NBER WORKING PAPER SERIES

IMMIGRANT-NATIVE SUBSTITUTABILITY: THE ROLE OF LANGUAGE ABILITY

Ethan G. Lewis

Working Paper 17609
http://www.nber.org/papers/w17609

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2011

Immigrant-Native Substitutability: The Role of Language Ability
Ethan G. Lewis
NBER Working Paper No. 17609
November 2011
JEL No. J24,J31,J61

ABSTRACT

Wage evidence suggests that immigrant workers are imperfectly substitutable for native-born workers with similar education and experience. Using U.S. Censuses and recent American Community Survey data, I ask to what extent differences in language skills drive this. I find they are important. I estimate that the response of immigrants' relative wages to immigration is concentrated among immigrants with poor English skills. Similarly, immigrants who arrive at young ages, as adults, both have stronger English skills and exhibit greater substitutability for native-born workers than immigrants who arrive older. In U.S. markets where Spanish speakers are concentrated, I find a "Spanish-speaking" labor market emerges: in such markets, the return to speaking English is low, and the wages of Spanish and non-Spanish speakers respond most strongly to skill ratios in their own language group. Finally, in Puerto Rico, where almost all workers speak Spanish, I find immigrants and natives are perfect substitutes. The implications for immigrant poverty and regional settlement patterns are analyzed.

Ethan G. Lewis
Department of Economics
Dartmouth College
6106 Rockefeller Hall
Hanover, NH 03755
and NBER
ethan.g.lewis@dartmouth.edu

Studies have found that the massive flow of immigrants into the U.S. in the past few decades has had little negative impact on the average wages of native-born workers (reviews include Borjas, 1994, and Friedberg and Hunt, 1995; a more recent article is Ottaviano and Peri, forthcoming, hereafter "OP"). However, many of these same studies tend to find that the new arrivals substantially depressed the wages of previous immigrant arrivals (Card, 2001; OP).[1] The fact that immigrants seem, by this evidence, not to fully compete in the native labor market may contribute to the relatively high rates of poverty among immigrants, and understanding why they do not may inform policies aimed at reducing immigrant poverty.

A plausible explanation for why immigrants do not fully compete is that their limited English skills restrict them to occupations where English skills are less important. The ability to speak the native language is associated with higher wages (e.g., Chiswick and Miller, 1995, 2007; Carliner, 1996; Dustmann and Fabbri, 2003; Bleakely and Chin, 2004; Ferrer, Green and Riddell, 2006). In addition, less-skilled immigrants in the U.S. appear to specialize in occupations which require relatively little communication (Peri and Sparber, 2009), although factors besides immigrants' relative lack of English skills (such as a greater willingness to work in manual jobs) could contribute to this.

In this paper, I ask how immigrants' language skills affect how closely they compete with native-born workers. How closely two different groups of workers (in this case, immigrants and natives) compete in the labor market is revealed by how closely their wages move together in response to changes in the size of one group relative to the other. Two groups of workers are "perfect substitutes" if their wages move in sync in response to changes in relative numbers: from the labor market's perspective they are identical. In contrast, two groups are "imperfect substitutes" if an increase in the size of one group relative to a second group lower the wages of the first group

[1] In OP's preferred specification, for example, the largest estimated impact of immigration since 1990 they can find on the average wages of native-born high school dropouts is -0.1 percent, compared to -8.1 percent for immigrants. Much of the more recent literature (e.g., Card, 2001, 2009) does not examine impacts on average wages directly, but instead focuses on the wage gaps between groups of native workers. Even measured this way, however, the impact on less-skilled natives is generally found to be small.

relative to the second. The more responsive relative wages are to relative supply, the less directly these two groups compete in the labor market. In the U.S., immigrants and natives with similar education and experience are imperfect substitutes (OP).

Put another way, then, this paper asks whether immigrants with strong English language skills are closer substitutes for natives than immigrants with poor English language skills. Specifically, it asks if the relative wages of immigrants (that is, relative to natives' wages) who speak English well respond more to changes in the relative total hours worked by immigrants than do the relative wages of immigrants who speak English poorly. To estimate this relationship, I rely primarily on variation across metropolitan areas, using data from the 2000 U.S. Census of Population and American Community Surveys from 2007-2009 (Ruggles et al., 2010). All comparisons are made within broad education groups (high school or less, more than high school) that previous research has found represent distinct labor types (Card, 2009; Goldin and Katz, 2008). These data contain self-reports of English language skills. As language skills may be correlated with other factors that affect substitutability with natives (such as legal status and access to job networks), I also make comparisons across age at arrival and years in the U.S., exploiting the fact that immigrants who arrive in the U.S. as children tend to have far better English language skills as adults (Bleakley and Chin, 2004), and, similarly, that English skills tend to rise with time in the U.S.

