Immigrants' Genes: Genetic Diversity and Economic Development in the US

Ager, Philipp and Brückner, Markus

University of Southern Denmark, National University of Singapore

4 December 2013

Immigrants' Genes:

Genetic Diversity and Economic Development in the US

Philipp Ager and Markus Brückner*

December 2013

Abstract

We examine the effect of genetic diversity on economic development in the United States. Our estimation strategy exploits that immigrants from different countries of origin differed in their genetic diversity and that these immigrants settled in different regions. Based on a sample of over 2250 counties, we find that increases in genetic diversity of US counties that arose due to immigration during the 19th century had a significant positive effect on US counties' economic development. We also detect a significant positive long-run effect of 19th century immigrants' genetic diversity on contemporaneous measures of income.

Keywords: Economic Growth, Genetic Diversity, Immigration, Melting Pot.

JEL: J11, O51, Z13

*Ager: University of Southern Denmark, Department of Business and Economics, Campusvej 55, 5230 Odense M, Denmark; Brückner: National University of Singapore, AS2 Level 6, 1 Arts Link, Singapore 117570. Contact e-mail: phag@sam.sdu.dk; ecsbm@nus.edu.sg. We are grateful to Casper Worm Hansen and Lars Lønstrup for helpful comments and suggestions. All remaining errors are our own.

I. Introduction

Genetic diversity has been attributed an important role to explain why there still exist persistent disparities in economic development across the globe (Spolaore and Wacziarg, 2013). In particular, genetic distance between countries has been shown to be associated with significantly lower levels of trust as well as lower levels of bilateral trade (Guiso et al., 2009). It has also been argued that genetic distance to the world technology frontier acts as a barrier to the diffusion of technology (Spolaore and Wacziarg, 2009). Ashraf and Galor (2013a) carefully address the endogeneity of genetic diversity to countries' economic prosperity by exploiting geographic distance to East Africa as an instrument for genetic diversity. Their estimates suggest that genetic diversity has a significant causal effect on countries' long-run economic development.

In this paper we supplement the cross-country literature by examining the link between genetic diversity and economic development at the subnational level. Our analysis exploits that, historically, the United States was a melting pot of immigrants. People from different countries of origin and thus with different genes migrated to the US. During the first wave of mass migration (1850-1920) the bulk of immigrants – about 30 million people – came mainly from Europe. The British Isles, Germany, Scandinavia, the Austro-Hungarian Empire, Russia, Italy and Greece constituted the most important emigration areas throughout this period (e.g. Hatton and Williamson, 1998, 2006; O'Rourke and Williamson, 1999). These immigrants tended to settle in regions based on network effects (see e.g. Hvidt, 1975; Wegge, 1998; Ager and Brückner, 2013). As a consequence, immigrants significantly changed the demographic composition of regions in the US, and with it their genetic diversity.

Based on a sample of over 2250 counties, our main finding is that genetic diversity of US counties during the period 1870-1920 was conducive to economic development. Controlling for GDP per capita in 1870, a one standard deviation increase in genetic diversity in 1870 was associated with an increase in GDP per capita growth during the 1870-1920 period of around 20 percent (equivalent to around 0.2 standard deviations). The significant positive effect of genetic diversity was present in both relatively rich and poor counties. We also document that our finding of a significant positive effect of genetic diversity on US counties' economic development is robust to using various other measures of prosperity, such as urbanization, wages, and manufacturing value added. Furthermore, we show that our finding is robust to using 10-year non-overlapping panel data and controlling for county fixed effects which, in turn, allows us to identify the impact of genetic diversity on economic growth from the within-county variation of the data.

It is also remarkable that our analysis suggests a persistent effect of genetic diversity on economic development. There is a positive and statistically significant association between immigrants' genetic diversity in 1870 and US counties' income per capita in 2010. Quantitatively, the estimated effects suggest that a one standard deviation increase in genetic diversity in 1870 was associated with a 10 percent higher income per capita today. Hence, US counties' historical genetic diversity during the 19th century has a positive and significant effect on contemporaneous income per capita. We also document that there exists a significant positive effect on contemporaneous income per capita when using the much sparser data on immigrants' genetic diversity in 1790.

Ashraf and Galor (2013a) argue that intermediate levels of genetic diversity are the most beneficial for economic development, i.e. there is an optimal level of genetic diversity.[1] It is important to note that immigrants flocking into the US during the 19th century were predominantly from European countries (e.g. Hatton and Williamson, 1994, 1998). In other words, immigrants that settled in the US during the 19th century brought with them intermediate levels of genetic diversity. Our finding of a significant positive average effect of genetic diversity in the 19th century US on subsequent economic growth is hence not add odds with the findings in Ashraf and Galor that there exists an optimal level of genetic diversity beyond which further increases in genetic diversity are detrimental for growth.

Our paper also contributes to the recent literature on the role of historical European settlements across the globe for long-run economic development. Putterman and Weil (2010), Easterly and Levine (2012) and Spolaore and Wacziarg (2013) have demonstrated that countries with a larger population of European ancestry display higher levels of economic development today. Importantly, as noted in Spolaore and Wacziarg (2013), the above findings suggest that early historical development matters, but the mechanism is through the intergenerational transmission rather than through fixed geographic factors. Our findings agree with that position: genetic diversity is one of the various components European ancestors carried with them when they settled in the US. We complement the above studies by showing that the genetic diversity of European immigrants as a part of European ancestry fostered economic development in the US. One advantage of our empirical analysis compared to previous studies is that we can tell the positive effect of European immi-

[1] According to these authors, at low levels of genetic diversity marginal increases in genetic diversity have substantial (positive) effects on the production possibility frontier that outweigh the negative effects of diversity arising from disarray and mistrust. On the other hand, when genetic diversity is already high, increasing diversity has negative effects for economic development as the costs from social disarray and mistrust outweigh the benefits of greater diversity that enables to expand the production possibility frontier.

grants' genetic diversity on economic development apart from the aggregated effect of European ancestry by directly controlling for (foreign-born) nationality shares.

Relative to existing cross-country regressions, our subnational estimation approach has several advantages for identifying causal effects of genetic diversity on economic development. In cross-country regressions a concern is that historical factors, related, for example, to colonization, affect both countries' genetic diversity and their economic development directly. The county level data enables us to circumvent this bias by controlling for US state fixed effects. These fixed effects absorb, in addition to nation-wide historical factors, time invariant factors, such as legal origin, that may vary at the US state level. Another advantage of our data is that the panel structure allows us to deal with time-invariant omitted factors like geography by controlling for county-specific fixed effects. This enables us to examine whether a significant effect of changes in genetic diversity on GDP per capita growth holds also in the time-series, i.e. whether GDP per capita growth changes following a preceding change in genetic diversity. To further address potential endogeneity concerns of immigrants' genetic diversity to their country of origin's economic prosperity we use Ashraf and Galor's (2013a,b) measure of countries' genetic diversity that is based on the geographic distance to East Africa.

Our findings demonstrate that a significant effect of genetic diversity on economic growth does not imply that economic development outcomes are deterministic. Episodes of mass immigration, as experienced in the United States during the 19th century, can significantly change the genetic diversity of countries; and by doing so, they affect countries' development path.

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 explains our baseline estimation strategy. Section 4 presents the main results. Section 5 presents robustness checks. Section 6 concludes.

