Labor Market Performance of Immigrants in Early Twentieth-Century America

Gihoon Hong[1] and Soyoung Lee[2]

Abstract

This paper examines the labor market performance of immigrants during the early twentieth-century America. In particular, we consider the immigrants from Europe, Asia, and South America who were present in America between 1900 and 1930 to investigate the degree of human capital transferability across the regions of origin. Accounting for the changes in unobserved immigrant quality across cohorts, we estimate the returns to work experience in the American labor market. The estimation results show that European immigrants assimilated into the American labor market fairly well while immigrants from Asia and South America did not, which is suggestive of imperfect human capital mobility. In addition, the estimation results obtained from synthetic cohorts indicate that the quality of immigrants was systematically different across years of arrival.

1   Introduction

One feature that characterizes early twentieth century America is the massive inflow of foreign-born workers. Borjas (1994) reports that 25.8 million persons entered the country during the First Great Migration between 1881 and 1924.Consequently, the labor market performance of immigrants in this period in the U.S. has received much attention.
Existing studies on immigrant assimilation generally find that recent immigrants who arrived in the U.S. during the 1960's or after fared quite a low wage at first,although their wage grew fast enough to show that they assimilated into the U.S. labor market relatively well (Chiswick, 1978). However, economists do not seem to agree on the labor market performance of the immigrants in earlier periods. For example, Higgs (1971) finds that

---

[1]Indiana University South Bend, USA.
[2]Indiana Wesleyan University, USA.

there was little difference in wages between foreign-born and native workers in the early 1900's using data collected by the U.S. Immigration Commission. On the other hand, using the state-level survey data for Michigan and California in late nineteenth century,Hanes (1996) reports that immigrants from the northwestern Europe experienced slower growth in wages than native-born workers.Minns (2000) uses the U.S. census data of 1900 and 1910 and considers the labor market performance of immigrants, and by distinguishing occupational sectors between blue-collar occupations and white-collar occupations, he concludes that foreign-born workers in this period experienced even faster growth in earnings than native-born Americans did in the same occupational sector. There are a number of reasons why existing studies provide radically different implications for immigrant assimilation in this period. First, the data used in these studies is limited because rich data sets that include earnings and enough observations on foreign workers are rare. Also, the lack of panel data sets makes it difficult to control for observed and unobserved person specific heterogeneity, possibly leading to biased estimates. Borjas (1994) shows that estimating the age-earning profile using a single cross-sectional data would generate biased estimates if the unobserved qualities of immigrants were systematically different across cohorts. Finally, the existing studies do not consider the possibility that immigrant's nativity may affect the transferability of human capital acquired abroad. As Chiswick (1978) noted, knowledge and skills are not perfectly mobile across countries, and language barriers or different cultural backgrounds may reduce the value of human capital more in the case of immigrants from non-European countries.

Using individual data from IPUMS, we estimate new specifications to study the immigrant assimilation process in the early twentieth century America. In order to assess immigrant labor market performance in this period, we use the Duncan Socioeconomic Index (SEI) scores as a measure of occupational status.[3] Our contribution to the literature is twofold. First, we extend the sample to include all immigrants from South America and Asia as well as European immigrants. By considering the nativity of immigrants as a source of variation, wetest whether human capital transferability varies across countries. In order to obtain sufficient observations for the non-European immigrants, our data set is constructed from the four census years in IPUMS ranging from 1900 to 1930. In addition, following Borjas (1994), we perform a syntheticcohorts analysisto control for the cohort effects. By comparing the results from both cross-sectional and cohort regressions, we are able to detect whether there existed any trends in the quality of newly arriving immigrants across cohorts.

The key results of the paper are summarized as follows: First, we show that the labor market performance of immigrants in the early twentieth century varies considerably across regions of origin. The results show that European immigrants assimilated in the American labor market fairly well as shown by the coefficient estimate of aging effects very close to its counterpart for American natives. In contrast, we find that Asian and South American immigrants did not experience any significant growth in occupational status over time during this period. The results also point to imperfect human capital mobility in that human capital accumulated in Asia or South America was valued less

---

[3]Hanes (1996) also use the same variable to analyze immigrants' labor market performance. However, he considers the single census year of 1900 and only European immigrants are included in his empirical analyses.

during the period than that in Europe. This indicates that language barriers or different cultural backgrounds negatively affected the degree of human capital transferability for non-European immigrants. Finally, estimating returns to host-country experience for the synthetic cohorts, we provide evidence that the unobserved quality or skills of European immigrants improved, while that of Asian immigrants declined over the same period.