The above discussion is written as if all communication must take place in English, but in some parts of the U.S., Spanish is also an important language. An extreme case is Puerto Rico, where both "immigrants" (who include foreign-born from Latin American countries, as well as U.S. born ethnic Puerto Ricans) and natives speak Spanish. As a further test of the language hypothesis, I ask whether wage movements reveal Puerto Rican immigrants and natives to be perfect substitutes. Since Puerto Rico is usually considered to be a single labor market (e.g., Borjas, 2008), it is not possible to exploit geographic variation in the size of the immigrant population. Instead, this

analysis relies on variation over time and across detailed education and experience groups in the relative size of the immigrant population, similar to OP, using Puerto Rican Censuses from 1970 to 2000 along with the 2007-2009 Puerto Rican Community Surveys (Ruggles et al., 2010). I also perform a parallel aggregate analysis of the continental U.S.

In less extreme cases than Puerto Rico, Spanish can be a dominant language, like in Miami and Los Angeles, or prevalent if not dominant, like in Chicago and San Francisco. I also consider whether the density of Spanish speakers affects how substitutable Spanish speaking immigrants are for natives. Under what conditions can a separate "Spanish-speaking" labor market develop? A key hypothesis from theory is that it requires not just a large number Spanish speakers overall, but a sufficiently rich distribution of skills among Spanish speakers. This, for example, distinguishes Miami -- where majority of both highly educated and less educated workers speak Spanish -- from Chicago, where Spanish speakers are disproportionately less educated.

Nearly all of the results are consistent with the view that language skills are an important source of imperfect substitutability between immigrants and natives. The estimates imply that the increase in the labor supplied by immigrants (relative to natives) since 1990 in the average metropolitan area is associated with about a 6 percent decline in the hourly wage of immigrants with poor English language skills, but only a 2 percent decline in the wages of immigrants with strong English language skills (each relative to natives with similar education and work experience). I also find that the wages of immigrants who arrived as young children and long ago are less sensitive to immigrant relative supply than those who arrived and older ages and recently. Also consistent with language driving imperfect substitutability, in Puerto Rico immigration is not associated with a decline in the wages of Puerto Rican immigrants relative to Puerto Rico-born workers.

Within the U.S., several pieces of evidence suggest that Spanish-speaking immigrants behave as if they have their own labor market in markets where they are numerous. First, the wage premium

for speaking English, for Spanish speakers, is small in markets with a heavy Spanish-speaking presence. Second, an influx of less-educated Spanish speakers pushes down the wages of less-educated Spanish-speaking immigrants more than other less-educated immigrants. Finally, immigration of educated Spanish-speakers appears to mitigate the impact of immigration on less-educated Spanish speakers' wages. Despite a heavy Cuban presence in Miami, for example, the relative wage of less-educated Miami Cubans is about the same as less educated Spanish-speaking immigrants elsewhere. The heavy presence of educated Cubans in Miami contributes to this, according to the estimates. Put differently, Spanish and non-Spanish speakers' wages respond most strongly to skill ratios in their own language group.

The findings refine our understanding of the forces which affect immigrant poverty. Previous research already suggested that the wage impacts of immigration were borne disproportionately by immigrants themselves, and these results says that they are borne particularly strongly by immigrants with poor English language skills, who are among the poorest immigrants (in recent U.S. data, 29 percent of immigrants who do not speak English were in poverty, compared to 15 percent of immigrants overall), and by Spanish-speaking immigrants. The estimates here are consistent with the increase in immigration since 1990 contributing an additional one to two percentage points to the poverty rate of less-educated low-English immigrants, and an additional two to four percentage points to the poverty rates of less-educated Spanish-speaking immigrants. However, immigrant poverty rates actually fell over this period, so this result only says that immigrant poverty rates would have fallen more quickly without additional immigration.[2]

[2] For example, poverty rates among less educated non-English speaking immigrants fell from 35 to 30 percent between 1990 and 2008.

