NBER WORKING PAPER SERIES

# IDENTITY, PAROCHIAL INSTITUTIONS, AND OCCUPATIONAL CHOICE: LINKING THE PAST TO THE PRESENT IN THE AMERICAN MIDWEST

Kaivan Munshi
Nicholas Wilson

Working Paper 13717
http://www.nber.org/papers/w13717

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2008

We are grateful to Daron Acemoglu, Abhijit Banerjee, Roland Benabou, Sandra Black, Leah Platt Boustan, Esther Duflo, Raquel Fernandez, Andrew Foster, Chris Foster, Vernon Henderson, Dennis Hogan, Michael Kremer, Ashley Lester, Glenn Loury, Nancy Luke, David Meyer, David Weil, and Michael White for helpful discussions and suggestions. Seminar participants at Georgia Tech, Harvard-MIT, Indiana University, NBER, and UCLA provided many helpful comments. Ben Feigenberg provided excellent research assistance. Research support from the National Science Foundation through grant SES-0617847 is gratefully acknowledged. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Identity, Parochial Institutions, and Occupational Choice: Linking the Past to the Present in the American Midwest

Kaivan Munshi and Nicholas Wilson

NBER Working Paper No. 13717
January 2008
JEL No. J24,J61,N11,R23

## ABSTRACT

This paper documents the presence of non-economic career motivations in the U.S. labor market, explores reasons why such motivations could arise, and provides an explanation for why they might have persisted across many generations. The analysis links ethnic (migrant) labor market networks in the American Midwest when it was first being settled, the local identity or attachment to place that emerged endogenously to maintain the integrity of these networks, and occupational choice today. While fractionalization may adversely affect the performance of secular institutions, ethnic competition in the labor market could at the same time have strengthened within-group loyalty and parochial institutions. These values and their complementary institutions, notably the church, could have mutually reinforced each other over many overlapping generations, long after the networks themselves had ceased to be salient. Counties with greater ethnic fractionalization in 1860 are indeed associated with steadily increasing participation in select religious denominations historically dominated by the migrants all the way through the twentieth century. Complementing this result, individuals born in high fractionalization counties are significantly less likely to select into geographically mobile professional occupations and, hence, to migrate out of their county of birth, despite the fact that these counties are indistinguishable from low fractionalization counties in terms of local public good provision and economic activity today.

Kaivan Munshi
Professor
Department of Economics
Brown University
Box B/ 64 Waterman Street
Providence, RI 02912
and NBER
munshi@brown.edu

Nicholas Wilson
Brown University
Nicholas_Wilson@brown.edu

# 1 Introduction

The textbook economic model assigns individuals to occupations on the basis of their ability, the opportunities they have to invest in human capital, and the types of jobs that are subsequently made available to them in the labor market. Outside of economics, social scientists have long argued that non-economic factors can also determine career choices. Weber's (1930) writings on the Protestant ethos are the classic reference in this literature. Contemporary sociologists have also described, for example, how working class communities instill values in their members that shape their aspirations and restrict occupational mobility (Gans 1962). Our objective in this paper is to document the presence of such non-economic motivations in the U.S. labor market, to explore the reasons why such motivations could arise, and to understand why they might persist across many generations.

We start with the idea that social groups can endogenously instill values in their members that increase loyalty and reduce the propensity to exit, particularly when they are vulnerable to acculturation (Bisin and Verdier 2000) or in competition with other groups (Sumner 1906). The negative consequences of population heterogeneity, typically measured by fractionalization along various economic and social dimensions, are well documented in the economics literature.[1] While ethnic or racial fractionalization may adversely affect the performance of *secular* institutions, we argue that such fractionalization can at the same time strengthen within-group loyalty and *parochial* institutions. These values and their complementary institutions tend to be persistent and can shape career choices long after the economic forces that gave rise to them have ceased to be salient.

In the setting that we consider – the American Midwest – the labor market was the primary domain in which ethnic groups interacted. Migrant workers had to bear most of the fluctuations in labor demand in the early stages of industrialization in the United States (Hoerder 1991). This is precisely the situation in which labor market networks are most useful and there is a wealth of historical evidence documenting the role played by ethnic migrant networks in finding jobs for their members at this time. With the arrival of the railroads around 1850, the Midwest in particular witnessed a large influx of European migrants in response to the new occupational opportunities that became available. Historians describe the efforts made by migrant communities to establish "occupational beachheads" and "toe-holds" in new locations, and to subsequently work hard to maintain these coveted positions once they were established. We will show that the cost to the network from the exit of one of its

[1] Alesina, Baqir, and Easterly (1999), for example, document the negative relationship between racial fractionalization and public good provision in the United States.

members and, hence, efforts to instill a sense of loyalty to the local community would have been greatest in fractionalized labor markets with many ethnic networks competing for coveted permanent jobs.

The migrant church was historically central to the maintenance of the ethnic network, providing a domain in which information could flow, commitments could be enforced, and loyalty to the local community could be instilled (Gjerde 1991). Outside of the labor market, the church provided other services to its members as well, including social activity and support when they were sick or infirm. Labor market networks and civic institutions based on a common European ancestry are largely irrelevant in the Midwest today (Gans 1979, Alba 1990). However, it would be incorrect to assume that these ethnic institutions disappeared without a trace. While the labor market networks might have disappeared, the church continues to provide important forms of mutual assistance, which require, in turn, a relatively high level of commitment in terms of time and effort from the congregation. We expect churches to have made a particular effort to instill a sense of group-loyalty among their members in fractionalized markets with competing ethnic groups at the time of initial settlement. Such loyalty and commitment to the local community would have led, in turn, to well functioning churches. If members' inputs in the church are complements, as suggested by Iannaccone (1998), and if there is sufficient overlap across generations, then the incumbent members of these well functioning churches would have had a greater incentive to instill an identity that discouraged geographical mobility in the generation that followed. This process would have repeated itself from one generation to the next, in the church and in other institutions that formed around it, linking initial settlement patterns to local identity or a loyalty to place that is independent of ethnicity today.

Loyalty to place or local identity (Hunter 1975, Guerson, Stueve and Fischer 1977, Hummon 1990) is well described by the following quote from a resident of Bloomington, Indiana who moved to that city to live with his wife who was born nearby, "... my chief ambition, I discovered during our early years in Bloomington, was not to make a good career but to make a good life. And such a life as I came to understand it, meant being a husband and a father first, and an employee second; it meant belonging to a place rather than to a profession ... So as I came to recognize my children's need and my own need for a firm home place, I came to understand my community's need for citizens who stay put. Most of what I valued in Bloomington was the result of efforts by people who loved the place, either because they grew up here and chose to stay, or because they landed here and chose to remain" (Sanders 2007: 67-68).

Two aspects of local identity emerge from the preceding quote. First, local identity is associated with values that stress loyalty to a specific place. As in Akerlof and Kranton (2000), we expect that individuals born and socialized in communities with a strong local identity will incur a utility cost when they deviate from prescribed behavior and leave their communities. In the U.S. labor market, this implies that individuals endowed with a strong sense of local identity will be less likely to choose professional occupations, which are associated with significantly greater work-related moves than non-professional occupations. Second, we see that the core value associated with local identity, notably a commitment to stay, complements local institutions. Strong identity and well functioning institutions are mutually reinforcing and, as discussed above, one cannot persist without the other (Bénabou and Tirole [2006] make the same argument to explain the persistence of mutually reinforcing collective beliefs and political equilibria).

The homogenization of production that is characteristic of the Midwest only began around 1880. Thus, between 1850 and 1880 there was a window of time during which this region was rapidly settled and during which there was substantial fluctuation in labor market conditions and, hence, ethnic rivalry across local areas. These initial conditions are crucial because once local institutions and a level of individual commitment were established, these mutually reinforcing effects would have remained in place over many generations in the future. We consequently exploit variation in initial conditions *within* the Midwest to identify the effect of historical networks, and the identity and parochial institutions they engendered, on occupational choice and geographical mobility long after the networks themselves had disappeared. The framework we have outlined generates two testable predictions. The first prediction is that areas with greater ethnic rivalry at the time of settlement should be associated with stronger local identity, which implies that individuals born (and socialized) in those areas should be less likely to enter a more mobile professional career today. The second prediction is that greater ethnic rivalry should also have given rise to better functioning *migrant* churches and, hence, greater religious participation in select denominations historically dominated by the migrants, over many generations in the future.

To test the first prediction we combine data from the U.S. census and the National Longitudinal Survey of Youth 1979 (NLSY79). A major advantage of situating the analysis in the Midwest is that individual level data on occupations and ethnicity can be obtained from the census going back to the time when this region was starting to develop. We focus on the 1860 census when rapid growth was just commencing and construct a measure of ethnic fractionalization, one minus the Herfindahl

concentration index, within each broad occupational category in each county in that census year. Averaging over all occupational categories in each county we arrive at a measure of ethnic fractionalization in 1860 that we expect is positively correlated with ethnic rivalry in the labor market at that time and, hence, with the strength of local identity in the county as it subsequently emerged. Matching the county-level measure of ethnic fractionalization with individual data from the NLSY79, we find that individuals born in high fractionalization counties are significantly *less* likely to be employed in professional occupations and are significantly *less* likely to have migrated out of their birth county in 2000.

The values that we believe are responsible for the career choices described above could not have persisted in a local community over so many generations without institutional support. The second prediction of our framework is that high fractionalization counties should have been associated with better functioning migrant churches and related parochial institutions. If these institutional features persisted over time, supporting and being supported by local identity, then we would expect to see greater participation in religious denominations that were historically dominated by migrants, notably the Lutherans and the Catholics, today. Using data from the Census of Religious Bodies, available at roughly ten-year intervals from 1860 to 2000, we successfully verify this important prediction. The share of Lutherans and Catholics in the population is significantly larger in high fractionalization counties by 1870 and, most importantly, this gap grows steadily wider over the course of the twentieth century. In contrast, participation in other denominations is significantly *lower* in the high fractionalization counties over the entire period.

Our interpretation of these striking results is that local institutions and individual values, determined endogenously by the labor market when the Midwest was first being settled in 1860, have subsequently persisted over time and continue to shape occupational choices in the county one hundred and forty years later. Alternative explanations assume that the ability endowment in the population, investments in human capital, or access to particular types of jobs today vary with fractionalization in 1860. For example, although ethnic networks and European ancestry are no longer directly relevant in the Midwest economy, ethnic fractionalization in 1860 could potentially be correlated with particular features of the economy at that time, which had persistent effects and are correlated with the *demand* for professional labor today. Similarly, the well documented negative correlation between fractionalization and public good provision could have given rise to poorly functioning schools in high fractionalization counties.

If identity is salient and the current demand for professional labor is uncorrelated with historical fractionalization, then we show that there will be a mismatch in the labor market as long as there is *ex ante* uncertainty in the demand for different types of jobs. High fractionalization counties will supply too few professionals and too many non-professionals. Professional labor must move into these counties and non-professional labor must move out of them *ex post* in competitive equilibrium. It follows that individuals *residing* in high fractionalization counties will be just as likely to hold professional jobs as individuals residing in low fractionalization counties, once the labor market clears. It is only individuals *born* in high fractionalization counties who should be less likely to be professionals. The alternative hypothesis, based on differences in the demand for professional labor across counties, predicts that individuals *born and residing* in high fractionalization counties should be less likely to be professionals. We will later verify that labor flows systematically with the level of historical fractionalization as described above and, more importantly, that there is no correlation between occupational choice and the historical level of ethnic fractionalization in the county of residence, ruling out the most obvious alternative explanation for our results.

Our results could also be obtained if the incoming migrants in the high fractionalization counties had lower ability, and this ability differential persisted across generations. Human capital transmission is less of a concern with our analysis since we are going back at least five generations; even if the ability distribution of the incoming migrants varied systematically with fractionalization in 1860, this heterogeneity would have long since disappeared, given the high rates of inter-county migration (close to 60 percent on average) that we document below.[2] Moreover, given the dynamic nature of the local economy at the time of initial settlement, fractionalization was not very persistent; the correlation between 1860 and 1870 fractionalization is just 0.3 and the correlation between 1860 and 1900 fractionalization is just 0.16. We thus do not expect to find a strong association between our measure of fractionalization in 1860 and the local heterogeneity that could have determined formative investments in the school system many decades later in the twentieth century, as documented by Goldin and Katz

[2]The empirical strategy that we employ is related to the strategy adopted by Fernandez and Fogli (2007) in a very interesting recent paper. Fernandez and Fogli establish that female labor force participation in the origin country affects labor force participation and fertility rates among second-generation female migrants in the United States. They rely on the assumption that cultural traits will be transmitted across space and so cultural heterogeneity will continue to manifest itself in an economic environment – the U.S. labor market – that does not distinguish between social groups. The obvious alternative explanation, which they take care to rule out, is that human capital rather than culture is being transmitted across generations. With our strategy, the chief concern is that there is something about the *place*, other than identity and its complementary institutions, that has persisted over time, which the preceding test on the county of residence, rather than the county of birth, helps rule out.

(1999). The NLSY provides information on AFQT scores and educational attainment. Information on local public good provision, including education expenditures, can also be obtained at the county level in 1990. As expected, we find no relationship between historical fractionalization and current investments in schooling or educational attainment.

Complementing recent studies that investigate the relationship between culture and growth (Barro and McCleary 2003, Tabellini 2005, Fernandez and Fogli 2007), we find that local identity and its accompanying institutions, notably the church, have an important effect on occupational choice. A one standard deviation decline in 1860 fractionalization would increase the proportion of individuals holding professional jobs in 2000 from 9 percent to 15 percent. Internal migration clears the labor market in our U.S. setting and so we do not expect identity to have a substantial effect on local growth rates. Nevertheless, the fact that non-economic factors have such a large and persistent effect on important economic choices in a region as homogenous as the American Midwest suggests that cultural effects may be even stronger across countries with very different histories.

# 2 The Institutional Setting

## 2.1 The Settling of the Midwest

The Midwest first began to be settled in the early nineteenth century with the expansion of the national canal system. The Erie Canal linking the Hudson to Lake Erie was completed in 1825 and numerous inter-regional and intra-regional canals were built over the next two decades (Fishlow 2000). However, it was only with the arrival of the railroad that the Midwest took off on a steeper growth trajectory. Before 1850 the Midwest had less than one thousand miles of track, but almost ten thousand were added by 1857 (Meyer 1989).

Improved rail transportation stimulated industrialization and the Midwest's share of national manufacturing increased rapidly between 1860 and 1920, with almost half of this increase occurring in the 1860's (Meyer 1989). This increase in economic activity led, in turn, to an increase in the demand for labor. In 1810, approximately 6 percent of the labor force (outside the southern states) resided in the Midwest. My 1860, this share had increased to 41 percent, with a further increase to 51 percent by 1880, after which regional growth converged to the national average (Margo 1999).

In this paper, the Midwest is comprised of the states of Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin (Missouri, the only pre-Civil War slave state in the Midwest, is excluded from the

analysis). Using county-level census data we see that the number of incorporated counties increases sharply from 1850 to 1860 and then flattens out by 1880 in Figure 1. Information on railroads, obtained from the Historical Map Archive at the University of Alabama, indicates that the number of these counties with a railroad also increases steeply over the 1850-1870 period, growing thereafter at a slower rate.[3] The rapid expansion of the railroad system and the economic activity that accompanied it led to a steep increase in the population of the Midwest as well as an influx of foreign migrants. Using county-level census statistics, the total population in our seven Midwestern states grew from less than 5 million in 1850 to 20 million in 1900. The number of foreign-born migrants nearly tripled between 1850 and 1860, reaching close to 20 percent of the population.