The rest of the paper is structured as follows. Section 2 discusses the data. The empirical model and results are presented in Section 3. We conclude in Section 4.

2   Data

Our main data set comes from the U.S. Censuses provided by the IPUMS project at the Minnesota Population Center ofthe University of Minnesota.[4] This data set provides individual-level data on personal characteristics, labor market status and immigration status. In particular, we consider census years from 1900 to 1930, and only individuals between sixteen years and sixty five years old are included in the sample.[5]We choose to focus on the early 20th century rather than the late 18th century because this is the period in which the immigration flows from South America and Asia increased. The descriptive statistics of main variables is given in Table 1.

The description of each variable follows here.[6]LITERACY is a dummy that indicates whether the respondent could read and in any language. We usethe variable to proxy for the educational attainment, because schooling is not available. URBAN and LABFORCE are dummy variables that indicate the location and the labor market status of the respondent, respectively. YRSUSA reports how long a person who was born in a foreign country or U.S. outlying area had been living in the United States. ENGLISH is a dummy thatindicates whether the respondent was able to speak English, and SEI is a constructed variable that assigns a Duncan Socioeconomic Index (SEI) score to each occupation. SEI is a measure of occupational status based upon the income level and educational attainment associated with each occupation in 1950, and the score was derived by using median income and education levels for men in 1950 to predict prestige assessments from a 1947 survey (of a select group of occupations).[7] Since the wage data was not collected until 1940 Census, we use SEI to proxy for wage in estimation.

---

[4]The data are available at https://usa.ipums.org/usa/.

[5]In order to maximize the number of observations used in estimation, we used the 5% samples from the U.S. Census when available (1900, 1930) and the 1% samples, otherwise.

[6]Immigrants from Africa or Oceania were omitted due to the small number of observations.

[7]See Duncan (1961) for information on the construction of the variable.

Table 1: Descriptive Statistics

	Natives	All Immigrants	Europeans	South Americans	Asians
AGE (mean)	34.31	39.10	39.52	34.41	36.90
MALE	78.06%	84.44%	85.04%	84.23%	93.64%
MARRIED	53.75%	61.78%	63.15%	55.88%	48.16%
URBAN	51.02%	73.40%	74.78%	58.18%	58.53%
LITERACY	93.59%	90.22%	90.81%	73.56%	80.87%
SEI (mean)	27.24	24.25	24.31	15.22	20.74
YEARS IN THE U.S. (mean)	N/A	19.74	20.10	13.61	16.05
ENGLISH	96.19%	89.42%	90.85%	55.27%	69.39%
(YEAR==1900)	29.74%	34.48%	35.14%	11.02%	22.83%
(YEAR==1910)	7.75%	9.50%	9.79%	5.65%	10.68%
(YEAR==1920)	8.76%	9.29%	9.45%	10.14%	9.52%
(YEAR==1930)	53.76%	46.73%	45.63%	73.19%	56.96%
Observations	3568471	814955	678304	28356	24066

From the Table 1, we can make a few comments on the composition of immigrants in early 20th century. First, even if the characteristics of natives andall immigrants groups are generally similar, we see that the characteristics differ greatly across the immigrant's place of origin. For example, European immigrants look similar to natives in most characteristics except location. South American and Asian immigrantsappear to have lived in more rural areas relative to Europeans, and their education level measured by LITERACY seems to be quite lower than natives or European immigrants. This group also possessed poor English skills and SEI variable shows that their labor market performance was not as strong as natives. In addition, South American and Asian immigrants differ from European immigrantsin that most of them were single male and they arrived in the U.S. more recently (the average duration of residence were around 13-16 years as opposed to 20 years in the European immigrants group). This suggests that the immigration flow from South America and Asia is a recent phenomenon, relative to immigration from other regions.

3   Empirical Specification and Results

3.1 Ordinary Least Square Regression

The main regression equations are specified in this section. We start with the following generic models that have been used in the literature.

$$SEI_{ij} = X_j \varphi_i + \delta_{1i} AGE_j + \delta_{2i} AGE_j^2 + \alpha y_i + \beta C_j + \gamma_i year_j + \varepsilon_{ij} \qquad (1)$$
$$SEI_{nl} = X_l \varphi_n + \delta_{1n} AGE_l + \delta_{2n} AGE_l^2 + \gamma_n year_l + \varepsilon_{nl} \qquad (2)$$

