Ethnic Diversity and the Provision of Local Public Goods: Evidence from U.S. History

(A preliminary and incomplete draft. Comments are welcome.)

Sun Go

Chung-Ang University

sungo@cau.ac.kr

November 2018

Abstract

Did the influx of immigrants help the early development of public education in the USA? Or did it hinder the expansion of universal public schooling? This old question is reexamined in scrutiny using a combination of county-level statistics and complete count population data from the 1850 census. The empirical result reveals that the dense presence of immigrants affected the supply of and demand for public elementary education differently. The relative size of the immigrant population had a negligible effect on the supply of common school education. On the other hand, the individual demand for public education was positively associated with the local density of the immigrant population. The positive impact of immigrants on individual demand for common schooling at the county level was found for most countries of the origin at different magnitudes.

Keywords: immigration, public education, common school, school tax, enrollment, attendance

JEL Classifications: N31, H75, J15

I. Introduction

Did immigration help the rise of U.S. public schooling from the mid-19th to early 20th centuries? Or, was immigration a barrier to the early development of public education? Would the onset of universal free public education in the U.S. have come earlier if the U.S. was ethnically, racially, and culturally more homogeneous? This paper studies the effect of immigration on the rise of American public schooling at the elementary and secondary level from 1850 to 1930, when mass immigrants entered the USA while common schools, and later public high schools, spread across the country.

In the 19th and early-20th centuries, the USA was a global leader in both public education and immigration admittance. The universal free public education system was first established at the elementary level in early-19th century New England, and spread to the South and the West after then. The growth of public high schools followed the common school movement from the late-19th to the early-20th centuries. (Cubberley, 1919; Goldin and Katz, 2010) During the same period, immigrants came to the USA in a large scale, mostly from Europe. The massive influx of immigrants peaked in 1914, and continued until 1930. (Carter et al., 2006)

The simultaneous rise of universal free public education and the massive influx of immigrants, as well as a rich cross-sectional variation across regions, states, and counties, offers a good chance to study the impact of immigration on the provision of public schooling. This question is significant not only for understanding the social and economic history of the USA but also for testing the modern economic theory of providing local public goods using American historical experience. Finding the positive or negative effect of immigration on the supply of public funding also speaks to today's concerns about how the surge of refugees into rich countries, and the backlash against it, will affect the availability of public support for the needy native population. Will the arrival of immigrants mean less schooling and other public goods for those already here? Or will the populist reaction take the form of "welfare chauvinism," with enough discrimination in favor of the native-born to protect their public subsidy levels? American localities' experience between 1850 and 1930 offers a rich testing ground of differential attitudes toward immigrants and different support for schooling.

Public economics provides theory to explain both the negative and positive effects of the immigration influx on the growth of local public education. The impact of increasing immigrants on the local political economy is not clearly predicted. The influx of immigrants made a locality more heterogeneous, which can lead to both the rise and fall of the investment in local public education. Greater community homogeneity, or social affinity, facilitates the collective decision making of raising taxes to supply local public goods. Thus, immigration-induced local heterogeneity could be a barrier to the investment in local schools. (Easterly and Levine, 1997; Alesina, Baqir, and Easterly, 1999; Lindert, 2004; Banerjee and Somanathan, 2007; Baldwin and Huber, 2010; Lindert, 2014) However, a culturally, ethnically, and linguistically more diverse community will be more interested in investing in local education, a chief social institution for assimilation and community identity. (Alesina and La Ferrara, 2005; Mirel, 2010; Alesina, Harnoss, and Rapoport, 2016) American history offers a good chance to empirically test and evaluate the two competing theories of public economics about the role of immigration in the development of the public education system.

The existing literature has studied the mass influx of immigrants in the USA and the rise of American common school and high school education in depth, respectively. Different from literature about the impact of mass immigration on public finance and economic development in the contemporary world, studies in economics and economic history about the mass influx of immigration in U.S. history have focused on its impact mostly on the labor market or total output, paying little attention to its relation to the provision of local public goods. (Ager and Brüker, 2013; Abramitzky and Boustan, 2017) Previous economic history research on the rise of American public elementary and secondary schooling have considered international and interstate migration as one of the many factors that affected the provision of local public education. However, the limited availability of data has prevented researchers from finding a solid result by using econometric analysis. (Go and Lindert, 2010; Goldin and Katz, 2010) This paper extends the existing literature to the investigation of the causal link between the two historic events by analyzing newly collected and digitized data. Previous research using statistical methods has mostly used aggregate-level data. Instead, more detailed individual and household level data are used here, thanks to the recently published digital version of the U.S. full-count census data from 1850 to 1930.

