People and Machines A Look at The Evolving Relationship Between Capital and Skill In Manufacturing 1850-1940 Using Immigration Shocks*

Jeanne Lafortune† José Tessada‡ Ethan Lewis §

December 2013

Abstract

Workers of the nineteenth and early twentieth century United States were buffeted by shocks derived both from major innovations in manufacturing production technology and large waves of immigration. This paper investigates these phenomena together, in a framework that allows us to study the response of production technology to immigration-induced changes in skill mix. This response reveals the impact technology has on the demand for workers of different skill levels because of the relative complementarity between technology and skills. Using a merge of public-use tabulations of the U.S. Manufacturing Censuses from 1850 and 1940, detailed by industry and county/city, with Census of Population data, we ask how the change in the use of manufacturing technologies in a locality responded to local immigration-induced changes in skill mix. In our study we exploit the fact that the available technologies changed over time and thus look at different period-relevant technologies, from factory production to electrification, taking into account the fact that the adoption and penetration of these technologies responds to the relative availability of workers of different skill levels. Our results show that in urban counties immigration significantly changed the skill ratios, thus modifying the market for workers and firms. We also find that capital stock, output, and average wages responded at the industry and aggregate levels to the immigration induced changes in the skill mix in a manner consistent with the view that technology and skill were *substitutes* in the nineteenth century. Furthermore, we find initial support for a shift in the production technology around the turn of the century, coincident with the spread of electricity, a result that is in line with the historical view that technology and skill became *complements*. Finally, we find no evidence that industry mix shifts were an important adjustment mechanism for absorbing immigration-induced skill mix shocks.

*We thank participants in the First EH/Clio Lab UC conference for their comments. Tessada thanks Fondecyt (Grant Iniciación #11110101) and Conicyt Programa de Investigación Asociativa SOC1102. Lafortune thanks Conicyt Programa de Investigación Asociativa SOC1102. The usual disclaimer applies. We wish to thank Claudia Allende, Francisco Muñoz and Sofía Garcés for excellent research assistance and Gérard Lafortune for careful data entry.

†PUC Chile and EH-Clio Lab UC. Email: jlafortune@uc.cl.

‡PUC Chile and EH-Clio Lab UC. Email: jtessada@gmail.com.

§Dartmouth College and NBER. Email: ethan.g.lewis@dartmouth.edu.

1 Introduction

Workers of the nineteenth and early twentieth century were buffeted by shocks derived from major innovations in production technology (e.g., Goldin and Katz, 1998; Jerome, 1934), large waves of immigration (e.g., Hatton and Williamson, 1998; Hatton and Williamson, eds, 1994), and, later, severe restrictions on immigration. Such changes likely generated winners and losers: industrialization simultaneously may have made artisanal skills less valuable while greatly expanding job opportunities for low skill workers; in contrast, low-skill immigration at the turn of the twentieth century may have raised the wages of artisans relative to laborers (Goldin, 1994).

While there is evidence about the likely impact of these shocks, credible identification is challenging because adoption of technology and the level of immigration are both likely to be at least partly confounded by economic forces which are difficult to fully control for, especially in historical data. The impact of technological change on skill demand has been studied both using variation across regions (e.g., Gray, 2013) –which may be confounded with differences in industry mix (Jerome, 1934) – variation across industries (e.g., Goldin and Katz, 1998; James and Skinner, 1985) and variation across plant sizes (e.g., Atack, Bateman and Margo, 2004) – which may be correlated with other non-technological determinants of skill demand.[1] Identifying the labor market impact of immigration using variation across regions is well known to be challenging (Borjas, 1994), and it is made even more so in historical data where it is not even possible to control for immigration's pure compositional impact, as microdata on wages largely do not exist.

In this project, we use new data and a new identification strategy to simultaneously uncover both the impact new manufacturing technologies of the nineteenth and twentieth century had on skill demand, and how local labor markets absorbed the massive waves of immigration during that same time period. A critical concept for this approach, outlined in the theory section below, is that credible estimates of the impact of skill supply on production technology reveal the "reverse," that is, how production technology affects skill demand. Both are really two sides of the same coin, deriving from the complementarity between technology and skills.[2]

Immigration-derived variation can help produce credible estimates of the impact of changes in skill mix, using an instrumental variables strategy which has been used successfully in modern immigration research (e.g., Card, 2001; Cortes, 2008; Lewis, 2011, among many others), but until recently, has seen little application in historical data. The instrument takes advantage of the fact that immigrants tend to persistently cluster into regional "enclaves" by country of origin, and that different origin groups have different mixes of skills. The instrument is essentially a prediction of the impact immigration would have had on skill mix if all new immigrants (by

[1] For example, if skill demand is measured with average wages, the plant size-wage correlation may be confounded by other forces, such as the greater productivity at large plants due to economies of scale.

[2] For another application of this idea in modern data, see Lewis (2011); and in historical data but in the agricultural sector, see Lafortune, Tessada and Gonzalez-Velosa (2013).

country of origin) continued to apportion themselves to enclave locations in the same way their fellow countrymen had long in the past (in our case, in 1850).

Our approach is also strengthened by a new dataset that we have put together into electronic format, tabulations of historical Manufacturing Censuses by the interaction of both industry and geography (county and/or city). These data allow us to examine regional changes in measures of production technology (among other things, capital intensity, horsepower, electrification) within - that is to say, holding constant - industry. This is critical because changes in the industry mix are a potential confounder of changes in production technology.[3]

Our initial analysis, which uses a combination of U.S. Population Census data from 1850-1940 and manufacturing censuses tabulations tabulated to the county/city x industry level for 1860-1940 (excluding 1890 and 1900)is shown below. To supplement this, we also present some results with a microdata sample covering 1850-1880 (originally created by Atack and Bateman, 1999), and we plan an analysis with a richer county-level dataset. Our key finding so far is that capital and skilled labor were substitutes relative to capital and unskilled labor in nineteenth century manufacturing, a finding which sharply contrasts with a consistent finding of capital-skill complementarity in modern U.S. manufacturing data. We show in a theory section below that this response of capital may have helped mute the local labor market impact of immigration on the wage structure. In contrast, our findings provide very little evidence that adjustments in industry mix helped absorb immigrant inflows, which is consistent with the modern literature on the labor market impact of immigration.

1.1 Background

Immigrants have shaped the U.S. manufacturing sector throughout its history. From Samuel Slater memorizing and bringing the plans for textile machines to the U.S., to the skilled British and other European artisans of the nineteenth century, and finally to the masses of less-skilled immigrant labor filling factories, immigrants have consistently played a prominent role in U.S. manufacturing (e.g., Berthoff, 1953), particularly during the Great Migration era of the nineteenth and early twentieth century. Interestingly, a prominent contemporaneous account of early twentieth century manufacturing states that its main initial motivation was to investigate how well mechanization had allowed the manufacturing sector to adapt to the severe immigration restrictions of the mid-1920s (Jerome, 1934).[4] The study's purpose was later shifted to include an investigation

[3] For example, without a control for industry, a shift towards more capital-intensive industries appears at an aggregate level to be a rise in the capital-intensity of production. We will also, as a part of this project, directly evaluate the impact of immigration-induced skill mix changes on industry mix, which is interesting for evaluating the predictions of open economy models.