where SEI_{ij} and SEI_{nl} are the Duncan occupation index of immigrant j and native l; X includes socioeconomic variables such as MALE, MARRIED, URBAN, LITERACY, etc.; $y$is the number of years that the immigrant has resided in the U.S.; $C$is the calendar year of arrival in the U.S.; *year* is a vector of dummy variables, where *year*varies from 1910 to 1930, that indicates if the observation is drawn from the specific Census.; and ε captures individual specific time shocks to the labor market performance that is not explained by other variables in the model. Also, because information on the work experience is hard to obtain for the period of interest, we include AGE and AGE^2 to capture the effect of general human capital on the labor market performance. However, Borjas (1994) shows that this specification suffers from the usual collinearity problem, because the YEARS IN THE U.S. variable, denoted by y_i, is a linear combination of the period effect, captured by γ, and the cohort effect, captured by β.Therefore, weneed to make an additional restriction to ease the identification of key parameters that the period effects are the same for immigrants and natives, or:

$$\gamma_i = \gamma_n = \gamma. \tag{3}$$

Equation (3) shows that our empirical specification controls for the period effects using time fixed effects, which implies that the relative labor market performance of immigrants and natives is independent of secular trend. Then, the age-occupation status profiles of immigrants and natives converge if$\delta_i + \alpha > \delta_n$,assuming immigrants were employed in the occupation with lower SEI than natives at the time of arrival. The OLS estimation results from equations (1) and (2)are presented in Table 2.

Table 2: The labor market performance of immigrants by nativity (OLS model)

VARIABLES	(1) Natives SEI	(2) All Immigrants SEI	(3) Europeans SEI	(4) South Americans SEI	(5) Asians SEI
AGE(δ)	0.577***	0.0600***	0.00732	-0.0359	0.187***
	(0.00341)	(0.00749)	(0.00828)	(0.0297)	(0.0459)
AGE squared	0.00986***	0.00332***	0.00286***	0.00170***	0.00434***
	(6.96e-05)	(0.000140)	(0.000154)	(0.000618)	(0.000893)
MALE	-4.113***	0.776***	1.466***	-2.260***	-1.470***
	(0.0277)	(0.0628)	(0.0698)	(0.244)	(0.525)
MARRIED	0.921***	3.106***	3.241***	0.776***	5.704***
	(0.0261)	(0.0512)	(0.0564)	(0.194)	(0.277)
URBAN	12.83***	9.069***	9.110***	5.574***	12.22***
	(0.0225)	(0.0496)	(0.0549)	(0.188)	(0.263)
LITERACY	12.03***	6.988***	6.927***	3.925***	4.307***
	(0.0452)	(0.0774)	(0.0861)	(0.217)	(0.343)
(YEAR==1910)	1.628***	1.628	1.628	1.628	1.628
	(0.0436)	(--)	(--)	(--)	(--)
(YEAR==1920)	2.392***	2.392	2.392	2.392	2.392
	(0.0417)	(--)	(--)	(--)	(--)
(YEAR==1930)	2.750***	2.75	2.75	2.75	2.75
	(0.0253)	(--)	(--)	(--)	(--)
ENGLISH		5.334***	4.675***	3.758***	5.592***
		(0.0768)	(0.0888)	(0.196)	(0.299)
YEAR OF IMMIGRATION		-0.0285***	-0.0205***	-0.207***	-0.0188*
		(0.00161)	(0.00177)	(0.00871)	(0.0103)
YEARS IN THE U.S. (α)		0.216***	0.248***	-0.0943***	0.228***
		(0.00281)	(0.00309)	(0.0132)	(0.0193)
CONSTANT	4.511***	53.07***	37.90***	402.0***	33.41*
	(0.0550)	(3.072)	(3.373)	(16.73)	(19.76)
Observations	3,568,471	814,182	678,236	28,356	23,393
R-squared	0.148	0.094	0.089	0.109	0.176

Notes: Standard errors are in parentheses. The parameter estimates of the year dummies for all immigrants, Europeans, South Americans, and Asians were restricted, not estimated, for identification. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 2 shows that most coefficients estimates are highly significant and the effects of key variables on the labor market performance are consistent with expectations. Most importantly, the aging effect for the all immigrants group, $\delta_{1i} + \alpha = 0.06 + 0.216 = 0.276$, is smaller than the aging effect for natives, $\delta_{1n} = 0.577$, which implies that the labor market performance of immigrants was not converging to that of natives in this period. This is consistent with the findings in the previous literature. However, the coefficient estimate of YEARS IN THE U.S. (= 0.216) for the all immigrants sample is positive and statistically significant. This suggests that immigrants accumulated host-country-specific human capital as they resided longer in the U.S., which lead to promotions in the occupation status over time. To investigate this further, we run the OLS regressions again by allowing the coefficients to vary over the place of origin of immigrants: Europe, South America, Asia.