Studies in economic growth, development economics, and cultural economics have also noticed the relationship between community diversity and the provision of local public goods. Most of the studies have analyzed cross-country variation to find a causal impact of ethnic, linguistic, and cultural diversity on the level of investment in public goods. (Easterly and Levine, 1997; Alesina, Baqir, and Easterly, 1999; Alesina and Giuliano, 2015) This paper tries to contribute to the literature by adding a novel analysis of the natural experiment in American history in the age of mass immigration.

This paper empirically studies the following two research questions using historical data.

(1) What was the effect of immigration influx on the rise of public elementary and secondary schooling in the USA from 1850 to 1930?

(2) Does the ethnic, linguistic, and cultural homogeneity of a community help or impede the provision of tax-financed local public goods?

The first research question seeks to uncover the role of mass immigration influx in the 19th and early-20th century development of American society and economy. It is also related to better understanding and explaining the early rise of public education, and human capital investment, in the USA, which significantly contributed to the fast growth of the American economy in the later period.

The second question focuses on testing the theories developed from public economics today using historical data. American history provides a possible quasi-experiment to identify the causal link from community heterogeneity to the provision of local public goods.

This research is significant at least in three ways. First, it contributes to the economic history literature on immigration and public education in the 19th and early-20th USA, and to help us better understand the causal effect of the influx of immigrants on the rise of early public elementary and secondary schooling. Second, it extends the public economics literature on local public finance by analyzing the historical quasi-experiment in the USA, marked by the simultaneous rise of the immigration influx and public schooling, to identify the causal effect of ethnic, cultural, and linguistic heterogeneity on the provision of local public goods. Finally, this research is related to today's concerns about the impact of community diversity, driven by the influx of immigrants, on the welfare of the native-born population, thus providing reference evidence for policy makers.

II. Methodology

I shall use various testing models to address different hypotheses that have been proposed and used by the existing literature studying ethnic, linguistic, and birthplace diversity and their empirical relation to economic development. (Alesina et al., 2003; Michalopoulos, 2012; Michalopoulos and Papaioannou, 2013; Alesina, Harnoss, and Rapoport, 2016; Alesina, Michalopoulos, and Papaioannou, 2016) Starting with tests on readily available county-level data, I will apply a core baseline estimating equation like the following:

$$y_{cst} = m_{cst}\beta_m + X_{icst}\beta_X + W_{cst}\beta_X + \rho_s + \tau_t + \varepsilon_{icst}$$

The subscripts i, c, s, and t represent individual, county, state and year, respectively. The dependent variable y_{cst} is a county level variable showing the provision of local public schooling, such as the number of public schools, teachers, and the amount of school revenue at the elementary and secondary level, all divided by local school population. The public school enrollment rate will be included in the set of the dependent variables. The right-hand side of the estimating equation has vectors of immigration-related variables (m_{cst}), individual and household-level control variables (X_{icst}), county-level control variables (W_{cst}), state fixed effects (ρ_s), and year fixed effects (τ_t), and the error term (ε_{icst}).

The key regressors are immigration-related variables, including measures of ethnic heterogeneity and the population shares of ethnic groups at the local level. Two issues arise here. One is to determine how to measure the community heterogeneity driven by the influx of immigrants. A candidate is the Herfindahl-Hirschman Index that is frequently used to measure market concentration. If the population share of an ethnic group a in county c and in year t is s_{act}, the Herfindahl-Hirschman Index of county c in year t is defined as follows.

$$\mathrm{HHI}_{ct} = \sum_a (s_{act})^2$$

A higher HHI represents an ethnically homogeneous county, and a lower index means that a county is more heterogenous in the ethnic composition. As suggested by Mauro (1995), the HHI index can be easily transformed into an index of ethnic fractionalization (FRAC_{ct}) as follows.

$$\mathrm{FRAC}_{ct} = 1 - \mathrm{HHI}_{ct} = 1 - \sum_a (s_{act})^2$$

Another measure of interest is the index of ethnic polarization, proposed by Reynal-Querol (2002) and Montalvo and Reynal-Querol (2005). The index is defined as follows.

$$\mathrm{POL}_{ct} = 1 - \sum_a \left(\frac{1/2 - s_{act}}{1/2}\right)^2 s_{act}$$

The polarization index measures how close the distribution of ethnic groups to the bipolar distribution. The index is maximized if the county has only two ethnic groups, and their population shares are the same as 1/2. Besides, there could be other measures of ethnic homogeneity. In this research, various measures will be used in the analysis and closely examined to uncover the role of ethnic homogeneity in the provision of public schooling.