[4] On page 3, Jerome states "Our survey had its origin in the hectic years of the post-War decade as an inquiry into the extent to which the effects of immigration restriction upon the supply of labor were likely to be offset by an increasing use of labor-saving machinery".

of the contribution of technological change to unemployment. This was of heightened concern during the Great Depression, when the study was completed, but it comes up continually and is being raised again in today's relatively high unemployment environment (Brynjolfsson and McAffee, 2011).[5]

The two motivations for Jerome's study are really two sides of the same coin: new technologies have different skill requirements, and immigration (or its restriction) can shift the set of skills available. Many have argued the arrival of factories reduced demand for skilled artisan labor while at the same time it raised demand for less-skilled production workers performing simple, repetitive tasks. For example, Atack et al. (2004) found using 1850-80 data that larger manufacturing plants –an indicator of factory (non-artisanal) production– paid lower wages –an indicator of lower average skill. On the flip side, it is the availability of less-skilled labor to fill factories that enabled the adoption of factory production. In particular, Goldin and Sokoloff (1984) argue that such labor was only readily available in Northern U.S. in the mid-nineteenth century, which is why the north industrialized first.[6] Kim (2007) shows that in 1850-1880, U.S. counties with higher immigrant density had larger manufacturing establishments. Chandler (1977) argues that modern manufacturing required professional management, and you also see evidence of a shift to more "white collar" jobs in the late nineteenth century (Katz and Margo, 2013).

After the switch to factory production from an artisan system, manufacturing is thought to have begun, perhaps somewhere around the turn of the twentieth century, a switch to continuous production system relying increasingly on electricity and large (more recently, automated) machinery, a system which Jerome called "mechanization."[7] The exact timing may have differed by industry, and of particular interest to us, location.[8] Goldin and Katz (1998) argue and provide evidence that the latter change is associated with greater skill and capital requirements, and so capital and skill became complementary by the early twentieth century, as they continue to be in modern times (e.g., Griliches, 1969; Lewis, 2011). They show that industries with greater capital- and electricity intensity had higher average production wages in 1919 and 1929, and had more educated workers in 1939. There are some different, or perhaps more nuanced, views of what mechanization did to skill requirements. Gray (2013) found that states which electrified more saw large relative increases in the employment of non-production workers, but among production workers decreases in the proportion of jobs requiring "dexterity" –which includes

[5]For a general view of immigration and labor markets during the period under study in this paper see Hatton and Williamson, eds (1994).

[6]Women and children initially filled such factories; in the South, in contrast, women and children's labor was already demanded by agriculture.

[7]Goldin and Katz (1998) present a slightly richer evolution in which the assembly line is another step between factories and mechanized continuous production.

[8]As an example of cross-industry heterogeneity, Berthoff (1953) describes how machines for weaving cotton textiles were developed much earlier than those for weaving woolen textiles. Similarly, Jerome's surveys suggest that steel and iron adopted mechanized production methods earlier than other industries. In terms of regional heterogeneity, Jerome (1934) found considerable cross-state variation in industrial power use, which is also the variation that Gray (2013) relies on.

craftsman– relative to those requiring manual labor. She argues the overall effect was to "polarize" labor demand, as craftsmen were likely in the middle of the wage distribution. In contrast, Jerome (1934) argued that conveyor belts and other handling technologies may have reduced demand for manual labor.

Under the previous factor system, Goldin and Katz (1998) argue that capital and skill would have been substitutes. Factory output substituted for the less capital-intensive artisanal production. Though this is a sensible view, the evidence for it is quite limited. One exception is James and Skinner (1985), who show that in 1850 capital and labor are more substitutable in manufacturing sectors that appear to be more skill-intensive than in sectors that are less skill-intensive.

Many of the studies above use variation in some technology-use measure - the right-hand side variable - to estimate the response of skill measures. We examine the other side of the coin: how immigration-induced changes in skill mix are associated with adjustments in various measures of technology use. As the theory section will describe, both approaches should reveal the nature of the complementarity between technology and skills. Our approach will also give insight in the ability of the economy to "absorb" large immigrant inflows, as adjustments to technology can help mitigate the impact of immigration on the wages of native-born workers (Lewis, 2013).

There is another way in which the economy may have absorbed immigrants: immigrants may shift the industry mix, as Heckscher-Ohlin (HO) trade theory would suggest. In early twentieth century agriculture, for example, Lafortune et al. (2013) find evidence that immigration shifted the mix of crops towards more labor-intensive ones. This is interesting per se because, in the extreme case where HO fully holds, an economy can adjust to skill mix changes without any long-run impact on the wage structure; more generally, such adjustments mitigate the wage impact of immigration. In addition, changes in industry mix may confound changes in production technology: to the extent that production technology differs across industries, an impact of immigration on industry mix may make it (spuriously) appear that production technology has shifted at the aggregate level. The solution is to examine changes in production technology within detailed industries - in other words, to hold industry constant - a purpose which motivates our data collection that is described below.

2 Theoretical Framework

Our work starts from a simple framework that considers a single (aggregate) production function with three production factors: capital (K), high skilled labor (H) and low skilled labor (L), which is a common formulation both in the immigration and the technology adoption literatures (see for example Lewis, 2011, 2013). We assume the production function is constant returns to scale and satisfies standard quasi-concavity constraints. Throughout we also assume that the capital is supplied elastically to that production method and that the interest rate is fixed at the economy level. Under these assumptions, the capital stock adjusts to maintain equality between its marginal product and the cost of capital, which implies that in equilibrium $d \ln \left(\frac{\partial Y}{\partial K} \right) = 0$, where Y is aggregate output. Under constant returns to scale, this translates into,

$$d \ln K = \frac{L \frac{\partial^2 Y}{\partial K \partial L}}{H \frac{\partial^2 Y}{\partial K \partial H} + L \frac{\partial^2 Y}{\partial K \partial L}} d \ln L + \frac{H \frac{\partial^2 Y}{\partial K \partial H}}{H \frac{\partial^2 Y}{\partial K \partial H} + L \frac{\partial^2 Y}{\partial K \partial L}} d \ln H \tag{1}$$

We can then derive the following expression, which describes the impact of a change in the endowment of high-to-low-skilled workers on the capital-to-low-skilled labor ratio,

$$d \ln K - d \ln L = \frac{H \frac{\partial^2 Y}{\partial K \partial H}}{L \frac{\partial^2 Y}{\partial K \partial L} + H \frac{\partial^2 Y}{\partial K \partial H}} (d \ln H - d \ln L) \tag{2}$$

The denominator in equation (2) is positive if the production function displays decreasing returns to capital, which was assumed. Therefore, the sign of the numerator indicates input complementarity with high skill labor: capital and high skill labor are "q-complements" if $\frac{\partial^2 Y}{\partial K \partial H} > 0$ and "q-substitutes" if $\frac{\partial^2 Y}{\partial K \partial H} < 0$. One can also derive a symmetric expression for the complementarity between capital and low skill labor from the response of the capital-to-high-skill labor ratio to changes in the relative endowment of high skill workers.[9] One shortcoming of this approach, however, is that it is not robust to mismeasurement of who is high and low skill, which is a serious concern in the economic census data we will use (as it contains only crude cuts of "skill"). To see this, note that if our empirical definition of "L" in the left-hand side of (2) (and in equation (4) shown later in this section) included some high skill workers, what we would get instead is a weighted average of the complementarity between capital and high and capital and low skill labor.

As Lewis (2013) emphasizes, for many purposes, we may anyway care more about the *relative* complementarity between capital and high skill and capital and low-skill labor, which, for example, determines the impact of capital deepening on returns to skill (shown below). As he shows, this relative complementarity is positive if and only if capital-labor ratios respond more positively than output-labor ratios to increases in the relative endowment of high skill workers. The response of output-to-low-skill workers is given by:

$$d \ln Y - d \ln L = \frac{(\alpha + \beta) H \frac{\partial^2 Y}{\partial K \partial H} + \alpha L \frac{\partial^2 Y}{\partial K \partial L}}{H \frac{\partial^2 Y}{\partial K \partial H} + L \frac{\partial^2 Y}{\partial K \partial L}} (d \ln H - d \ln L) \tag{3}$$

where $\alpha = H(\partial Y / \partial H)/Y$ is high-skill labor's output share and $\beta = L(\partial Y / \partial L)/Y$ is low-skill's share. If high skill and low skill labor are both q-complementary with capital, the output per low-skill labor ratio would increase in response to a shock to high-to-low-skilled endowment

[9] See Lafortune et al. (2013).

ratio. If one labor type is q-complementary and the other is not, the response is ambiguous.