Where did these migrants come from? Individual-level data, including characteristics such as age, sex, occupation, and country of birth, are publicly available from the Population Census each decade from 1850 to 1930. We use the 1-in-100 sample from the 1860-1900 IPUMS to study changes in the migrant population in these critical early decades in Table 1. The English (13 percent), the Irish (25 percent), and the Germans (32 percent), dominated the migrant population in the Midwest in 1860, just after the first wave of migration described above, with no other ethnic group accounting for more than a 3 percent share of the migrants in that year. Subsequently, the English and the Irish were displaced by the Germans and the Scandinavians over the 1860-1900 period. Notice that the Italians, Poles, and Slavs continue to be insignificant in 1900, although they would display a substantial presence in Midwestern cities such as Cleveland, Chicago, and Pittsburgh by the first quarter of the twentieth century.

What jobs did the migrants occupy? Table 2 reports the occupational distribution of the migrants from the IPUMS sample in 1860, 1880, and 1900. Although agriculture was the dominant sector in this period, the share of farm employment declines from 62 percent in 1860 to 48 percent in 1900, with manufacturing operatives and laborers accounting for much of the increase in non-farm employment. These trends are consistent with the growth of the manufacturing sector described above and they are similar for the foreign-born migrants and the native workers (not reported).

The apparent similarity in the occupational distribution for migrants and natives masks differences in the type of jobs they had access to within each occupational category. Labor markets in the nineteenth century could be divided into three segments: a stable segment with permanent employment, an

[3]Railroad maps were used to construct a county-level binary variable indicating whether any part of a railroad ran through the county in a given year. Railroad maps were unavailable in some census years in which case we used maps that were closest in vintage to those census years (the discrepancy never exceeded three years).

unstable segment with periodic short-term unemployment, and a marginal but highly flexible segment characterized by spells of long-term and short-term employment (Gordon, Edwards, Reich 1982). Migrants being newcomers to the U.S. market typically ended up in the unstable and marginal segments, where the uncertain labor demand naturally provided an impetus for the formation of ethnic networks that helped their members find jobs (Hoerder 1991). We would expect these networks to have been particularly important during the rapid expansion phase around 1860, with the influx of migrants and the opening up of new labor markets in both agriculture and manufacturing. The initial conditions in the empirical analysis, measuring ethnic rivalry in the labor market at the time of settlement, will consequently be measured using data at this point in time.

### 2.2 Ethnic Labor Networks in the Midwest

Accounts by contemporary observers and a rich social history literature indicate that friends and kin from the origin community in Europe played an important role in securing jobs for migrants in the Midwest in the nineteenth century and the first quarter of the twentieth century. Early historical studies used census data, which provides fairly detailed occupational and ethnic information, to identify ethnic clusters in particular locations and occupations. Gordon, Edwards, and Reich (1982) note that although foreign-born workers comprised just over 20 percent of the labor force in 1870, they accounted for 43 percent of the iron and steel operatives, 43 percent of the woolen mill workers, and 63 percent of the miners. Nearly a quarter of railroad workers were Irish, a third of the miners were British, and about half the workers in the baking and confectionary business were German. While such clustering suggests that underlying ethnic networks were channelling their members into particular occupations, it could simply reflect the fact that migrants arrived with specific skills. Hutchinson's (1956) analysis of 1950 census data, however, indicates that clustering continued even among the migrants' children, with the concentration of particular ethnic groups in some industries actually increasing from the first to the second generation.

Although census data are a useful source of information, they do not provide details of the migration process and its connection to ethnic networks in the United States. Over the past four decades, however, social historians have linked parish registers and county data in specific European sending communities to census and church records in the United States to construct the entire chain of migration from those communities as it unfolded over time. This remarkable research effort has documented the formation of new settlements in the Midwest by pioneering migrants, the subsequent channelling of migrants

from the origin community in Europe to these settlements, as well as the movement of groups from the original settlement to new satellite colonies elsewhere in the United States. As with the census data, this research identifies occupational and geographic clustering, but at a dissagregate level. Over 45 percent of the Swedish emigrants from the parish of Rätvik eventually settled in Isanti County, Minnesota (Ostergren 1976). Two-thirds of the emigrants from Balestrand located in Norway Grove, Wisconsin in the first decade of migration from that Norwegian community (Gjerde 1985). And one-third to one-half of the German emigrants from Westerkappeln settled in Duden County, Missouri (Kamphoefner 1987). Although less detailed origin-country information is available for southern European migrants, similar ethnic clustering in particular neighborhoods of Midwestern cities has been documented for Polish, Italian, and Slovak immigrants from specific sending regions (Alexander 1991, Bodnar, Simon, and Weber 1982).

A possibly stronger indicator of the importance of migrant networks is the maintenance of ethnic ties over successive moves *within* the United States. Italians moved from Southern Illinois to the "Italian Hill" in St. Louis when coal mining operations were reduced in the 1920's and Slavs moved from mines in Western Pennsylvania to Detroit's growing automobile industry in the same decade (Bodnar 1985). Norwegians from Balestrand initially settled in Norway Grove, Wisconsin, but over time they established six satellite settlements in Wisconsin, Minnesota, Iowa, and Illinois (Gjerde 1985). A similar pattern has been documented for Norwegians immigrants from Fortun, who initially settled in Vernon and Crawford Counties, Missouri, but later established satellite communities throughout the Midwest (Gjerde 1997).

While the preceding descriptions of ethnic clustering are informative, ethnic concentration within specific departments or firms in a local industry provides possibly the strongest evidence that labor networks were active. Nearly all three thousand employees of the Peninsular Gas Company in Detroit in 1900 were Polish, and Croatians held only three jobs in Indiana's oil refineries: stillman helper, fireman, and still cleaner (Bodnar 1985). Italians in Pittsburgh's steel industry dominated the carpentry, repair, and rail shops. And, relying on friends and relatives, Poles established occupational niches at the Jones and Laughlin and Oliver Mills on Pittsburgh's Southside, Heppenstalls and the Pennsylvania Railroad in Lawrenceville, and at the Armstrong Cork Company and the H.J. Heinz Plant. As John K. a Polish immigrant put it, "The only way you got a job [was] through somebody at work who got you in" (Bodnar, Simon, and Weber 1982: 56).

What kept ethnic networks in place so far from their origin locations? It has been argued that

"[migrants] from varying regions [in the origin country] formed a community based on common nationality and religion centered on the central cultural institution – the church" (Gjerde 1991: 176). The building of a church was one of the first organized actions in the migrant community once it arrived in an area (Barton 1975, Bodnar, Simon, and Weber 1982). Churches provided both economic and social support to their members. Information about jobs and potential land transactions flowed within the congregation and the church also served as a public arena in which members who had reneged on their obligations could be sanctioned.

Given the variety of economic opportunities in the United States, individuals and small groups drawn from the same origin community in Europe often had an incentive to move and seek employment elsewhere. The stability of the *local* community in the United States was thus essential for the viability of the labor market network. One strategy to maintain stability would have been to instill a sense of loyalty to this community. The discussion that follows will discuss how efforts to engender a sense of local identity by the church and the institutions that formed around it might have varied across Midwestern counties when the first wave of migrants arrived in the region.

### 2.3 Ethnic Fractionalization Across the Midwest

"You take in the erection department – it was mostly all Slavs ... Not Slovaks, it was Polish ... We didn't have Lithuanians there and the Russians were not involved there ... Now if a Russian got his job in a shear department ... he's looking for a buddy, a Russian buddy. He's not going to look for a Croatian buddy. And if he see the boss looking for a man he says, 'Look, I have a good man,' and he's picking out his friends." (Polish immigrant in Pittsburgh, quoted in Bodnar, Simon, and Weber 1982:62).

Numerous historians have described the efforts made by ethnic groups to establish a "toe-hold" (Thistlethwaite 1991) or a "beachhead" (Bodnar, Simon, and Weber 1982) in particular industries or establishments when they first settled in an area. The preceding quote suggests, in addition, that once a network had established a presence in the labor market, it was essential to *maintain* that presence. The discussion that follows will describe the labor market conditions under which migrant communities would have had the greatest incentive to restrict exit and, hence, instill a sense of local identity among their members. An explanation for the persistence of this identity over multiple generations, based on the complementary role of the church and other related institutions, is postponed to the next section.

Consider a market with $N$ migrant workers drawn from $M$ communities. The number of workers in

each community is exogenously determined and we do not require the labor market to be in equilibrium. The workers are competing for a fixed number of coveted permanent jobs, which provide a total surplus $R$. Tullock's (1980) canonical model of rent seeking can be conveniently adapted to this setting to describe the share of total surplus captured by community $i$

$$S_i = \frac{n_i^\alpha}{n_i^\alpha + \sum_{j=1}^{M-1} n_j^\alpha}$$

where $n_i$ is the number of workers belonging to community $i$, $n_j$ is the corresponding number of workers from each community $j \neq i$, and $\alpha > 0$.[4] The cost to community $i$ from the exit of of its members at the margin can be described by the expression

$$R\frac{dS_i}{dn_i} = R\frac{\alpha n_i^{\alpha-1} \sum_{j=1}^{M-1} n_j^\alpha}{\left(n_i^\alpha + \sum_{j=1}^{M-1} n_j^\alpha\right)^2}.$$

Assuming that all networks are of equal size, $n_i = n_j = N/M$,

$$R\frac{dS_i}{dn_i} = \frac{R\alpha}{N}\left(1 - \frac{1}{M}\right),$$

which is increasing in $M$. Holding $N$ and $R$ constant, an increase in the number of communities is associated with a greater marginal cost of exit regardless of whether the returns to network size are increasing ($\alpha > 1$) or decreasing ($\alpha < 1$) at the margin. It follows that communities in labor markets with greater ethnic competition should have invested more in instilling a sense of local identity and in building well functioning parochial institutions.

To measure ethnic competition we assume that networks were most active in the migrant population and that they were organized on the basis of the country of origin. Recall that the migrant church brought individuals from different communities in the origin country together into a large and stable network. Numerous accounts of occupational mobility in the nineteenth century indicate that although networks supported the movement of their members across establishments and even across industries, migrants could rarely change the type or skill-level of their jobs. Competition between networks is consequently assumed to occur within the broad occupational categories specified in Table 2.

Because networks vary in size in practice, we use a standard measure of fractionalization, defined as one minus the Herfindahl index of ethnic concentration, to measure competition.[5] The IPUMS

[4] Tullock's specification is identical to the equation above except that the number of workers is replaced by the investment in rent seeking.

[5] The Herfindahl index of ethnic concentration is computed as the sum of the squared share of each ethnicity in the occupational category. It is easy to verify that for the special case with networks of equal size, the fractionalization measure is equal to $1 - N^2/M$, which is increasing in the number of communities $M$, for a given number of migrant workers $N$, as above.

provides the country of birth and the occupation (where relevant) for each sampled individual. Using these data, we compute ethnic fractionalization in each occupational category in each Midwest county in 1860. The weighted average of the occupation-specific statistics, where the weight is measured by the share of migrants in the occupation, then provides us with an overall measure of ethnic fractionalization in the county. Figure 2 plots the fractionalization measure, which has a mean of 0.5 and a standard deviation of 0.2, across the seven Midwest states. Counties that were not incorporated and those without foreign-born migrants in 1860 are unshaded in the Figure. Although there is substantial variation in the fractionalization measure, which is useful for the statistical tests that follow, no spatial clustering is evident in Figure 2.

While we focus on the effect of ethnic fractionalization in strengthening within-group solidarity, fractionalization could at the same time undermine the functioning of secular institutions. Public services were provided at a rudimentary level in 1860 and the first investments in secondary schooling, for example, occurred more than 50 years later (Goldin and Katz 1999). Given the weak serial correlation in the fractionalization variable, we do not expect 1860 fractionalization to have been strongly correlated with heterogeneity in the community many decades later, when it may have mattered for the provision of local public goods. As expected, ethnic fractionalization in 1860 will be shown to be uncorrelated with ethnic (and racial) fractionalization in 1990. We will also verify that 1860 fractionalization is uncorrelated with local public good provision, including education expenditures, in 1990.

### 2.4 Historical Fractionalization and Current Economic Conditions

To test the hypothesis that local identity and its complementary parochial institutions shape career choices, fractionalization must be uncorrelated with the demand for professionals, intrinsic ability in the population, and local public good provision. We show in this section that ethnic fractionalization in 1860 is uncorrelated with measures of economic activity, racial and ethnic fractionalization, and expenditures on local public goods in 1990 once a few important characteristics of the 1860 economy are controlled for. More stringent tests ruling out a link between 1860 fractionalization and current economic conditions are discussed in Section 5.

Ethnic labor networks are no longer active in the American Midwest. Nevertheless, fractionalization in 1860 could have been correlated with particular features of the economy at the time of initial settlement that had persistent effects. To explore this possibility, we proceed to understand what

determined fractionalization in the first place. In a rapidly expanding Midwest economy, some of the variation in fractionalization across counties was no doubt a consequence of accidental initial settlement by ethnic groups in particular locations, which fueled the arrival of more migrants as networks crystallized. At the same time, fractionalization would have been determined by the demand for labor, with more ethnic groups attracted to rapidly growing areas. We have already discussed the importance of transportation links in the development of the Midwest and Table 3 consequently investigates the effect of railroads and distance to canals and a Great Lakes harbor on fractionalization in 1860.[6] Counties with a railroad running through them and counties that are closer to a canal or harbor have significantly higher fractionalization in Table 3, Column 1. Counties close to a harbor had a greater proportion of the workforce engaged in manufacturing and a smaller proportion in agriculture in 1860 in Table 3, Columns 2-3. Improved transportation, more generally, is associated with a larger population in 1860 in Column 4.

Notice that the pattern of coefficients in Table 3, Column 1 with fractionalization as the dependent variable matches perfectly with the corresponding pattern in Column 4 with county population as the dependent variable. Counties with superior transportation infrastructure, which were more populated and presumably growing more rapidly, were more fractionalized in 1860. The population of the county in 1860 could have determined subsequent agglomeration in economic activity, with long-term implications for the growth of the local economy. Not surprisingly, we see in Table 4, Columns 1-3 that while 1860 fractionalization has an insignificant effect on 1990 agriculture share and manufacturing share, it is significantly and positively associated with current county population as well as population density (not reported). More urban counties will be more racially and ethnically diverse, and it follows that 1860 fractionalization is significantly and positively correlated with racial fractionalization in 1990 in Column 5.[7]

We expect the fractionalization effect to disappear once important features of the local economy that would directly determine economic activity today, such as manufacturing share, agriculture share,

[6]Data on the distance to the nearest canal (or navigable river) and the nearest Great Lakes harbor is obtained from Jordan Rappaport's website at the Kansas City Federal Reserve Bank. The distance is computed in each case from the county centroid.

[7]The manufacturing share in 1990 is defined as the share of the civilian labor force employed in manufacturing in that year. The agriculture share in 1990 is computed using the farm population and the total population in the county in that year. All these statistics, as well as the area of each county used to compute the population density, are obtained from the 1994 County Data Book, compiled by the U.S. Bureau of the Census. Racial fractionalization is computed from the 1990 IPUMS as one minus the Herfindahl index of racial concentration, using the same five racial groups as in Alesina, Baqir, and Hoxby (2004). Ethnic fractionalization is computed from the 1990 IPUMS as one minus the Herfindahl index of racial concentration, using the same 16 white ethnic groups as in Alesina, Baqir, and Hoxby.

and particularly population in 1860, are accounted for. The fractionalization coefficient is indeed small and insignificant in Columns 7-11 of Table 4, once the county controls are included. In contrast, agricultural counties in 1860 continue to be agricultural in 1990 and more populated counties in 1860 remain larger in 1990. Notice, however, that manufacturing in 1860 does not have a persistent effect. Although the factory system began to replace artisan shops by 1820, production continued to be largely organized in workshops managed by labor contractors who hired their own employees until 1870 (Gordon, Edwards, and Reich 1982). The heavy manufacturing that characterized the Midwest economy in the twentieth century, with its emphasis on the iron and steel industry, only came at the turn of the century (Meyer 1989). Recently it has been argued that the surge of foreign immigration in the second half of the nineteenth century provided the impetus for the factory system and the subsequent industrialization of the Midwest (Kim 2007). Whatever the explanation, it is clear that the pattern of manufacturing around 1860, spread throughout the Midwest in small towns, had little connection with the heavy manufacturing, concentrated in large cities, that followed in the twentieth century. This is presumably why 1860 characteristics have such little power in predicting the share of manufacturing in the county in 1990.