Another issue is to define each ethnic group. The basic unit of the ethnic group is the immigrants' country of origin. However, treating each country of origin identically could be potentially problematic. For example, a county where 1/3 of the population are American natives, another 1/3 are immigrants from England, and the remaining 1/3 are immigrants from Ireland will be quite different from a county where 1/3 of the population are American natives, 1/3 are immigrants from China, and the rest are immigrants from Portuguese Angola. If we calculate the measures of ethnic homogeneity simply by the immigrants' country of origin, the two counties will be identified as counties with a same degree of ethnic heterogeneity. Thus, in the research, the ethnic group will be defined by a union of countries proximate to each other, while the proximity between countries is measured by their geographical and linguistic proximity.

Measuring the population share of a country of origin can be improved by the use of full count census micro sample data. Previous literature constructed a measure of ethnic and cultural diversity using a population share of each ethnic group, identified by the country of origin from the county-level aggregate reports of the decennial censuses. I improve this measure using the individual full count data of the decennial censuses. The individual full count data help us identify the diversity of a county at the household level, which is a more accurate measure. I can also construct the measures by the household and individual characteristics, such as age, sex, and occupation. Another merit of using individual-level data is additional information about the family. For example, a US-born native, whose father is a European migrant, would have been different from other US-born native whose father was also born in the U.S. The ethnicity of the spouse, and the ethnicity of the spouse of a sibling could have also effectively changed the ethnic and cultural diversity in a locality.

A remaining issue is possible endogeneity between the influx of immigrants and the provision of public goods. For example, a statistical relationship between the two variables is not one's causal impact, but could be driven by a hidden confounding variable. I consider at least two ways to address this endogeneity issue. One is to use a railway network as an instrument for the influx of immigrants into a county, as proposed by Sequeira, Nunn, and Qian (2017). Another is to use the path dependence of the immigration movement. Immigrants tended to settle down in a place where people from the same country of origin lived. Then, the county-level variation in the number of the immigrants from a country each period can be quasi-random, modeled by local pulling and pushing factors, as pioneered by Burchardi, Chaney, and Hassan (2018). The newly constructed panel data will be a good ground for improving such identification strategy, offering detailed information from the full-count micro census data and various historical reports of local statistics.

Additionally, using the census sample micro data, the effect of the county-level ethnic diversity on local children's school attendance can be also analyzed in the individual level. A school-age child's school attendance is available in the 1% or 5% census sample data. The individual-level school attendance analysis can be useful for understanding the role of ethnic diversity in the demand side, whereas the county level analysis presents its impact on the local supply of public schooling. Literacy is also reported in some years for those older than 20. When linking the ethnic diversity of the birthplace of an adult to the literacy in the adulthood, an indirect test of its effect on human capital accumulation can be performed.

III. Data

The county-level index of ethnic diversity is constructed using complete count census data. Currently, complete count data are digitized for the 1850, 1880, 1900, 1910, 1920, 1930, and 1940

U.S. decennial census. The complete count data contain variables showing each individual's birthplace. Using individuals' or their householders' information, various versions of the county-level ethnic diversity indexes can be constructed.

The county-level data about public education were collected from the U.S. census social statistics schedule for 1850, and state reports of public education for 1880 and later. The number of public schools by level (elementary vs. secondary), school teachers, and school income variables are collected, and divided by the school population.

Additionally, individual-level data are also analyzed. The census sample data contains variables showing the school attendance of the school-age children. Linking county-level variables related to ethnic diversity, and controlling for variables regarding school supply and demand, an individual's school attendance is analyzed.

IV. Preliminary Results

I first present preliminary results using 1850 data. Two ethnic diversity indexes, HHI1 and HHI2, are calculated in the form of the Herfindahl-Hirschman Index. HHI1 is calculated based on the birthplaces defined by the U.S. state or foreign country level, whereas HHI2 is calculated using a more broadly defined regional-level birthplace data. Because there was also significant migration across states in the USA in the 19th century, I also treated all the U.S. states equivalent to a foreign country in calculating HHI1. The regions to calculate HHI2 are defined by US non-South, US South, US other, Other North America, Central and South America, Northern Europe, Western Europe, Southern Europe, Central and eastern Europe, Russian Empire and Baltic states, Other Europe, Asia, Africa, Others.