As was already mentioned, the relative size of the two cross-derivatives is revealed by whether the response in (2) or (3) is larger; that is, differencing (3) from (2), the response of the capital-to-output ratio. A revealing way to write this response is in terms of the response of relative wages:

$$d\ln K - d\ln Y = Y\alpha\beta \frac{\frac{\partial \ln(W_H/W_L)}{\partial K}}{H\frac{\partial^2 Y}{\partial K \partial H} + L\frac{\partial^2 Y}{\partial K \partial L}}(d\ln H - d\ln L) \tag{4}$$

The numerator of (4) contains the response of high-skill relative wages (with $W_H = \partial Y/\partial H$ and $W_L = \partial Y/\partial L$), assuming workers are paid their marginal product, to capital, which has the same sign as the response of capital-output ratios to increases in high-skill relative supply. For example, if capital and high skill labor are more complementary than capital and low-skill labor, then capital-to-output ratio should rise in response to an increase in the relative endowment of high skill labor. Equation (4) reminds us that complementarities work in both directions: the estimated response also reveals how capital adoption (our model of technological change) affects skill demand. This is useful, as actual measures of the wage structure are quite crude.

Indeed, our estimates of (2) and (4) could be used to learn something about the likely magnitude of the response of relative wage to changes in skill endowments. One can show that

$$\frac{\partial \ln(\frac{W_H}{W_L})}{\partial \ln(\frac{H}{L})} = -C_{HL} + \frac{\partial \ln(\frac{W_H}{W_L})}{\partial \ln(K)} \frac{H\frac{\partial^2 Y}{\partial K \partial H}}{H\frac{\partial^2 Y}{\partial K \partial H} + L\frac{\partial^2 Y}{\partial K \partial L}} \tag{5}$$

where $-C_{HL}$ represents the short-run (capital fixed) wage adjustment to a change in relative skill supply. Note that this expression implies that the long-run relative wage impacts of, say, an immigration-induced change in skill mix may be smaller or larger than this depending on the sign of the responses estimated in (2) and (4). For example, if capital complements skilled labor relative to unskilled labor –if the response in (4) is positive– then the long-run response of relative wages to immigration is diminished.[10] Relative wage impacts are larger than this when capital is skill neutral.

Two specific contrasting examples of prominently used production functions may be helpful in delineating this last point. It is common for studies of the modern-day labor market impact of immigration to model labor demand using an aggregate production function featuring constant elasticity of substitution (CES) and separable capital, like $K^\alpha \left(H^\sigma + L^\sigma\right)^{\frac{1-\alpha}{\sigma}}$. In such a setup, rK/Y is fixed at α in the long run, and

[10]Note that in this three-factor setup capital is always an absolute q-complement of skill ($\partial^2 Y/\partial K \partial H > 0$) whenever it is a relative complement of skill (that is, if $\partial \ln(\frac{W_H}{W_L})/\partial \ln(K) > 0$).

$$\frac{\partial \ln(\frac{W_H}{W_L})}{\partial \ln(\frac{H}{L})} = (\sigma - 1) \tag{6}$$

Put differently, the response of relative wages is treated as an estimate of the "elasticity of complementarity" (Hamermesh, 1993) between H and L ($-C_{HL} = \sigma - 1$), which is constant and, more the point, unaffected by the adjustment of capital. At another extreme, in the CES production function featuring capital-skill complementarity in Autor, Levy and Murnane (2003), $((K+L)^{\sigma} + H^{\sigma})^{\frac{1}{\sigma}}$, it remains the case that $-C_{HL} = \sigma - 1$ but it turns out that $\frac{\partial \ln(\frac{W_H}{W_L})}{\partial \ln(\frac{H}{L})} = 0$ as skill mix changes are entirely absorbed by adjustments in capital in the long run.

Extending the model: Changes in modes of production Up to now we have worked under the assumption that we can represent the economy with an aggregate production function. However, this is not necessarily the only way to model the adjustment to the changes in the relative endowment of high-to-low-skilled labor. In particular, as Beaudry and Green (2003) suggest, if there are two modes of production, each of them characterized by different intensities of use of the factors, then the economy can respond to the changes in the relative endowments choosing a different mode of production rather than just moving along the same isoquant as before.

To see how this works, consider the case where in the economy we can produce the same final good Y with two different modes of production: 1 and 2, and denote with by Y_i the amount of the good produced using mode i, and assume that for any set of factor prices mode 2 is low-skilled labor- and capital- intensive vis-a-vis mode 1, which is how Goldin and Katz (1998) model the difference between artisanal (mode 1) and industrial (mode 2) production, and that factor prices are determined in the economy. In this case, if we start from a high-skilled abundant situation and there is an increase in the supply of low-skilled labor, the new equilibrium will be characterized by a switching to mode 2. This final equilibrium will show a smaller effect on the relative wage of the low-skilled workers, and, more importantly, could be confused with a different level of complementarity between capital and both types of labor (in a single aggregate production function). In the context of the period where we have some new technologies being adopted, this is another mechanism we will explore by examining the response of indicators of production mode, such as plant size, to changes in skill mix.

Multiple Sectors To be written.

3 Empirical Methodology

3.1 Baseline equation

Following the main results from our model, we want to estimate the following equation

$$y_{cit} = \phi \ln \left(\frac{H}{L}\right)_{ct} + \beta X_{ct} + \nu_c + \eta_t + \mu_i + \epsilon_{cit} \tag{7}$$

where y_{cit} corresponds to an outcome of interest in industry i in county c at time t (the outcomes we can measure right now are capital-to-low-skilled labor, wage bill, value of output per low-skilled worker, horsepower per low-skilled worker), $(H/L)_{ct}$ is the high-to-low-skilled labor ratio in the county c at time t, X_{ct} is a vector of time varying county-level controls and ν_c, η_t, and μ_i represent country, time and industry fixed effects. In practice, because it is thought that production techniques may have been quite different in the nineteenth and twentieth century (Chandler, 1977; Jerome, 1934), we allow the fixed effects to vary by century.[11]

The interpretation of the coefficient ϕ depends on the relevant outcome that is being estimated (as shown by the equations (2), (3), and (4)). In equation (4), for example, where $ln(K/Y)$ is the outcome, it captures the complementarity between capital and skill relative to capital and low-skill: ϕ will be positive if capital complements skilled labor relative to unskilled labor ($\phi > 0$ implies that $\partial \ln(W_H/W_L)/\partial K > 0$).

An extension of this equation would allow the coefficient ϕ itself to be a function of sector characteristics and/or of time. This specification can then capture changes in the impact of changes in relative endowments on the outcomes. Given that our model relates these impacts to the relative complementarity of capital and both types of labor, these differences in the value of ϕ can capture changes in technology and/or modes of production. This is an important question to explore given that some of the technological innovations that emerged in manufacturing around these decades might have significantly affected the relation between skill and capital.

We also explore whether county- or city-wide (aggregate) outcomes are influenced by estimating the following equation, which corresponds to equation (7) but using data aggregated at the geographic level,

$$y_{ct} = \phi \ln \left(\frac{H}{L}\right)_{ct} + \beta X_{ct} + \nu_c + \eta_t + \epsilon_{ct} \tag{8}$$

In this specification y_{ct} is an outcome variable -capital to low-skilled worker, wage (bill), (value of) output per low-skilled worker, number of establishments per worker- measured at the county level. In this case we can explore how the county as whole adjusts to the changes in the skill-

[11] Another reason to do this is that the construction of our instrument, described below, uses different data in the nineteenth and twentieth centuries.

mix of workers. Estimates of (8) may suffer from aggregation bias: shifts in output mix towards industries that use a different production technology could confound the results. This is why the industry-city data, which allow us to estimate (7) instead, are critical.