I. *Motivation and Background*

It is useful to begin with some basic facts about immigrants' and natives' language skills and wages. Panel A of Table 1 shows average English skills for immigrants and natives. It was constructed by combining data from the 2007, 2008 and 2009 American Community Surveys (Ruggles et al., 2010) for the working-age residents of 136 high immigration metropolitan areas described below.[3] These metropolitan areas contain over 80 percent the immigrants living in the U.S. Table 1 also shows separate means by whether or not the respondent has any college education. Throughout the paper, those with a high school degree or less will be referred to as "less educated" and will be a focus of the analysis, as they make up a disproportionate share of those in poverty.

The first line of Panel A shows that only 46 percent of immigrants speak English "only" or "very well." In contrast, among the native-born, this number is over 98 percent (lower half of table). To a useful first approximation, U.S. natives are fluent in English, whereas only half of immigrants are. The latter rises to 68 percent if you include immigrants who say they speak English "well" (but not "very well"). By this broader measure of English proficiency, about half of immigrants without college are proficient in English, and 90 percent of immigrants with college education are proficient in English (columns 2 and 3).

Spanish speakers will also be a focus of this study, and Panel B shows the English skills of just those immigrants and natives who report speaking Spanish at home. Spanish-speaking immigrants have below average English skills, with only 26 percent fluent in English; even within education category their English skills are below average. This is mostly driven by Mexicans who are the largest immigrant group and who have poor English skills. (74 percent of Mexican immigrants in this sample report speaking English not well or not at all.) Even native-born Spanish speakers, who

[3] "Working age" is defined as being between age 16 and 65 and with at least one year of "potential work experience" which means being old enough to have spent time outside of school given normal progression through school.

make up 7.7 percent of the working-age population, have imperfect English. Only 82 percent claim to be fluent in English, and almost 7 percent say they do not speak English well or at all.

The analysis below also exploits variation in English skills by immigrants' age at arrival and time in the U.S. Figure 1a shows average language skills by these characteristics. The share who speak English well has a positive monotonic association with time in the U.S., rising from half of those who arrived in the past five years to 90 percent of those who have been in the U.S. at least 40 years (a pattern which may partly reflect cohort differences in English skills). Figure 1a also shows the sharp decline in the share of immigrants who speak English well (as adults) in age at arrival: it is higher among those who arrived before age 10 compared to those who arrived at older ages. This fact was exploited in Bleakley and Chin (2004) (BC) to study the effect of language skills on wages. They argued that there is a "critical period" at young ages when children are able to easily learn English. Below, I ask if there is a similar kinked relationship in the substitutability of immigrants for natives by age at arrival.[4] Figure 1b shows that this kinked relationship is mainly present among less-educated immigrants; among more educated immigrants, the relationship is smoother.[5]

Suggestive evidence that English skills are important for economic well-being is shown in Table 2, which shows mean log hourly wages and poverty by nativity, education, and English and Spanish language skills. The first two rows of Panel A show that even within education category, immigrants tend to earn less than natives. For example, less educated immigrants earn on average 18 log points (about 18 percent) less than less educated natives. The next row shows, pertinent to the idea that language skills might matter for this wage gap, the wages of immigrants who are fluent

[4] Note that I am not fully implementing BC's methods here. A key thing I am not doing, but they did, is differencing out age-at-arrival patterns among immigrants from English-speaking countries to account for other factors, besides language, associated with age-at-arrival. Wage response estimates for the small number of immigrants from English-speaking countries were too unreliable to exploit this approach.

[5] Splitting the sample by age at arrival and education is somewhat problematic given BC's finding that arriving as a young child tends to raise educational attainment. The appropriate approach in light of this is to aggregate together more- and less-educated workers. Because the rest of the analysis is split by education, however, I will split it this way for the estimates by age-at-arrival.

in English are very similar to that of natives, including among less-educated immigrants and natives. There is also a steep wage gradient in self-reported English, shown below that. These wage differences translate to difference in poverty as well, shown in panel B. Whereas 30 percent of less-educated immigrants (and even 23 percent of those with college education) who do not speak English are in poverty, only 14-16 percent of less-educated immigrants who speak English well or very well are in poverty, similar to poverty rates among less-educated natives. Finally, the bottom rows of Table 2 show that Spanish speakers are worse off than the typical immigrant, something their poor English skills (Table 1) likely contributes to. Immigrants who speak only Spanish, shown in the bottom row of the table, have wages and poverty rates about the same as the typical non-English speaker.