II. Data

This section discusses the data used to examine the impact of genetic diversity on GDP per capita. Our main data source is the Inter-University Consortium for Political and Social Research (ICPSR) 2896 data file (Haines, 2010).[2] The ICPSR 2896 file contains detailed decennial US county and state level data on

[2]More information about the data set, such as, scope of study, data collection and data source can be found at http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/02896.

demographic, economic, and social variables, which were collected by the US Bureau of the Census. One key advantage of the ICPSR 2896 file is that it enables us to exploit the underlying cultural heterogeneity in the United States at the county level. In particular, the ICPSR 2896 file comprises – from 1870 onwards – detailed information about the country of origin of foreign-born, which is necessary to calculate the genetic diversity index described below.

We construct a time-varying, US county-specific index of immigrants' genetic diversity as:

$$GeneticDiversity_{cs,t} = \sum_{n=1}^{N} Immigrants_{ncst} \times pdiv_n \tag{1}$$

where $Immigrants_{ncst}$ is the stock of immigrants of nationality (i.e. country of origin) n in county c of US state s in period t and $pdiv_n$ is Ashraf and Galor's (2013a) migratory-distance based predicted genetic diversity of nationality n.[3]

We use Ashraf and Galor's (2013a) genetic diversity values that are predicted by the migratory distance from East Africa to circumvent endogeneity concerns between observed genetic diversity and economic development. Since Ashraf and Galor's (2013) genetic diversity values represent the genetic diversity of the country of origin's population, it is important that we have sufficiently large random samples of people drawn out of their country of origin's population at hand, such that an immigrant group's genetic diversity and country of origin's genetic diversity are alike (law of large numbers). Appendix Table 2 lists the total number of foreign born by country of origin in our sample. For the majority of nationalities the numbers are in the thousands. Thus, it appears reasonable to assume that the genetic diversity of an immigrant group is alike to the genetic diversity prevailing in the country of origin.

III. Estimation Methodology

Following the empirical growth literature (see e.g. Durlauf et al., 2005), our baseline estimating equation takes the following functional form:

[3] A description of all other variables used in the empirical analysis is available in the Data Appendix. Appendix Table 1 shows summary statistics.

$$[ln(GDP)_{cs} - ln(GDP)_{cs,t_0}] = \varphi ln(GDP)_{cs,t_0} + \theta GeneticDiversity_{cs,t_0} + \Gamma Immigration_{cs,t_0} + \epsilon_{cs} \quad (2)$$

where $[ln(GDP)_{cs} - ln(GDP)_{cs,t_0}]$ is the change in the natural logarithm of GDP per capita between 1870 and 1920 in county c of US state s; $ln(GDP)_{cs,t_0}$ is the natural logarithm of county c's initial level of GDP per capita in 1870; $GeneticDiversity_{cs,t_0}$ is a measure of county c's genetic diversity that is due to the stock of immigrants (see Section 2) in 1870. $Immigration_{cs,t_0}$ is a set of control variables that includes county c's nationality shares by country of origin, the share of African-Americans, the share of Native Americans as well as county c's overall population size and area in 1870. The error term ϵ_{cs} is clustered at the state level in order to correct for arbitrary spatial correlation across counties within US states.

A key feature of neoclassical growth models is convergence. Convergence requires that $|\lambda| \equiv |\varphi + 1| < 1$. In that case, the above model specification implies that a change in genetic diversity has an effect on transitional GDP per capita growth; there is no effect on the long-run GDP per capita growth rate. The long-run effect on the level of GDP per capita is $\theta/(-\varphi)$. This can be directly seen by noting that equation (2) is equivalent to:

$$ln(GDP)_{cs} = \lambda ln(GDP)_{cs,t_0} + \theta GeneticDiversity_{cs,t_0} + \Gamma Immigration_{cs,t_0} + \epsilon_{cs} \quad (3)$$

The control for immigration in equation (3) implies that θ captures the effect of differences in genetic diversity among immigrant groups, and not the effect that immigration had on economic growth, which is captured by Γ. Given recent work by Ashraf and Galor (2013a,b) that has questioned the exogeneity of genetic diversity we use values of genetic diversity that are based on distance from Addis Adaba to construct a measure of US counties' predicted genetic diversity that arises from immigration (see Section 2 for details on how this measure is constructed).

We furthermore note that our estimating equation uses initial values of immigration. Current immigration could be a function of counties' contemporaneous income (for example, richer counties may have a greater demand for a diverse workforce). This, in turn, would imply that least squares regressions that enter current immigration in the estimating equation suffer from endogeneity bias. However, this endogeneity bias is circumvented by using initial immigration and controlling for initial GDP per capita as we do in equation (2). The reason is that by controlling for initial GDP per capita, we clean the error term of any changes in counties' GDP per capita that are due to initial GDP per capita.

There are several reasons why we use in our baseline estimating equation 1870 as the initial year. Ideally, we would like to go back as far as possible in history to study the long-run (persistent) effect that genetic diversity has on economic development. Unfortunately, this is not possible due to data limitations. The year 1870 is chosen because it provides us with a relatively large sample of observations (around 2250) for GDP per capita growth between the 1870-1920 period (i.e. half a century) and avoids the inclusion of the 1861-1865 American Civil War. As a robustness check we also present the results using 1850 and 1860 as initial years.[4]

We stop in 1920 for our baseline sample, on which we will also conduct a ten-year non-overlapping panel analysis during 1870-1920, as this was the year when the period of free immigration ended.[5] The second half of the 19th century was a period of mass immigration (see Appendix Figure 1), characterized by an inflow of predominantly European immigrants. As a robustness check, we will present estimates based on data that: (i) cover longer time-spans of economic growth (from 1870 up to 2010); (ii) using an index of immigrants' genetic diversity in 1920 to examine the effect on post-1920 growth; (iii) and using (the very limited) cross-county data on foreign-born that exist for 1790 to examine the effect of immigrants' genetic diversity on income over a period covering more than two centuries.

IV. Main Results

Table 1 presents our baseline estimates of the effect that immigrants' genetic diversity had on US counties' GDP per capita growth during the 1870-1920 period. Column (1) shows least squares estimates where the right-hand-side variables in the estimating equation are US counties' 1870 GDP per capita and 1870 genetic diversity only. In columns (2)-(4) we add as control variables counties' population size and area, state fixed effects, the share of Native and African-Americans and nationality shares by country of origin.[6] The point

[4]The US Bureau of the Census included in 1850 for the first time questions about individuals' birthplace by country of origin. Compared to the 1870 Census, digitized records for the years 1850 and 1860 are only available as a 1 percent random sample from the Integrated Public Use Microdata Series (IPUMS), see Ruggles et al. (2010).

[5]As Goldin (1994, page 223) notes: "With the passage of the Emergency Quota in May 1921 the era of open immigration to the United States came to an abrupt end."

[6]The full set of estimates on the population and immigrant control variables is available in Appendix Table 3.

estimates on genetic diversity are positive and significantly different from zero at the 1 percent level in all regressions; quantitatively, they range between 0.08 to 0.11. Figure 1 illustrates the estimated positive relationship between immigrants' genetic diversity and economic growth graphically.[7]

The estimated effect of immigrants' genetic diversity on economic growth is not only statistically significant, it is also quantitatively sizable. For example, the coefficient of 0.09 in column (4) suggests that, on average, a one unit increase in genetic diversity increased US counties' GDP per capita during the 1870-1920 period by around 10 percent. An alternative interpretation is that a one standard deviation increase in immigrants' genetic diversity increased GDP per capita growth during the 1870-1920 period by around 21 percent (equivalent to around 0.25 standard deviations; or alternatively, 0.4 percent per annum).