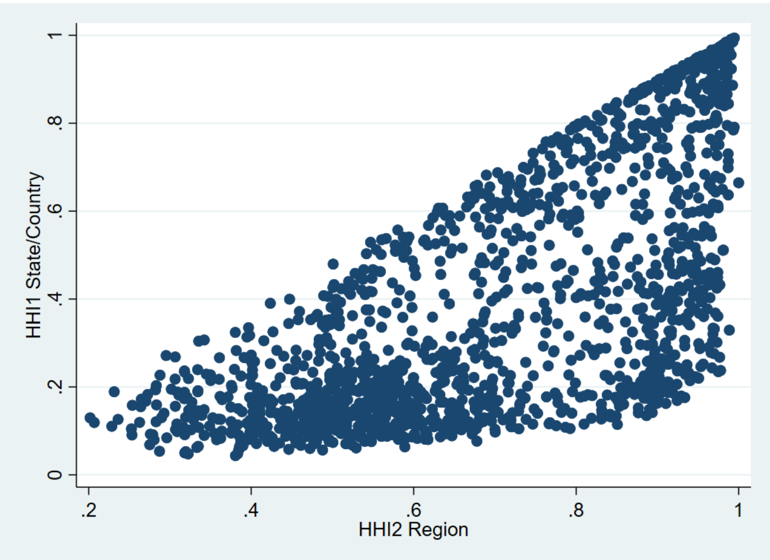

Figure 1. The relationship between HHI1 and HHI2, 1850

The relationship between HHI1 and HHI2 is depicted in Figure 1. The Herfindahl-Hirschman Indexes are the measures of ethnic concentration. Thus, a greater HHI1 or HHI2 represents an ethnically more homogeneous county. Theoretically, HHI2 cannot be greater than HHI1. When the county residents are from countries in different regions, HHI1, the country-level homogeneity, will be exactly same as HHI2, the region-level homogeneity. But if some people of the county are from different countries in a region, HHI1 will be greater than HHI2. Figure 1 shows that there is wide variation between HHI1 and HHI2, implying that they actually measure ethnic diversity in a differential way.

The preliminary county-level regression result is reported in Table 1. The dependent variables are the number of common schools, teachers, school income, school tax revenue, and private schools, all divided by the school population of ages 5-14. The effect of mass immigration and ethnic diversity on the local provision of public education is captured by two groups of key regressors. The first group consists of the Herfindahl-Hirschman Indexes, HHI1 and HHI2. The second is a group of variables showing the percentage of householders in a county, who were born in a specific region. The second group variables capture the effect of an ethnic group on local public schooling. The theoretical background of this ethnicity-specific effect if unclear, and a possible explanation would be a cultural heritage of the ethnic group.

In the regressions, variables to control for the local characteristics that might have affected the local provision of public schooling are included. Such variables are the farm value per school-age child, the percentage of the farm households, the percentage of the female householders, the percentage of the black householders, and a natural logarithm of the population. All the regressions also control for the state fixed effects.

An interesting result is the conflicting effects of HHI1 and HHI2 on the supply of common schools. HHI1, the country-level homogeneity, had a negative effect on the supply of common schools, implying that ethnic diversity measured by the state and country level is good for the local provision of public education. However, HHI2, the region-level ethnic homogeneity showed an opposite effect, which means that the cross-regional ethnic heterogeneity might have hindered the development of public education at the local level. Most "home region" percentages did not show a clear effect, except the negative impact of the U.S South and Canada. The negative Southerners' effect on the rise of common schooling is consistent with the findings of the literature, such as Cubberley (1919) and Go and Lindert (2010). The percentage of the cross-state migrants from the South also had a negative impact on the supply of common school teachers.

The effect of ethnic diversity on public school income and tax revenue were not clear. The common school enrollment rate was negatively associated with HHI1 and the percentage of the Southerners. The percentage of immigrant householders from Western Europe, and also "Russia and Baltic," showed a positive correlation with the enrollment rate. HHI1 was also negatively associated with the provision of private schools. HHI2, the region-level measure of ethnic homogeneity did not show a statistically significant impact on either the supply of private schools or the public school enrollment rate.