3.2 Identification strategy

The skill-mix variable is likely to be endogenous as workers are likely to locate geographically where their skills are most valued (or to alter their skill acquisition decisions in response to local labor market conditions). We use immigration shocks as the source of variation to identify changes in the relative endowment of high-to-low-skilled labor. However, since immigrants are also likely to elect their location based on economic conditions that may affect technology choices, we construct an instrument given by

$$\ln(pred\ ratio)_{ct} = ln\left(\frac{\sum_j \left(\frac{N_{jc0}}{N_{j0}} HS\_imm_{jt}\right) + HS\_nat_{c0}\frac{HS\_nat_t}{HS\_nat_0}}{\sum_j \left(\frac{N_{jc0}}{N_{j0}} LS\_imm_{jt}\right) + LS\_nat_{c0}\frac{LS\_nat_t}{LS\_nat_0}}\right) \tag{9}$$

where j represents each country of birth, c (US) county, and t period; N is the stock of immigrants (not broken out by skill); HS_imm_{jt} and LS_imm_{jt} are the *national* stocks of high-skill and low-skill immigrants from each country in each period, respectively; HS_nat_{c0} and LS_nat_{c0} are the stock of natives by skill in some base year, 0; and $\frac{HS\_nat_t}{HS\_nat_0}$ and $\frac{LS\_nat_t}{LS\_nat_0}$ are the *national* growth rates of high and low-skill native-born populations from the base year to t. Note that the first term in the numerator and denominator includes $\frac{N_{jc0}}{N_{j0}}$, which represents the share of immigrants from j living in c as of some base year census. (In practice, as will be described below, we will have two base years: 1850 for 1860-1880 and 1880 for 1900-1940.) This is used to apportion the current stocks of immigrants by country to locations within the U.S. Thus, the first term in the numerator and denominator represents the number of high- and low-skill immigrants, respectively that would be living in c if immigrants were still apportioned across counties in the same manner as they were in the base year. This style of instrument has been widely used to study modern-day immigration impacts (see, for example Card, 2001; Cortes, 2008; Lewis, 2011) but until recently has seen limited application in this historical context. It attempts to circumvent the problem of endogenous location choice by allocating immigrants to counties based on the location of immigrants from one's country of birth in previous waves. We use the previous location of all immigrants instead of allowing high- and low-skilled individuals from a given country to be distributed in a distinct way such that these shares are less likely to capture economic conditions particularly suitable for a given skill level. Lafortune and Tessada (2013) provided significant evidence of ethnic network's role in the determination of the first location of immigrants arriving to the U.S., which supports the validity of the instrument.[12]

[12] Just a few papers have used this type of instrument, for example Goldin (1994); Kim (2007); Lafortune et al. (2013).

We modify this instrument according to the approach taken in Smith (2012) to add the predicted skills of natives to the instrument. We predict the latter from the lagged location of high- and low-skill natives interacted with the national growth rate of skills among native-born workers. Thus, the instrument represents the predicted skill ratio given the initial locations of immigrants and natives and *national* changes in the country mix of immigrants and the skill mix of immigrants and natives.

Thus our instrument represents a predicted skill ratio based on the interaction of initial conditions and national changes in the skill and country-composition of workers. Because it is structured like the *actual* skill ratio, a first stage coefficient of one means that predicted immigration-driven changes in skill mix have a one-for-one impact on the actual skill ratio; coefficients different than one imply that the actual skill mix is offset by either native migratory response or other offsetting demographic changes (for example, if trends in native-born literacy differed in high- and low-immigration markets).

4 Data and Descriptive Statistics

This project requires information regarding a variety of outcomes, most specifically, population flows and stocks and inputs/outputs from manufacturing industries. In particular, given our objective of measuring the impact of factor availability in local markets, we need this information at a city or urban county level over a long period of time. Given that most manufacturing activity was concentrated in urban zones, we propose to focus only on the most urban counties in our baseline period. We do this so that our sample does not suffer from sample selection, as counties that became urban during the period may have become so because of a change in their labor endowment. Furthermore, this simplifies the matching process in terms of county/city boundaries for such a long time period. This section details where the data comes from and the fraction of it that we would need to input.

First, information regarding the number of high and low-skill individuals in a given locality can be obtained in each decade from IPUMS data (Ruggles, Alexander, Genadek, Goeken, Schroeder and Sobek, 2010) from 1850 to 1940 (except in 1890). There are really two options for defining "skill" in these data: occupation or literacy.[13] An advantage of literacy is that it is something close to a pre-labor market skill, whereas occupation-derived measures are a match between workers' skills and local labor market demand conditions. A disadvantage is that literacy does not necessarily have a strong relationship with the skill requirements of different kinds of manufacturing jobs (other than the small number of white collar jobs).[14] Also, literacy rates in

[13]Completed education is not available until 1940; only measures of school enrollment for youth are available prior to that time.

[14]Literacy is also not available in 1940. In its place we define as "illiterate" anyone who reports fewer than two

the U.S. were quite high even in the mid-nineteenth century, so it may not be a very relevant skill margin. So in addition to literacy, in future drafts we will consider occupation-based divisions of skills.

Secondly, we will use immigration as a shock to factor endowment of local labor markets that immigration generates over the period 1850 to 1940. This is a period of great potential for this purpose as immigration flows were very large. It also includes periods of slower immigration driven by potentially exogenous factors (Civil War, First World War) and by a dramatic change in the legal environment (1924's Johnson Act). We propose to use an instrumental variable approach as detailed above in equation (9). To construct this instrument, we first need a reliable estimate of the location of immigrants of different origins in a "base year" (the $\frac{N_{jc0}}{N_{j0}}$ in (9)). We actually use two base years for this purpose: 1850, which we apply to nineteenth century years (1860-1880), and for which we obtained a 100% sample by querying ancestry.com; and 1880, which we apply to the twentieth century years in our data (1910-1940), and for which a 100% sample is available from IPUMS. We use these 100% tables to alleviate concerns of small-cell biases (see Aydemir and Borjas, 2010). There is also an aggregate portion of the instrument. In principle, are several ways we could have constructed the national number of high and low-skill immigrants arriving after 1850. We chose to measure the with the stocks of each types of migrant from each country in 1850 to 1940 by aggregating IPUMS data. From 1900-1930 we could have used the Census question regarding the year of entry; we chose not to use this because it is only available in these years.[15]

Finally, our outcome variables focus on the adjustment mechanisms in the manufacturing sector over this period. Our conceptual framework calls for data at the level of the labor market x industry. These can be obtained from published Manufacturing Census tabulations. Conveniently for our analysis, manufacturing censuses occurred roughly concurrently with the Census

years of education.

[15] Another option is to use an administrative source on immigrant flows. One which we have considered is the flows of immigrants as documented in the Report of the Immigration Commissioner of the period (from 1899-1932) and for some additional periods previous to that. In those documents, the skill level of each immigrant is based on their pre-immigration occupations instead of their current one, akin to the data used by Friedberg (2001), which is likely to be less correlated with local economic conditions. Furthermore, immigrants include not just the net stock but the total flow which may be more exogenous than the number who eventually stay in the United States (Angrist, 2002). However, the fact that the data is, for some years, reported at the ethnicity level and for others at the level of the country of last residence, may introduce more noise in the variable, making the first stage weaker. So for the moment, we have chosen to rely entirely on the IPUMS source for the "aggregate" part of the instrument, although we have already entered these "administrative" data and may use them in the future. We also prefer this administrative source to alternatives available to us. The Ellis Island data set, which includes all passengers who arrived to the port of New York (Bandiera, Rasul and Viarengo, 2013), does not include any variable that would allow us to classify immigrants by their skill level. In addition, given our focus on geographical dispersion of immigrants over the period, using only immigrants arriving to New York may bias our estimates significantly. Finally, it is unclear how many of the passengers in this database were actually immigrants, that is not simply visiting the United States. Ship manifests could also be used but not all of them include occupation at origin or other variables to separate high- from low-skill migrants. Furthermore, this would have required a data-entry process that was not necessary for our current project, given that our focus is on aggregates and not on individuals' characteristics. We are aware of no other data sources than the ones detailed here that would permit us to pursue our hypothesis.

