Revisiting the Slowdown in the Economic Assimilation of Immigrants: Limitations to the Language Based Model of Human Capital

Lee Huang, Albert Kuo, Anthony Pan, Leila Safavi

April 9, 2016

Abstract

This paper examines immigrant earnings and how the rate of economic assimilation for immigrants in the United States has changed over time not only in their entry wages, but also their rate of wage growth. Several factors influence the observed cohort effects on the immigrant wage gap, including immigration period, national origin, education level, state of residence, economic shocks, and English language proficiency. While previous studies have singled out English proficiency as the main driver behind the immigrant wage gap, we found that the importance of English proficiency on wage growth varies over time. Moreover, other factors that reflect the changing composition of immigrants also play an important role in the wage gap.

1 Introduction

Immigration to the United States has increased significantly since the 1970's – the number of immigrants rose from 9.6 million in 1970 (4.7% of the US population) to 40 million in 2010 (12.9% of the US population).[1]

The assimilation of immigrants has been explored extensively in economic literature. In particular, Borjas (2015) showed that both immigrant entry wages and the rate of wage growth in the US differ based on the time of immigration: people who immigrate to the US in more recent cohorts have lower entry wages and a lower rate of wage growth. Furthermore, he argued that these differences are not due to changes in macroeconomic conditions, changing national origin composition of the immigrant population, geographic settlement of newer immigrant waves, or labor supply effects of the Great Recession. Instead, the decline in the rate of wage growth is caused by a reduction in the rate of human capital accumulation, namely English language proficiency.

Our paper verifies and extends Borjas' findings by examining the role of English language proficiency over time. In light of the limited impact of high proficiency language skills on wages, we then investigate alternative hypotheses that may explain recent cohorts of immigrants' slower rate of assimilation.

[1] http://www.migrationpolicy.org/article/frequently-requested-statistics-immigrants-and-immigration-united-states

2 Data and Methodology

2.1 Data

Our data consists of 2.75 million observations drawn from the 1960, 1970, 1980, 1990, and 2000 censuses, including samples from the 2009-2011 American Community Survey which serve as a proxy for the 2010 census. Because the census data lacks certain country indicators, we merged education and income data from Barro and Lee, Penn World Tables, and the World Bank with our census data. We divided the immigrant population into ten cohorts by their entry date into the United States and created a cohort for non-immigrants included in the sample. Each cohort consists of males between the ages of 18-65 actively employed in the labor force and not enrolled in school. Immigrant cohorts only include those who arrived at or after age 18. We also created five cross-sections within our data for the years 1970, 1980, 1990, 2000 and 2010, following the decennial census. After the selection criterion, the total sample size that was used for analysis consisted of 515,713 individuals, of which 37,084 are immigrants.

Because data for the exact number of weeks each individual works per year had many missing values, we imputed the missing values with another variable that describes the number of weeks an individual works per year in intervals. We took the mean of the interval that corresponded with each individual – for example, if an individual responded that they worked between 40 and 47 weeks, we estimated that they worked 43.5 weeks per year.

2.2 Methodology

Similar to Borjas, we used a fixed effects model in order to estimate the cohort effect on wages – namely, that later cohorts of immigrants demonstrate lower earnings potential than earlier cohorts. We utilized the following regression model to estimate the cohort fixed effects:

$$\log w_{l\tau} = \phi_{c\tau} + X_{l\tau}\beta_\tau + \epsilon_{l\tau}$$

where $\log w_{l\tau}$ represents weekly earnings of a person l in cross-section τ, X is a vector of age as a third-order polynomial, and $\phi_{c\tau}$ is a vector of fixed effects, particularly each cohort's entry year for the first analysis as seen later in Table 1. The fixed effects model allows us to identify the difference between native and immigrant workers while demeaning each cohort, since we expect individual cohorts to have different earnings potentials.

After replicating Borjas' study (2015, Table 1), we focused on the language proficiency portion of Borjas' research. We analyzed the differences in acquisition of English proficiency between each immigrant cohort using a logit model. Furthermore, we ran a regression model to examine the role of English proficiency on wages with a second order factor.

Finally, we ran a linear regression model across the entire sample with log weekly earnings as the dependent variable and cohort group as the main explanatory variable. The model controlled for year of the census, the age, English proficiency, and race of the individual, and the GDP, fertility rate, and population of their origin country; the significance of these factors, separate and combined, were analyzed.