Table 1. 1850 county-level regression result

VARIABLES	Common Schools per 5-14 child	Teachers per 5-14 child	School income per 5-14 child	School tax revenue per 5-14 child	Private schools per 5-14 child	Enrollment Rate (Age 5-14)
Farm value per 5-14 child	0.0001	-0.0003	0.4221**	0.4347**	0.0002**	0.0002*
	[0.0002]	[0.0004]	[0.0410]	[0.0293]	[0.0001]	[0.0001]
% Farm	0.0116**	-0.0106**	-0.4301	-0.3311	-0.0011*	-0.0023**
	[0.0015]	[0.0033]	[0.3974]	[0.2725]	[0.0005]	[0.0006]
% Female householder	0.0123	0.0159	8.5966**	0.0231	0.0128**	0.0206**
	[0.0078]	[0.0171]	[2.0511]	[1.4230]	[0.0026]	[0.0033]
% Black householder	-0.0009	0.0428*	-2.6823	0.2892	0.0096**	0.0118**
	[0.0077]	[0.0169]	[2.0300]	[1.4075]	[0.0025]	[0.0032]
ln(Total population)	0.0009**	0.0026**	0.1915**	0.0621	0.0000	0.0003**
	[0.0003]	[0.0006]	[0.0675]	[0.0464]	[0.0001]	[0.0001]
% US householder	-0.0120	-0.0745+	2.6429	-2.2245	0.0066	0.0102
	[0.0175]	[0.0384]	[4.5100]	[3.2021]	[0.0058]	[0.0073]
% US South householder	-0.0142**	-0.0120**	-0.0527	0.0450	-0.0014+	-0.0037**
	[0.0021]	[0.0046]	[0.5594]	[0.3857]	[0.0007]	[0.0009]
% Canadian householder	-0.0381*	-0.0858*	1.4040	-3.9432	0.0032	0.0041
	[0.0192]	[0.0420]	[4.9198]	[3.5070]	[0.0064]	[0.0080]
% Latin American householder	-0.0182	-0.1066**	-2.4701	-3.1346	0.0033	0.0002
	[0.0184]	[0.0404]	[4.8187]	[3.3710]	[0.0061]	[0.0077]
% Northern European	-0.0163	-0.0878*	3.9745	-0.1371	0.0050	0.0084
	[0.0176]	[0.0386]	[4.5383]	[3.2246]	[0.0058]	[0.0074]
% Western European	-0.0280	-0.0953*	1.1048	0.2888	0.0098	0.0155+
	[0.0201]	[0.0442]	[5.1576]	[3.6872]	[0.0067]	[0.0084]
% Southern European	-0.0104	-0.1821+	27.3955*	4.9920	-0.0102	-0.0099
	[0.0474]	[0.1040]	[12.1866]	[8.6769]	[0.0157]	[0.0199]
% Central and Eastern European	-0.0191	-0.0879*	1.7330	-2.4197	0.0055	0.0078
	[0.0175]	[0.0384]	[4.5120]	[3.2067]	[0.0058]	[0.0073]
% Russian and Baltic	-0.5254	0.7347	-30.2755	-25.6109	0.6992**	0.6976**
	[0.5367]	[1.1770]	[140.8373]	[98.2199]	[0.1778]	[0.2241]
% Asian	-0.1144	-0.3904	-50.4413	8.3674	0.0984	0.0722
	[0.3123]	[0.6837]	[80.8088]	[57.0611]	[0.1033]	[0.1315]
% African	0.7045	-2.0777+	-44.5527	-157.1322+	0.1611	-0.1041
	[0.5019]	[1.1007]	[130.8352]	[91.8662]	[0.1663]	[0.2096]
HH index (state level)	-0.0046*	-0.0013	0.1129	0.2181	-0.0025**	-0.0030**
	[0.0021]	[0.0047]	[0.5526]	[0.3893]	[0.0007]	[0.0009]
HH index (regional level)	0.0085**	-0.0099	-0.2460	-0.1839	-0.0005	-0.0015
	[0.0028]	[0.0061]	[0.7384]	[0.5059]	[0.0009]	[0.0012]
Constant	0.0103	0.0658+	-2.8375	2.7547	-0.0040	-0.0090
	[0.0175]	[0.0389]	[4.5604]	[3.2043]	[0.0059]	[0.0073]
Observations	1,508	1,517	1,429	1,516	1,517	1,514
R-squared	0.4641	0.0892	0.3749	0.4135	0.1971	0.2731

Standard errors in brackets. ** p<0.01, * p<0.05, + p<0.1. All models control for the state fixed effects.