For our purpose, what is important is that once we control for a few features of the nineteenth century economy, fractionalization in 1860 has no effect on characteristics of the economy today, such as the share of manufacturing and urbanization (measured by total population or by population density), that are associated with the demand for professional labor. All of the regressions that follow will consequently include these important features of the 1860 economy as controls. Notice, however, that more ethnically fractionalized counties in 1860 have significantly *lower* religious fractionalization in 1990, with and without the controls in Column 6 and Column 12.[8] This result, which stands conspicuously apart from the other regressions reported in Table 4 will be clarified in the discussion on the church and the persistence of identity that follows in Section 3.

We complete this section by describing the relationship between local public good provision in 1990 and ethnic fractionalization in 1860 in Table 5. Local government expenditures per capita in 1990 are regressed on ethnic fractionalization, manufacturing share, agriculture share, and county population in 1860 in Column 1. Retaining the same set of regressors, the dependent variable is the share of expenditure allocated to education, health, police, roads, and welfare in Columns 2-6.[9] The expenditure

[8]Religious fractionalization is computed from the 1990 Census of Religious Bodies as one minus the Herfindahl index of religious concentration, using the same 18 religious denominations as in Alesina, Baqir, and Hoxby (2004).

[9]These data are obtained from the Annual Survey of Governments, 1990. Alesina, Baqir, and Easterly (1999) use data

shares for these five public goods add up to 0.79, with the education share being as large as 0.6. The other shares range from 0.03 to 0.07. Given that the standard deviation of the fractionalization variable is 0.2, it is evident from the point estimates that 1860 fractionalization has a negligible effect on public good provision today. Using the same shares, Alesina, Baqir, and Easterly (1999) document a negative and significant relationship between contemporaneous racial fractionalization and the allocation of resources to certain public goods, particularly education, roads, and welfare. We have already shown that 1860 ethnic fractionalization is uncorrelated with 1990 racial fractionalization in Table 4, and not surprisingly we see in Table 5 that historical ethnic fractionalization has no effect on current local public good provision, including education expenditures.

# 3 The Church and the Persistence of Identity

In a recent paper, Akerlof and Kranton (2005) describe how firms and other economic organizations can instill a sense of identity or loyalty among their workers to solve agency problems. In our setting, the natural organization to instill loyalty in the community would have been the local church. The church was among the first institutions to be established when immigrants arrived in an area (Hoerder 1991, Barton 1975, Bodnar, Simon, and Weber 1982). The church congregation provided many forms of mutual assistance including credit, insurance, job referrals, business information, and social support (Gjerde 1985, 1997, Alexander 1991). Indeed, it has been argued that immigrants participated in church communities to benefit from these economic and social services, instead of being drawn to the church by a particular belief or ideology (Bodnar 1985).

Despite these material attractions, exit from the church and the local labor market was always a threat in nineteenth century America, especially in areas that had been recently settled. There was a tension between a "folk society" centered around the church parish and a competitive "individualistic society" with its many opportunities (Ostergren 1976). In addition, migration to the United States, and subsequent internal migration, was often organized under the auspices of relatively small communities from the origin country. These migrant networks were too small to maintain a viable church congregation and labor network when they first arrived in an area and so churches in the Midwest typically brought together many regional groups with the same national origin. Inter-regional conflict was common, and there was always the possibility that groups within the church would move again to take advantage of new opportunities that became available elsewhere (Gjerde 1997). Under these

drawn from the same source.

circumstances, the church had to make a special (costly) effort to instill a sense of loyalty to the local community among its members that transcended individual, kin, and regional affiliations.

If local identity played a role in reducing exit in the nineteenth century, how and why did it persist long after the ethnic labor networks it supported ceased to be salient? Our explanation for this persistence is based on the observation that churches continue to provide important forms of social support to their members. Church activities include Sunday school service, youth groups, pot-lucks, informal home parties, and food, visits, and other forms of support when members of the congregation are ailing or infirm. The church also lies at the center of a cluster of inter-linked civic institutions, including the school and various voluntary organizations. Life in a Midwest community revolves around these institutions, which bring families and multiple generations together on a regular basis (Elder and Conger 2000). We focus on the church in the analysis that follows with the understanding that individuals participate in and benefit from a wider range of related institutions; well functioning churches will in general be associated with tight-knit well functioning communities.[10]

There are complementarities associated with church inputs; if the rest of the congregation commits time and effort to the church, then the returns to the individual's inputs would increase as well (Iannaccone 1998). The presence of these complementarities introduces two problems. First, self-interested individuals will devote a level of inputs that is sub-optimal because they do not internalize the benefits that the rest of the congregation derives from their actions. Second, a coordination problem can arise since individuals will only make career choices that are associated with a high level of commitment to the church if they expect other members of their generation, who are simultaneously making these choices, to do likewise. A strong sense of local identity reduces each of these problems, serving as a commitment device and increasing the level of church inputs in equilibrium. Given that church inputs are complements, it is easy to verify that the older generation in a church with a committed congregation will be more willing to bear the cost of instilling a strong sense of local identity in the generation that follows, resulting in a higher level of inputs. Local identity and church inputs can reinforce each other in this way over many overlapping generations.[11]

[10]The number of civic and social associations in the county and the corresponding number of religious organizations can be obtained from the County Business Patterns, 1990. The number of not-for-profit organizations in the county can be obtained from National Center for Charitable Statistics Core Files, 1990. Controlling for manufacturing share, agriculture share, and county population in 1860, ethnic fractionalization in that year has no effect on any of these variables. Note, however, that these statistics are based on the *number* of organizations, whereas what we require are measures of *participation*. For example, we will see below that 1860 fractionalization has a significant and positive effect on religious participation in 1990 despite the fact that it is uncorrelated with the number of religious organizations.

[11]The persistence in local identity that we describe is related to a paper by Bénabou and Tirole (2006) linking invest-

Local identity and church inputs cannot be observed directly. However, we do observe the number of church participants across counties and over time. Assuming heterogeneity in the payoff from non-participation in the population, it follows that well functioning churches will have larger congregations. The testable prediction from the preceding discussion is that high fractionalization counties should begin with stronger local identity, a higher level of church inputs and, hence, greater church participation at the time of settlement. More importantly, the effect of fractionalization on church participation should grow over time, reflecting the underlying divergence in church performance that is necessary to sustain differences in local identity. This prediction distinguished our model, with identity as a commitment device, from the model of religious cults proposed by Iannaccone (1992) in which self-sacrifice and strict norms of behavior are used to screen out free-riders and ensure higher levels of participation among those that remain. While this alternative mechanism may induce high levels of participation at the intensive margin, it results in small congregations and is typically short-lived.

Up to this point we have not distinguished between different religious denominations. Based on the country of origin of the incoming migrants, reported in Table 1, most of the migrant churches would have been Lutheran or Catholic. Regressing the population-share of different denominations, computed with data from the Census of Religious Bodies in 1860, on the share of migrants in that year, counties with a greater share of migrants are indeed disproportionately Lutheran and Catholic. A strong test of our framework, which relies on the role of migrant churches in shaping and maintaining local identity, is that the predictions derived above should apply to those migrant denominations only. In particular, higher fractionalization counties should be associated with a greater share of Lutherans and Catholics in the population around the time of initial settlement and this effect of fractionalization should grow stronger over time. In contrast, the model has no prediction for the relationship between fractionalization and participation in other denominations, providing us with a useful falsification test.[12]

ments in human capital at the household level with the political equilibrium at the macro level. In their model, children choose a level of effort (schooling) based on their belief about the returns to this effort. If a sufficiently large number of households invest in effort, they will form a pivotal voting block and set a low tax rate, generating, in turn, a high return to effort that reinforces the initial beliefs. Thus, two political equilibria can arise; a low-effort equilibrium with substantial redistribution (high taxes) and a high-effort equilibrium with little redistribution (low taxes). Bénabou and Tirole fix the equilibrium by assuming that children have imperfect willpower and so will underinvest in effort if left to themselves. Their parents, who provide them with information about the returns to effort, will consequently systematically inflate these returns. This "ideological" position leads parents in the high-effort equilibrium to (optimally) ignore negative signals about the returns to effort, allowing particular equilibria and the collective beliefs that support them to persist over many generations. In this paper we use local identity rather than ideology to fix the level of church inputs and sustain a social equilibrium over many generations.

[12]The implicit assumption underlying these predictions is that the mapping from church inputs to participation should

The relationship between ethnic fractionalization and religious participation can be tested with data from the Census of Religious Bodies (CRB), which has been conducted at roughly ten-year intervals from 1860 to 2000. This census was conducted as part of the population census from 1860 to 1890, with census enumerators collecting information from individual churches in each county. Subsequently, the U.S. Bureau of the Census conducted the CRB separately from the population census in 1906, 1916, 1926, and 1936. Starting from 1952, the National Council of Churches of Christ undertook the responsibility of conducting the CRB, with subsequent census rounds in 1972, 1980, 1990 and 2000.

The 1860-1890 census rounds collected information on the number of church seats by denomination in each county. From 1890 onwards, information was collected on the number of members directly, and from 1972 onwards the number of adherents was collected as well. Despite these changes in the management of the CRB and the measure of religious participation, we uncover clear changes in the mix of denominations as well as the effect of 1860 fractionalization on religious participation over time.

Table 6 reports changes in the mix of denominations and overall participation rates in our Midwestern counties over the 1860-2000 period. To take account of the fact that the measure of participation was changing over time, we report statistics in the first and the last census-year that each measure was used. Thus, participation is measured by the number of church seats from 1860 to 1890, by the number of members from 1890 to 1952, and by the number of adherents from 1972 to 2000.[13] The participation rate is then computed as the number of participants divided by the contemporaneous population in the county. Five denominations – Baptist, Catholic, Lutheran, Methodist, and Presbyterian – account for roughly 80 percent of church participants over the 1860-2000 period. Among these denominations, the Lutherans grow rapidly in popularity over the 1860-1890 period and the 1890-1952 period, remaining stable thereafter. In contrast, the Methodists and the Presbyterians decline steadily over time. There is no clear trend among the Baptists and the Catholics.

The inability of the Baptist church to increase its share of church participants contrasts with the surge in popularity of this denomination elsewhere in the United States, as documented by Finke and Stark (1992). The Midwest stands apart from the rest of the country in that the "traditional"

not vary with fractionalization. Fractionalization in 1860 was correlated with characteristics of the local economy at that time and, hence, with the characteristics of the incoming migrants, potentially violating this assumption. Given the dynamic nature of the economy at the time of initial settlement, however, 1860 fractionalization would soon have been uncorrelated with population characteristics and we expect the predictions for participation to hold up within a few decades of 1860.

[13] Although the number of members was also collected in the 1972-2000 census rounds, this statistic is not available for Catholics, a major denomination in our Midwestern counties, in these rounds.

denominations, particularly the Catholics and the Lutherans, continue to dominate and we will connect this observation to the link between migrants churches, local identity, and religious participation below. Even without the Baptist surge, religious participation increased steadily over time in the Midwest, rising over the 1860-1890 and 1890-1952 periods. Based on the statistics in the most recent 1972-2000 period, 55-60 percent of the population in our Midwestern counties are church adherents.

We now proceed to estimate the relationship between ethnic fractionalization in 1860 and religious participation in the county in each round of the CRB. The religious participation regression includes the same set of county-level controls as the regressions in Table 4, Columns 7-12 and Table 5. Because the regression is estimated over many census years, we simply report the 1860 fractionalization coefficient, together with the 95 percent confidence band, in each census year in Figure 3. This coefficient is less precisely estimated in the early census years, but grows steadily larger, while remaining statistically significant, all the way through to 2000.[14]

A stronger test of our framework is that its predictions should only apply to participation in migrant denominations. We consequently proceed to estimate two separate regressions in each census year; the first regression has the share of Lutherans and Catholics in the population as the dependent variable and the second regression has the share of all other denominations as the dependent variable. The 1860 fractionalization coefficient, with the corresponding 95 percent confidence band, is reported for each regression in each census year in Figure 4. We now see that the share of Lutherans and Catholics in the population is significantly greater in high fractionalization counties by 1870 and that the fractionalization effect gets steadily larger over time, as predicted. Although we do not report results separately by denomination, this pattern is obtained for both the Catholics and the Lutherans. A one standard deviation decline in 1860 fractionalization would increase the population share of Lutherans and Catholics in the county by four percentage points (22 percent) in 2000. In contrast, the fractionalization coefficient is *negative* and significant, and very stable over time, in the companion regression. These results, taken together, provide a simple explanation for the negative correlation

[14]The 1860 fractionalization coefficient with standard error in parentheses is -0.66(0.39). This outlying coefficient is omitted from Figure 3 and Figure 4 that follows to clarify changes in the fractionalization coefficient over time. As noted, church seats and the number of members are both available in the 1890 CRB. We use the first statistic to measure church participation in Figure 3 and Figure 4 because it is more in line with trends in the fractionalization coefficient over time. Although not reported, the coefficient on 1860 agriculture share is also positive and significant in the religious participation regressions and grows larger over time. Unlike ethnic fractionalization, which soon ceased to be directly relevant, recall from Table 4 that agricultural counties in 1860 remained disproportionately agricultural in 1990. It is well known that farming communities tend to be more religious and community-oriented and so the persistent effect of the 1860 agriculture share may simply reflect the persistence of agricultural activity in particular areas over time.

between ethnic fractionalization in 1860 and religious fractionalization in 1990 that we reported in Table 4.

Our explanation for the pattern in Figure 4 for the migrant denominations is that the gap in church inputs between low and high fractionalization counties is widening over time, giving rise to an accompanying increase in the participation-gap. An alternative explanation for this pattern is that there are increasing returns to scale in church performance; Lutheran and Catholic churches were large to begin with in the high fractionalization counties and grew larger over time. The time-trend in the fractionalization coefficient for the migrant and non-migrant denominations in Figure 4 would appear to rule out this explanation; we should otherwise have observed a (symmetric) declining trend for the non-migrant denominations. Nevertheless, we experimented with augmented specifications that included (i) the share of Lutherans and Catholics, and (ii) a full set of 32 ethnic shares (which sum up to the share of migrants in the population), in 1860 by county, as regressors. The initial share of Lutherans and Catholics, as well as some of the ethnic shares do have persistent effects. However, the fractionalization coefficient remains stable over the course of the twentieth century with both the augmented specifications. The strong and persistent relationship between 1860 fractionalization and church participation contrasts with the absence of any effect of this variable on current economic conditions in Table 4 and Table 5, once particular features of the local economy are controlled for.[15] We cannot observe local identity, but to the extent that well functioning churches require committed congregations, this result sets the stage for the analysis of occupational choice with local identity that follows.

# 4 Occupational Choice with Local Identity

If local identity shapes occupational choice and fractionalization is uncorrelated with the demand for professional labor, then individuals born in high fractionalization counties should be less likely to choose professional careers but individuals subsequently residing in those countries should be as likely to be professionals once the market clears. The model developed in this section and described in greater detail in the Appendix derives this joint hypothesis more formally. It also explains why individuals born in high fractionalization counties would shade their choices towards less mobile non-professional occupations, giving rise to a surplus of non-professional labor in those counties on average, even though

[15]The pattern of coefficients reported in Figure 4 would also be obtained if the county controls were omitted from the participation regression, although the point estimates would actually be somewhat smaller.

the labor market must ultimately clear. The key insight is that this result will be obtained as long as there is *ex ante* uncertainty in the demand for different types of jobs across locations.