of Population over this entire period. Unfortunately, the tabulations are available only in paper format but we have digitalized them.[16]

One issue in covering such a long time series is that the unit of geography reported in these tables changes over time. In 1860 and 1870, the data is available only available by county while in 1880 and later, the main geographic tabulations are for largest cities, occasionally supplemented by tabulations for selected urban counties. Because of this change of geography, and because, with rare exception, cities are within county boundaries, we have chosen to make "county" the unit of analysis, aggregating each city's data in later years to the main county in which it is located.[17]

In each year, a number of variables of interest are reported for different industries (e.g., capital, output, employment); in later years there is a minimum "cell size" to be included (often, at least 5 establishments) while in 1860 and 1870, it appears that almost all establishments were tabulated.[18] However, even with these reporting restrictions, there is "balancedness" in the sense that the industries detailed for each city often repeat, allowing us to use panel methods as detailed in the empirical methods section.[19]

The variables that we will be able to measure using this method are shown in Table 1. We have finer definitions of some of the aggregates in a few years (like the number of salaried/wage earners, female/male workers, detailed capital or expenses categories) and value-added starting in 1910. To compensate the absence of capital after 1920, we do observe horsepower as a proxy.

Currently we have entered data for 1860-1880 and for 1910-1940. Unfortunately we are therefore missing data on the middle of our period, which may be a critical period for a shift in production technology. Nevertheless, when merged altogether we have a fairly rich and detailed panel, covering over 20,000 area x industry x year cells, comprised of 126 cities and 429 industries. These area cover more than half of the U.S. immigrant population at both the beginning at the end of our sample, and the industry division is very detailed. The means of our sample are shown in Table 2.

These data have some limitations. Greater coverage, and in some cases more detailed variables, can be obtained by examining aggregates at the county-level; so we will use as an additional source the digitized tables compiled by Michael Haines and co-authors (now publicly

[16]We will eventually provide a complete set of references in a data appendix; scanned versions of each set of census tabulations are available on the census.gov website. Occasional missing and poorly scanned pages from these tabulations were obtained from physical copies of the volumes.

[17]The only significant exception to this is New York City, which spans multiple counties and whose county composition changes over time. We therefore construct New York City to cover the five "boroughs" (counties) that make it up at the end of the period throughout the entire 1860-1940 period. This aggregates together Brooklyn and New York City, which reported as separate cities in earlier years.

[18]Home industries, which may have been important in these early years, were not included; there was also a sales threshold for inclusion.

[19]Industries were matched by hand by the authors, aggregating where necessary to create consistency over time. Details will eventually be provided in a data appendix.

available in the NHGIS database; see (Minnesota Population Center, 2011)). However, as alluded to above, the responses we measure in this latter dataset may be confounded by shifts in industry mix, which we will not be able to control for. A final detailed source, firm-level data, has been compiled for a sample for the period 1850-1880 by Atack and Bateman (1999), which we will also use.

5 Preliminary Results

5.1 First stage

Our identification strategy relies on the impact regional clustering of immigrants has on skill ratios as the origin composition of immigrants shifts over time, an approach which seen a lot of use in modern studies of the labor market impact of immigration. While far from unchallengeable as a source of exogenous variation, it is a demanding instrument for a number of reasons. First, we are allocating immigrants (both high and low-skill) using the county of residence of ALL previous residents, no matter what their occupation. If there is any correlation between occupations and location (as shown in Lafortune and Tessada, 2013), this is more likely to be exogenous but also costly in terms of power. Second, we allocate immigrants arriving over using fixed location shares. This requires a fair amount of stability in the location choice of immigrants. Finally, this instrument also relies on the skill mix of immigrants differing substantially from natives.

Before turning to the first stage results, it is worth considering in more specific detail the components of variation in the instrument over this period. A primary source is the differences in the distribution of immigrant groups across locations, (the $\frac{N_{jc0}}{N_{j0}}$ in (9). In other words, where were the enclaves? For the 1850 base year, which we apply the nineteenth century data, the top locations of the six largest immigrant groups are shown in Table 3. Although New York is the top locations for all groups (or close to it for Canadians), and port cities are common for all groups, the pattern of destinations other than New York tends to differ across groups. Note that Italians and Russians had already begun to cluster in San Francisco long before the big wave of Italian and Russian migration.

A second sources of variation in the instrument is the of over time in the country composition of immigrants, shown in Figure 1 for the same six groups (see below):

Irish immigration peaks early in the period, German in the middle, and Italian and Russian/Polish immigration latest. A third source of variation is the skills of the different immigrant groups compared to the native-born population. That will depend on the particular market under study, but this Figure 2 shows it in the aggregate (see below):

Figure 2 shows the conventional wisdom: German and English immigrants were high skill, so