3 Analysis

3.1 Age-Adjusted Linear Regression Model

We begin by recreating Borjas' analysis of age-adjusted wages and immigration cohort. Table 1 reports cohort fixed effects with cohorts based on 10 year periods of immigrant arrival. Typically, the longer a cohort remains in the US, the more economically assimilated they become. This is presented as narrowing differentials between immigrant and native weekly earnings over time. The results from Table 1 indicate that immigrant cohorts after 1985 appear to have lower rates of economic assimilation relative to previous cohorts.

In comparison to the results from Borjas (2015), our results appear similar, which is expected since we used the same regression model: $\log w_{l\tau} = \phi_{c\tau} + X_{l\tau}\beta_\tau + \epsilon_{l\tau}$. There are small differences in our outputs due to our use of a smaller dataset. However, the trend of wage differential reduction over time remains similar for each of the cohorts. Since we used a much smaller dataset than Borjas did, the estimated differentials in Table 1 have larger standard errors [2].

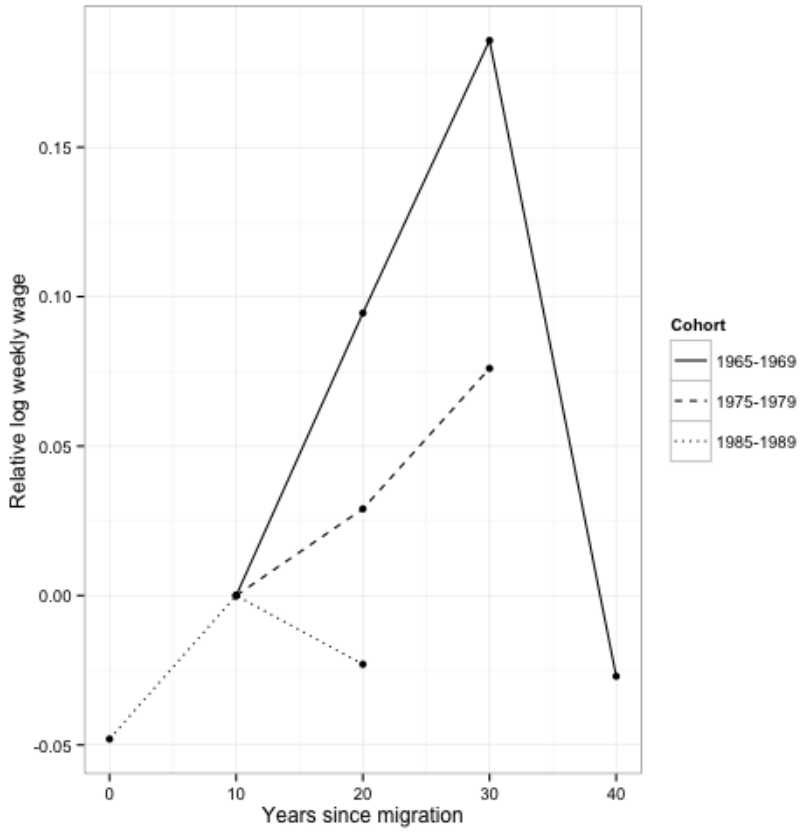

Figure 1: Aging profiles of relative wage of immigrants by cohort, obtained from Table 1. The relative log weekly wage is normalized to 0 at 10 years after entry, due to missing values in the data.

[2]Borjas reported standard errors of less than .01; our table reports standard errors up to -16.76.

Table 1: Age-Adjusted Relative Weekly Earnings of Immigrant Cohorts, by Census Cross Section