The baseline estimates show evidence of significant convergence in counties' GDP per capita during the 1870-1920 period. For example, in column (4) the AR coefficient is around 0.13 and has a standard error of around 0.05. The implied per annum convergence rate is approximately 4 percent; and the implied half-life is around 17 years. Hence, as predicted by neoclassical growth models, initially poorer counties experienced subsequently higher GDP per capita growth. Figure 2 illustrates this cross-county convergence in GDP per capita graphically.[8]

A fundamental question in the economic growth and development literature is how persistent are the effects of major economic shocks on economic development over time (see e.g. Nunn, 2013). Genetic traits are transmitted from one generation to the next (see e.g. Spolaore and Wacziarg, 2013). It is therefore interesting to examine whether the genetic diversity of immigrants brought to the US during the age of mass migration still matters for current economic development.

Table 2 documents that there is a persistent effect of 1870's genetic diversity on contemporaneous measures of income per capita. In column (1) of Table 2 the dependent variable is income per capita in 2010. The estimated coefficient on 1870 genetic diversity is around 0.04 and significantly different from zero at the 1 percent level. In the subsequent columns (2)-(5) the dependent variable is income per capita in 2000, 1990, 1980 and 1970, respectively. The estimated coefficients on genetic diversity in 1870 are in that case around 0.05 to 0.06; and all of these are significantly different from zero at the 1 percent level. In Appendix Table 4 we show that similar results are obtained if we use the 1920 value of genetic diversity instead of 1870.

[7] Appendix Figure 2 illustrates the estimated positive relationship between immigrants' genetic diversity and economic growth at the state level graphically

[8] Appendix Figure 3 illustrates the cross-state convergence in GDP per capita graphically

Table 3 shows that there exists also a significant positive effect on contemporaneous measures of economic development when using immigrants' genetic diversity as far back as two centuries. We construct a measure of immigrants' genetic diversity in 1790, using equation (1). This is the farthest possible year that we can go back in time given the availability of US county data.[9] Unfortunately, no data exist on counties' GDP per capita in 1790. We therefore use urbanization in 1790 as a proxy for counties' initial economic development.

In columns (1) of Table 3 the dependent variable is income per capita in 2000. The estimated coefficient on 1790 genetic diversity is around 3.1 in column (1) and statistically significant at the 1 percent level. Quantitatively, the estimated coefficient on immigrants' 1790 genetic diversity suggests that a one standard deviation higher genetic diversity in 1790 was associated with a higher value of income per capita in the year 2000 of around 0.3 standard deviations. Hence, immigrants' genetic diversity has a significant effect on economic development over a period as long as two centuries. The remaining columns of Table 3 show that there exists also a significant positive effect of immigrants' 1790 genetic diversity if the dependent variable is income per capita in 1970 or GDP per capita in 1940, 1910, 1880, and 1850, respectively.

V. Robustness

In this section we discuss the sensitivity of our baseline results to various robustness checks, such as removing observations that could be deemed as outliers; using alternative measures of counties' economic development; differences between rich and poor counties; non-linearities in the effect of genetic diversity; panel data estimation; alternative measurements of immigrants' genetic diversity; and alternative time periods.

A. Excluding Outliers

Table 4 documents that our findings are robust to excluding outliers. We use the Hadi (1992) method for detecting observations that could be deemed as outliers. The Hadi method is a multivariate procedure based on the Mahalanobis distance. This method for detecting outliers has been commonly used in the empirical growth literature (see, for example, Easterly et al., 2004). Following the literature, we impose a cut-off value of 5 percent for detecting outliers. For the sample that excludes these observations, the estimated coefficients

[9] The aggregated county statistics of the US Census in 1790 report the number of persons by nationality (English and Welsh, Scots, Irish, Dutch, French, German, Hebrew and other nationalities); see the ICPSR 2896 file for further information. We use the information on nationality shares to construct the genetic diversity index in 1790.

on genetic diversity continue to be positive and significant at the 1 percent level. Quantitatively, they are around 0.08 with a standard error of around 0.02. Moreover, removing outliers continues to yield significant cross-county convergence in GDP per capita.

B. Alternative Measures of Economic Development

There is also evidence of immigrants' genetic diversity being beneficial for economic development when considering alternative measures of US counties' prosperity. In Table 5 we report estimates where the dependent variable is the log of the urbanization rate. In the economic history literature, urbanization rates are commonly used as a proxy for economic development (see e.g. Acemoglu et al., 2002). The estimated coefficients on genetic diversity are all positive and significantly different from zero. Again, there is evidence for convergence: less urbanized US counties in 1870 experienced significantly higher urbanization growth rates during the 1870-1920 period. Quantitatively, the estimated coefficient on the genetic diversity index suggests that a one standard deviation increase in immigrants' genetic diversity was associated with a 7 percent increase in urbanization growth during the 1870-1920 period (equivalent to around 0.47 standard deviations).

Column (1) of Table 6 shows that there exists also a significant positive effect of immigrants' genetic diversity when the output measure is limited to manufacturing value added. The estimated coefficient of 0.03 suggests that a one standard deviation increase in immigrants' genetic diversity was associated with a 0.1 standard deviation increase of the 1870-1920 manufacturing value added growth rate. Columns (2)-(4) of Table 6 show that immigrants' genetic diversity in 1870 had also a significant positive effect on the 1870-1920 growth rates of the manufacturing share, wages in the manufacturing sector as well as the number of manufacturing establishments per capita.

C. Rich vs Poor Counties

Genetic diversity had a significant positive effect on 1870-1920 per capita GDP growth in both rich and relatively poor US counties. This is shown in Table 7. In columns (1)-(4) we report estimates for the sample of US counties with below median GDP per capita in 1870; in columns (5)-(8) we report estimates for US counties with above median 1870 GDP per capita. The estimated coefficient on genetic diversity is around 0.05 for the former with a standard error of around 0.025; for the latter the coefficient is around 0.11 with a standard error of around 0.04. Hence, genetic diversity was beneficial for economic growth in both rich and poor counties, although relatively richer counties benefited more. In terms of convergence, there is evidence of poor counties growing faster than relatively rich counties in both sub-samples. For example, according to column (4) the per annum convergence rate is around 4 percent in the group of counties with below median 1870 GDP per capita; whereas according to column (8) the per annum convergence rate is around 3 percent in the group of counties with above median 1870 GDP per capita.

D. Non-Linear Effects of Genetic Diversity

Table 8 shows that there is no evidence of a significant squared effect of genetic diversity on economic growth. On the other hand, the linear effect of genetic diversity continues to be positive and significantly different from zero at the 1 percent level. Ashraf and Galor (2013a) documented for a world sample of countries that intermediate levels of genetic diversity associated with European and Asian population have a significant positive effect on economic development; however, extremely high degrees of genetic diversity found in African populations have been detrimental for these countries' economic development. It is important to note that in 1870 immigrants to US counties were almost exclusively from Europe (e.g. Hatton and Williamson, 1998). Hence, the absence of a significant squared effect in our sample is consistent with Ashraf and Galor's finding that the marginal effect of genetic diversity is positive except for extremely high values of diversity, which are absent in our sample.

E. Panel Data Estimation

There exists also a significant positive effect of genetic diversity on GDP per capita if we use 10-year non-overlapping panel data. The relevant results are reported in Table 9. The panel covers the 1870-1920 period; it is balanced with 2261 cross-section units (counties) and 5 time periods. Columns (1)-(4) of Table 9 are structured in exactly the same way as the previous tables. The estimated coefficients on the genetic diversity variable range between 0.04 to 0.06.

An advantage of the panel approach is that it allows us to control for county fixed effects. The results from panel regressions with county fixed effects are shown in columns (5) and (6) of Table 9. Regardless of whether we use least squares or system-GMM estimation, the estimated coefficient on genetic diversity is positive and significantly different from zero at the 1 percent level. Quantitatively, the coefficient on the genetic diversity variable is 0.05 in column (5) and 0.07 in column (6).