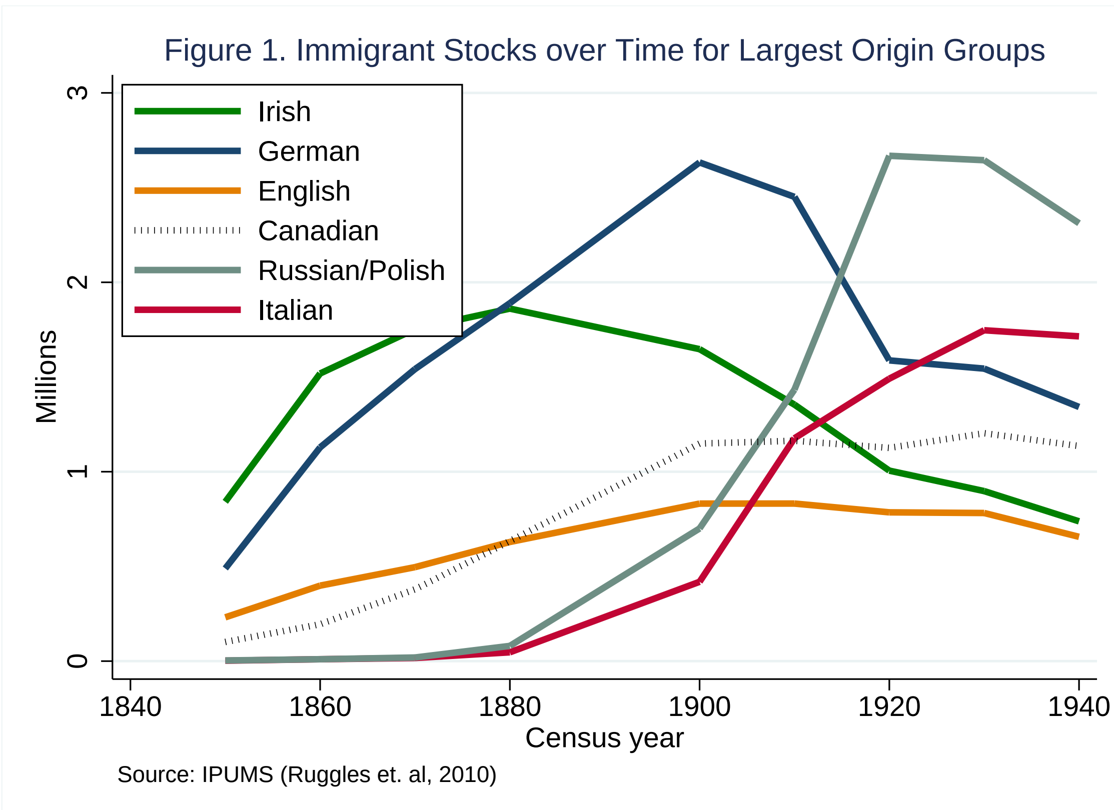

Figure 1: Stocks by group (of countries) of origin

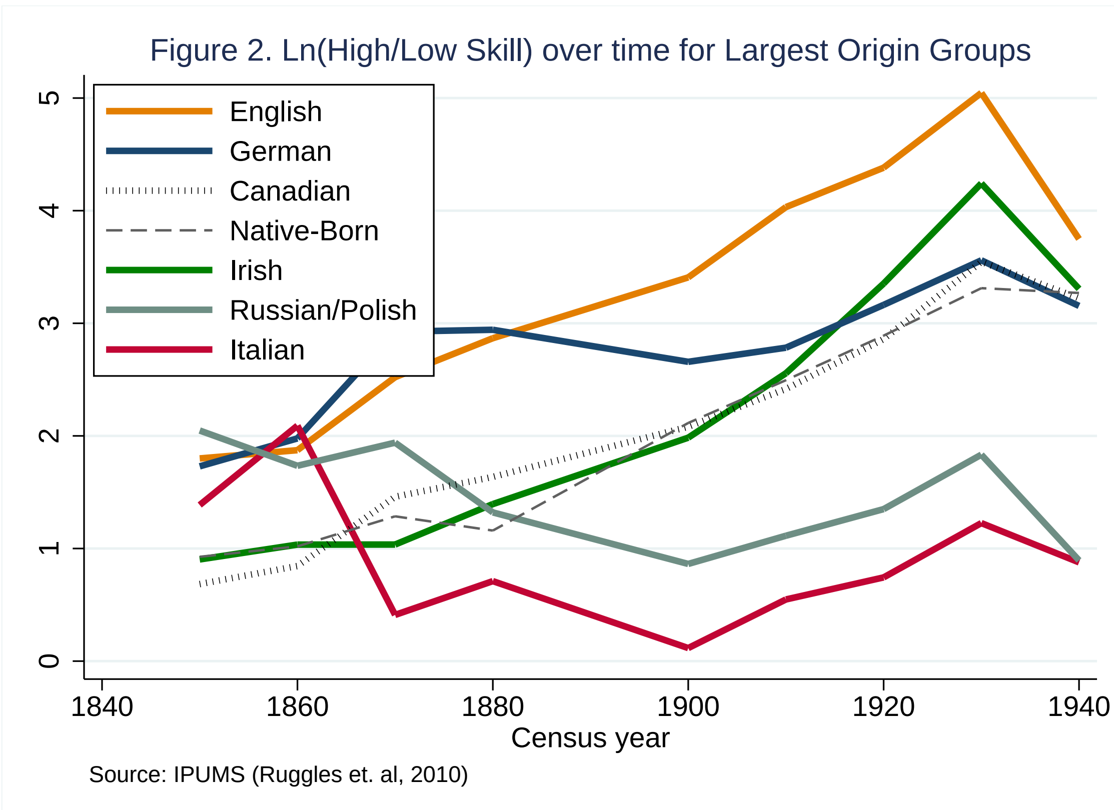

Figure 2: Skill ratio by group (of countries) of origin

concentrations of them would tend to raise the average skills of workers in an area. In contrast, by the time of the wave of Russian and Italian immigration, these groups had very low literacy skills compared to the native-born population.

Table 4 shows the first stage regressions estimated in the industry x county level data. The first column shows the estimate for our full sample. To address that there are multiple "copies" of county within a year (and over time), we cluster standard errors on county. In addition, we weight by the inverse of the number of industries represented in a county (to give each county equal weight).[20] The first stage coefficient is 0.9, which is not significantly different from one. Recall that one is what you would expect if "predicted" immigration had a one-for-one impact on skill mix (that is, if native migration or other realized demographic changes did not offset the impact of predicted immigration on skill ratios). In addition, the first stage is strong, with an F-stat around 15.

The remaining columns of Table 4 examine this first stage relationship in various subsamples. The first stage is strong in the years/cells of data on which we have information on capital stocks (column 2; see also Table 1); a bit weaker in the sample with data on horsepower (column 3). Usefully, the first stage is significant in both the nineteenth and twentieth centuries. Unsurprisingly, it is a bit weaker in the twentieth century data, a period of declining immigration. In order to address the weakness of the first stage in some subsamples, we will present reduced form F-statistics in parallel with all instrumental variables regressions presented below. These reduced form F-stats should give correctly sized tests even in the presence of a weak first stage.

5.2 Adjustments of Technology

Table 5 shows instrumental variables (IV) and ordinary least squares (OLS) estimates of the relationship between skill mix and manufacturing production outcomes. IV estimates are in panel A. In columns (1) - (3) this is estimated in the "capital subsample," that is, using cells 1860-1920 on which we have data on capital stocks. Column (1) shows that capital per worker has a small and insignificant relationship with skill ratios. However, in this same sample, skill ratios have a significant positive relationship with output per worker: it should not be surprising that more skilled markets tend to produce more output per worker within a given industry. Differencing columns (1) and (2), therefore, skill ratios are significantly *negatively* related to capital-output ratios (column 3). Recall from the theory section that this implies that unskilled labor is a q-complement of capital relative to skilled labor. Note that this contrasts with a large amount of modern evidence, including some from a period that overlaps with ours Goldin and Katz (1998) which suggests that capital and skill are complements. At least some of the modern evidence is built on (potentially) incorrectly interpreting the response of capital-labor ratios Lewis

[20]The standard errors are larger if we do not make this weighting adjustment, but the F-stat remains above 10.

(2013), or else is more reliant on descriptive evidence (e.g., cross-industry correlations). However, even a study which uses the same style of identification strategy and outcome finds evidence of capital-skill complementarity in manufacturing in modern data Lewis (2011). Another difference, however, is that we are looking deeper into the nineteenth century than even Goldin and Katz (1998), a period in which other evidence finds some support for capital-unskilled complementarity (e.g., James and Skinner, 1985), evidence which Goldin and Katz (1998) note as well. We will return to this point below when we examine the results separately by century.

Column (4) reveals that establishment size does not appear to significantly confound these changes in capital usage, an issue raised in Katz and Margo (2013); though size is positively related to skills, the relationship is not significant. Column (5) shows a response of average wages to increases in skill share that is stunningly similar in magnitude to the response of output per worker, consistent with workers being paid close to their marginal product. Skills are negative related to horsepower intensity, a variable which is only available at the end of our sample period, but the relationship is not significant.

OLS estimates, shown in panel B, are generally much smaller in magnitude. This, in part, might reflect the much greater precision of these estimates, but the point estimate for the capital/output outcome is not significantly different from zero, unlike the IV estimate. A standard story would be that OLS estimates are attenuated by measurement error. This seems a plausible contributor to bias in this context, with a crude self-reported measure of skill and conditional on a large number of fixed effects. However, other unobserved differences might also bias some of the OLS coefficients towards zero. A key unobservable might be the local outside option of low-skill workers. For instance, to take a Goldin and Sokoloff (1984) type of story, certain areas may have very productive agricultural land. In such areas, low-skill workers might drawn to the area but away from manufacturing, which could reduce the adoption of capital- and low skill-intensive production techniques.[21]

In the OLS estimates we do see a significant relationship between skills and plant size, but the relationship with horsepower, while still negative, remains insignificant.

As noted above, some research, including James and Skinner (1985) and Katz and Margo (2013), suggest that the nineteenth and twentieth century production technologies might have been quite different in their relationship between skills and capital usage. Table 6 explores this possibility by separating the sample into pre-1900 and post-1900 periods. Columns (1) and (2) examine capital per unit output. The IV estimates, in panel A, suggest that the capital-unskilled relative complementarity revealed in Table 5 was confined to the nineteenth century: there is a negatively and highly significant relationship between capital/output ratios and skill ratios in the 1860-80 period. This disappears by the 1910-20 data. To be fair, the twentieth relationship is very imprecise, consistent with either capital-skill complementarity (as previous research has

[21] It might be possible to evaluate this specific story empirically, something we could pursue in the future.

found) capital-unskilled complementarity, or capital neutrality. OLS estimates, in panel B, are also uninformative. In addition, we are missing data from a potentially key period of transition: between 1880 and 1910. So for the moment, are results do not say much about capital-skill complementarity in early twentieth century manufacturing (although the weight of evidence from before our study supports it).