	Dependent variable: Log Weekly Wages				
	1970	1980	1990	2000	2010
1950-59 arrivals	0.042^{***}	0.014^{***}	0.125^{***}	0.056^{***}	
	(0.001)	(0.003)	(0.004)	(0.006)	
1960-64 arrivals	-0.053^{***}	-0.034^{***}	-0.009^{***}	0.049^{***}	
	(0.0002)	(0.002)	(0.003)	(0.005)	
1965-69 arrivals		-0.157^{***}	-0.062^{***}	0.029^{***}	-0.184^{***}
		(0.001)	(0.003)	(0.004)	(0.004)
1970-74 arrivals	-0.236^{***}	-0.261^{***}	-0.147^{***}	-0.112^{***}	-0.074^{***}
	(0.001)	(0.001)	(0.002)	(0.004)	(0.005)
1975-79 arrivals			-0.214^{***}	-0.185^{***}	-0.138^{***}
			(0.001)	(0.003)	(0.006)
1980-84 arrivals		-0.270^{***}	-0.294^{***}	-0.200^{***}	-0.143^{***}
		(0.001)	(0.001)	(0.002)	(0.006)
1985-89 arrivals			-0.304^{***}	-0.256^{***}	-0.279^{***}
			(0.001)	(0.001)	(0.004)
1990-94 arrivals			-0.284^{***}	-0.229^{***}	-0.249^{***}
			(0.003)	(0.001)	(0.002)
1995-99 arrivals				-0.223^{***}	-0.193^{***}
				(0.003)	(0.002)
2000-04 arrivals				-0.208^{***}	-0.315^{***}
				(0.004)	(0.003)
2005-09 arrivals					-0.320^{***}
					(0.004)
Observations	61047	110393	119100	128931	26715
R^2	0.179	0.133	0.153	0.116	0.131

Note: Standard errors are in parentheses and are clustered at the cohort level. The age-adjusted wage differentials between each immigrant cohort and natives are calculated from a regression estimated separately in each cross section. The dependent variables is log weekly wages with adjustments made for worker age. *p<0.1; **p<0.05; ***p<0.01.

Figure 1 depicts the general trend that earlier cohorts have a faster increase in relative weekly wage and assimilate more quickly. However, at 40 years after immigration, the cohort that immigrated between 1965-1969 experiences a sharp drop in weekly wage. This is not seen in Borjas' analysis, which suggests that it is a result of statistical variation due to sampling rather than deeper socioeconomic factors we might be interested in. Only 2319 individuals from the 1965-1969 immigrant cohort have weekly wages and of these, only 68 were in the 2010 census; thus, we have limited power to estimate these coefficients accurately.

3.2 English Proficiency

Our paper further establishes the importance of English proficiency. As Borjas (2015) highlights, the odds of having high English proficiency is much higher for earlier generations of immigrants. We measure English proficiency on a scale of 1 to 5, with 1 being the lowest level of proficiency (does not speak English) and 5 being the highest level of proficiency (speaks only English).

Figure 2 below plots the change in proportion of English fluency within each cohort group over time. The earliest cohorts had a very high proportion of immigrants who spoke English very well, and a very low proportion of immigrants who did not speak English well. However, as time progresses, the proportion of immigrants who speak English very well decreases while the proportion of immigrants who do not speak English well increases. The graph reveals a clear trend that immigrants in later cohorts have a lower English speaking ability than immigrants in earlier cohorts.