F. Alternative Measurements of Genetic Diversity

Panel A of Table 10 shows that estimates change very little if we limit our measure of immigrants' genetic diversity to European immigrants. The coefficient on the measure of immigrants' genetic diversity that captures the differential impact on 1870-1920 economic growth of genetic diversity brought along by European immigrants is around 0.08. The standard error associated with this estimate is around 0.03 so that we can reject the hypothesis that the estimated coefficient is equal to zero at the 1 percent significance level. As the bulk of immigrants during the 19th century were from Europe, the similarity in coefficients reported in Table 10 and our baseline estimates reported in Table 1 is not surprising.

In Panel B of Table 10 we present estimates from regressions that use an alternative index of immigrants' genetic diversity based on the binary criteria of whether an immigrant group was represented in a county. The index is constructed as:

$$GD\ Indicator_{cs,t} = \sum_{n=1}^{N} D_{ncst} \times pdiv_n \tag{4}$$

where D_{ncst} is an indicator variable for the presence of immigrants of nationality (i.e. country of origin) n in county c of US state s in period t; and $pdiv_n$ is Ashraf and Galor's (2013a) migratory-distance based predicted genetic diversity of nationality n. The estimated coefficients on this alternative index of immigrants' genetic diversity are also positive and significantly different for all four specifications. Quantitatively, the estimated coefficient of 1.0 in column (4) of Panel B in Table 10 suggests that a one standard deviation increase in this index was associated with an increase in GDP per capita growth during the 1870-1920 period of around 9 percentage points (0.1 standard deviations).

G. Using Initial Years before the American Civil War

Table 11 presents estimates of the effect that immigrants' genetic diversity had on US counties' GDP per capita growth, using instead of 1870 as initial year the years 1850 and 1860. Compared to the 1870 Census, digitized records for the years 1850 and 1860 are only available as a 1-percent random sample from the

Integrated Public Use Microdata Series (IPUMS). Columns (1)-(4) of Table 11 show the results for the 1850-1920 period and in columns (5)-(8) of the same table we report the results for the 1860-1920 period. For the 1860-1920 period the point estimates on the effect of immigrants' 1860 genetic diversity on 1860-1920 GDP per capita growth are positive and statistically significant at least at the 5 percent significance level. Compared to the 1870-1920 period, we lose more than 25 percent of the observations when we use 1850 as initial year. Still we obtain a positive and statistically significant effect of immigrants' 1850 genetic diversity on 1850-1920 GDP per capita growth in two out of four specifications.

VI. Conclusion

During the age of mass migration the US experienced a historically unprecedented inflow of immigrants. Whereas a large literature exists examining the effects of this immigration inflow on economic outcomes in the US (e.g. Ferrie, 1997; Hatton and Williamson, 1998; Abramitzky et al., 2012), no study has examined yet the differential effects of immigrants' genetic diversity. This paper examined these effects by exploiting that immigrants from different countries of origin, and thus with different genetic diversities, settled in different regions in the US.

Based on a sample of over 2250 counties, our estimates showed that counties' with initially higher levels of immigrants' genetic diversity experienced subsequently faster (transitional) GDP per capita growth: a one standard deviation increase in 1870 immigrants' genetic diversity increased 1870-1920 GDP per capita growth by around 0.3 standard deviations. This significant positive effect was present both in the cross-section of counties and at the within-county level. We also found that immigrants' genetic diversity in the 19th century had a significant positive effect on contemporaneous measures of income per capita.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson, "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 2012, *102* (5), 1832–56.

Acemoglu, Daron, Simon Johnson, and James A. Robinson, "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics*, 2002, *117* (4), 1231–94.

Ager, Philipp and Markus Brückner, "Cultural diversity and economic growth: Evidence from the US during the age of mass migration," *European Economic Review*, November 2013, *64*, 76–97.

Ashraf, Quamrul and Oded Galor, "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 2013a, *103* (1), 1–46.

_ and _ , "Genetic Diversity and the Origins of Cultural Fragmentation," *American Economic Review, Papers & Proceedings*, 2013b, *103* (3), 528–33.

Durlauf, Steven N., Paul A. Johnson, and Jonathan R.W. Temple, "Growth Econometrics," in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1A, Amsterdam: North-Holland, 2005, chapter 8, pp. 555–677.

Easterly, William and Ross Levine, "The European Origins of Economic Development," NBER Working Paper 18162. 2012.

_ , _ , and David Roodman, "Aid, Policies, and Growth: Comment," *American Economic Review*, 2004, *94* (3), 774–780.

Ferrie, Joseph P., "The Entry of Antebellum European Immigrants Into the U.S. Labor Market, 1840-60.," *Explorations in Economic History*, 1997, *34* (3), 295–330.

Goldin, Claudia, "The political economy of immigration restrictions in the United States, 1890 to 1921," in Claudia Golding and Gary Libecap, eds., *The Regulated Economy. An Historical Approach to Political Economy*, University of Chicago Press, 1994, pp. 223–258.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Cultural Biases in Economic Exchange?," *Quarterly Journal of Economics*, 2009, *124* (3), 1095–1131.

Hadi, Ali S., "Identifying Multiple Outliers in Multivariate Data," *Journal of the Royal Statistical Society*, 1992, *54* (761-771).

Haines, Michael R., "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002. ICPSR02896-v3.," Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21. doi:10.3886/ICPSR02896.v3 2010.

Hatton, Timothy J. and Jeffrey G. Williamson, "What Drove the Mass Migrations from Europe in the Late Nineteenth Century?," *Population and Development Review*, 1994, *20* (3), 533–559.

_ **and** _ , *The Age of Mass Migration: Causes and Economic Impact*, New York: Oxford University Press, 1998.

_ **and** _ , *International Migration in the Long Run: Positive Selection, Negative Selection, and Policy*, New York: Springer, 2006.

Hvidt, Kristian, *Flight to America: The Social Background of 300,000 Danish Immigrants*, New York: Academic Press, 1975.

Nunn, Nathan, "Historical Development," in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth (forthcoming)*, Vol. 2, Amsterdam: North-Holland, 2013.

O'Rourke, Kevin H. and Jeffrey G. Williamson, *Globalization and History: The Evolution of a Nineteenth-Century Atlantic Economy*, Cambridge, MA: MIT Press, 1999.

Putterman, Louis and David N. Weil, "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, 2010, *125* (4), 1627–82.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Roland Goeken, Matthew B. Schroeder, and Matthew Sobek., "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]," Minneapolis: University of Minnesota. 2010.

Spolaore, Enrico and Roman Wacziarg, "The Diffusion of Development," *Quarterly Journal of Economics*, 2009, *124* (2), 469–529.

_ **and** _ , "How Deep Are the Roots of Economic Development?," *Journal of Economic Literature*, 2013, *51* (2), 325–369.

Wegge, Simone A., "Chain Migration and Information Networks: Evidence From Nineteenth-Century Hesse-Cassel," *Journal of Economic History*, 1998, *58* (4), 957–986.