The estimated relationship between skills and establishment size, shown in columns 3 and 4, is largely uninformative, though there is a positive, significant relationship in OLS in the nineteenth century. This is the opposite of what has been found in the past, e.g., in Katz and Margo (2013), although previous findings tend to be estimated at the plant, rather than market level. The relationship between skills and average wages (columns 5-6) strengthen across the centuries in both IV and OLS estimates, though not significantly so. This strengthening relationship between skills and wages is suggestive of a widening wage gap between skilled and unskilled workers over this period.

Data on horsepower usage is not available in the nineteenth century tabulations of the manufacturing censuses, however it is available in the original census records. We have found using similar regression in plant level data that horsepower is significantly negatively related to our skill measure in the the nineteenth century. This is consistent with the idea that power complemented unskilled labor relative to skilled labor in nineteenth century manufacturing.

We have also estimated Table 6 without controlling for industry fixed effects.[22] The coefficients are quite similar. For example, the coefficient (standard error) in the capital/output ratio regression is -0.410(0.149) in the nineteenth century and 0.026(0.319) in the twentieth century data. This suggests two things: (1) skill mix had very little impact on the location of industry; and, therefore (2) aggregate-level regressions – that is using county-level rather than county x industry data – should be relatively unbiased by the lack of controls for industry mix. The first point is consistent with evidence in modern data that industry mix is not a major source of adjustment to immigration-induced changes in skill mix (See Card and Lewis, 2007; Gonzales and Ortega, 2011; Lewis, 2003), although it is in contrast with findings in agricultural data from this same historical period, where crop-mix adjustments appear to dominate areas' adjustment to immigrant inflows (Lafortune et al., 2013). This finding thus helps to clarify that Heckscher-Ohlin trade theory, or at least Rybczynski-effects driven by the mix of workers' skills, seems not to hold very strongly in the U.S. manufacturing sector since at least 1860.

The second conclusion from this result allows us to turn to county-level data, which are richer geographically and in terms of variable detail, for additional analysis, without much concern that our results will be contaminated by endogenous shifts in industry mix. This is something we plan to do in future versions of the analysis.

[22] We thank Osbourne Jackson for this suggestion.

6 Conclusions

Our preliminary analysis suggests that immigration between 1860 and 1940 was a sufficiently important shock to the local labor force to alter skill ratios in urban counties. It also suggests that the capital stock, output, and average wages all responded to immigration-induced changes in skill ratios, a relationship which holds *within* industry as well as in the aggregate. These estimated responses provide strong support for the notion that capital and skill were *substitutes* in nineteenth century manufacturing, something which appears to have dramatically changed around the turn of the century. In our estimates we find evidence of this change but we are missing data in the key period of transition (1880- 1910), and so we are not currently able to pin down the precise timing of this change; this is something we will explore in the future as we continue assembling a richer dataset for that specific period. In any case the change appears to happen exactly around the time that innovations and technologies such as the factory and electricity started spreading across the US. Finally, we find no support for the idea that shifts in industry mix helped absorb immigrant inflows during the nineteenth and early twentieth centuries, although the response of capital to immigrant-induced skill mix changes could have helped mute the impact of immigration on the wage structure (something which we do not observe directly).

The current version of our analysis suffers from some limitations. First, we have examined a very crude measure of skill composition based on literacy. Not only might this not be a very relevant skill margin – especially towards the end of our period – but it may obscure more subtle relationship between skills and technology, such as the notion that technological advancements throughout this period were raising demand for skills as the "poles" of the skill distribution relative to the middle (see the arguments in Gray, 2013; Katz and Margo, 2013). Our measure of capital stocks is also very broad, though the same is true of many of the existing U.S. historical studies on manufacturing. Finally, we have not completed the process of data entry and verification, including the aforementioned lack of data at the turn of the century.

References

Angrist, Josh, "How Do Sex Ratios Affect Marriage And Labor Markets? Evidence From America's Second Generation," *The Quarterly Journal of Economics*, August 2002, *117* (3), 997–1038.

Atack, Jeremy and Fred Bateman, "Nineteenth-Century U.S. Industrial Development through the Eyes of the Census of Manufactures," *Historical Methods*, 1999, *32* (4), 177–188.

_ , _ , and Robert A. Margo, "Skill Intensity and Rising Wage Dispersion in Nineteenth-Century American Manufacturing," *Journal of Economic History*, March 2004, *64* (1), 172–192.

Autor, David H., Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*, November 2003, *118* (4), 1279–1334.

Aydemir, Abdurrahman and George J. Borjas, "Attenuation Bias in Measuring the Wage Impact of Immigration," Working Paper 16229, NBER July 2010.

Bandiera, Oriana, Imran Rasul, and Martina Viarengo, "The Making of Modern America: Migratory Flows in the Age of Mass Migration," *Journal of Development Economics*, 2013, *102* (C), 23–47.

Beaudry, Paul and David Green, "Wages and Employment in the United States and Germany: What Explains the Differences?," *American Economic Review*, June 2003, *93* (3), 573–602.

Berthoff, Rowland Tappan, *British Immigrants in Industrial America*, Cambridge: Harvard University Press, 1953.

Borjas, George J., "The Economics of Immigration," *Journal of Economic Literature*, December 1994, *32* (4), 1667–1717.

Brynjolfsson, Erik and Andrew McAffee, *Race Against the Machine: How the Digital Revolution is Accelerating Innovation and Irreversibly Transforming Employment and the Economy*, Lexington: Digital Frontier Press, 2011.

Card, David, "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, January 2001, *19* (1), 22–64.

_ **and Ethan Lewis**, "The Diffusion of Mexican Immigrants During the 1990s: Explanations and Impacts," in George J. Borjas, ed., *Mexican Immigration to the United States*, Univ 2007, pp. 193–228.

Chandler, Alfred D. Jr., *The Visible Hand: The Managerial Revolution in American Business*, Harvard University Press, 1977.

Cortes, Patricia, "The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data," *Journal of Political Economy*, June 2008, *116* (3), 381–422.

Friedberg, Rachel, "The Impact of Mass Migration on the Israeli Labor Market," *Quarterly Journal of Economics*, November 2001, *106* (4), 1373–1408.

Goldin, Claudia, "The Political Economy of Immigration Restriction in the U.S.," in Claudia Goldin and Gary Libecap, eds., *The Regulated Economy*, University of Chicago Press Chicago 1994.

— **and Kenneth Sokoloff**, "The Relative Productivity Hypothesis of Industrialization: The American Case, 1820 to 1850," *Quarterly Journal of Economics*, August 1984, *99* (3), 461–487.

— **and Lawrence F. Katz**, "The Origins of Technology-Skill Complementarity," *Quarterly Journal of Economics*, August 1998, *113* (3), 693–732.

Gonzales, Libertad and Francesc Ortega, "How do very open economies absorb large immigration flows? Recent evidence from Spanish regions.," *Labour Econonomics*, January 2011, *18* (1).

Gray, Rowena, "Taking Technology to the Task: The Skill Content of Technological Change in the Early Twentieth Century United States," *Explorations in Economic History*, July 2013, *50* (3), 351–367.

Griliches, Zvi, "Capital-Skill Complementarity," *Review of Economics and Statistics*, November 1969, *51* (4), 465–468.

Hamermesh, Daniel, *Labor Demand*, Princeton University Press, 1993.

Hatton, Timothy J. and Jeffrey G. Williamson, *The Age of Mass Migration*, Oxford: Oxford University Press, 1998.

— **and —, eds**, *Migration and the International Labor Market 1850-1939*, New York, NY: Routledge, 1994.

James, John A. and Jonathan S. Skinner, "The Resolution of the Labor-Scarcity Paradox," *Journal of Economic History*, September 1985, *45* (3), 513–540.

Jerome, Harry, *Mechanization in Industry*, Cambridge, MA: National Bureau of Economic Research, 1934.

Katz, Lawrence F. and Robert A. Margo, "Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective," Working Paper 18752, NBER February 2013.