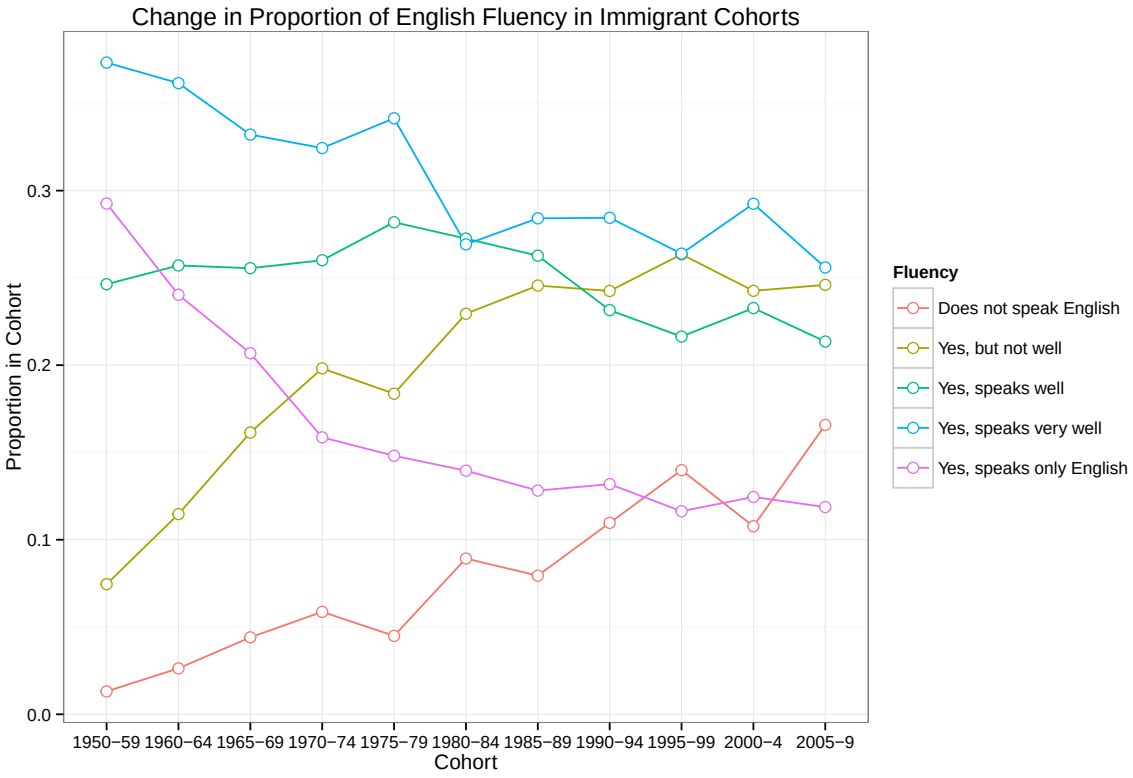

Figure 2: Histograms of English proficiency levels by census date: the proportion English proficiency declines in later cohorts; early cohorts had higher levels of English proficiency.

In the following model, we analyze the differences in the acquisition of high English proficiency between cohorts to determine whether the trend in the probability of language acquisition correlates with the slower economic assimilation we observe in later immigrant cohorts.

Table 2: Impact of English Proficiency on Age-Adjusted Relative Weekly Earnings of Immigrants, by Census Cross Section

	Dependent variable: High English Proficiency			
	(1)	(2)	(3)	(4)
	1980	1990	2000	2010
1960-64 arrivals	0.5892***	0.7120***	0.5868***	
	(0.0124)	(0.0213)	(0.0123)	
1965-69 arrivals	0.4250***	0.4034***	0.5917***	
	(0.0199)	(0.0173)	(0.0299)	
1970-74 arrivals	0.3149***	0.3084***	0.3783***	0.6489***
	(0.0253)	(0.0159)	(0.0275)	(0.0410)
1975-79 arrivals		0.2565***	0.2991***	0.5564***
		(0.0128)	(0.0289)	(0.0707)
1980-84 arrivals	0.2569***	0.1822***	0.2670***	0.4517***
	(0.0249)	(0.0085)	(0.0309)	(0.0691)
1985-89 arrivals		0.1596***	0.2115***	0.3065***
		(0.0074)	(0.0280)	(0.0512)
1990-94 arrivals		0.1881***	0.2201***	0.4026***
		(0.0090)	(0.0298)	(0.0712)
1995-99 arrivals			0.1919***	0.3304***
			(0.0247)	(0.0602)
2000-04 arrivals			0.2253***	0.3292***
			(0.0278)	(0.0586)
2005-09 arrivals				0.3140***
				(0.0540)
Observations	5529	8355	11939	3175

Note: The dependent variable, English proficiency, is a binary variable equal to 1 if the survey reported that they "speak English very well" or that they "speak only English" and 0 otherwise. Data from the 1970 census is omitted as language data was not collected at that time. The odds ratio are reported. Standard errors are in parentheses and are clustered at the cohort level. *p<0.1; **p<0.05; ***p<0.01.

Table 2 shows a logistic regression model, using pre-1960's immigrants as a baseline in all census years except in 2010 where the 1960-64 cohort is used as a baseline. The logistic regression explores the effect of being in each cohort on the probability of having a high English proficiency. We find that at each census, later generations of immigrants have much lower odds of reporting that they "speak English very well" or that they "speak only English." Our results are highly significant for each cohort, indicating that there are distinctions between cohorts in terms of language proficiency. Since we also see distinctions between cohorts in economic assimilation, this suggests that they

may be caused by differences in language proficiency.