The wage gap between less educated immigrants and natives is analyzed directly in multivariate regressions in Table 3. Column (1) repeats the finding in from column (2) of Table 2, that less-educated natives earn about 18 percent less than less-educated natives. Column (2) of Table 3 shows that a single control variable -- a dummy for speaking English only or very well -- can account for most of this gap. The coefficient on this control suggests that there is a 21 percent wage premium to speaking English fluently, a finding which is consistent with previous estimates of the returns to speaking English in the U.S. labor markets (e.g, Chiswick and Miller, 1995, and Carliner, 1996). This likely overstates the causal effect of English speaking ability, however. Column (3) shows that the addition of simple demographic and skill controls reduces the magnitude of this coefficient. Estimates in BC's study exploiting age at arrival are, in fact, consistent with no causal effect of English language skills on wages. They could account for wage gaps across immigrants with varying English language skills entirely with education differences across these groups.

The last three columns of Table 3 focus on the large minority of Americans -- both immigrants and natives -- who speak Spanish at home. Column (4) shows that there is a return to English fluency,

albeit smaller, in this subpopulation as well, and that, conditioning on English fluency, there is no immigrant-native wage gap. Might English skills matter less when there are large numbers of other Spanish speakers in the same labor market? To find out, column (5) adds an interaction between the English fluency dummy and the share of the metro area's population who speak Spanish at home. The coefficient on this interaction is negative and significant, consistent with the idea that English skills become less valuable (for Spanish speakers) as the size of the Spanish-speaking population increases.[6] Column (6) splits the Spanish-speaking share into shares among the more- and less-educated populations. Consistent with what was hypothesized in the introduction -- that it requires a sufficiently rich distribution of skills to create a Spanish-speaking labor market -- it is only the Spanish-speaking share among more educated workers that is associated with a diminished importance of English language skills among less educated workers. To take an extreme example, in markets, like Miami, where a majority of college educated workers speak Spanish (see Appendix Table A1 for other examples) these estimates predict there will be no premium to English fluency.[7] Below, I reassess how the relative wages of Spanish-speaking immigrants is affected by the density of Spanish and English speakers in the labor market, using the estimation framework I describe now.

II. Theory and Derivation of an Estimation Equation

A starting point for a simple theory of how language skills matter in the U.S. labor market is the notion that among those with otherwise similar skills, those who cannot communicate in English well (hereafter, "speak" English for short) imperfectly substitute for workers who do; the former

[6] Interestingly, this control raises the coefficient on English fluency dummy back to the level it is in the entire sample; that is to say, the fact that Spanish speakers are geographically clustered fully accounts for their lower average return to speaking English.

[7] Though this result is quite preliminary, it does provide some counterweight to the evidence that the tendency for immigrants to geographically cluster is bad for them (e.g., Cutler, Glaeser, and Vigdor, 2008). To be fair to those authors, though, their analysis is at the neighborhood, rather than market level.

might not be very effective in occupations which require a lot of communication, for example (Peri and Sparber, 2009). For simplicity, imagine workers can be sharply divided into those who can and cannot speak English, indexed with j=1 and j=0, respectively. If these two types are imperfectly substitutable, the wage premium to being able to speak English will decline in the relative number of workers who speak English and who do not, which I capture with the following relationship:

$$(1) \quad \ln(w_0/w_1) = a - b\ln(L_0/L_1)$$

L_j is the number of workers and w_j is the wage of language type j workers, and a and b are positive constants. b measures the degree of imperfect substitutability: it will be zero if those who can and cannot speak English are perfect substitutes.[8] In principle, (1) could be estimated using variation across labor markets and over time in the relative number of workers who speak English "only," "very well" or "well" as a proxy for L_1 and the remaining workers as a proxy for L_0. In order to be consistent with prior estimates of the impact of immigration on wages, however, it is useful to translate (1) into something which directly involves immigrants and natives.[9]