The model is set up so that two types of jobs are available: professional and non-professional. Individuals who are *ex ante* identical live for two periods, working in the second period of their lives. Those individuals who choose to occupy more productive professional jobs must invest in training in the first period of their lives. Individuals who expect to end up in non-professional jobs incur no such cost. There are two locations in this economy with $N$ individuals born in each location in each period. On average, $sN$ professional jobs and $(1-s)N$ non-professional jobs are demanded in each location in each period. However, these locations also face positive and negative demand shocks with equal probability, which separately shift the demand for professional and non-professional labor by $\epsilon sN$ and $\epsilon(1-s)N$ respectively, but leave the total number of professional jobs $2sN$ and non-professional jobs $2(1-s)N$ constant across the two locations in all periods.[16]

Once demand shocks are introduced, labor must flow across locations at the beginning of each period to clear the market. In addition, we assume that professional jobs have an individual-specific component to them. The opportunity to enhance overall productivity by switching a professional working in a particular location with a professional working in the other location arrives with probability $P$ in each period. This additional dimension of mobility in professional occupations is based on the idea that the labor market for a school teacher or a secretary tends to be local. In contrast, professionals such as university professors or management consultants are continually re-sorting across local and regional labor markets as new opportunities arise and fresh cohorts of workers enter. A professional can advance her career considerably with such a move if it becomes available and we will later see that professionals are indeed much more likely to migrate from their county of birth.

Individuals dislike moving, particularly those with a strong local identity. Let the cost of moving be $C_1$ for individuals born in location 1 and $C_2 < C_1$ for individuals born in location 2 with weaker local identity. This is the only difference between workers in the two locations. We assume that all workers are employed and that professional workers always take advantage of the productivity enhancing career opportunities when they arise, in which case total output in this economy remains constant across all states of the world. Although wages for professionals and non-professionals will adjust across the two locations to clear the labor market in practice, the competitive labor allocation can in that case

[16] Within a location, demand shocks for professionals and non-professionals could be positively or negatively correlated. We could also allow the size of the shocks to vary, $\bar{\epsilon}$ for the professionals and $\underline{\epsilon}$ for the non-professionals, without changing the results reported below.

be conveniently derived as the solution to the Central Planner's problem of minimizing training and moving costs across both locations.

Let the supply of professional labor in location 1 be $x_1$. We show in the Appendix that expected cost is a piece-wise linear function of $x_1$, as described in Figure 5, which is minimized at $x_1^* = sN - \epsilon sN$ for $P < 1/2(C_1 + C_2)/(C_1 - C_2)$ and at $x_1^{**} = sN - \epsilon sN - \epsilon N(1-2s)$ for $P > 1/2(C_1 + C_2)/(C_1 - C_2)$. The equilibrium supply of professional labor in location 1, with stronger local identity, falls short of the expected demand $sN$ for all values of $P$. Moreover, the share of professionals supplied by location 1 is strictly less than the corresponding share in location 2.[17]

The intuition for this result is that as long as there is *ex ante* uncertainty in the type of jobs demanded, some individuals will have to move in both locations with positive probability to clear the market. By allocating a surplus of non-professionals to location 1, the Central Planner increases the probability of such moves, but trades this off against the lower probability of having individuals born in location 1 move *ex post* on the job.[18] We show in the Appendix that the supply of professionals in location 1 could also fall short of the expected demand without uncertainty, but only if $P > (C_1 + C_2)/(C_1 - C_2)$.

In our set up, individuals *born* in location 1 are less likely to be professionals because they incur a greater cost when they move ($C_1 > C_2$). However, this result could also be obtained if $C_1 = C_2$ and the demand for professional labor is lower in location 1, $s_1 N < s_2 N$. The second prediction of the model, which allows us to rule out this alternative demand-side explanation, is that once the labor market has cleared, the share of professionals *residing* in the two locations should on average be the same ($sN$). In contrast, if the demand for professional labor is lower in location 1, then individuals born and subsequently residing in that location should be less likely to be professionals.

# 5 Empirical Analysis

## 5.1 Individual Data

To test the predictions from the previous section we need information on the individual's career choice, county of birth, and county of residence (post-employment). The National Longitudinal Survey of

[17] For $P < 1/2(C_1 + C_2)/(C_1 - C_2)$, the share of professionals is $s - \epsilon s$ in location 1 and $s + \epsilon s$ in location 2. For $P > 1/2(C_1 + C_2)/(C_1 - C_2)$, the corresponding shares are $s - \epsilon(1-s)$ and $s + \epsilon(1-s)$.

[18] The Central Planner could make adjustments on other dimensions as well, by allowing for some level of unemployment or by letting professionals born in location 1 forego career opportunities with finite probability. The *ex ante* career choices that we are interested in would nevertheless continue to vary systematically across locations as described.

Youth 1979 (NLSY79) is the only large-scale data set that we are aware of that includes this information. The NLSY consists of a nationally representative sample of American high school seniors in 1979 who were subsequently interviewed annually from 1979 to 1994 and biennially thereafter. The survey collects basic information on the respondent's age, gender, race and, most importantly, county of birth. The Armed Forces Qualification Test (AFQT), which is designed to provide an unbiased measure of the individual's intelligence, was administered to all respondents in 1979. Subsequent survey rounds collected contemporaneous information on educational attainment, employment, occupation, income, and county of residence. We will study occupational choice and other outcomes related to that economic decision at two points in time – 1994 and 2000 – when the respondents were old enough to be settled in their careers and to have made some job-related moves. Occupational choices from the NLSY in these years will be matched to census data on historical fractionalization, both in the individual's county of birth and the contemporaneous county of residence, to test the predictions of the model.

Table 7 reports descriptive statistics for the individuals in our sample, who were on average 18 years old in 1979 and so around 33 years old in 1994 and 39 years old in 2000. Occupational categories in the NLSY (up to the 2000 round) are based on the 1970 codes from the census. Professional occupations are defined to include relevant codes listed under the Professional, Technical, and Kindred Workers category [1-196]. Job-related geographical mobility is the chief property that distinguishes professional and non-professional occupations in our model. We consequently exclude technical occupations and other occupations where career moves are unlikely to be important from this category.[19]

Based on this occupational classification, 9 percent of the respondents hold professional jobs, with little change from 1994 to 2000. 56 percent of all respondents had migrated out of their birth-county by 1994, with an increase to 59 percent by 2000. Consistent with the assumption that professional occupations are associated with greater mobility, 75 percent of the professionals and 53 percent of the non-professionals had migrated out of their birth-county by 1994 (these differences are significant at the 5 percent level and similar in 1994 and 2000).

Individuals in the sample are 33 years old by 1994 and should be established in the labor market. Nevertheless, employment levels continue to increase over time, from 81 percent in 1994 to 92 percent in 2000. Conditional on being employed, annual income (in 2000 dollars) also increases from 28 thousand in 1994 to 33 thousand in 2000. These changes in employment and income are presumably life-cycle

[19]These occupations include Nurses, dieticians, and therapists [74-76], Religious workers [86,90], Social and recreation workers [100,101], Teachers, except college and university [141-145], Technicians [150-174], Writers, artists, and entertainers [175-196].

effects, but they could, in principle, be due to selective attrition since this is a longitudinal survey. Notice, however, that racial composition and the proportion of women in the sample are very stable over the 1994-2000 period. Thus, we do not observe selective attrition from the sample, at least with respect to two important demographic characteristics that are associated with income and employment.

### 5.2 Fractionalization in the County of Birth and Occupational Choice

The first prediction from the model of occupational choice incorporating local identity is that individuals born in counties with greater ethnic fractionalization in 1860 should be less likely to hold professional jobs. Including race, gender, and age as regressors (although their omission would not affect the results) we see in Table 8, Columns 1-2 that individuals born in high fractionalization counties are indeed less likely to hold professional jobs in the 1994 and 2000 rounds of the NLSY. The individuals in our sample are drawn from 150 of the approximately 400 Midwestern counties that were incorporated and had attracted foreign migrants by 1860. While it thus seems unlikely that a few outlying counties are driving the results, we nevertheless report nonparametric estimates of the relationship between occupational choice and historical fractionalization in Figure 6. We see that the probability that the individual is a professional declines steadily with fractionalization, both in 1994 and in 2000, verifying the robustness of our results.[20]

Our test of the hypothesis that identity shapes occupational choice relies on the assumption that fractionalization is uncorrelated with the demand for professional workers in the county. Recall from Table 4, however, that historical fractionalization had a positive and significant effect on current economic characteristics, measured by population in the county as well as racial and ethnic fractionalization, but that this effect disappeared once important characteristics of the 1860 economy were included as regressors. More urban counties will have a greater demand for professional workers, shifting *up* the coefficient on the fractionalization variable in Table 8, Columns 1-2. The regressions that follow will thus include the same 1860 characteristics as in Tables 4 and 5 – population, manufacturing share and agriculture share – to control for variation in the *demand* for professional workers across counties.

The prediction that high fractionalization counties should supply fewer professional workers is derived conditional on the total *surplus* that was historically available to migrants in the labor market

[20]The nonparametric regressions are estimated using the Epanechnikov kernel smoothing function. Less than 5 percent of the observations in Table 8, Columns 1-2 are drawn from counties with fractionalization below 0.45 and so too much weight should not be placed on the extremely steep initial decline in Figure 6. Although all the parametric regressions that follow will use the full set of counties, we verified that the fractionalization coefficient is unchanged when the sample is restricted to individuals drawn from counties in the 0.45-0.8 range.

$R$ and the total number of migrant workers $N$. Everything else equal, an increase in $R$ would increase the cost of exit at the margin, encouraging efforts to instill local identity historically and lowering the propensity to enter professional occupations today. Superior transportation infrastructure in 1860 was associated with high fractionalization and a larger county population in Table 3 and so we expect fractionalization to be positively correlated with $R$. It follows that exclusion of $R$ from the occupation choice regression in Table 8, Columns 1-2 would have shifted the coefficient on fractionalization *down*. The same 1860 characteristics that we described above could be used to control for $R$, but since the professional demand effect and the surplus effect work in opposite directions, the effect of their inclusion on the fractionalization coefficient would now be ambiguous. By the same argument, we can no longer predict the sign of the coefficients on the 1860 characteristics when they are included in the occupation choice regression. For example, larger 1860 population is associated with urbanization today and, hence, an increased demand for professional jobs. At the same time, a larger 1860 population is associated with a larger $R$, stronger identity and, hence, a reduced supply of professional workers. Finally, we would want to account for the number of migrants $N$ when estimating the effect of fractionalization on occupational choice. This variable is highly correlated with county population (the correlation is 0.9) and so will be omitted from the regressions that follow. 1860 population captures the professional demand effect, the surplus effect, and the effect of $N$ on occupational choice.

We see in Table 8, Columns 3-4 that the coefficient on 1860 fractionalization becomes more negative and is more precisely estimated once the additional county-level controls are included in the occupational choice regressions. To allow for the possibility that individuals with particular ethnic ancestry continue to be concentrated in specific occupations today, we include the population share of each ethnicity in the 1860 census, computed at the level of the county, in Columns 5-6. Given the high rate of inter-county migration, we do not expect the ethnic shares in 1860 to be relevant today and not surprisingly the fractionalization coefficient is hardly affected by the inclusion of these additional regressors. Based on the estimates in Column 4, a one standard deviation decline in ethnic fractionalization would increase the probability of holding a professional job from 9 percent to as much as 15 percent. Individuals born in counties with a greater share of manufacturing and agricultural jobs in 1860 are also less likely to hold professional jobs. Population in 1860, in contrast, has no effect on occupational choice in both the 1994 and the 2000 rounds. Finally, women and non-whites are significantly less likely to hold professional jobs in Columns 1-6.

We complete the description of the occupational choice regressions by discussing a number of

robustness tests (not reported). First, we computed ethnic fractionalization in 1860 with men only. Both men and women participated in the workforce in the nineteenth and early twentieth centuries, with ethnic networks channeling women into jobs as well. Bodnar (1980), for example, cites a 1930 study of two thousand foreign-born women, most of whom reported that they had secured their first job through social connections. Nevertheless, we might expect labor networks to have been organized along gender lines within ethnic groups, with male networks occupying a dominant position in the labor market and in the communities they were drawn from. The coefficient on the alternative fractionalization measure (computed using men alone) continues to be negative and significant, both in 1994 and 2000, although it is smaller in size.

Inspection of the occupation categories in Table 2 indicates that farmers made up 50 percent of the workforce in 1860. Ethnic competition may have been less relevant in this category and so we recomputed the county-level fractionalization statistic placing zero weight on fractionalization among the farmers as a second robustness test.[21] The fractionalization coefficient becomes slightly larger in absolute magnitude and is more precisely estimated, both with and without the 1860 controls. In a related exercise, we explored the possibility that ethnic competition occurred outside the labor market entirely by computing fractionalization without regard to occupational category or occupational status. This alternative measure of fractionalization, in contrast, had an insignificant effect on occupational choice.

As a third test, we separately replaced 1860 characteristics with 1850 and 1870 characteristics to assess the stability of the results in other census years around the time of initial settlement. The coefficient on 1850 fractionalization is negative and significant, albeit smaller in magnitude. The coefficient on 1870 fractionalization is also negative, but not significant at conventional levels. Although we focus on economic characteristics at the time of initial settlement in this paper, we do not rule out the possibility that relevant measures of population heterogeneity and other economic characteristics later on had persistent effects as well. All that we require is that 1860 fractionalization should have a direct and independent effect on identity and church inputs today. The clean trend in the church

[21] We include all individuals who report being employed, regardless of their age, when computing the fractionalization statistic since there were no age restrictions on employment at that time. Apart from the 11 broad occupational categories in Table 2, some individuals in the census were also assigned to an undefined occupational category. Women were disproportionately represented in this category, which presumably covers home production and other informal activities. As with the farmers, we assume that networks were less relevant in this category and assign it zero weight when computing ethnic fractionalization in all the regressions that we report. Nevertheless, we verified that assigning a weight to this category based on its share of the migrant workforce in the county had no effect on the estimated fractionalization coefficient in the occupational choice regression.

participation regressions reported in Figure 4, starting from 1870 and stretching uninterrupted all the way through to 2000, provides strong support for this claim.

Fourth, we replaced total population with the migrant population and, separately, with the workforce, in the occupational choice regression. These variables are all highly correlated, and not surprisingly the results were unchanged. Fifth, we included the share of migrants in the workforce, and the interaction of this variable with ethnic fractionalization, as additional regressors. This specification allows for the possibility that ethnic fractionalization had a larger effect in counties where migrants made up a larger share of the total population. However, both these variables had an insignificant effect on occupational choice, leaving the (uninteracted) fractionalization coefficient unchanged.

Sixth, we expanded the set of professional occupations by including individuals assigned to the "managers and administrators" code [245] within the broader Managers and Administrators, except Farm category [201-245]. Most of the specific occupations listed in this broad category do not conform to our definition of a professional occupation and the "managers and administrators" classification does not provide much information on the actual nature of jobs that it covers. However, 74.7 percent of the individuals assigned to that category migrated out of the county of their birth, which is close to the level of migration for the other occupations that we include in the professional category. Inclusion of the managers and administrators in the professional classification increased the share of professionals in the sample to 17.7 percent. Although these results need to to interpreted with caution since the additional professional occupations are selected on the basis of an outcome (migration) rather than their fundamental characteristics, we nevertheless verified that the fractionalization coefficient in the occupational choice regression with this expanded definition of professional activity remained negative and significant. Finally, we ran the occupational choice regression in other years – 1993, 1996, 1998 – generating a pattern of fractionalization coefficients very similar to the point estimates obtained in 1994 and 2000.