Appendix: Data Appendix

Dependent Variables

VARIABLE	YEARS	DESCRIPTION
Output growth	1870 - 1920	Total output in per capita terms is formed as the sum of manufacturing value added and agricultural value added (Source: ICPSR 2896 file). The growth variable is calculated as the change in logarithmic units.
Manufacturing Value Added	1870 - 1920	We use manufacturing value added in per capita terms. We calculate manufacturing value added as the difference between manufacturing output and the cost of materials used in manufacturing (Source: ICPSR 2896 file). We had to impute the manufacturing data for the year 1910 because no manufacturing data were reported in the 1910 Census at the county level. See the ICPSR 2896 codebook for more details.
Agricultural Value Added	1870 - 1920	Agricultural value added is in per capita terms. We calculate agricultural value added as the difference between agricultural output and the cost of inputs used in agriculture (Source: ICPSR 2896 file). We use the variable *farmout*, which contains the estimated value of farm products, as measure for agricultural output for the years 1870 - 1900. For 1910 - 1920 we use as agricultural output the sum of values of crops, value of dairy products, value of chickens and eggs produced, value of animals slaughtered (only available 1910), value of honey and wax produced and the value of wool produced. As a proxy of the input costs in agriculture, we use expenditure for fertilizer (available 1880 - 1920) and for feed (available 1910 - 1920). See the ICPSR 2896 codebook for more details.
Income per capita	1970 - 2010	County level per capita income is from the Bureau of Economic Analysis (Regional Data); http://www.bea.gov/index.htm.
Urban growth	1870 - 1920	Urban growth is calculated as the change in logarithmic units. The Census declared a county population as urban, if at least 2500 inhabitants lived in urban places (Source: ICPSR 2896 file).

Dependent Variables (CONTINUED)		
Manufacturing Value Added Growth	1870 - 1920	The growth rate of manufacturing value added per worker in the manufacturing sector is calculated as the change in logarithmic units (Source: ICPSR 2896 file). For further information see the variable description of manufacturing value added above.
Growth of Manufacturing Share	1870 - 1920	We use the occupation classification variable *OCC* from the IPUMS to construct the manufacturing share based on the US labor force. Individuals with occupation code 130-266 (1870) and 70-598 (1920) are classified as manufacturing workers. Individuals with occupation code 1-12 (1870) and 1-69 (1920) are classified as agricultural workers. The manufacturing share is the fraction of manufacturing workers out of manufacturing and agricultural workers and the growth rate is calculated as the change in logarithmic units.
Wage Growth	1870 - 1920	The growth rate of manufacturing wages per worker in the manufacturing sector is calculated as the change in logarithmic units (Source: ICPSR 2896 file).
Growth of Manufacturing Establishments	1870 - 1920	The growth rate of manufacturing establishments per capita is calculated as the change in logarithmic units (Source: ICPSR 2896 file).

ADDITIONAL CONTROLS

VARIABLE	YEARS	DESCRIPTON
Population	1870	Total population in US counties in 1870 (Source: ICPSR 2896 file).
Share of Native Americans	1870	Share of Native Americans (declared by the historical US Census as indian population) out of the total population in 1870 (Source: ICPSR 2896 file).
Share of African-Americans	1870	Share of African-Americans (declared by the historical US Census as negro population) out of the total population in 1870 (Source: ICPSR 2896 file).
Share of Foreign-Borns	1870	Share of foreign borns out of the total population in 1870 (Source: ICPSR 2896 file).
County Area	1870	County Area in square miles measured in 1880 (Source: ICPSR 2896 file).

Figures and Tables

Figure 1:
Ln Output Growth and Genetic Diversity

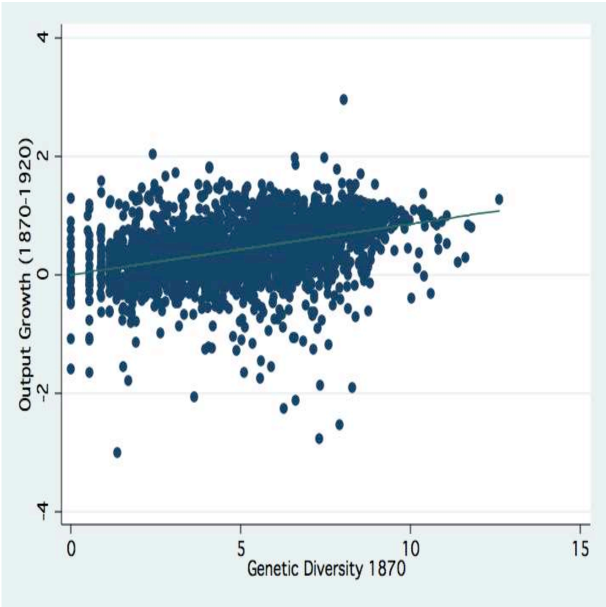


Figure 2:
Ln Output Growth and Initial Ln Output per capita

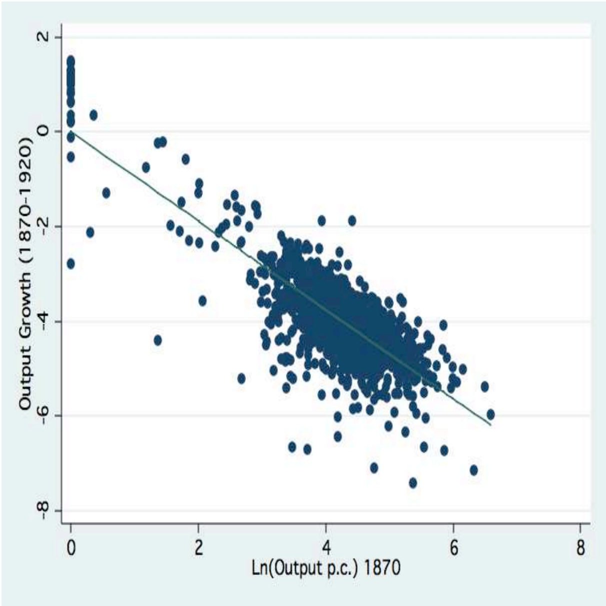


TABLE 1
THE EFFECT OF GENETIC DIVERSITY ON OUTPUT GROWTH

Dependent Variable: Δln(Output p.c.)				
	(1)	(2)	(3)	(4)
Genetic Diversity 1870	0.0854***	0.114***	0.0768***	0.0905***
	(0.0143)	(0.0259)	(0.0251)	(0.0256)
Ln(Output p.c.) 1870	-0.939***	-0.875***	-0.873***	-0.872***
	(0.0742)	(0.0556)	(0.0547)	(0.0494)
Observations	2261	2261	2261	2261
R^2	0.658	0.689	0.750	0.768
State FE	no	no	yes	yes
Population Controls	no	yes	yes	yes
Nationality Shares	no	no	no	yes

The dependent variable is the change of ln output per capita between 1870 and 1920. The method of estimation is least squares. Genetic Diversity (1870) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Further initial control variables (1870) are population size, county area, the share of African-Americans, the share of Native Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1.

TABLE 2

THE EFFECT OF GENETIC DIVERSITY ON INCOME PER CAPITA TODAY

Dependent Variable:	ln(Income p.c.)	ln(Income p.c.)	ln(Income p.c.)	ln(Income p.c.)	ln(Income p.c.)
Year	2010	2000	1990	1980	1970
Genetic Diversity 1870	0.0426***	0.0512***	0.0526***	0.0549***	0.0639***
	(0.00551)	(0.00674)	(0.00655)	(0.00616)	(0.00713)
Ln(Output p.c.) 1870	-0.00563	-0.00314	-0.00450	0.000687	0.00402
	(0.00740)	(0.00897)	(0.00880)	(0.00640)	(0.00892)
Observations	2227	2227	2227	2227	2227
R^2	0.453	0.468	0.525	0.540	0.623
State FE	yes	yes	yes	yes	yes
Population Controls	yes	yes	yes	yes	yes
Nationality Shares	yes	yes	yes	yes	yes

The dependent variable is ln income per capita for the years 1970 to 2010. The method of estimation is least squares. Genetic Diversity (1870) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Further initial control variables (1870) are population size, county area, the share of African-Americans, the share of Native Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE 3
THE EFFECT OF GENETIC DIVERSITY OVER TWO CENTURIES

