. M., & Baker, T. G. (1987). An examination of farm sector real asset dynamics: 1910–85. *American Journal of Agricultural Economics*, *69*(3), 532–546.
Feigenbaum, J. (2016). JAROWINKLER: Stata module to calculate the Jaro-Winkler distance between strings. *Statistical Software Components*.
Fernández-Huertas Moraga, J. (2011). New evidence on emigrant selection. *The Review of Economics and Statistics*, *93*(1), 72–96.
Fernihough, A. (2017). Human capital and the quantity–quality trade-off during the demographic transition. *Journal of Economic Growth*, *22*(1), 35–65.
Fernihough, A., Ó Gráda, C., & Walsh, B. M. (2015). Intermarriage in a Divided Society: Ireland a Century Ago. *Explorations in Economic History*, *56*, 1–14.
Ferrie, J. P. (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, D. (1980a). Irish emigration in the later nineteenth century. *Irish Historical Studies*, *22*(86), 126–143.
Fitzpatrick, D. (1980b). The disappearance of the Irish agricultural labourer, 1841-1912. *Irish Economic and Social History*, *7*, 66–92.
Fitzpatrick, D. (1984). *Irish Emigration 1801-1921* (Vol. 1).
Gaeltacht Commission. (1926). *Gaeltacht Commission: Report*. Dublin: The Stationery Office.
Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). Do institutions cause growth? *Journal of Economic Growth*, *9*(3), 271–303.
Guinnane, T. (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* (pp. 261–282). NBER.
Hatton, T. J., & Williamson, J. G. (1993). After the Famine: Emigration from Ireland, 1850–1913. *The Journal of Economic History*, *53*(03), 575–600.
Juif, D. (2015). Skill Selectivity in Transatlantic Migration: The Case of Canary Islanders in Cuba. *Revista de Historia Económica/Journal of Iberian and Latin American Economic History*, 1–34.
Kosack, E., & Ward, Z. (2014). Who Crossed the Border? Self-Selection of Mexican Migrants in the Early Twentieth Century. *The Journal of Economic History*, *74*(04), 1015–1044.
Long, J., & Ferrie, J. (2013). British, American, and British-American Social Mobility: Intergenerational Occupational Change Among Migrants and Non-Migrants in the Late 19th Century. *Working Paper*.
Magee, G. B., & Thompson, A. S. (2006). The global and local: explaining migrant remittance flows in the English-speaking world, 1880–1914. *The Journal of Economic History*, *66*(1), 177–202.
Massey, C. 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, R. E. (1901). *Special Report on Surnames in Ireland: Together with Varieties and Synonymes of Surnames and Christian Names in Ireland*. Dublin: H.M. Stationery Office.
Miller, K. A. (2008). *Ireland and Irish America: Culture, Class, and Transatlantic Migration*. Field Day Publications.

Mokyr, J. (1983). *Why Ireland starved: a quantitative and analytical history of the Irish economy, 1800-1850*. Routledge.
Mokyr, J., & Ó Gráda, C. (1982). Emigration and poverty in prefamine Ireland. *Explorations in Economic History*, *19*(4), 360–384.
Moulton, B. 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.
Ó Gráda, C. (1975). A note on nineteenth-century Irish emigration statistics. *Population Studies*, *29*(1), 143–149.
Ó Gráda, C. (1980). Primogeniture and Ultimogeniture in Rural Ireland. *Journal of Interdisciplinary History*, *10*(3), 491.
Ó Gráda, C. (1994). *Ireland: a new economic history 1780-1939*. Clarendon Press. Retrieved from http://ideas.repec.org/b/oxp/obooks/9780198205982.html
Plantinga, A. J., Lubowski, R. N., & Stavins, R. N. (2002). The effects of potential land development on agricultural land prices. *Journal of Urban Economics*, *52*(3), 561–581.
Rapoport, H., & Docquier, F. (2006). The economics of migrants' remittances. *Handbook of the Economics of Giving, Altruism and Reciprocity*, *2*, 1135–1198.
Rodríguez-Pose, A., & von Berlepsch, V. (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), 628–651.
Royal Commission on Labour. (1893). *The Agricultural Labourer, Assistant Commissioners' Reports on the Agricultural Labourer* (Vol. IV). London: H.M. Stationery Office.
Schrier, A. (1958). *Ireland and the American emigration, 1850-1900*. Dufour Editions.
Spitzer, Y. (2013). Pogroms, Networks, and Migration. Retrieved from http://www.gmu.edu/centers/publicchoice/wed%20seminars/papers%20spring%2014/4_9%20Spitzer.pdf
Spitzer, Y., & Zimran, A. (2015). *Self-Selection of Immigrants on the Basis of Living Standards: Evidence from the Stature of Italian Immigrants at Ellis Island 1907–1925*. unpublished manuscript. Retrieved from http://www.eh.net/eha/wp-content/uploads/2013/12/Proposal.pdf
Stark, O., & Bloom, D. E. (1985). The new economics of labor migration. *The American Economic Review*, *75*(2), 173–178.
Stark, O., & Taylor, J. E. (1989). Relative Deprivation and International Migration. *Demography*, *26*(1), 1–14.
Taylor, E. J. (1999). The new economics of labour migration and the role of remittances in the migration process. *International Migration*, *37*(1), 63–88.
Turner, M. (2002). *After the famine: Irish agriculture, 1850-1914*. Cambridge University Press.
van Leeuwen, M., & Maas, I. (2011). *Hisclass: A Historical International Social Class Scheme*. Universitaire Pers Leuven.
Waldinger, R. (2013). Engaging from abroad: The sociology of emigrant politics. *Migration Studies*, *2*(3), 319–339.
Wegge, S. A. (1998). Chain migration and information networks: Evidence from nineteenth-century Hesse-Cassel. *The Journal of Economic History*, *58*(4), 957–986.
Wegge, S. A. (2002). Occupational self-selection of European emigrants: Evidence from nineteenth-century Hesse-Cassel. *European Review of Economic History*, *6*(3), 365–394.
Wegge, S. A. (2008). Network strategies of nineteenth century Hesse-Cassel emigrants. *The History of the Family*, *13*(3), 296–314.
Wegge, S. A., Anbinder, T., & Ó Gráda, C. (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.