### 5.3 Fractionalization and Outcomes Related to Occupational Choice

An individual born in a county with higher ethnic fractionalization in 1860 is less likely to select into a professional occupation in 1994 and 2000. Our interpretation of this result is that individuals born in high fractionalization counties identify strongly with their local communities and so wish to avoid the spatial mobility that comes with professional occupations. The regression results reported in Table 9, Columns 1-2 indicate that individuals from high fractionalization counties are indeed significantly less

likely to migrate from the county of their birth. On average, around 58 percent of the individuals in the sample migrate from the county of birth. The point estimates indicate that a one standard deviation increase in fractionalization reduces migration by 8 percentage points (a 14 percent decline). Among the other regressors, none of the 1860 county-level controls significantly affect migration, but Whites are significantly more likely to move. Age is also (mechanically) positively associated with migration.

Does the effect of fractionalization on occupational choice and migration that we have just described have economic consequences? The results in Table 9, Columns 3-4 indicate that while individuals born in counties with a greater share of manufacturing and agriculture in 1860 are significantly more likely to hold a job by 2000, employment levels do not vary significantly with ethnic fractionalization in 1994 or 2000.[22] Whites and males are, not surprisingly, significantly more likely to be employed, although the importance of these individual characteristics declines over time.

In contrast with the results for employment, the income regressions reported next in Table 9, Columns 5-6 indicate that high fractionalization is associated with significantly lower income (in 2000), conditional on being employed. Average annual income in 2000 was 33,000 dollars, and so our estimates indicate that a one standard deviation increase in fractionalization would have reduced income by 2,300 dollars (a 7 percent decrease).[23] A greater share of manufacturing and agriculture in 1860 is associated with lower income in 2000, with Whites and males earning significantly more in 1994 and 2000.

### 5.4 Fractionalization in the County of Residence and Occupational Choice

The second prediction of the model is that historical fractionalization in the county of residence should have no effect on occupational choice. This implies that non-professional workers must flow out of the high fractionalization counties, while professional workers flow in to clear the labor market. This prediction rules out the possibility that differences in the demand for professional labor across counties are driving our results.

To test the prediction that non-professional workers flow out of the high fractionalization counties we divide the sample into individuals who stay and individuals who out-migrate from their county

[22] Although we focus on *ex ante* career choices, we noted in Section 3 that the Central Planner could also reduce movement costs by allowing for some unemployment or by reducing career-related moves for professionals born in counties with strong local identity. We see no evidence of such adjustments in employment. Restricting the sample to professionals, we do find that individuals born in high fractionalization counties are less likely to migrate, but once again the fractionalization coefficient is insignificant at conventional levels.

[23] This result rules out an alternative explanation for the excess supply of non-professionals in high fractionalization counties based on the idea that labor networks in those counties generate disproportionately high wages in non-professional occupations. As an additional check, we restricted the sample to individuals in non-professional occupations and verified that individuals born in high fractionalization counties did not earn significantly higher incomes.

of birth (by 1994 and 2000) in Table 10, Columns 1-4. Matching the results in Table 8 with the full sample, fractionalization in the county of birth lowers the probability that the individual will be employed in a professional occupation for both stayers and out-migrants in Table 10, Columns 1-4. While the coefficient for the two groups is comparable in 1994, it is substantially larger – roughly twice as large – for the out-migrants in 2000. Some of the excess supply of non-professional labor in the high fractionalization counties is evidently reduced through the selective out-migration of workers from those counties. To test the accompanying prediction that professional workers will flow into the high fractionalization counties, we restrict the sample to individuals who move into our Midwest counties from anywhere in the United States. Regressing occupational choice on the county of residence of the in-migrants in Table 10, Columns 5-6 we see that although the fractionalization coefficient is insignificant in 1994, it is positive and significant in 2000 as expected.

The labor flows that we have described all work towards reducing the mismatch between the supply and the demand for professional and non-professional labor across counties. Measuring historical fractionalization in the county of residence in Table 10, Columns 7-8 we see that the fractionalization coefficient is no longer significant and actually switches signs in 2000. It is only individuals born in high fractionalization counties who are less likely to be professionals. Once the market has cleared, individuals residing in those counties are just as likely to hold professional jobs, as predicted by the model.[24]

One remaining concern is that the mismatch between the supply and the demand for professional labor in the high fractionalization counties could have been driven by unexpected changes in the demand for professionals in those counties rather than a strong local identity. Suppose, for example, that the demand for professionals was systematically lower in high fractionalization counties historically and then increased in the 1980's, with the restructuring of the U.S. economy, after our NLSY cohort had made their career choices. To rule out this possibility, we regressed manufacturing share, agriculture share, and county population in 1970, obtained from the 1972 County Data Book, on the same 1860 characteristics as in Table 4, Columns 7-9. 1860 fractionalization has no effect on economic conditions in 1970, just as we saw earlier with the same conditions in 1990.

[24]The second prediction of the model that fractionalization in the county of residence should be uncorrelated with occupational choice is based on the assumption that the entire labor market can clear. The test that we implement, instead, is based on a restricted NLSY cohort. Note, however, that labor supply and demand should match in each cohort in steady-state equilibrium.

### 5.5 Fractionalization and Individual Ability

Although individuals are assumed to be *ex ante* identical in the model of occupational choice, we could easily allow for heterogeneity in ability within each location. If the cost of training is declining in ability, then individuals above an ability-threshold would select into the professional occupation, with a higher threshold in location 1 with stronger local identity. Suppose, instead, that identity is irrelevant and the ability distribution in location 2 dominates the corresponding distribution in location 1. The number of professionals supplied by location 2 would still exceed the number of professionals supplied by location 1, and this alternative explanation would match all the results in Tables 8-10.

We have already ruled out variation in the demand for professional jobs as an alternative explanation, but could variation in individual ability across counties as described above have generated these results instead? Variation in ability could arise due to differences in innate talent or the quality of the school system. High fractionalization counties had superior transportation infrastructure in 1860 and so were most likely growing relatively rapidly at that time. It is entirely possible that particular types of migrants were attracted to those counties, although we would expect migrants drawn to areas with many competing ethnic groups to have been positively selected on innate talent. Given the high rates of inter-county migration, even in the Midwest, we would in any case expect few families to have maintained an unbroken line of descent to the present day in the same county. Moreover, we noted in Table 5 that 1860 fractionalization was uncorrelated with education expenditures in 1990. Given the weak serial correlation in our fractionalization measure, it is unlikely that it was correlated with relevant measures of heterogeneity early in the twentieth century when important public investments were being made. Consistent with this view, we see in Table 11, Columns 1-3 that fractionalization, and all the 1860 variables for that matter, have no effect on AFQT scores, high school completion, and college completion.[25] In contrast, individual characteristics have a strong effect on AFQT scores and college completion.

Even if innate talent and the quality of the school system do not vary systematically with fractionalization, we might still expect individuals from high fractionalization counties to invest less in the college education that is necessary to obtain a professional job. One reason why we do not see this effect in Column 3 may be that there is consumption value to higher education. Ideally, we would

[25]The fact that fractionalization is uncorrelated with education also shuts down an alternative channel through which fractionalization could affect religious participation (see Sacerdote and Glaeser [2001] for an analysis of the complex relationship between education and religious participation).

want to estimate the effect of fractionalization on the choice of college major, but this information is unavailable in the NLSY. Nevertheless, notice that individuals born in high fractionalization counties are less likely to have attended college out of state (conditional on having attended) in Column 4, although this effect misses being significant at conventional levels. This result may reflect a preference for being close to home or a recognition that the quality of the education these individuals receive is less relevant for their future careers. Collecting the results with education expenditures in Table 5 and test scores and educational attainment in Table 11, there is no indication that individuals born in high fractionalization counties are less capable or less prepared to invest in the human capital that is necessary to secure professional jobs.

# 6 Conclusion

This paper draws a connection between ethnic labor networks in the American Midwest when it was first being settled, the local identity that emerged endogenously to support these networks and then persisted over many generations, and occupational choice today. Individuals born in counties with greater ethnic fractionalization in 1860, which we expect to be associated with stronger local identity and better functioning parochial institutions, are significantly less likely to hold professional jobs, which come with greater geographical mobility, in 2000.

The results in this paper are relevant to the ongoing debate on "economic institutions" (Acemoglu, Johnson and Robinson [2001, 2002]) versus "culture" (Barro and McCleary 2003, Tabellini 2005, Fernandez and Fogli 2007) as determinants of growth. We find that local cultural traits generate significant variation in occupational choice across what appears to be a relatively homogenous region – the American Midwest – and we conjecture that these cultural effects might be even larger across countries with very different histories. At the same time, culture cannot be sustained without institutional support. Social institutions such as the church and the family help sustain cultural traits, which in turn keep these institutions alive. As long as these institutions continue to be useful, cultural traits can persist long after the economic circumstances that gave rise to them have ceased to be relevant. The economics literature has focussed much of its attention on economic and political institutions. Our results suggest that social institutions, with their complementary cultural traits, might have important effects on growth as well.

# 7 Appendix: Labor Market Equilibrium

## 7.1 Population and Production Technology

Two types of jobs are available in this economy: professional and non-professional. Individuals who are *ex ante* identical live for two periods, working in the second period of their lives. Those individuals who choose to occupy professional jobs incur a training cost $C_e$ in the first period of their lives. Individuals who expect to end up in non-professional jobs incur no such cost. The expected output obtained from a professional worker who takes advantage of career opportunities when they arise is $\overline{\theta}$ and the output obtained from an non-professional worker with certainty is $\underline{\theta} < \overline{\theta}$.

There are two locations in this economy with $N$ individuals born in each location in each period. On average, $sN$ professional jobs and $(1-s)N$ non-professional jobs are demanded in each location in each period. However, these locations also face demand shocks, with two states of the world occurring with equal probability:

**State 1**: $sN + \epsilon sN$ professional and $(1-s)N + \epsilon(1-s)N$ non-professional jobs in location 1.
$sN - \epsilon sN$ professional and $(1-s)N - \epsilon(1-s)N$ non-professional jobs in location 2.
**State 2**: $sN - \epsilon sN$ professional and $(1-s)N - \epsilon(1-s)N$ non-professional jobs in location 1.
$sN + \epsilon sN$ professional and $(1-s)N + \epsilon(1-s)N$ non-professional jobs in location 2.

Notice that these demand shocks are skill neutral, in the sense that the probability of receiving a shock is the same for professional and non-professional workers within each location. The shocks are also positively correlated for professional and non-professional workers within a location. We could relax each of these assumptions without changing any of the results that follow.[26] In addition, the opportunity to enhance overall productivity by switching a professional working in a particular location with a professional working in the other location arrives with probability $P$ in each period.

Let the cost of moving be $C_1$ for individuals born in location 1 and $C_2 < C_1$ for individuals born in location 2 with weaker local identity. We assume that all workers are employed and that professional workers always take advantage of the productivity enhancing career opportunities when they arise, in which case total output in this economy remains constant across all states of the world: $2N\left[s\overline{\theta} + (1-s)\underline{\theta}\right]$. The competitive labor market equilibrium can be obtained in that case as the solution to the Central Planner's problem of minimizing training and moving costs across both locations.

[26] Specifically, we allowed shocks to be perfectly negatively correlated for professional and non-professional workers within each location. We also allowed the size of the shocks to vary; $\overline{\epsilon}$ for professionals and $\underline{\epsilon}$ for non-professionals.

### 7.2 The Central Planner's Problem

Let the supply of professional labor in location 1 be $x_1$. From the structure of the demand shocks it then follows that the supply of professional labor in location 2 will be $2sN - x_1$ and that the supply of non-professional labor in location 1 will be $N - x_1$. We derive $x_1$ as the solution to the Central Planner's cost minimization problem. Expected cost turns out to be a piece-wise linear function of $x_1$ and so it will be convenient to solve what is essentially a linear programming problem in three regimes:

**Regime 1**: $x_1 \in [sN - \epsilon sN, sN + \epsilon sN]$

The supply of professional labor in each location is sufficient to satisfy the minimum demand in that location but does not exceed the maximum demand.

The labor flow in each state of the world can then be derived as:

**Flow in state 1**: $sN + \epsilon sN - x_1$ professional labor from location 2 to location 1.

$(1-s)N + \epsilon(1-s)N - (N - x_1)$ non-professional labor from location 2 to location 1.

**Flow in state 2**: $x_1 - [sN - \epsilon sN]$ professional labor from location 1 to location 2.

$(N - x_1) - [(1-s)N - \epsilon(1-s)N]$ non-professional labor from location 1 to location 2.

The Central Planner chooses $x_1$ to minimize expected cost

$$E(C) = 2sNC_e + \frac{1}{2}\epsilon N\,(C_1 + C_2) + Px_1C_1 + P(2sN - x_1)C_2, \tag{1}$$

where the second term on the right hand side is the cost associated with movement at the start of the period and the last two terms reflect movement of professional labor during the period. Because $C_1 > C_2$ it is easy to verify that $E(C)$ is increasing linearly in $x_1$ and will be minimized at $x_1^* = sN - \epsilon sN$. Labor flows with $x_1 = x_1^*$ are then obtained as:

**Flow in state 1** ($x_1 = x_1^*$): $2\epsilon sN$ professional labor from location 2 to location 1.

$\epsilon N(1-2s)$ non-professional labor from location 2 to location 1.[27]

**Flow in state 2** ($x_1 = x_1^*$): No flow of professional labor.

$\epsilon N$ non-professional labor from location 1 to location 2.

**Regime 2**: $x_1 \in [sN - \epsilon sN - \epsilon N(1-2s), sN - \epsilon sN]$

[27] To generate a positive labor flow we require $s \leq 1/2$. This is a reasonable assumption since just a small fraction of jobs (9 percent in our data) are professional. For the case with asymmetric shocks, $\bar{\epsilon}$ for the professionals and $\underline{\epsilon}$ for the non-professionals, the corresponding condition is $s \leq 1/(1 + \bar{\epsilon}/\underline{\epsilon})$.

We now reduce the supply of non-professional labor in location 2, but at most to the point where no non-professional labor flows to location 1 in state 1. In our set up, any reduction in non-professional labor supply in location 2 must lead to a reduction in professional labor in location 1 by the same amount. It then follows that the supply of professional labor in location 1 will no longer be sufficient to meet even the minimum demand in that location, while the supply of professional labor in location 2 will exceed the maximum demand in that location. Labor flows at the beginning of the period will necessarily increase, with an accompanying increase in moving costs, but we will see that this may be outweighed by the reduced cost of relocation for professional workers from location 1.

The labor flow in each state is derived as:

**Flow in state 1**: $sN + \epsilon sN - x_1$ professional labor from location 2 to location 1.

$(1-s)N + \epsilon(1-s)N - (N - x_1)$ non-professional labor from location 2 to location 1.

**Flow in state 2**: $sN - \epsilon sN - x_1$ professional labor from location 2 to location 1.

$(N - x_1) - [(1-s)N - \epsilon(1-s)N]$ non-professional labor from location 1 to location 2.

The expected cost can then be expressed as:

$$E(C) = 2sNC_e + \frac{1}{2}\left[\epsilon N + sN(1-\epsilon) - x_1\right](C_1 + C_2) + Px_1C_1 + P(2sN - x_1)C_2. \tag{2}$$

Collecting terms, $E(C)$ is declining linearly in $x_1$ and continues to be minimized at $x_1^* = sN - \epsilon sN$ if $P < 1/2(C_1+C_2)/(C_1-C_2)$. However, the local minimum is obtained at $x_1^{**} = sN - \epsilon sN - \epsilon N(1-2s)$, if the sign of the inequality is reversed. This will be the case if $P$ and $C_1 - C_2$ are sufficiently large. Substituting $x_1^{**}$ in equation (2) above, it is easy to verify that the term in square brackets is greater than $\epsilon N$, the term corresponding to it in equation (1), which implies that moving costs at the beginning of the period increase when going from $x_1^*$ to $x_1^{**}$. However, moving costs during the period decrease with the reduction in $x_1$, and this effect dominates under the conditions on $P$ and $C_1 - C_2$ derived above. Labor flows with $x_1 = x_1^{**}$ are obtained as:

**Flow in state 1** ($x_1 = x_1^{**}$): $\epsilon N$ professional labor from location 2 to location 1.

No flow of non-professional labor.