In addition, we find that the odds of not knowing English well decrease as time passes from the first year of entry. In general, the odds of reporting high English proficiency in comparison to the baseline group increases across each cohort for each census year; the longer one stays in the US, the more likely they are to have higher rates of English proficiency.

While Borjas (2015) suggests that slower rates of English language acquisition can account for later cohorts' decreased rates of assimilation, he does not directly assess the impact of language acquisition on wages. In Table 3, we confirm the significance of English proficiency for economic assimilation. We run a fixed effects regression of the log of weekly wages, adjusted for age, on a constructed English proficiency variable and its quadratic term. We adapted the SPEAKENG variable to an ordinal variable LANGRUN where we rank five levels of proficiency from 1 as "Does Not Speak English" to 5 as "Speaks Only English." Our results show that weekly earnings of immigrants increases as English proficiency increases, lending more credence to Borjas' hypothesis. Our proficiency output shows a significant, positive trend in wage growth. Meanwhile, our quadratic proficiency output is negative, indicating that wage growth tapers off as immigrants master the language.

Table 3: Impact of English Proficiency on Age-Adjusted Relative Weekly Earnings of Immigrants, by Census Cross Section

	Dependent variable: Log Weekly Wages			
	(1)	(2)	(3)	(4)
	1980	1990	2000	2010
Proficiency	0.2767***	0.3502***	0.4115***	0.4005***
	(0.0211)	(0.0339)	(0.0495)	(0.0716)
Proficiency^2	-0.0176**	-0.0227**	-0.0295**	-0.0153
	(0.0027)	(0.0048)	(0.0072)	(0.0109)
Observations	5528	8355	11939	3175
R^2	0.1063	0.1346	0.1371	0.2066

Note: The dependent variable, English proficiency, is an ordinal variable 1-5 which ranks how well an respondent speaks English. 1 corresponds to "Does Not Speak English" and 5 corresponds to "Speaks Only English." Data from the 1970 census is omitted as language data was not collected at that time. Standard errors are in parentheses and are clustered at the cohort level. *p<0.1; **p<0.05; ***p<0.01.

While Borjas' analysis establishes that immigrants in later cohorts have slower rates of English adoption, he fails to investigate the magnitude of the benefits for English-speaking immigrants. We developed a cohort fixed-effects model with clustered standard errors to assess whether immigrants who rate themselves as either "speaking English very well" or "speaking only English" are more likely to make wages closer to non-immmigrants for a given census year. Following Borjas' analysis, if English language skills are a key determinant of lagging economic assimilation, we would expect smaller negative wage impacts from immigrant status, particularly for older generations who have higher rates of English proficiency.

Table 4: Year-Adjusted Relative Weekly Earnings of Immigrant Cohorts with High English Proficiency

	Dependent variable: Log Weekly Wages			
	(1)	(2)	(3)	(4)
	1980	1990	2000	2010
1950-59 arrivals	0.0145***	0.1354***	0.0874***	
	(0.0007)	(0.0011)	(0.0035)	
1960-64 arrivals	-0.0174***	0.0286***	0.1213***	
	(0.0004)	(0.0009)	(0.0027)	
1965-69 arrivals	-0.0936***	0.0043***	0.0684***	-0.3715***
	(0.0004)	(0.0006)	(0.0020)	(0.0043)
1970-74 arrivals	-0.1608***	-0.0908***	-0.0095***	0.0969***
	(0.0004)	(0.0003)	(0.0018)	(0.0035)
1975-79 arrivals		-0.1367***	-0.0624***	-0.0800***
		(0.0005)	(0.0015)	(0.0032)
1980-84 arrivals	-0.1493***	-0.1902***	-0.0958***	-0.0544***
	(0.0004)	(0.0005)	(0.0008)	(0.0027)
1985-89 arrivals		-0.1707***	-0.1457***	-0.1888***
		(0.0004)	(0.0005)	(0.0018)
1990-94 arrivals		-0.1063***	-0.0889***	-0.0582***
		(0.0003)	(0.0009)	(0.0009)
1995-99 arrivals			-0.0020	-0.0224***
			(0.0012)	(0.0016)
2000-04 arrivals			0.0289***	-0.1289***
			(0.0014)	(0.0021)
2005-09 arrivals				-0.1346***
				(0.0020)
Observations	107802	115291	123432	25348
R^2	0.1321	0.1515	0.1135	0.1294

Note: The age-adjusted wage differentials between each immigrant cohort and natives are calculated from a regression estimated separately in each census cross section. The dependent variable is log weekly earnings adjusted for age. High English proficiency. Standard errors are in parentheses and are clustered at the cohort level. *p<0.1; **p<0.05; ***p<0.01.