Columns (3) – (5) in Table 2 reveal that after considering immigrants' home region in the regression, there is considerable variation in the labor market performance over time as well as in the transferability of human capital and the quality of immigrants across regions. First of all, the results in Column (3) show that European immigrants in this period have experienced aconsiderable rise in their labor market status after they arrive.However, the insignicant estimate of AGE suggests that their source-country human capital was not valued well, which caused them fail to catch up with the native workers in the American labor market. This empirical finding is not consistent with prior expectation given the cultural and linguistic similarity between European countries and America in this period. Moreover, a comparison of the aging effect for South Americans and Asians also reveals some interesting facts regarding the assimilation process in early immigrants from non-European countries.The largest aging effect found in Column (5) indicates that immigrants from Asian countries in the early twentieth-centuryAmerica did not suffer much from the loss of human capital much. In addition, the large and significant duration effect (0.228), as denoted by α, implies that Asian immigrants actively acquired host-country-specific human capital. Combined with the pure age effect (δ), an additional year of stay in the U.S. caused Asian immigrants to experience the fastest growth among all the immigrant groups in terms of the Duncan occupation index. On the other hand, the results in Column (4) seem to suggest that South American immigrants suffered the most from imperfect human capital transferability and, as a result, failed to assimilate well.

The overall message from the OLS regression results in Table 2 is that immigrants in the early twentieth century America did not assimilate well. However, there still exists substantial variation in the aging effects across regions of immigrant origin: European immigrants experienced moderate improvement over time in terms of occupation status, although they still failed to catch up with their native counterparts. While Asians seem to have enjoyed the largest returns to labor market experience in America, South American immigrants were shown to have struggled with assimilation. There are a couple of possible explanations for the sharp differences across regions of origin in the labor market experience of early immigrants.The first hypothesis is that immigrants are favorably self-selected for labor market success (Chiswick, 1999). This implies that due to the substantial costs associated with migration, only those with higher human capital (in the form of either higher education or innate ability) would find immigration a worthwhile investment. Then, assuming that the distributions of human capital are similar across different source countries, this type of positive selectivity would be more intense as the

cost of migration rises. Considering the geographic distance between the source and the host country is a good proxy for the migration costs, one may conclude that Asian immigrants in this period are the most favorably selected. Although the positive selection theory provides a prediction that is consistent with the OLS results in Table 2, it is not feasible to empirically test the hypothesis because it requires having data on the source country labor markets.

An alternative explanation for the results in Table 2 is that the cross-sectional nature of the regression may yield erroneous results ifthere exist systematic differences in productivity across immigrant cohorts. Borjas (1994) shows how the cross-section relationship between the labor market performance of immigrants and years since migration may produce biased estimates in the presence of cohort effects. Given the lack of panel data on immigrants in this period, the literature has addressed this concern by creating synthetic cohorts by tracking specific immigrant waves across the decennial Censuses (Borjas, 1994). We now turn to the regression analysis using the synthetic cohorts model to account for the cohort effects.

3.2 Age-occupation Status Profile

In this section, following the literature, we construct synthetic cohorts to explicitly control for cohort effects and period effects. One way to do this is to consider immigrants who were in their 20s and came to the U.S. 5 years ago or less as of 1900. Then, we follow this specific immigrant group by tracking individuals who were in their 30s and spent more than 5 years but less than 15 years in the U.S. in 1910 Census, as well as others who were in their 40s and spent more than 15 years and less than 25 years in the U.S. in 1920 Census, and so on. This way, we can ease the problem that arises when estimating age-earning profile using one cross-sectional data set. The results are presented in Table 3.

First of all, it is evident that the total aging effect ($\alpha + \delta$) is greater than the OLS estimate as found in Table 2 (0.52 vs. 0.27). This suggests that the quality of immigrant cohorts are systematically different in this sample, causing the cross-sectional study to produce spurious estimates. In addition, the aging effect for the European immigrant sample also turns out to be much larger now, even close to that of the native group (0.517 vs. 0.577). Again, the sharp difference between the regression results provide evidence of cohort effects. In particular, the fact that the results from the synthetic cohort analysis produces a greater aging effect indicates that the unobserved quality of European immigrants had improved over time, which causes the cross-sectional regression to understate the returns to host-country labor market experience.