**Flow in state 2** ($x_1 = x_1^{**}$): $\epsilon N(1-2s)$ professional labor from location 2 to location 1.

$2\epsilon N(1-s)$ non-professional labor from location 1 to location 2.

**Regime 3**: $x_1 \in [0, sN - \epsilon sN - \epsilon N(1-2s)]$

We now reduce the supply of professional labor in location 1, with an accompanying increase in non-professional labor, even further so that professional labor flows from location 2 to location 1 and non-professional labor flows in the opposite direction in both states of the world.

Labor flows are now derived as:

**Flow in state 1**: $sN + \epsilon sN - x_1$ professional labor from location 2 to location 1.

$(N - x_1) - [(1 - s)N + \epsilon(1 - s)N]$ non-professional labor from location 1 to location 2.

**Flow in state 2**: $sN - \epsilon sN - x_1$ professional labor from location 2 to location 1.

$(N - x_1) - [(1 - s)N - \epsilon(1 - s)N]$ non-professional labor from location 1 to location 2.

The corresponding expected cost expression is obtained as:

$$E(C) = 2sNC_e + \frac{1}{2}\left[2(sN - x_1)\right](C_1 + C_2) + Px_1C_1 + P(2sN - x_1)C_2. \tag{3}$$

It is straightforward to verify that $E(C)$ is unambiguously decreasing in $x_1$ and, hence, is minimized at $x_1^{**} = sN - \epsilon sN - \epsilon N(1 - 2s)$.

### 7.3 Equilibrium Labor Allocation

As described in Figure 5 and derived above, $E(C)$ is increasing in $x_1$ to the right of $sN - \epsilon sN$ and decreasing in $x_1$ to the left of $sN - \epsilon sN - \epsilon N(1 - 2s)$ for all values of $P$. For $x_1 \in [sN - \epsilon sN - \epsilon N(1 - 2S), sN - \epsilon sN]$, $E(C)$ is declining in $x_1$ for $P < 1/2(C_1 + C_2)/(C_1 - C_2)$, whereas $E(C)$ is increasing in $x_1$ when the sign of the inequality is reversed. The global minimum is consequently obtained at $x_1^* = sN - \epsilon sN$ for $P < 1/2(C_1 + C_2)/(C_1 - C_2)$ and at $x_1^{**} = sN - \epsilon sN - \epsilon N(1 - 2s)$ for $P > 1/2(C_1 + C_2)/(C_1 - C_2)$. The supply of professional labor in location 1 falls short of the expected demand $sN$ for all values of $P$. Without uncertainty, the supply of professionals in location 1 could still fall short of the expected demand, but only if $P > (C_1 + C_2)/(C_1 - C_2)$.[28]

[28] Without uncertainty in labor demand, $sN$ professional and $(1 - s)N$ non-professional jobs are available in each location in each period. Let $x_1 \in [0, sN]$ measure the supply of professional workers in location 1. It then follows that $sN - x_1$ professional workers would flow from location 2 to location 1 and $(N - x_1) - (1 - s)N$ non-professional workers would flow in the opposite direction at the beginning of each period. Using the same notation as above, the Central Planner chooses $x_1$ to minimize

$$E(C) = 2sNC_e + (sN - x_1)(C_1 + C_2) + Px_1C_1 + P(2sN - x_1)C_2.$$

It follows that $x_1 = sN$ if $P < (C_1 + C_2)/(C_1 - C_2)$. $x_1 = 0$ and professional labor is under-supplied in location 1 if the sign of the inequality is reversed.

## References


[1] Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5): 1369-1401.

[2] Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2002. Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *Quarterly Journal of Economics*, 117(4): 1231-1294.

[3] Akerlof, George A. and Rachel E. Kranton. 2000. Economics and Identity, *Quarterly Journal of Economics* 105(3): 715-53.

[4] Akerlof, George A. and Rachel E. Kranton. 2005. Identity and the Economics of Organizations, *Journal of Economic Perspectives* 19(1): 9-32.

[5] Alba, Richard D. 1990. *Ethnic Identity: The Transformation of White America*. New Haven and London: Yale University Press.

[6] Alesina, Alberto, Reza Baqir, and William Easterly. 1999. Public Goods and Ethnic Division, *Quarterly Journal of Economics* 114(4): 1243-84.

[7] Alesina, Alberto, Reza Baqir, and Caroline Hoxby. 2004. Political Jurisdictions in Heterogeneous Communities, *Journal of Political Economy* 112(2): 348-396.

[8] Alexander, June Granatir. 1991. Moving into and out of Pittsburgh: Ongoing Chain Migration, in *A Century of European Migrants, 1830-1930*, eds. Rudolph J. Vecoli and Suzanne M. Sinke, Urbana: University of Illinois Press.

[9] Barro, Robert J. and Rachel M. McCleary. 2003. Religion and Economic Growth Across Countries, *American Sociological Review* 68: 760-781.

[10] Barton, Josef J. 1975. *Peasants and Strangers: Italians, Rumanians, and Slovaks in an American City, 1890-1950*. Cambridge, Massachusetts: Harvard University Press.

[11] Bénabou, Roland and Jean Tirole. 2006. Belief in a Just World and Redistributive Politics, *Quarterly Journal of Economics* 121(2): 699-746.

[12] Bisin, Alberto and Thierry Verdier. 2000. "Beyond the Melting Pot": Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits, *Quarterly Journal of Economics* 115(August): 955-988.

[13] Bodnar, John. 1980. Immigration, Kinship, and the Rise of Working-Class Realism in Industrial America, *Journal of Social History*. 45-65.

[14] Bodnar, John. 1985. *The Transplanted: A History of Immigrants in Urban America*. Bloomington: Indiana University Press.

[15] Bodnar, John, Roger Simon, and Michael P. Weber. 1982. *Lives of Their Own: Blacks, Italians, and Poles in Pittsburgh, 1900-1960*. Urbana: University of Illinois Press.

[16] Elder Jr., Glen H. and Rand D. Conger. 2000. *Children of the Land: Adversity and Success in Rural America*. Chicago and London: University of Chicago Press.

[17] Fernandez, Raquel and Alessandra Fogli. 2007. Culture: An Empirical Investigation of Beliefs, Work, and Fertility. New York University Department of Economics mimeograph.

[18] Finke, Roger and Rodney Stark. 1992. *The Churching of America, 1776-1990*. New Brunswick, NJ: Rutgers University Press.

[19] Fishlow, Albert. 2000. Internal Transportation in the Nineteenth and Early Twentieth Centuries, in *The Cambridge Economic History of the United States*, eds. Stanley L. Engerman and Robert E. Gallman, Cambridge, England: Cambridge University Press.

[20] Gans, Herbert. 1962. The Urban Villagers: Group and Class in the Life of Italian-Americans. New York: The Free Press of Glencoe.

[21] Gans, Herbert. 1979. Symbolic Ethnicity: The Future of Ethnic Groups and Cultures in America, *Ethnic and Racial Studies* 2(January): 1-20.

[22] Gerson, Kathleen, C. Ann Stueve, and Claude S. Fischer. 1977. Attachment to Place, in *Networks and Places: Social Relations in the Urban Setting*, eds. Claude S. Fischer et al., New York: The Free Press.

[23] Gjerde, Jon. 1985. *From Peasants to Farmers: The Migration from Balestrand, Norway to the Upper Middle West*. Cambridge, England: Cambridge University Press.

[24] Gjerde, Jon. 1991. Chain Migrations from the West Coast of Norway, in *A Century of European Migrants, 1830-1930*, eds. Rudolph J. Vecoli and Suzanne M. Sinke, Urbana: University of Illinois Press.

[25] Gjerde, Jon. 1997. *The Minds of the West: Ethnocultural Evolution in the Rural Middle West, 1830-1917*. Chapel Hill: University of North Carolina Press.

[26] Goldin, Claudia and Lawrence Katz. 1999. Human Capital and Social Capital: The Rise of Secondary Schooling in America, 1910-1940, *Journal of Interdisciplinary History* 29 (March): 683-723.

[27] Gordon, David M., Richard Edwards, and Michael Reich. 1982. *Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States*. Cambridge England: Cambridge University Press.

[28] Hoerder, Dirk. 1991. International Labor Markets and Community Building by Migrant Workers in the Atlantic Economies, in *A Century of European Migrants, 1830-1930*, eds. Rudolph J. Vecoli and Suzanne M. Sinke, Urbana: University of Illinois Press.

[29] Hummon, David M. 1990. *Commonplaces: Community Ideology and Identity in American Culture*. Albany: State University of New York Press.

[30] Hunter, Albert. 1975. The Loss of Community: An Empirical Test Through Replication, *American Sociological Review* 40(5): 537-552.

[31] Hutchinson, E. P. 1956. *Immigrants and Their Children 1850-1950* New York: John Wiley and Sons, Inc.

[32] Iannaccone, Laurence R. 1992. Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and Other Collectives, *Journal of Political Economy* 100(2): 271-297.

[33] Iannaccone, Laurence R. 1998. Introduction to the Economics of Religion, *Journal of Economic Literature* 36: 1465-1496.

[34] Kamphoefner, Walter D. 1987. *The Westfalians: From Germany to Missouri*. Princeton: Princeton University Press.

[35] Kim, Sukkoo. 2007. Immigration, Industrial Revolution and Urban Growth in the United States, 1820-1920: Factor Endowments, Technology and Geography, *NBER Working Paper* No. 12900.

[36] Margo, Robert A. 1999. Regional Wage Gaps and the Settlement of the Midwest, *Explorations in Economic History* 36: 128-143.

[37] Meyer, David R. 1989. Midwestern Industrialization and the American Manufacturing Belt in the Nineteenth Century, *Journal of Economic History* 49(4): 921-937.

[38] Ostergren, Robert C. 1976. *Rätvik to Isanti: A Community Transplanted*. Ph.D. dissertation, University of Minnesota.

[39] Sacerdote, Bruce and Edward L. Glaeser. 2001. Education and Religion, *NBER Working Paper* No. 8080.

[40] Sanders, Scott Russell. 2007. To Stay or Move On: The Artist's Dilemma. *Bloom*, February/March: 65-68.

[41] Sumner, William G. 1906. *Folkways*. New York: Ginn.

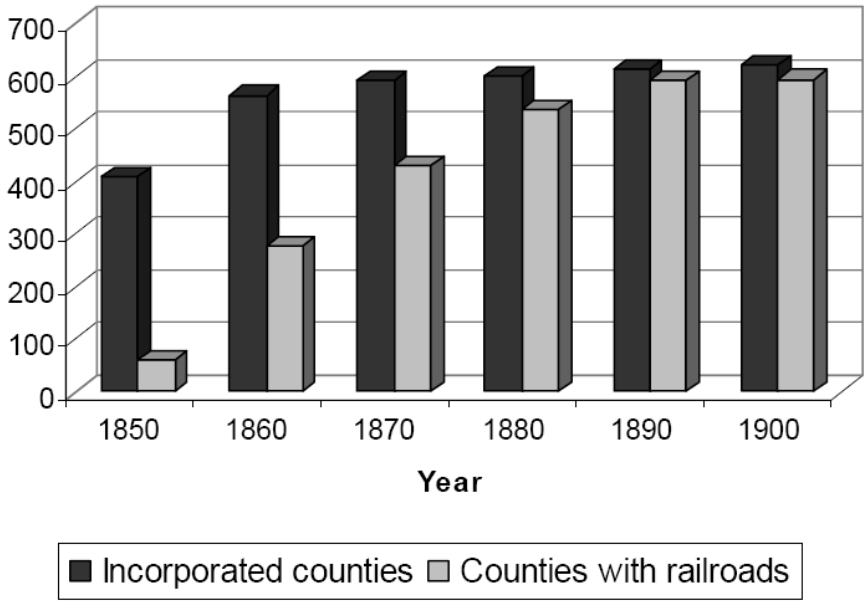
[42] Tabellini, G. 2005. Culture and Institutions: Economic Development in the Regions of Europe, CESifo Working Paper No. 1492.

[43] Thistlethwaite, Frank. 1991. Migration from Europe Overseas in the Nineteenth and Twentieth Centuries, in *A Century of European Migrants, 1830-1930*, eds. Rudolph J. Vecoli and Suzanne M. Sinke, Urbana: University of Illinois Press.

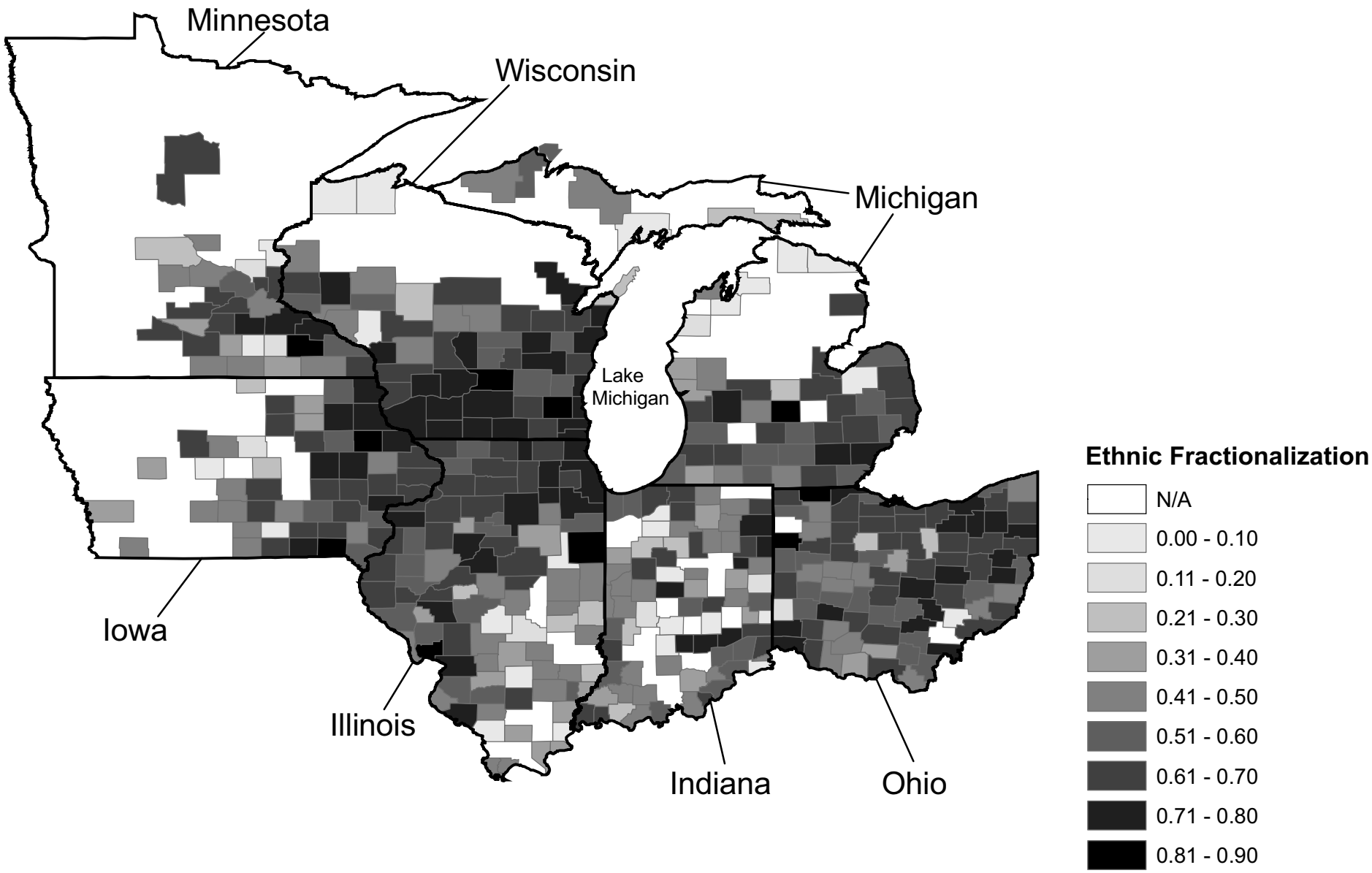
[44] Tullock, Gordon. 1980. Efficient Rent Seeking, in *Toward a Theory of the Rent-Seeking Society*, eds. James M. Buchanan, Robert D. Tollison, and Gordon Tullock, College Station: Texas A&M Press.