Table 2. 1850 individual-level regression result

VARIABLES	Attendance	Attendance	Literacy	Literacy
Age	0.0256**	0.0255**	-0.0007**	-0.0007**
	[0.0007]	[0.0007]	[0.0001]	[0.0001]
Female	-0.0098*	-0.0100*	-0.0506**	-0.0505**
	[0.0041]	[0.0041]	[0.0020]	[0.0020]
Rural	-0.0583**	-0.0461**	-0.0217**	-0.0070*
	[0.0078]	[0.0081]	[0.0030]	[0.0030]
# Children in the household	-0.1129	-0.1153	-0.0032**	-0.0026**
	[0.0729]	[0.0724]	[0.0005]	[0.0005]
Farm value per 5-14 child	0.0319**	0.0285**	0.0282**	0.0220**
	[0.0055]	[0.0057]	[0.0033]	[0.0033]
Common Schools per 5-14 child	3.7234**	3.1861**	0.6165**	0.0573
	[0.3460]	[0.3568]	[0.1706]	[0.1778]
School income per 5-14 child	0.0093**	0.0100**	0.0083**	0.0092**
	[0.0022]	[0.0023]	[0.0013]	[0.0013]
School tax revenue per 5-14 child	-0.0077*	-0.0094**	-0.0111**	-0.0121**
	[0.0033]	[0.0033]	[0.0016]	[0.0016]
HHI1 (state/country)	-0.2674**	-0.3027**	-0.0638**	-0.0934**
	[0.0222]	[0.0233]	[0.0109]	[0.0118]
HHI2 (region)	0.1636**	0.2494**	0.0048	-0.0079
	[0.0260]	[0.0391]	[0.0126]	[0.0187]
US householder	0.1341+	0.1615*	0.0295	0.0074
	[0.0747]	[0.0776]	[0.0271]	[0.0297]
US South	-0.0932**	-0.0604**	-0.0951**	-0.0614**
	[0.0081]	[0.0085]	[0.0050]	[0.0052]
Canadian	0.0093	0.0379	-0.1138**	-0.1201**
	[0.0784]	[0.0812]	[0.0313]	[0.0332]
Latin American	0.1254	0.0763	-0.2195**	-0.2338**
	[0.1068]	[0.1080]	[0.0471]	[0.0492]
Northern European	0.0926	0.1508	-0.0722+	-0.0886*
	[0.1335]	[0.1322]	[0.0378]	[0.0344]
Western European	0.0678	0.1107	0.0437	0.0246
	[0.1357]	[0.1342]	[0.0386]	[0.0352]
Central and Eastern European	0.0749	0.1258	0.0405	0.0134
	[0.1337]	[0.1325]	[0.0377]	[0.0344]
Russian and Baltic	-0.0076	0.0567	0.0672+	0.0633+
	[0.2745]	[0.2733]	[0.0392]	[0.0360]
Asian	-0.6671**	-0.6335**	-0.0826	-0.0849
	[0.0757]	[0.0784]	[0.1173]	[0.1230]
African	-0.1391	-0.1395	-0.0582	-0.0771
	[0.2184]	[0.2287]	[0.1630]	[0.1647]
Constant	0.0474	0.3868+	0.9576**	0.8838**
	[0.0922]	[0.2167]	[0.0327]	[0.0772]
County's share of ethnic groups	No	Yes	No	Yes
State FE	Yes	Yes	Yes	Yes
Observations	49,045	49,045	88,739	88,739
R-squared	0.1493	0.1531	0.0836	0.0923

Standard errors in brackets. ** p<0.01, * p<0.05, + p<0.1

Table 2 shows the regression result using the individual-level sample data. The dependent variables are the indicator variables, the school attendance of a child aged 5-14 and the literacy of an adult older than 20. All the county-level variables that are used in the county-level analysis are included in the regressions. Additional individual-level control variables are age, a female indicator, a rural residence indicator, and the number of children in the household to capture the intra-household resource competition.

The conflicting effects of HHI1 and HHI2 are found again in the individual-level regressions. The coefficient of HHI1 is estimated negative again, consistently implying that the county-level diversity measured by the county-level birthplaces was good for local public schooling, and possibly helped the parents send their children to commo schools. However, the effect of HHI2 on the children's school attendance was still positive. The cross-regional diversity had a negative impact on local public schooling. Adult literacy was also negatively associated with HHI1, although HHI2 was almost barely associated with adult literacy. Additionally, the negative effect of the percentage of the Southerners on children's school attendance was found again in the individual-level regressions.

Table 3. County ethnic homogeneity and the school attendance of children aged 5-14.

	1850	1880	1900	1910	1920	1930
HHI1 (state/country)	-0.0069	-0.0921**	-0.0921**	-0.0636**	-0.0938**	-0.0752**
	[0.0275]	[0.0138]	[0.0123]	[0.0105]	[0.0089]	[0.0085]
HHI2 (region)	-0.0547	0.1093**	0.1056**	0.0560**	0.0838**	0.0725**
	[0.0338]	[0.0147]	[0.0132]	[0.0104]	[0.0093]	[0.0093]
Age	0.0284**	0.0445**	0.0617**	0.0471**	0.0431**	0.0542**
	[0.0013]	[0.0005]	[0.0004]	[0.0003]	[0.0003]	[0.0003]
Female	-0.0101	0.0004	0.0093**	0.0104**	0.0076**	0.0058**
	[0.0074]	[0.0026]	[0.0021]	[0.0016]	[0.0014]	[0.0014]
Rural	-0.0055	-0.0057	-0.0282**	-0.0235**	-0.0361**	-0.0387**
	[0.0137]	[0.0039]	[0.0028]	[0.0019]	[0.0016]	[0.0016]
# Children in the household	-0.0899	-0.3076**	-0.2644**	-0.5047**	-0.1798**	-0.2686**
	[0.1051]	[0.0529]	[0.1020]	[0.1854]	[0.0450]	[0.0456]
Constant	0.0567	-0.1930**	0.0506	0.4026**	0.4941**	0.3379**
	[0.1390]	[0.0621]	[0.0452]	[0.0357]	[0.0249]	[0.0157]
Observations	15,639	122,366	167,196	189,155	219,155	244,435
R-squared	0.1249	0.1878	0.1979	0.1845	0.1654	0.2007

Robust standard errors in brackets. ** p<0.01, * p<0.05, + p<0.1. All models also include state fixed effects. The IPUMS 1% sample data of each year are used.