To do so, first recall that nearly all natives report speaking English fluently. The relative number of non-English speakers is driven almost entirely by immigration, and, in practice (demonstrated below) moves almost one-for-one (in percent terms) with the relative number of immigrants. Mathematically, $\ln(L_0/L_1) \approx c + \ln(L^F/L^N)$, where L^F and L^N are, respectively, the number of foreign-born and native-born workers. Imposing this linear approximation on (1), we have that:

$$(1') \quad \ln(w_0/w_1) = a' - b\ln(L^F/L^N)$$

[8] (1) falls out of a capital-neutral single-good nested CES production function representation of the economy, where the innermost nest contains workers with and without English skills in this case. In this interpretation, the coefficient b represents an inverse "elasticity of substitution" between language types, and "a" represents a demand shifter which is a function of factor share and productivity parameters embedded in the production function. Other, more general, functional forms are possible, but this CES-derived approach is common (e.g., OP; Raphael and Smolensky, 2008).
[9] Another potential reason to do so is evidence that self-reported measures of language skills are not reliable (Dustman and van Soest, 2001, 2002; Dustmann and Fabbri, 2003). Despite this, the results below are similarly strong when using self-reported English skills directly, suggesting that they are at least reliable enough (at the aggregate level) to be useful.

To translate the left hand side into the wages of immigrants and natives, first note that since all natives are assumed to speak English we can impose $w^N = w_1$. (In practice, below, natives who report imperfect English will be dropped from the wage sample.) As for immigrants' wages, the analysis will explore variation across groups of immigrant with varying English ability. In particular, I will look across the ten year "age-at-arrival" and five year "years in U.S." categories shown in Figure 1 (in addition to directly across the English language skill categories shown in Tables 1 and 2). So suppose *g* indexes these different categories of immigrants, and fraction φ_g of group *g* speaks English. I can write the mean log wage of group g immigrants as $\ln w^{Fg} = \varphi_g \ln w_1 + (1-\varphi_g)\ln w_0 + \varepsilon_g$, where ε_g represents sources of immigrant-native wage gaps other than English skills (for example, ethnic discrimination or legal status.) Translating this into a immigrant-native wage gap:

$$\begin{aligned}\ln\left(w^{Fg}/w^{N}\right) &= \left[\varphi_g \ln w_1 + \left(1-\varphi_g\right)\ln w_0 + \varepsilon_g\right] - \ln w_1 \\ &= \left(1-\varphi_g\right)\left(\ln w_0 - \ln w_1\right) + \varepsilon_g \\ &= \left(1-\varphi_g\right)\ln\left(w_0/w_1\right) + \varepsilon_g\end{aligned}$$

Substituting (1') into this produces

$$(2)\quad \ln\left(w^{Fg}/w^{N}\right) = a_g - \left(1-\varphi_g\right)b\ln\left(L^F/L^N\right) + \varepsilon_g$$

The intercept, a_g, is a combination of constants. (2) implies that among otherwise similar immigrants and natives, the sensitivity of the relative wages of foreign workers to changes in foreign relative supply is diminishing in the share of foreign workers with good language skills. For example, immigrants who arrive as children tend to have better English skills than those who arrive as adults, and so (2) implies that the wages of those who arrive as children should respond less to immigrant inflows than immigrants who arrive as adults.

This simple model leaves out several things. First, according to the model, immigrants with perfect English, $\varphi_g = 1$, are perfect substitutes for natives; their relative wages are insensitive to immigrant relative supply. An important simplification used to derive (2) was that other sources of immigrant-native wage gaps (ε_g) are unrelated to immigrant relative supply. If this is not the case, then even fluent immigrants' relative wages may be sensitive to the number of immigrants relative to natives. This may also bias estimate of (2), an issue which will be discussed further below.

Second, this model assumes that English skills are equally important for all jobs -- the "b" in (1) is a constant.[10] As it is plausible English skills are more important in high-skill jobs, the estimates below allow the effects to vary by the education level of the worker.

Finally, this simple model can also only partly accommodate the fact that in some parts of the U.S., Spanish, not English, is the dominant language. In Puerto Rico, where both immigrants and natives speak Spanish, (2) applies and implies immigrants' relative wages should not respond to the relative number of immigrants. In markets with a mix of Spanish-speaking and English-speaking workers, it is not clear what will happen, but some theories are suggestive. According to Lang's (1986) theory of language discrimination, wages are lower for Spanish-speaking (or generally, non-English speaking) immigrant laborers than natives because they bear the cost of training a bilingual supervisor; however, where Spanish is spoken by a majority of workers (as it is in some U.S. markets), the sign of wage gap with natives reverses. In Peri and Sparber (2009), less-educated immigrants are segregated into manual occupations because of their inferior (English) communication skills. Though they did not discuss it in the paper, one might imagine that with a large enough density of fellow Spanish speakers, it might be possible for Spanish speakers to have

[10] This would be of great concern only if the variation in the wage response across immigrant groups were driven by variation in the importance of English rather than the English skills of the immigrants in the group -- that is, variation in "b" not in φ_g.

access to a full range of occupations.[11] In both theories, having enough skilled (Lang's "supervisors") or educated Spanish speakers would be important for a separate "Spanish" labor market to emerge, which is supported by preliminary evidence in Table 3. This will be evaluated below by adding terms to (2) measuring the size of the Spanish-speaking labor pool by education.