[45] Weber, Max. 1930. *The Protestant Ethic and the Spirit of Capitalism*. London, England: Allen and Unwin.

## Figure 1: Early Growth in the Midwest



# Figure 2: Ethnic Fractionalization in 1860



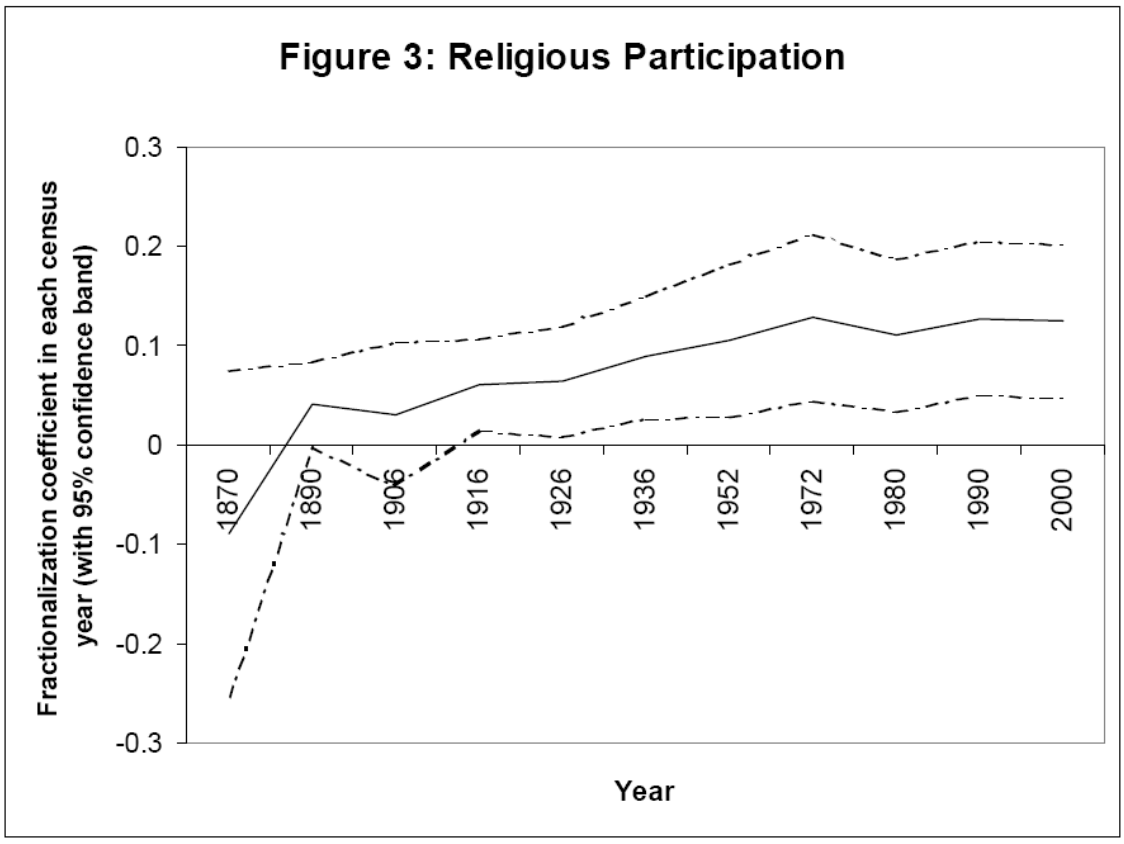
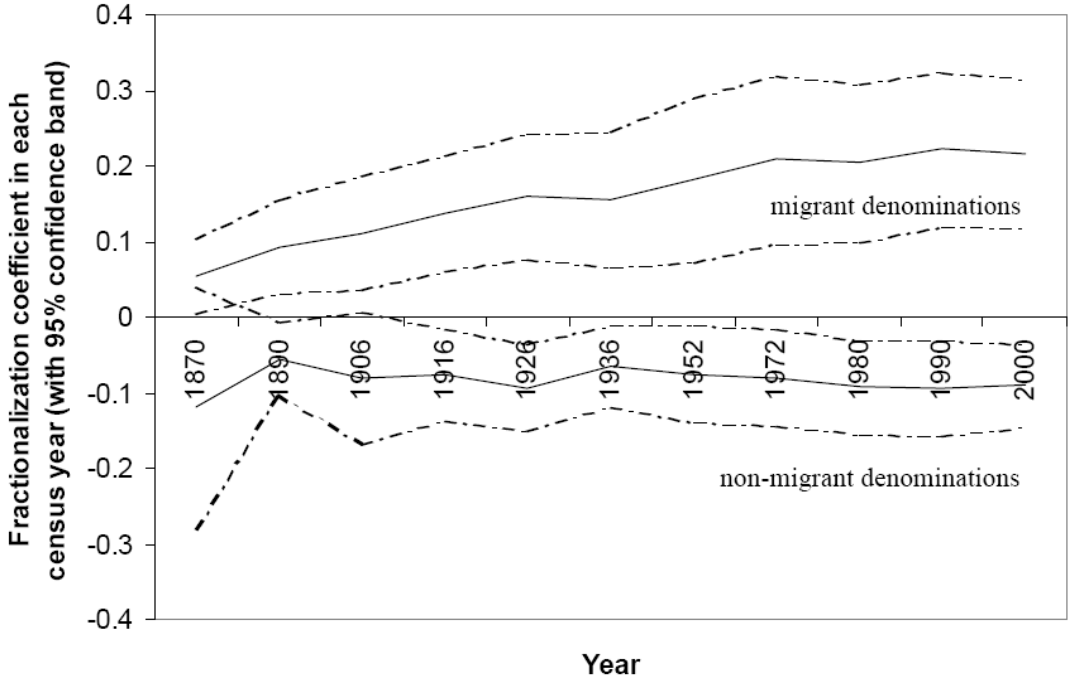


Figure 3: Religious Participation
Fractionalization coefficient in each census year (with 95% confidence band)
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
1870
1890
1906
1916
1926
1936
1952
1972
1980
1990
2000
Year

**Figure 4: Religious Participation by Denomination**



## Figure 5: Labor Supply in Equilibrium

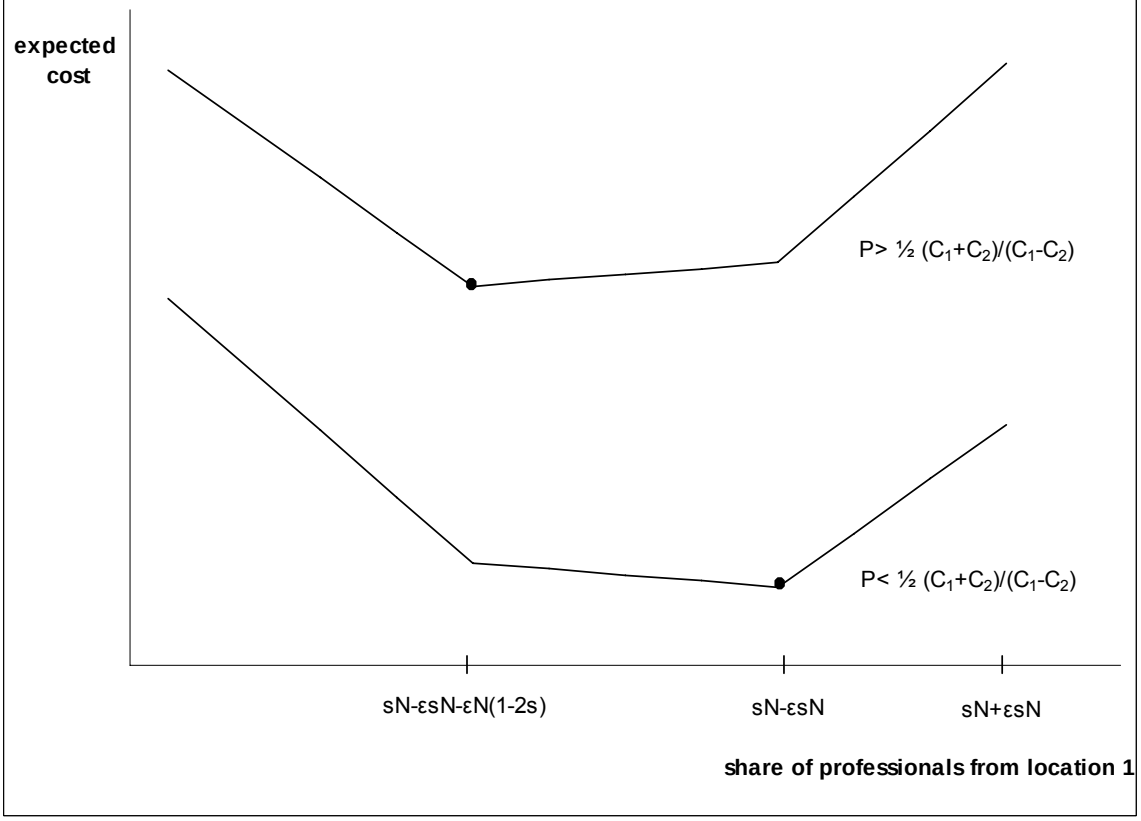


Figure 6: Historical Fractionalization and Current Occupational Choice

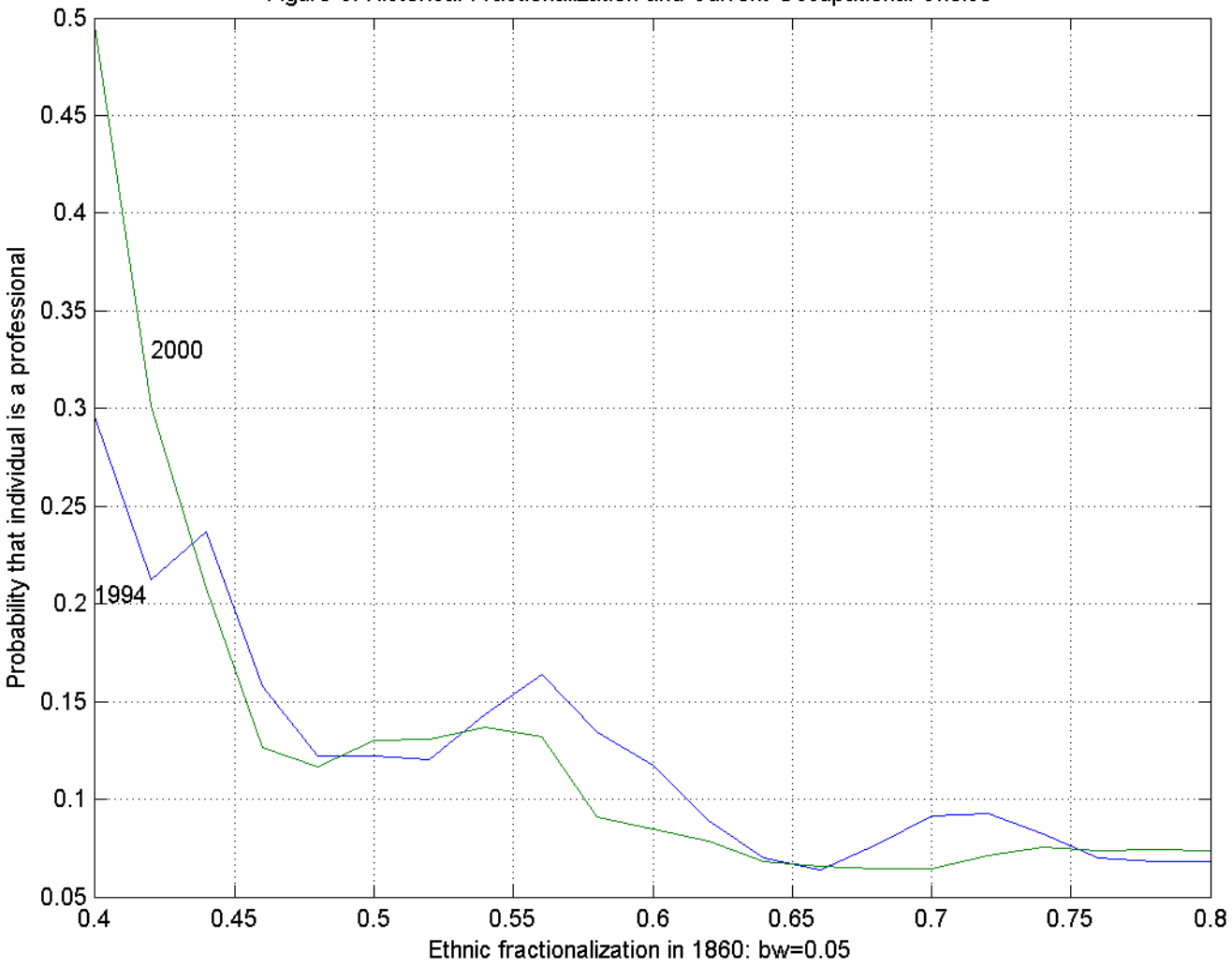


**Table 1: Ethnic Distribution, 1860-1900**

| Census year: | 1860 | 1880 | 1900 |
|---|---|---|---|
| | (1) | (2) | (3) |
| *Scandinavia* | | | |
| Danish | 0.01 | 0.02 | 0.02 |
| Finish | 0.00 | 0.00 | 0.01 |
| Norwegian | 0.03 | 0.07 | 0.07 |
| Swedish | 0.02 | 0.06 | 0.10 |
| *British Isles* | | | |
| English | 0.13 | 0.11 | 0.09 |
| Irish | 0.25 | 0.19 | 0.11 |
| Scottish | 0.03 | 0.03 | 0.02 |
| Welsh | 0.01 | 0.01 | 0.01 |
| *Western Europe* | | | |
| Belgian | 0.00 | 0.01 | 0.01 |
| Dutch | 0.01 | 0.01 | 0.01 |
| French | 0.03 | 0.02 | 0.01 |
| German | 0.32 | 0.37 | 0.41 |
| Italian | 0.00 | 0.00 | 0.01 |
| Swiss | 0.02 | 0.02 | 0.02 |
| *Eastern Europe* | | | |
| Czech | 0.00 | 0.01 | 0.02 |
| Hungarian | 0.00 | 0.00 | 0.00 |
| Polish | 0.00 | 0.01 | 0.02 |
| USSR | 0.00 | 0.00 | 0.01 |
| *Other* | 0.14 | 0.05 | 0.04 |
| *Total* | 1.00 | 1.00 | 1.00 |

Source: IPUMS 1:100 sample, including all foreign-born individuals.

**Table 2: Occupational Distribution, 1860-1900**

| Census year: | 1860 | 1880 | 1900 |
|---|---|---|---|
| | (1) | (2) | (3) |
| *White collar* | | | |
| Professional | 0.04 | 0.04 | 0.05 |
| Manager | 0.04 | 0.04 | 0.06 |
| Clerical | 0.00 | 0.01 | 0.02 |
| Sales | 0.01 | 0.02 | 0.03 |
| *Farm* | | | |
| Farmer | 0.50 | 0.41 | 0.31 |
| Laborer, Farm | 0.12 | 0.17 | 0.17 |
| *Blue collar, nonfarm* | | | |
| Craftsman | 0.10 | 0.08 | 0.09 |
| Operative | 0.05 | 0.08 | 0.09 |
| Household Service | 0.05 | 0.05 | 0.05 |
| Service | 0.00 | 0.01 | 0.02 |
| Laborer, Non-Farm | 0.09 | 0.10 | 0.12 |
| *Total* | 1.00 | 1.00 | 1.00 |

Source: IPUMS 1:100 sample, including all foreign-born individuals who report that they are employed and report an occupational category.