Table 4. County ethnic homogeneity and the school attendance of children aged 15-18.

	1850	1880	1900	1910	1920	1930
HHI1 (state/country)	-0.0137	-0.1910**	-0.1448**	-0.1221**	-0.0338+	-0.0832**
	[0.0495]	[0.0243]	[0.0214]	[0.0208]	[0.0198]	[0.0177]
HHI2 (region)	-0.0241	0.4488**	0.4702**	0.4170**	0.2647**	0.2281**
	[0.0587]	[0.0256]	[0.0231]	[0.0217]	[0.0221]	[0.0199]
Age	-0.0999**	-0.0987**	-0.1194**	-0.1383**	-0.1594**	-0.1736**
	[0.0057]	[0.0019]	[0.0016]	[0.0015]	[0.0014]	[0.0013]
Female	-0.0990**	-0.0281**	0.0389**	0.0431**	0.0550**	0.0135**
	[0.0128]	[0.0044]	[0.0037]	[0.0034]	[0.0034]	[0.0030]
Rural	0.1701**	0.1383**	0.0614**	0.0503**	0.0103*	-0.0597**
	[0.0208]	[0.0061]	[0.0047]	[0.0043]	[0.0041]	[0.0036]
# Children in the household	-0.0770**	-0.1242**	-0.1527**	-0.0663**	-0.0945**	-0.2470**
	[0.0212]	[0.0063]	[0.0095]	[0.0096]	[0.0096]	[0.0093]
Constant	1.9080**	1.6150**	1.8280**	2.6503**	3.0203**	3.3453
	[0.2469]	[0.1062]	[0.0575]	[0.0658]	[0.0495]	[90.9579]
Observations	5,183	40,001	60,679	73,715	75,892	92,467
R-squared	0.1209	0.1367	0.1280	0.1431	0.1544	0.1913

Robust standard errors in brackets. ** p<0.01, * p<0.05, + p<0.1. All models also include state fixed effects. The IPUMS 1% sample data of each year are used.

The regression results reported in Table 3 and Table 4 reconfirms the previous result using the 1850 samples. Table 3 is the result of the analysis using the sample of the children aged 5-14, which is assumed to be the school ages at the elementary level. The dependent variable is the school attendance indicator of each children. The regressions now have a smaller number of control variables due to the limited availability of the data. Consistent with the previous result, the state and country level homogeneity index was negatively associated with the individual-level school attendance, implying that the positive impact of ethnic diversity on local public education at the elementary level. Table 4, the results using the sample of the children aged 15-18, the high school age group, shows a similar result. Also, the coefficient estimates of the region-level county ethnic homogeneity were mostly positive, same as the previous results. That is, the region-level ethnic diversity was negatively associated with the school attendance, both at the elementary and secondary levels. This was found in all the regression results for 1880, 1900, 1910, 1920, and 1930.

Table 5. Ethnic homogeneity and school attendance, a 1850-1930 pooled sample

	5~14 State FE	5~14 County FE	15-18 State FE	15-18 County FE
HHI1 (state/country)	-0.0488**	0.0595**	-0.1200**	-0.0998**
	[0.0043]	[0.0217]	[0.0081]	[0.0280]
HHI2 (region)	0.0594**	-0.0716**	0.3262**	0.0717+
	[0.0046]	[0.0271]	[0.0089]	[0.0376]
Age	0.0500**	0.0499**	-0.1439**	-0.1438**
	[0.0001]	[0.0006]	[0.0007]	[0.0013]
Female	0.0068**	0.0068**	0.0271**	0.0270**
	[0.0008]	[0.0009]	[0.0016]	[0.0023]
Rural	-0.0260**	-0.0160**	0.0292**	0.0073+
	[0.0009]	[0.0021]	[0.0019]	[0.0039]
# Children in the household	-0.2445**	-0.2397**	-0.1381**	-0.1365**
	[0.0280]	[0.0288]	[0.0041]	[0.0047]
y1850	-0.2226**	-0.2251**	-0.2076**	-0.2103**
	[0.0039]	[0.0102]	[0.0069]	[0.0117]
y1880	-0.2622**	-0.2620**	-0.2245**	-0.2452**
	[0.0015]	[0.0063]	[0.0028]	[0.0091]
y1900	-0.2043**	-0.2059**	-0.2063**	-0.2195**
	[0.0013]	[0.0035]	[0.0024]	[0.0080]
y1910	-0.0037**	-0.0064*	-0.0167**	-0.0303**
	[0.0011]	[0.0025]	[0.0023]	[0.0081]
y1920	0.0359**	0.0346**	-0.0343**	-0.0403**
	[0.0010]	[0.0018]	[0.0023]	[0.0055]
Constant	0.3881**	0.3836**	2.6395**	2.9207**
	[0.0185]	[0.0138]	[0.0378]	[0.0346]
Observations	957,946	957,946	347,937	347,937
R-squared	0.2355	0.1940	0.1669	0.1501
Number of county		3,120		3,109

Robust standard errors in brackets. ** p<0.01, * p<0.05, + p<0.1. "State FE" models use a pooled 1850-1930 census sample data, and control for the state fixed effects. "County FE" models use the same data and control for the state-country fixed effects.