Dependent Variable:	ln(Income p.c.)	ln(Income p.c.)	ln(Output p.c.)	ln(Output p.c.)	ln(Output p.c.)	ln(Output p.c.)
Year	2000	1970	1940	1910	1880	1850
Genetic Diversity 1790	3.147***	2.108***	6.746**	4.994*	5.968***	5.514***
	(1.151)	(0.725)	(3.161)	(2.792)	(2.000)	(1.241)
Ln(Urbanization) 1790	0.439	0.353*	-0.302	-0.246	-0.294	0.534**
	(0.339)	(0.197)	(0.796)	(0.778)	(0.816)	(0.260)
Observations	139	139	149	149	149	148
R^2	0.466	0.618	0.475	0.467	0.651	0.582
State FE	yes	yes	yes	yes	yes	yes
Slave Share	yes	yes	yes	yes	yes	yes
Nationality Shares	yes	yes	yes	yes	yes	yes

The dependent variable is ln income per capita for the years 2000 and 1970 and ln output per capita for the yeas 1940, 1910, 1880 and 1850. The method of estimation is least squares. Genetic Diversity (1790) measures county c's genetic diversity that is due to the stock of immigrants; see section 2 for further details. Further initial control variables (1790) are the share of African-Americans, the share of Whites and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the county level: *** p<0.01, ** p<0.05, * p<0.1.

TABLE 4
REMOVING OUTLIER

	Dependent Variable: Δln(Output p.c.)			
	(1)	(2)	(3)	(4)
Genetic Diversity 1870	0.0687***	0.0822***	0.0735***	0.0864***
	(0.0130)	(0.0253)	(0.0206)	(0.0222)
Ln(Output p.c.) 1870	-0.686***	-0.672***	-0.670***	-0.689***
	(0.0537)	(0.0436)	(0.0419)	(0.0386)
Observations	2219	2217	2217	2217
R^2	0.370	0.410	0.523	0.548
State FE	no	no	yes	yes
Population Controls	no	yes	yes	yes
Nationality Shares	no	no	no	yes

Observations identified by the Hadi (1992) outlier procedure at the 5 percent level are excluded from the sample. The dependent variable is the change of ln output per capita between 1870 and 1920. Genetic Diversity (1870) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Further initial control variables (1870) are population size, county area, the share of African-Americans, the share of Native Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1.

TABLE 5
THE EFFECT OF GENETIC DIVERSITY ON URBAN GROWTH

Dependent Variable: Δln(Urbanization)				
	(1)	(2)	(3)	(4)
Genetic Diversity 1870	0.0296***	0.0283***	0.0307***	0.0302***
	(0.00358)	(0.00354)	(0.00794)	(0.00747)
Ln(Urbanization) 1870	-0.425***	-0.407***	-0.444***	-0.466***
	(0.112)	(0.126)	(0.128)	(0.130)
Observations	2262	2262	2262	2262
R^2	0.188	0.198	0.272	0.289
State FE	no	no	yes	yes
Population Controls	no	yes	yes	yes
Nationality Shares	no	no	no	yes

The dependent variable is the change of ln urbanization between 1870 and 1920. The method of estimation is least squares. Genetic Diversity (1870) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Further initial control variables (1870) are population size, county area, the share of African-Americans, the share of Native Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1.

TABLE 6
THE EFFECT OF GENETIC DIVERSITY ON MANUFACTURING

Dependent Variable:	Δln(Mfg. VA)	ΔMfg. Share	Δln(Mfg. Wages)	Δln(Mfg. Estab.)
Genetic Diversity 1870	0.0285***	0.0238**	0.0328***	0.0689***
	(0.0104)	(0.0104)	(0.00919)	(0.0224)
Ln(Mfg. Value Added) 1870	-0.979***			
	(0.0196)			
Mfg. Share 1870		-0.733***		
		(0.0454)		
Ln(Mfg. Wages) 1870			-0.960***	
			(0.0130)	
Ln(Mfg. Establishments) 1870				-0.930***
				(0.0208)
Observations	2028	2157	2012	2030
R^2	0.667	0.427	0.874	0.731
State FE	yes	yes	yes	yes
Population Controls	yes	yes	yes	yes
Nationality Shares	yes	yes	yes	yes

The dependent variable is in column (1) the change of ln manufacturing value added per capita; in column (2) the change of the manufacturing share; in column (3) the change of ln manufacturing wages; in column (4) the change of ln manufacturing establishments per capita between 1870 and 1920. The method of estimation is least squares. Genetic Diversity (1870) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Further initial control variables (1870) are population size, county area, the share of African-Americans, the share of Native Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1.

TABLE 7
OUTPUT SAMPLE SPLIT: POOR VS. RICH COUNTIES

	Dependent Variable: Δ ln(Output p.c.)							
	Poor	Poor	Poor	Poor	Rich	Rich	Rich	Rich
Genetic Diversity 1870	0.0585**	0.0815***	0.0445*	0.0543**	0.0859***	0.103***	0.0970**	0.110***
	(0.0247)	(0.0260)	(0.0236)	(0.0242)	(0.00988)	(0.0295)	(0.0399)	(0.0396)
Ln(Output p.c.) 1870	-1.043***	-0.892***	-0.894***	-0.889***	-0.722***	-0.766***	-0.659***	-0.687***
	(0.0812)	(0.0641)	(0.0545)	(0.0573)	(0.0677)	(0.0729)	(0.0801)	(0.0780)
Observations	1130	1130	1130	1130	1131	1131	1131	1131
R^2	0.701	0.731	0.808	0.828	0.288	0.358	0.489	0.522
State FE	no	no	yes	yes	no	no	yes	yes
Population Controls	no	yes	yes	yes	no	yes	yes	yes
Nationality Shares	no	no	no	yes	no	no	no	yes

The dependent variable is the change of ln output per capita between 1870 and 1920. The method of estimation is least squares. The sample is split into counties with output per capita below (poor) and above (rich) the median. Genetic Diversity (1870) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Further initial control variables (1870) are population size, county area, the share of African-Americans, the share of Native Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1.

TABLE 8
NON-LINEAR EFFECT OF GENETIC DIVERSITY ON OUTPUT GROWTH

Dependent Variable: Δ ln(Output p.c.)				
	(1)	(2)	(3)	(4)
Genetic Diversity 1870	0.0785*	0.0809**	0.0499	0.0809**
	(0.0437)	(0.0321)	(0.0409)	(0.0330)
Genetic Diversity^2 1870	0.000679	0.00419	0.00348	0.00131
	(0.00345)	(0.00265)	(0.00386)	(0.00391)
Ln(Output p.c.) 1870	-0.939***	-0.876***	-0.873***	-0.872***
	(0.0740)	(0.0550)	(0.0543)	(0.0496)
Observations	2261	2261	2261	2261
R^2	0.658	0.690	0.750	0.768
State FE	no	no	yes	yes
Population Controls	no	yes	yes	yes
Nationality Shares	no	no	no	yes

The dependent variable is the change of ln output per capita between 1870 and 1920. The method of estimation is least squares. Genetic Diversity (1870) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Further initial control variables (1870) are population size, county area, the share of African-Americans, the share of Native Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1.