**Table 3: Transportation Infrastructure and County Characteristics, 1860**

| Dependent variable: | ethnic fractionalization | manufacturing share | agriculture share | population |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Railroad through county, 1860 | 0.049 | 0.001 | -0.013 | 0.102 |
| | (0.022) | (0.008) | (0.013) | (0.012) |
| Distance to canal, 1890 | -0.668 | 0.169 | -0.034 | -0.469 |
| | (0.191) | (0.097) | (0.125) | (0.102) |
| Distance to Great Lakes harbor | -0.252 | -0.066 | 0.100 | -0.136 |
| | (0.073) | (0.029) | (0.048) | (0.056) |
| Observations | 401 | 401 | 401 | 401 |

Note: Robust standard errors in parentheses.

Distance to canal and distance to Great Lakes harbor measured in thousands of kilometers.

Fractionalization is one minus the the Herfindahl index of ethnic concentration, averaged across occupational categories.

Manufacturing share and agriculture share in 1860 computed using IPUMS.

Population divided by 100,000.

## Table 4: Fractionalization in 1860 and County Characteristics in 1990

| Year: | 1990 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dependent variable: | agri share | manufac share | pop | ethnic frac | racial frac | religious frac | agri share | manufac share | pop | ethnic frac | racial frac | religious frac |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Fractionalization, 1860 | -0.022 | 0.024 | 2.590 | 0.045 | 0.072 | -0.080 | 0.003 | 0.018 | -0.269 | -0.027 | 0.011 | -0.124 |
| | (0.014) | (0.019) | (0.792) | (0.106) | (0.020) | (0.033) | (0.015) | (0.020) | (0.854) | (0.113) | (0.017) | (0.034) |
| Manufacturing share, 1860 | -- | -- | -- | -- | -- | -- | -0.090 | -0.056 | 1.636 | 0.176 | 0.084 | -0.257 |
| | | | | | | | (0.038) | (0.053) | (1.063) | (0.309) | (0.055) | (0.110) |
| Agriculture share, 1860 | -- | -- | -- | -- | -- | -- | 0.136 | 0.057 | -0.962 | 0.049 | -0.080 | -0.119 |
| | | | | | | | (0.027) | (0.037) | (0.746) | (0.225) | (0.030) | (0.074) |
| Population, 1860 | -- | -- | -- | -- | -- | -- | -0.087 | 0.030 | 11.716 | 0.326 | 0.246 | 0.124 |
| | | | | | | | (0.031) | (0.026) | (5.372) | (0.110) | (0.052) | (0.076) |
| Observations | 437 | 437 | 437 | 437 | 437 | 437 | 437 | 437 | 437 | 437 | 437 | 437 |

Note: Robust standard errors in parentheses.

Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.

Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.

Population is divided by 100,000.

Manufacturing share in 1990 defined as share of civilian labor force employed in manufacturing.

Agriculture share in 1990 is computed using farm population and total population in county.

Ethnic fractionalization in 1990 is one minus the Herfindahl index of (white) ethnic concentration based on 16 ethnicities.

Racial fractionalization in 1990 is one minus the Herfindahl index of racial concentration based on 5 racial groups.

Religious fractionalization in 1990 is one minus the Herfindahl index of religious concentration based on 18 denominations.

**Table 5: Fractionalization in 1860 and Local Government Expenditure in 1990**

| Year: | 1990 | | | | | |
|---|---|---|---|---|---|---|
| Dependent variable: | total expenditure per capita | education share | health share | police share | roads share | welfare share |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Fractionalization, 1860 | 0.143 | -0.055 | -0.011 | 0.014 | 0.019 | 0.006 |
| | (0.145) | (0.047) | (0.026) | (0.004) | (0.012) | (0.009) |
| Manufacturing share, 1860 | 0.432 | 0.065 | -0.019 | 0.004 | -0.050 | -0.034 |
| | (0.469) | (0.155) | (0.060) | (0.014) | (0.030) | (0.029) |
| Agriculture share, 1860 | -0.042 | 0.020 | 0.047 | -0.011 | -0.0001 | -0.025 |
| | (0.410) | (0.106) | (0.048) | (0.009) | (0.026) | (0.031) |
| Population, 1860 | 0.168 | -0.167 | -0.005 | 0.026 | -0.046 | 0.000 |
| | (0.160) | (0.047) | (0.016) | (0.007) | (0.020) | (0.010) |
| Observations | 437 | 437 | 437 | 437 | 437 | 437 |

Note: Robust standard errors in parentheses.

Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.

Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.

Population is divided by 100,000 and population density is measured in thousands per square mile.

Total expenditure per capita is measured in thousands of dollars (1990).

Shares in columns (2) through (6) are computed as fraction of total expenditure.

**Table 6: Distribution of Denominations and Religious Participation, 1860-2000**

| Measure of religious participation: | religious participation | | | | | |
|---|---|---|---|---|---|---|
| | church seats | | church members | | church adherents | |
| Census year: | 1860 | 1890 | 1890 | 1952 | 1972 | 2000 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Distribution of denominations | | | | | | |
| Baptist | 0.14 | 0.11 | 0.08 | 0.06 | 0.06 | 0.09 |
| Catholic | 0.11 | 0.13 | 0.29 | 0.27 | 0.30 | 0.33 |
| Lutheran | 0.05 | 0.12 | 0.14 | 0.22 | 0.21 | 0.20 |
| Methodist | 0.38 | 0.27 | 0.21 | 0.18 | 0.19 | 0.13 |
| Presbyterian | 0.14 | 0.08 | 0.06 | 0.05 | 0.05 | 0.03 |
| Other | 0.18 | 0.28 | 0.22 | 0.22 | 0.19 | 0.22 |
| *Total* | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Proportion religious | 0.59 | 0.67 | 0.30 | 0.50 | 0.58 | 0.54 |

Source: Census of Religious Bodies.

Proportion religious is computed as the number of church seats divided by the population 1860-1890, the number of church members divided by the population 1890-1952, and the number of church adherents divided by the population 1972-2000.

**Table 7: NLSY79 Descriptive Statistics**

| | year | |
|---|---|---|
| | 1994 | 2000 |
| | (1) | (2) |
| Professional | 0.09 | 0.09 |
| | (0.29) | (0.28) |
| Migrated out of county of birth | 0.56 | 0.59 |
| | (0.01) | (0.01) |
| Employed | 0.81 | 0.92 |
| | (0.01) | (0.01) |
| Income | 27.80 | 33.06 |
| | (0.51) | (0.58) |
| White | 0.79 | 0.79 |
| | (0.01) | (0.01) |
| Female | 0.50 | 0.51 |
| | (0.01) | (0.01) |
| Age | 33.36 | 39.38 |
| | (0.05) | (0.05) |

Note: Standard errors in parentheses.
Professional occupations are relevant codes in the Professional, Technical and Kindred Workers category.
All variables except income and age are binary.
Income is measured in thousands of dollars (2000).

## Table 8: Fractionalization in the County of Birth and Occupational Choice

| Dependent variable: | professional | | | | | |
|---|---|---|---|---|---|---|
| Year: | 1994 | 2000 | 1994 | 2000 | 1994 | 2000 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Fractionalization, 1860 | -0.146 | -0.163 | -0.253 | -0.316 | -0.215 | -0.395 |
| | (0.083) | (0.088) | (0.096) | (0.109) | (0.129) | (0.126) |
| White | 0.031 | 0.045 | 0.047 | 0.059 | 0.049 | 0.060 |
| | (0.022) | (0.019) | (0.020) | (0.020) | (0.021) | (0.025) |
| Female | -0.065 | -0.023 | -0.064 | -0.021 | -0.069 | -0.024 |
| | (0.020) | (0.017) | (0.020) | (0.016) | (0.021) | (0.017) |
| Age | 0.0003 | 0.003 | 0.001 | 0.003 | 0.001 | 0.003 |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.005) | (0.004) |
| Manufacturing share, 1860 | -- | -- | -0.213 | -0.521 | -0.092 | -0.482 |
| | | | (0.172) | (0.213) | (0.258) | (0.244) |
| Agriculture share, 1860 | -- | -- | -0.272 | -0.431 | -0.247 | -0.400 |
| | | | (0.144) | (0.160) | (0.157) | (0.157) |
| Population, 1860 | -- | -- | 0.018 | 0.011 | -0.017 | 0.044 |
| | | | (0.017) | (0.025) | (0.039) | (0.045) |
| Ethnic shares, 1860 | No | No | No | No | Yes | Yes |
| Observations | 1209 | 1122 | 1209 | 1122 | 1209 | 1122 |

Note: Standard errors in parentheses are clustered at the county level.

Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.

White, female, and age are individual-level characteristics.

Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.

Population divided by 100,000.

Professional is a binary variable indicating whether the individual is employed in a professional occupation.

Professional occupations are relevant codes in the Professional, Technical, and Kindred Workers category.

Ethnic shares, 1860 is a full set of 32 ethnic shares computed as the population of the ethnic group divided by the total population in the county.

## Table 9: Fractionalization and Outcomes Related to Occupational Choice

| Dependent variable: | migrated | | employed | | income | |
|---|---|---|---|---|---|---|
| Year: | 1994 | 2000 | 1994 | 2000 | 1994 | 2000 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Fractionalization, 1860 | -0.388 | -0.414 | 0.114 | 0.037 | 1.659 | -11.722 |
| | (0.155) | (0.138) | (0.108) | (0.065) | (4.804) | (6.183) |
| Manufacturing share, 1860 | -0.071 | -0.242 | 0.078 | 0.362 | -1.195 | -26.379 |
| | (0.305) | (0.260) | (0.230) | (0.136) | (10.268) | (12.231) |
| Agriculture share, 1860 | -0.171 | -0.317 | 0.031 | 0.253 | -8.197 | -21.016 |
| | (0.311) | (0.222) | (0.164) | (0.107) | (7.808) | (9.359) |
| Population, 1860 | 0.021 | 0.065 | -0.015 | 0.013 | -0.210 | 0.449 |
| | (0.039) | (0.040) | (0.026) | (0.015) | (1.041) | (1.691) |
| White | 0.232 | 0.284 | 0.153 | 0.043 | 5.773 | 4.791 |
| | (0.048) | (0.052) | (0.024) | (0.021) | (1.850) | (1.567) |
| Female | -0.014 | -0.014 | -0.155 | -0.059 | -11.986 | -14.476 |
| | (0.017) | (0.024) | (0.021) | (0.015) | (1.122) | (0.988) |
| Age | 0.013 | 0.015 | 0.001 | 0.003 | 0.764 | 0.211 |
| | (0.005) | (0.005) | (0.004) | (0.004) | (0.208) | (0.312) |
| Observations | 1598 | 1437 | 1614 | 1332 | 1251 | 1122 |

Note: Standard errors in parentheses are clustered at the county level.

Fractionalization is one minus the the Herfindahl index of ethnic concentration, averaged across occupational categories.

Fractionalization is measured in the county of birth.

Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.

Population is divided by 100,000.

White, female, and age are individual-level characteristics.

Migrated is a binary variable that indicates whether the individual resides outside the county of birth.

Employed is a binary variable that indicates whether the individual currently holds a job.

Income is measured in thousands of dollars (2000).

## Table 10: Fractionalization in the County of Residence and Occupational Choice

| Dependent variable: | professional | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sample: | stayers | | out-migrants | | in-migrants | | current residents | |
| Year: | 1994 | 2000 | 1994 | 2000 | 1994 | 2000 | 1994 | 2000 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Fractionalization, 1860 | -0.182 | -0.153 | -0.213 | -0.343 | -0.016 | 0.205 | -0.089 | 0.091 |
| | (0.105) | (0.104) | (0.124) | (0.146) | (0.094) | (0.088) | (0.074) | (0.067) |
| Manufacturing share, 1860 | -0.243 | -0.078 | -0.197 | -0.674 | -0.431 | -0.098 | -0.367 | -0.046 |
| | (0.194) | (0.225) | (0.221) | (0.260) | (0.167) | (0.237) | (0.131) | (0.172) |
| Agriculture share, 1860 | -0.225 | -0.159 | -0.277 | -0.480 | -0.254 | 0.045 | -0.248 | -0.001 |
| | (0.147) | (0.206) | (0.192) | (0.237) | (0.147) | (0.139) | (0.114) | (0.104) |
| Population, 1860 | 0.041 | -0.021 | -0.006 | 0.023 | 0.048 | 0.008 | 0.041 | -0.012 |
| | (0.023) | (0.018) | (0.028) | (0.046) | (0.034) | (0.024) | (0.017) | (0.016) |
| White | 0.054 | 0.026 | 0.016 | 0.061 | 0.086 | 0.062 | 0.072 | 0.048 |
| | (0.023) | (0.021) | (0.045) | (0.025) | (0.021) | (0.024) | (0.013) | (0.020) |
| Female | -0.026 | 0.028 | -0.097 | -0.057 | -0.095 | -0.076 | -0.065 | -0.031 |
| | (0.020) | (0.020) | (0.026) | (0.022) | (0.020) | (0.021) | (0.015) | (0.016) |
| Age | 0.004 | -0.005 | -0.005 | 0.006 | -0.008 | 0.001 | -0.002 | -0.001 |
| | (0.006) | (0.005) | (0.006) | (0.006) | (0.006) | (0.006) | (0.004) | (0.004) |
| Observations | 557 | 479 | 648 | 637 | 648 | 651 | 1205 | 1130 |

Note: Standard errors in parentheses are clustered at the county level.

Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.

Fractionalization and county characteristics are measured in the county of birth in Columns 1-4 and in the county of residence in Columns 5-8.

Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.

Population is divided by 100,000.

White, female, and age are individual-level characteristics.

Professional is a binary variable indicating whether the individual is employed in a professional occupation.

Professional occupations are relevant codes in the Professional, Technical, and Kindred Workers category.

"Stayers" are individuals whose Midwest county of birth and county of residence are the same.

"Out-migrants" are individuals who reside in a county outside of their Midwest county of birth.

"In-migrants" are individuals who reside in a Midwest county outside of their county of birth.

**Table 11: Fractionalization and Individual Ability**

| Dependent variable: | AFQT score | high school completion | college completion | college out of state |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Fractionalization, 1860 | -0.535 | 0.016 | -0.095 | -0.228 |
| | (7.457) | (0.100) | (0.104) | (0.145) |
| Manufacturing share, 1860 | 11.570 | 0.197 | 0.003 | -0.420 |
| | (17.828) | (0.160) | (0.241) | (0.311) |
| Agriculture share, 1860 | -3.015 | -0.017 | -0.261 | -0.535 |
| | (14.672) | (0.143) | (0.193) | (0.308) |
| Population, 1860 | -2.769 | -0.052 | -0.018 | -0.029 |
| | (1.981) | (0.023) | (0.023) | (0.030) |
| White | 22.608 | 0.010 | 0.083 | 0.008 |
| | (1.626) | (0.024) | (0.018) | (0.045) |
| Female | -2.728 | 0.000 | -0.039 | -0.002 |
| | (1.248) | (0.017) | (0.020) | (0.024) |
| Age | 2.880 | 0.006 | 0.005 | 0.025 |
| | (0.332) | (0.003) | (0.004) | (0.006) |
| Observations | 2187 | 2023 | 2023 | 1154 |

Note: Standard errors in parentheses are clustered at the county level.

Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.

Fractionalization is measured in the county of birth.

Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.

Population is divided by 100,000.

White, female, and age are individual-level characteristics.

AFQT is the score on the Armed Forces Qualification Test.

High school completion is a binary variable indicating whether the individual completed high school, including GED.

College completion is a binary variable indicating whether the individual completed a four-year college/university degree.

College out of state is a binary variable indicating whether the individual attended college out of state of birth.