Table 3: The labor market performance of immigrants by nativity
(Synthetic cohort model)

VARIABLES	(1) All Immigrants SEI	(2) Europeans SEI	(3) South Americans SEI	(4) Asians SEI
AGE(δ)	0.278***	0.253***	-0.150	0.0150
	(0.0531)	(0.0577)	(0.343)	(0.287)
AGE squared	-0.00643***	-0.00632***	-0.00285	-0.00214
	(0.000814)	(0.000875)	(0.00543)	(0.00425)
MALE	2.226***	2.507***	-0.668	0.704
	(0.253)	(0.277)	(1.750)	(2.051)
MARRIED	3.710***	3.737***	1.244	5.987***
	(0.207)	(0.223)	(1.312)	(1.002)
URBAN	7.832***	7.765***	6.875***	12.05***
	(0.192)	(0.208)	(1.305)	(0.915)
LITERACY	5.633***	5.378***	4.867***	2.421**
	(0.250)	(0.263)	(1.390)	(1.194)
ENGLISH	5.074***	4.372***	3.616***	6.673***
	(0.250)	(0.273)	(1.381)	(0.978)
YEAR OF IMMIGRATION	-0.0144	0.0655	-0.0883	-0.00930
	(0.0405)	(0.0436)	(0.271)	(0.222)
YEARS IN THE U.S. (α)	0.242***	0.264***	0.313	0.267
	(0.0307)	(0.0330)	(0.198)	(0.164)
CONSTANT	24.65	-127.1	176.8	17.75
	(76.70)	(82.46)	(512.5)	(420.0)
Observations	51,145	43,507	804	1,871
R-squared	0.119	0.118	0.123	0.180

Notes: Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

On the other hand, Columns (3) and (4) reveal that neither South American or Asian immigrants assimilate in this periods. Even though we cannot completely exclude the possibility that the results are driven by the small number of observations after constructing the synthetic cohorts, the statistically insignificant estimates of both the age effect and the duration effect seem to suggest that (relatively) more recent cohorts from Asian countries arrived with progressively lower human capital, which causes the previous results in Table 2 to overstate the returns to host-country experience. Clearly, immigrants from Asia in this period had suffered from the language barrier or the lack of

host country-specific human capital.

Although the synthetic cohorts may not be the representatives for the corresponding immigrant groups because of the small number of observations, Figure 1 reveals an interesting fact about the change in relative labor market performance of immigrants over time. After the cohort and the period effects are controlled,all immigrants suffer from the imperfect transferability of human capital upon arrival, but Europeans immigrants suffered significantly less than Asians or South Americans. Interestingly, no immigrant group was able to converge to the native group in terms of labor market performance in this period. This is in contrast with the findings from studies of postwar America.[8]

4   Conclusion

This paper presents an empirical analysis on the labor market performance of immigrants from various regions in the early twentieth century America. Considering a measure of occupational status as proxy for the labor market performance of immigrants, we investigate the transferability and the quality of human capital of immigrants across the regions.

The results from regression exercise suggest that labor market performance of immigrants in the early twentieth century significantly varied across different immigrants groups depending on the places of origin. While immigrants from European countries are shown to have experienced considerable improvement in their occupation status over time, similar to their American counterparts in this period, we find returns to host-country experience for Asian and South American immigrants to be statistically insignificant. The results point to imperfect transferability of human capital across regions, potentially arising from the cultural and linguistic differences between the source and host country.

In addition, constructing synthetic cohorts to account for unobserved cohort and period effects, we find that immigrant quality changed substantially across cohorts. In particular, the estimation results indicate that the unobserved quality of European immigrants improved over time while that of Asian immigrants declined during the sample period.

References

[1]   Borjas, G., The Economics of Immigration, *Journal of Economic Literature*, **32**, (1994), 1667-1717.

[2]   Chiswick, B., The Effect of Americanization on the Earnings of Foreign-Born Men, *Journal of Political Economy,* **86**, (1978),897-921.

[3]   Chiswick, B., Are Immigrants Favorably Self-Selected?, *American Economic Review,* **89**, (1999),181-185.

[4]   Duncan, O. D., A Socioeconomic Index for All Occupations. (O. D. Duncan, P. K. Hatt, C. C. North, & A. J. Reiss, Eds.) *Occupations and Social Status,* New York: Free Press of Glencoe, 1961.

---

[8]Using cross-sectional analysis, Chiswick (1978) shows that log earning of immigrants in the 1970s U.S. overtook that of natives within 15 years after migration.

[5]   Hanes, C.,Immigrants' Relative Rate of Wage Growth in the Late 19th Century. *Explorations in Economic History, 33*, (1996), 35-64.

[6]   Higgs, R., Race, Skill, and Earnings: American Immigrants in 1909. *Journal of Economic History, 31*, (1971), 420-438.

[7]   Minns, C., Income, Cohort Effects, and Occupational Mobility: A New Look at Immigration to the United States at the Turn of the 20th Century. *Explorations in Economic History, 37*, (2000), 326-350.