TABLE 9
THE EFFECT OF GENETIC DIVERSITY ON OUTPUT PER CAPITA
10-YEAR NON-OVERLAPPING PANEL ESTIMATION

Dependent Variable: ln(Output p.c.)						
	LS	LS	LS	LS	LS	SYS-GMM
Genetic Diversity $_{t-1}$	0.0587***	0.0478***	0.0352***	0.0373***	0.0497***	0.0690***
	(0.00754)	(0.00853)	(0.00958)	(0.00902)	(0.0141)	(0.00591)
Ln(Output p.c.) $_{t-1}$	0.523***	0.534***	0.462***	0.448***	0.106***	0.203***
	(0.0521)	(0.0491)	(0.0545)	(0.0543)	(0.0341)	(0.0284)
Observations	11305	11305	11305	11305	11305	11305
R^2	0.801	0.802	0.821	0.831	0.855	
Number of Counties					2261	2261
Year FE	yes	yes	yes	yes	yes	yes
State FE	no	no	yes	yes	no	no
County FE	no	no	no	no	yes	yes
Population Controls	no	yes	yes	yes	yes	yes
Nationality Shares	no	no	no	yes	yes	yes

The dependent variable is the change of ln output per capita between 1870 and 1920. In Columns (1)-(5) the method of estimation is least squares. In column (6) the method of estimation is system GMM. Genetic Diversity (1870) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Further initial control variables (1870) are population size, county area, the share of African-Americans, the share of Native Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE 10
DIFFERENT MEASURES OF GENETIC DIVERSITY

Dependent Variable: ln(Output p.c.)

PANEL A: EUROPEAN GENETIC DIVERSITY

	(1)	(2)	(3)	(4)
European Genetic Diversity 1870	0.0890***	0.122***	0.0908***	0.0848***
	(0.0120)	(0.0239)	(0.0206)	(0.0257)
Ln(Output p.c.) 1870	-0.948***	-0.880***	-0.876***	-0.872***
	(0.0707)	(0.0521)	(0.0535)	(0.0498)
Observations	2261	2261	2261	2261
R^2	0.661	0.695	0.753	0.767
State FE	no	no	yes	yes
Population Controls	no	yes	yes	yes
Nationality Shares	no	no	no	yes

PANEL B: GENETIC DIVERSITY INDICATOR

	(1)	(2)	(3)	(4)
Genetic Diversity Indicator 1870	1.824***	1.509***	0.881***	1.002***
	(0.338)	(0.377)	(0.290)	(0.286)
Ln(Output p.c.) 1870	-0.921***	-0.864***	-0.869***	-0.864***
	(0.0741)	(0.0565)	(0.0555)	(0.0506)
Observations	2261	2261	2261	2261
R^2	0.644	0.679	0.747	0.764
State FE	no	no	yes	yes
Population Controls	no	yes	yes	yes
Nationality Shares	no	no	no	yes

The dependent variable is the change of ln output per capita between 1870 and 1920. The method of estimation is least squares. In Panel A, European Genetic Diversity (1870) measures county c's genetic diversity that is due to the stock of European immigrants; see Section 2 for further details. See Section 5.F. for the construction of the Genetic Diversity Indicator (1870) in Panel B. Further initial control variables (1870) are population size, county area, the share of African-Americans, the share of Native Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1.

TABLE 11
GENETIC DIVERSITY BEFORE THE CIVIL WAR

	Dependent Variable: ln(Output p.c.)							
	1850-1920	1850-1920	1850-1920	1850-1920	1860-1920	1860-1920	1860-1920	1850-1920
Genetic Diversity 1850	0.0494***	0.0242***	0.00303	0.00118				
	(0.00650)	(0.00863)	(0.00609)	(0.00632)				
Ln(Output p.c.) 1850	-0.960***	-0.910***	-0.914***	-0.916***				
	(0.0579)	(0.0648)	(0.0631)	(0.0703)				
Genetic Diversity 1860					0.0489***	0.0344***	0.0133**	0.0171**
					(0.00794)	(0.00791)	(0.00532)	(0.00693)
Ln(Output p.c.) 1860					-0.902***	-0.788***	-0.890***	-0.903***
					(0.0543)	(0.0519)	(0.0611)	(0.0591)
Observations	1615	1615	1615	1615	2015	2015	2015	2015
R^2	0.690	0.711	0.772	0.778	0.524	0.571	0.664	0.682
State FE	no	no	yes	yes	no	no	yes	yes
Population Controls	no	yes	yes	yes	no	yes	yes	yes
Nationality Shares	no	no	no	yes	no	no	no	yes

The dependent variable is in columns (1)-(4) the change of ln output per capita between 1850 and 1920 and in columns (5)-(8) the change of ln output per capita between 1860 and 1920. The method of estimation is least squares. The method of estimation is least squares. Genetic Diversity in 1850 (1860) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Further initial control variables 1850 (1860) are population size, county area, the share of African-Americans, the share of Native Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix Table 1: Descriptive Statistics

Variables	Years	Obs	Mean	Std. Dev.
Output Growth	1870-1920	2261	1.3987	0.8429
Urban Growth	1870-1920	2261	0.1147	0.1530
Mfg. Value Added Growth	1870-1920	2027	1.2446	0.5345
Mfg. Share (Change)	1870-1920	2146	0.1667	0.2512
Mfg. Wages Growth	1870-1920	2011	1.5378	0.5874
Mfg. Establishment Growth	1870-1920	2029	-0.6746	0.7942
Ln(Output per capita)	1850	1615	4.0468	0.7449
Ln(Output per capita)	1860	2015	4.1501	0.5457
Ln(Output per capita)	1870	2261	4.3076	0.7752
Ln(Output per capita)	1880	2261	4.0238	0.5983
Ln(Output per capita)	1910	2261	5.0356	0.5365
Ln(Output per capita)	1920	2261	5.7063	0.5403
Ln(Output per capita)	1940	2261	5.0029	0.6782
Ln(Income per capita)	1970	2227	8.0477	0.2318
Ln(Income per capita)	1980	2227	8.9774	0.2151
Ln(Income per capita)	1990	2228	9.6072	0.2122
Ln(Income per capita)	2000	2228	10.0587	0.2207
Ln(Income per capita)	2010	2228	10.3822	0.2027
Genetic Diversity	1790	149	0.5435	0.0203
Genetic Diversity	1850	1615	3.1023	3.3222
Genetic Diversity	1860	2015	3.8116	3.4022
Genetic Diversity	1870	2261	5.1528	2.3519
European Genetic Diversity	1870	2261	5.0017	2.3361
Genetic Diversity Indicator	1870	2261	0.2533	0.0923
Ln(Urbanization Rate)	1790	149	0.0347	0.1232
Ln(Urbanization Rate)	1870	2261	0.0571	0.1412
Ln(Mfg. Value Added per capita)	1870	2068	6.4316	0.4954
Mfg. Share	1870	2159	0.1948	0.2362
Ln(Mfg. Wages)	1870	2049	5.2681	0.6480
Ln(Mfg. Establishment per capita)	1870	2068	-5.5442	0.8377
Ln(Population)	1870	2261	9.0424	1.3735
Ln(County Area)	1870	2261	6.4319	0.7712
Share Native American	1870	2261	0.0054	0.0487
Share African-Americans	1870	2261	0.1512	0.2155
Share Foreign-Born	1870	2261	0.1122	0.1497

Appendix Table 2: US Stock of Foreign Born

Country of Origin	1870	1920
Ireland	1855829	1036760
Germany	1690854	1684464
United Kingdom	770390	1125942
Canada	493464	2120880
Sweden	211577	625003
France	116401	150848
Switzerland	75076	116518
China	63254	0
Holland	43285	125641
Mexico	36644	467751
Bohemia	36479	359550
Denmark	23465	181797
Austria	16572	575439
Italy	7181	1609781
Poland	5840	1139515
Atlantic Island	2288	37120
Cuba	2250	9749
West Indies	1318	20274
Spain	1183	40291
Portugal	974	63462
Belgium	470	50084
Russia	89	1399843
Hungary	46	392987
Australia	21	4259
Bulgaria	0	1697
Finland	0	144564
Greece	0	175847
Norway	0	358327
Rumania	0	96165
Yugoslavia	0	167007
Lithuania	0	131185
Luxemburg	0	5563
Syria	0	45999
Armenia	0	29751
Central and South America	0	12074