Table 5 shows the regression result using the pooled sample of the 1850-1930 data. The results again reveal the positive association of the state-level ethnic diversity, and the negative association with the region-level ethnic diversity, with the school attendance of the elementary and high-school population.

V. Next Plan

I am currently working on expanding the data set from 1850 to 1930, which may cover the entire period of the mass immigration influx in the USA. This will allow me to utilize the panel structure of the data to better identify the effect of immigration, and ethnic diversity, on the provision of local public schooling. I will also do the robustness checks using instrumental variables proposed by Sequeira, Nunn, and Qian (2017) and Burchardi, Chaney, and Hassan (2018).

References

Abramitzky, Ran and Leah Boustan (2017), "Immigration in American Economic History," *Journal of Economic Literature,* 55(4): 1311-1345.

Ager, Philipp and Markus Brüker (2013), "Cultural Diversity and Economic Growth: Evidence from the US during the Age of Mass Migration," *European Economic Review,* 64(1): 76-97.

Alesina, Alberto, Reza Baqir, and William Easterly (1999), "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics,* 114(4): 1243-1284.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (2003), "Fractionalization," *Journal of Economic Growth,* 8(2): 155-194.

Alesina, Alberto and Paola Giuliano (2015), "Culture and Institutions," *Journal of Economic Literature,* 53(4): 898-944.

Alesina, Alberto, Johann Harnoss, and Hillel Rapoport (2016), "Birthplace Diversity and Economic Prosperity," *Journal of Economic Growth,* 21(2): 101-138.

Alesina, Alberto and Eliana La Ferrara (2005), "Ethnic Diversity and Economic Performance," *Journal of Economic Literature,* 43(3): 762-800.

Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou (2016), "Ethnic Inequality," *Journal of Political Economy,* 124(2): 428-488.

Baldwin, Kate and John D. Huber (2010), "Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision," *American Political Science Review*, 104(4): 644-662.

Banerjee, Abhijit and Rohini Somanathan (2007), "The Political Economy of Public Goods: Some Evidence from India," *Journal of Development Economics,* 82(2): 287-314.

Burchardi, Konrad, Thomas Chaney, and Tarek Hassan (2018), "Migrants, Ancestors, and Foreign Investments," *Review of Economic Studies,* forthcoming.

Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (2006), *Historical Statistics of the United States, Millennial Edition*, Cambridge University Press.

Cubberley, Ellwood P. (1919), *Public Education in the United States: A Study and Interpretation of American Educational History,* Houghton Mifflin Company.

Easterly, William and Ross Levine (1997), "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics,* 112(4): 1203-1250.

Go, Sun and Peter H. Lindert (2010), "The Uneven Rise of American Public Schools to 1850," *Journal of Economic History,* 70(1): 1-26.

Goldin, Claudia and Lawrence H. Katz (2010), *The Race between Education and Technology,* Harvard University Press.

Lindert, Peter H. (2004), *Growing Public: Social Spending and Economic Growth since the Eighteenth Century*, Volume I and II, Cambridge University Press.

Lindert, Peter H. (2014), "Private Welfare and the Welfare State," in Larry Neal and Jeffrey G. Williamson eds., *The Cambridge History of Capitalism,* Volume II, pp. 464-500.

Mauro, Paolo (1995), "Corruption and Growth," *Quarterly Journal of Economics,* 110(3): 681-712.

Michalopoulos, Stelios (2012), "The Origins of Ethnolinguistic Diversity," *American Economic Review,* 102(4): 1508-1539.

Michalopoulos, Stelios and Elias Papaioannou (2013), "Pre-colonial Ethnic Institutions and Contemporary African Development," *Econometrica,* 81(1): 113-152.

Mirel, Jeffrey E. (2010), *Patriotic Pluralism: Americanization Education and European Immigrants,* Harvard University Press.

Montalvo, Jose G. and Marta Reynal-Querol (2005), "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review,* 95(3): 796-816.

Reynal-Querol, Marta (2002), "Ethnicity, Political Systems, and Civil Wars," *Journal of Conflict Resolution,* 46(1): 29-54.

Sequeira, Sandra, Nathan Nunn, and Nancy Qian (2017), "Migrants and the Making of America: The Short- and Long-Run Effects of Immigration during the Age of mass Migration," NBER Working Paper No. 23289.