Kim, Sukkoo, "Immigration, Industrial Revolution and Urban Growth in the United States, 1820-1920: Factor Endowments, Technology and Geography," Working Paper 12900, NBER February 2007.

Lafortune, Jeanne and José Tessada, "Smooth(er) Landing? The Role of Networks in the Location and Occupational Choice of Immigrants," Mimeo, Instituto de Economía, Pontificia Universidad Católica de Chile September 2013.

__ , __ , **and Carolina Gonzalez-Velosa**, "More Hands, More Power? The Impact of Immigration on Farming and Technology Choices in US Agriculture in the Early Twentieth Century," Working Paper 431, Instituto de Economía, Pontificia Universidad Católica de Chile March 2013.

Lewis, Ethan, "Local, Open Economies Within the U.S.: How Do Industries Respond to Immigration?," Working Paper 04-01, Federal Reserve Bank of Philadelphia 2003.

__ , "Immigration, Skill Mix, and Capital-Skill Complementarity," *Quarterly Journal of Economics*, May 2011, *126* (2), 1029–1069.

__ , "Immigration and Production Technology," *Annual Review of Economics*, 2013, *5* (1), 165–191.

Minnesota Population Center, "National Historical Geographic Information System: Version 2.0," Technical Report, University of Minnesota, Minneapolis, MN 2011.

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

Smith, Christopher, "The Impact of Low-Skilled Immigration on the Youth Labor Market," *Journal of Labor Economics*, 2012, *30* (1), 55–89.

Table 1: Manufacturing Variable Availability by Year

Variables	Years
Capital	1850-1920
Wages	1860-1940
Establishments	1860-1940
Workers	1850-1940
Material Costs	1860-1940
Products Value	1850-1940
Horsepower	1910-1930
Fuel/Power Costs	1890, 1910-1930

Table 2: Descriptive Statistics on Area x Industry Sample

Variable	Mean	Standard Deviation	No. of Area x Ind Cells
ln(Capital/Worker)	7.034	1.089	20,102
ln(Output/Worker)			
...Capital Sample:	7.764	0.883	20,102
...Full Sample:	7.960	0.956	24,536
ln(Capital/Output)	-0.731	0.721	20,102
ln(Establishments/Worker)	2.481	1.261	25,653
ln(Wages/Worker)	6.313	0.671	24,528
ln(Horsepower/Output)	-8.198	1.279	7,495
Area Level Variables:			
ln(Skill Ratio)	2.500	0.979	25,715
Instrument	2.224	0.941	25,715

Unweighted means. Full sample comprised of 126 "areas" (counties, except for New York City), and 429 industries over the years 1860,1870,1880,1910,1920,1930, and 1940. Skill Ratio is literate/non literate population older than 15, except in 1940 when fewer than 2 years of education proxies for literacy.

Table 3: Top Locations of Immigrants in 1850, by Origin

County	Percent	County	Percent
English		German	
New York	11.0%	New York	10.9%
Philadelphia	5.4%	Hamilton, OH (Cincinnati)	6.6%
Oneida, NY (Utica)	1.8%	St. Louis, MO	4.9%
Hamilton, OH (Cincinnati)	1.7%	Philadelphia	3.9%
Allegheny, PA (Pittsburgh)	1.6%	Baltimore	3.9%
Irish		Italian	
New York	17.4%	New York	21.2%
Philadelphia	7.4%	New Orleans	18.2%
Boston	3.7%	Hamilton, OH (Cincinnati)	5.0%
Middlesex, MA (Cambridge+)	2.3%	San Francisco	4.3%
New Orleans	2.2%	Philadelphia	3.6%
Canadian		Russian/Polish	
Boston	2.8%	New York	29.0%
New York	2.8%	New Orleans	3.9%
Middlesex, MA (Cambridge+)	2.4%	St. Louis, MO	3.5%
Wayne, MI (Detroit+)	2.0%	San Francisco	3.5%
Worcester, MA	1.8%	Philadelphia	3.2%

Source: ancestry.com.

Table 4: First stage regressions

	Full Sample (1)	Capital (2)	Horsepower (3)	Before 1900 (4)	1900 and later (5)
Pred. ln(skill Ratio)	0.899***	1.207***	0.551**	1.599***	0.484**
	(0.226)	(0.302)	(0.253)	(0.442)	(0.188)
R-Squared	0.894	0.879	0.936	0.788	0.900
N	25,394	19,786	7,372	14,028	11,366
Fixed Effects:					
Industry x Century	Y	Y	Y	Y	Y
Area x Century	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y

Outcome is ln(literate/not literate) in the age 15+ population of the area. Standard errors in parentheses, calculated to be robust to arbitrary error correlation with area (=county, except New York City). Significance levels: * 0.1%, ** 0.05%, *** 0.01%.

Table 5: Manufacturing outcomes, All Available Years

	Capital per worker (1)	Output per worker (2)	Capital per output (3)	Workers per establishment (4)	Wages per establishment (5)	Horsepower/ output
			Panel A: Instrumental Variables			
ln(skill ratio)	-0.004	0.315***	-0.319**	0.154	0.283**	-1.954
	(0.112)	(0.107)	(0.133)	(0.152)	(0.125)	(1.543)
R-Squared	0.678	0.714	0.347	0.540	0.749	0.540
Root MSE	0.629	0.474	0.612	0.893	0.334	0.908
Red. Form F-stat	0.001	18.49***	5.38**	1.11	6.39**	1.86
			Panel B: Ordinary Least Squares			
ln(skill ratio)	0.083***	0.057***	0.027	0.101***	0.056***	-0.354
	(0.031)	(0.021)	(0.038)	(0.027)	(0.020)	(0.368)
R-Squared	0.679	0.727	0.378	0.540	0.765	0.592
Root MSE	0.628	0.464	0.598	0.893	0.323	0.856
N	20,102	20,102	20,102	25,653	24,528	7,495
Fixed Effects:						
Industry x Century	Y	Y	Y	Y	Y	Y
Area x Century	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y

All outcomes in logs. Right-hand side variable is ln(literate/not literate) in the age 15+ population of the area. Columns (4) and (5) include data for 1860 to 1940 while column (1)-(3) are restricted to 1860-1920. Standard errors in parentheses, calculated to be robust to arbitrary error correlation with area (=county, except New York City). Significance levels: * 0.1%, ** 0.05%, *** 0.01%.

Table 6: Manufacturing outcomes, by Century

	Capital/Output		Workers/Establishment		Wages/Worker	
	Pre-1900	Post-1900	Pre-1900	Post-1900	Pre-1900	Post-1900
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Instrumental Variables						
ln(skill ratio)	-0.366**	0.029	0.117	0.416	0.223*	0.391
	(0.153)	(0.316)	(0.145)	(0.373)	(0.119)	(0.327)
R-Squared	0.228	0.473	0.497	0.525	0.396	0.696
Root MSE	0.701	0.459	0.848	0.930	0.373	0.284
Red. Form F-stat	7.73***	0.01	0.56	1.99	6.65**	1.37
Panel B: Ordinary Least Squares						
ln(skill ratio)	0.041	-0.071	0.130***	0.012	0.039**	0.113
	(0.040)	(0.055)	(0.030)	(0.050)	(0.016)	(0.080)
R-Squared	0.285	0.474	0.497	0.529	0.428	0.711
Root MSE	0.674	0.458	0.848	0.925	0.363	0.277
N	14,024	5,762	14,024	11,308	14,020	10,188
Fixed Effects:						
Industry x Century	Y	Y	Y	Y	Y	Y
Area x Century	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y

All outcomes in logs. Right-hand side variable is ln(literate/not literate) in the age 15+ population of the area. Columns (4) and (5) include data for 1860 to 1940 while column (1)-(3) are restricted to 1860-1920. Standard errors in parentheses, calculated to be robust to arbitrary error correlation with area (=county, except New York City). Significance levels: * 0.1%, ** 0.05%, *** 0.01%.