APPENDIX TABLE 3
Displaying all Controls for Table 1

Dependent Variable: Δln(Output p.c.)				
	(1)	(2)	(3)	(4)
Genetic Diversity 1870	0.0854***	0.114***	0.0768***	0.0905***
	(0.0143)	(0.0259)	(0.0251)	(0.0256)
Ln(Output p.c.) 1870	-0.939***	-0.875***	-0.873***	-0.872***
	(0.0742)	(0.0556)	(0.0547)	(0.0494)
Ln(Population) 1870		-0.121***	-0.0854**	-0.0970***
		(0.0378)	(0.0337)	(0.0332)
Ln(County Area) 1870		-0.0895**	-0.00982	0.00481
		(0.0392)	(0.0379)	(0.0345)
Share Native Americans 1870		0.629*	-0.0896	0.00315
		(0.329)	(0.186)	(0.185)
Share African-Americans 1870		-0.208	-0.0553	-0.140
		(0.179)	(0.172)	(0.196)
Share Foreign-Borns 1870		-0.132	-0.450	
		(0.304)	(0.363)	
Share Chinese 1870				-1.449**
				(0.652)
Share Atlantic Island 1870				-52.88
				(53.25)
Share Australia 1870				778.2***
				(156.2)
Share Austro-Hungarian 1870				-5.928*
				(3.199)
Share Belgians 1870				-12.71*
				(6.727)
Share Bohemians 1870				-0.617*
				(0.336)
Share Cubans 1870				10.30
				(7.961)
Share Denmark 1870				0.497
				(0.393)
Share France 1870				-5.166**
				(2.285)
Share Germans 1870				0.183
				(0.264)
Share Netherlands 1870				1.556***
				(0.519)
Share Ireland 1870				0.665
				(0.467)
Share Italians 1870				-12.12***
				(1.873)
Share Mexicans 1870				-2.336***
				(0.245)
Share Polnish 1870				-25.14
				(29.83)
Share Portuguese 1870				36.85
				(171.0)
Share Russians 1870				-6.711
				(5.184)
Share Spanish 1870				-113.4***
				(9.344)
Share Swiss 1870				-2.710
				(2.734)
Share West Indies 1870				-12.86
				(8.210)
Share Canadians 1870				-0.217
				(0.527)
Share United Kingdom 1870				-1.550
				(1.153)
Share Sweden 1870				0.127
				(0.206)
Share Other Foreign-Born 1870				-21.32
				(57.83)
Observations	2261	2261	2261	2261
R^2	0.658	0.689	0.750	0.768
State FE	yes	yes	yes	yes
Population Controls	yes	yes	yes	yes
Nationality Shares	yes	yes	yes	yes

The dependent variable is ln income per capita for the years 1970 to 2010. The method of estimation is least squares. Genetic Diversity (1870) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1.

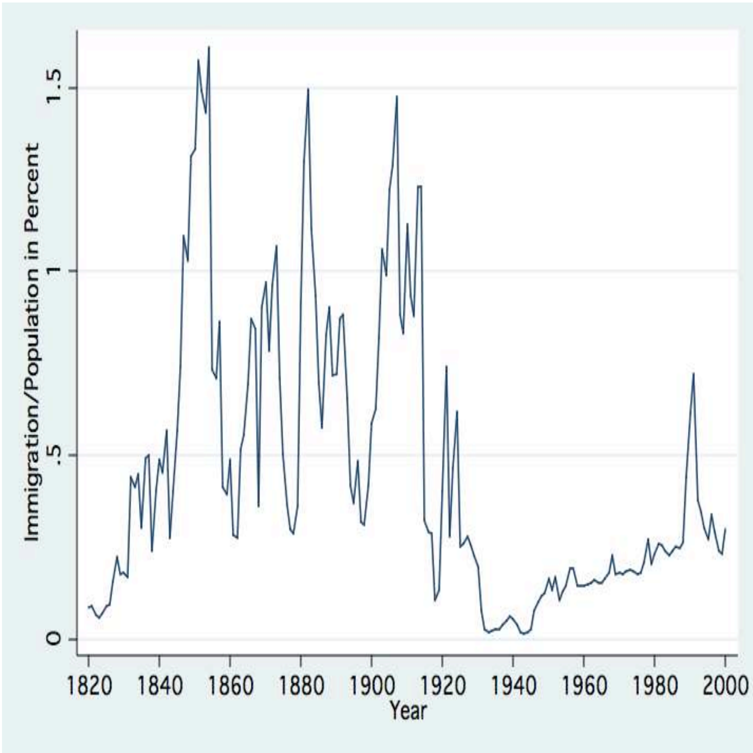
APPENDIX TABLE 4

THE EFFECT OF GENETIC DIVERSITY ON INCOME PER CAPITA TODAY

Dependent Variable:	ln(Income p.c.)	ln(Income p.c.)	ln(Income p.c.)	ln(Income p.c.)	ln(Income p.c.)
Year	2010	2000	1990	1980	1970
Genetic Diversity 1920	0.0210***	0.0223***	0.0256***	0.0307***	0.0319***
	(0.00388)	(0.00442)	(0.00417)	(0.00421)	(0.00470)
Ln(Output p.c.) 1920	0.0346**	0.0517**	0.0592***	0.0498***	0.0832***
	(0.0135)	(0.0195)	(0.0180)	(0.0159)	(0.0192)
Observations	2227	2227	2227	2227	2227
R^2	0.506	0.518	0.589	0.602	0.675
State FE	yes	yes	yes	yes	yes
Population Controls	yes	yes	yes	yes	yes
Nationality Shares	yes	yes	yes	yes	yes

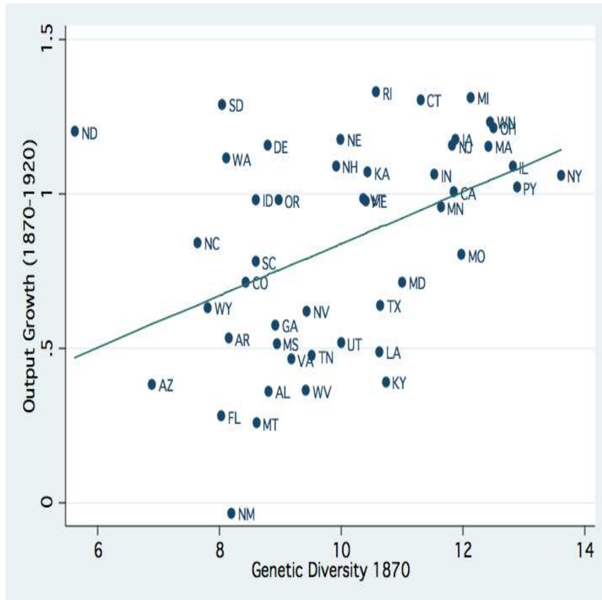
The dependent variable is ln income per capita for the years 1970 to 2010. The method of estimation is least squares. Genetic Diversity (1920) measures county c's genetic diversity that is due to the stock of immigrants; see Section 2 for further details. Further initial control variables (1920) are population size, county area, the share of African-Americans and the share of foreign-born by country of origin (estimates not reported in the table). Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix Figure 1:
Annual Immigration as a Fraction of the US Population



Source: Historical Statistics of the United States (Barde et al., 2006), Statistical Abstract of the US, eh.net database and Kim (Figure 3, 2007).

Appendix Figure 2:
Ln Output Growth and Genetic Diversity



Appendix Figure 3:
Ln Output Growth and Initial Ln Output per capita