Our results in Table 4 do not support this hypothesis. We find that despite having near fluency, the effects of belonging to a specific immigrant cohort remain negative and highly significant for all census cross-sections except for the 1995-99 cohort in the 2000 census. While our results in Table 3 suggest that higher rates of English proficiency have a strong positive effect on economic assimilation, the gains due to English language proficiency do not not overcome immigrant status even at near fluent levels. The results are all significant at the $p < 0.01$ level and negative with the exception of two values which are close to 0. Therefore while our paper does further Borjas' focus on language acquisition, we find Borjas' treatment of English proficiency as the central measure of human capital accumulation to be potentially misguided as even high levels of human capital fail to compensate for immigration status.

It is important to note, however, that the magnitude of the coefficients for our restricted sample are lower than those found in Table 1. While the absolute value of the cohort effects from the general pool ranges from 0.0417 to 0.320, the range of coefficients for the high proficiency speakers is much smaller, ranging from 0.0145 to 0.189, with a sole large value of 0.371. These results support our earlier findings that language acquisition does have a positive impact on age-adjusted wages. After accounting for language proficiency, the cohort effects are less strongly ordered in terms of negative effect. Earlier and later generations have similar cohort coefficients for a given census year, suggesting that English speaking skills are a key factor in preventing later cohorts from achieving economic assimilation.

3.3 Adjusted Linear Regression Model

Based on our earlier analysis, we found that cohort effects still influence the magnitude of the wage gap between non-immigrants and immigrants, even after accounting for English language proficiency. Our final extension on Borjas' analysis is to determine whether additional factors beyond language proficiency can account for the slowing rate of economic assimilation for later immigrant generations. We looked into a set of variables, including factors characterizing the immigrant, such as age, race, English proficiency, and level of education, and their origin country, such as the fertility rate and GDP.

Table 5: Summary Statistics for Controlled Variables

Statistic	N	Mean	St. Dev.	Min	Max
Age	515,713	39.587	11.988	18	64
English Proficiency	438,486	4.783	0.665	1	5
Level of Education	515,713	2.608	1.207	1	5
Fertility	508,852	2.268	16.584	1.150	2,415.880
GDP (Expenditure)	508,322	8,033,228	3,555,475	1,634.986	13,351,904
Population (in millions)	508,322	256.190	104.932	0.204	1,324.353

These controlled factors were selected for using step-wise AIC, which exhaustively searches for the best model. Other factors that were included but subsequently dropped during model selection include the Gini coefficient, which measures income inequality, and the average level of schooling in the origin country. Broadly speaking, we interpret GDP, population and fertility rate as a proxy for the rate of development in the origin country. We may expect immigrants from more economically and socially developed nations to have higher access to education, health care and other resources

which may increase stock of human capital and the likelihood that they are more accustomed to hiring practices in the US.

Other than English proficiency, there are only a few missing values for the controlled variables; thus, we include all the variables and drop missing observations. There is a wide range in fertility, GDP, and population, which implies that immigrant backgrounds are diverse and may therefore contribute to differences in relative wages among immigrant cohorts. In the following model, we analyze whether these differences contribute to the rate of economic assimilation. In particular, we examine whether the changing rate of economic assimilation over time can be explained by factors other than English proficiency.

In Table 6, all estimates of relative weekly earnings are adjusted for year, since wages increase over time. In column 1 of Table 6, we observe that later immigrant cohorts have a greater disadvantage in earnings compared to natives when adjusted for age, similar to our conclusions from Table 1. In Column 2, after adjusting for age and English proficiency, we observe that the differences between the immigrant cohorts in relative weekly earnings are diminished, but not entirely eliminated. This supports Borjas' argument that English proficiency is an important factor. However, other factors like GDP and race appear to play an equally important role. Column 3 controls for age, English proficiency and GDP of country of origin and we find that the cohort effects are no longer significant. After we further adjust for fertility rate and population in Column 4, we find that the cohort effects remain not significant. Interestingly our results in Columns 3 and 4 switch from positive to negative, suggesting that further investigation into the demographics of country of origin in conjunction with language acquisition may be warranted. In conclusion, the changing composition of immigrant groups over time are shown to play an important role alongside English proficiency in accounting for the observed differences in economic assimilation among cohorts.

4 Conclusion

We observed that economic assimilation rates for immigrants differ based on when they immigrated by using a fixed effects model with standard errors clustered by cohort. In particular, the immigrant wage gap is more slowly overcome in later immigrant cohorts. This can be partially explained by the lowered likelihood of acquiring high English proficiency for later immigrants. However, cohort effects remain among highly fluent immigrants as well, suggesting that language proficiency alone does not completely explain these differences.

Furthermore, while language proficiency is shown to have a significant impact, differences across the immigrant cohorts persist after accounting for language proficiency. In order to account for these differences, other factors reflecting the immigrant composition such as race, age and factors related to their origin country such as GDP, fertility rate, and population are included in our regression. Our linear regression model with clustered standard errors and control variables found that immigrant composition and country-of-origin differences also play a significant role in explaining the decreasing earnings potential of later cohorts.

Further research can examine the interactions between language proficiency and immigrant composition in terms of their impact on wage differentials. While our analysis shows that both sets of factors can account for cohort differences in economic assimilation, it remains unclear which of these factors are the most important and which, if any, are causal. Further research may also recreate our analysis for lower levels of English proficiency. While we have shown that high levels of English proficiency may not be enough to eliminate cohort effects, gains in lower levels of proficiency may matter more than gains in higher levels of proficiency. A future set of analyses would examine heterogenity in wage gains to proficiency.

Table 6: Year-Adjusted Relative Weekly Earnings of Immigrant Cohorts

	Dependent variable: Log Weekly Wages			
	(1)	(2)	(3)	(4)
1950-59 arrivals	0.065^{***}	0.169^{***}	0.246	-0.041
	(0.002)	(0.012)	(0.227)	(0.231)
1960-64 arrivals	0.023^{***}	0.164^{***}	0.257	-0.047
	(0.002)	(0.013)	(0.229)	(0.243)
1965-69 arrivals	0.013^{***}	0.128^{***}	0.239	-0.096
	(0.002)	(0.014)	(0.234)	(0.250)
1970-74 arrivals	-0.135^{***}	0.082^{***}	0.227	-0.112
	(0.001)	(0.014)	(0.239)	(0.254)
1975-79 arrivals	-0.125^{***}	0.101^{***}	0.260	-0.087
	(0.002)	(0.015)	(0.248)	(0.266)
1980-84 arrivals	-0.186^{***}	0.059^{***}	0.217	-0.139
	(0.0005)	(0.015)	(0.241)	(0.258)
1985-89 arrivals	-0.272^{***}	0.057^{***}	0.248	-0.117
	(0.001)	(0.015)	(0.247)	(0.272)
1990-94 arrivals	-0.259^{***}	0.083^{***}	0.304	-0.059
	(0.001)	(0.014)	(0.244)	(0.264)
1995-99 arrivals	-0.288^{***}	0.143^{***}	0.382	0.002
	(0.001)	(0.014)	(0.240)	(0.264)
2000-04 arrivals	-0.477^{***}	-0.011	0.259	-0.133
	(0.002)	(0.014)	(0.236)	(0.261)
2005-09 arrivals	-0.553^{***}	-0.057^{***}	0.236	-0.168
	(0.001)	(0.014)	(0.228)	(0.255)

Note: (1) Adjusted for age (2) Adjusted for age and English proficiency (3) Adjusted for age, English proficiency, and GDP of origin country (4) Adjusted for age, English proficiency, race, and GDP, fertility rate, population of origin country; *p<0.1; **p<0.05; ***p<0.01

References

[1] Borjas, George J. "The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again." *Journal of Human Capital 9.4*, 2015.

[2] Zong, Jie, and Jeanne Batalova. "Frequently Requested Statistics on Immigrants and Immigration in the United States." *Migrationpolicy.org.* Migration Policy Institute, 25 Feb. 2015. Web. 09 Apr. 2016.