III. Estimation and Identification

The main estimates of (2) will use variation across skill groups and metropolitan areas in the relative aggregate hours worked of the immigrant population, as follows:

$$(3)\ \ln\left(w_{ict}^{Fg}/w_{ict}^{N}\right)=\alpha_{git}+\beta_{g}\ln\left(H_{ict}^{F}/H_{ict}^{N}\right)+\varepsilon_{gict}$$

where *i* indexes two education groups (high school or less, more than college) and c indexes metropolitan areas, and H represents the aggregate hours of the specified group. β_g is the estimate of $-(1-\varphi_g)b$ in (2) which is expected to be negative and to be smaller in magnitude for immigrants with stronger English-language skills. (3) will be estimated both jointly and separately by education group.

All estimates of (3) include time-varying education group controls, α_{git}, which, like the slope, will be allowed to vary across immigrant groups, *g*. These controls capture economy-wide changes in the wage structure. The dependent variable will be computed as the difference in the mean log wages of immigrants in group *g* and natives with the same education, *i*, in the same metropolitan area, *c*, and year, *t*, or "cell *ict*" for short.[12] To reduce the influence of compositional differences between immigrants and natives on this estimated mean wage gap, natives' mean wages will be computed

[11] In particular, one might assume the output of Peri and Sparber's production function made with Spanish-speaking workers was perfectly substitutable for the output made with English-speaking workers.

[12] This uses the fact that $\ln\left(w_{ict}^{Fg}/w_{ict}^{N}\right)=\ln w_{ict}^{Fg}-\ln w_{ict}^{N}$.

using weights that give them, on average, the same education and experience as the immigrants in group *g* in cell *ict*.[13] (In practice, the raw wage gap produces similar results.) (3) will be estimated by ordinary least squares (that is, unweighted) and standard errors will be computed to be robust to arbitrary error correlation across observations on the same metropolitan area.

The error term in (3) captures other determinants of immigrants' relative wages. There are two broad reasons to expect that this error term will be correlated with immigrants' relative hours, which will lead estimates of β_g to be biased. Areas with a high relative demand for some immigrant subgroup would tend to simultaneously have high relative wages and hours for the workers in this group, thereby generating a positive correlation between the error term and the explanatory variable and leading to slope estimates less negative.[14] A standard way to address this is to use predictable variation in the size of immigrant inflows based on the labor markets in which immigrants from different parts of the world tend to cluster (Mexicans in LA, Russians in New York, etc.) This approach is described in the Appendix. In practice, it tends to produce similar results to the ordinary least squares estimates that are presented below, suggesting that this type of bias may not be large. In addition, in section IV.C., I will present estimates of (2) that rely only on aggregate variation -- that is, across education x experience groups, similar to OP -- rather than variation across labor markets. This also produces similar, if less precise, estimates.

(3) is also likely to be biased because immigrants with poor English language skills tend to be dissimilar from U.S. natives in other ways which affect their substitutability with natives. For example, the largest group of immigrants with poor language English language skills are Mexicans,

[13] In particular, those with high school or less are divided into high school completers and high school dropouts, and those with some college or more will be divided into those with and without four year degrees. Within these cells, workers are further divided into five-year potential experience bands up to 40. The mean of native log wages are computed weighted by p/(1-p), where p is the fraction of each detailed education x experience x metropolitan area x year cell that are group g immigrants, among group g immigrants and natives in that cell. (This weight is interacted with the ACS or Census sample weight.) Cells with no natives or no group g immigrants are dropped.

[14] These estimates are also well-identified only if the geographic units involved act as closed economies. U.S. evidence suggests that local labor markets behave like closed economies in the sense that immigrants do not appear to "displace" native-born workers with the same skills (e.g., Card and DiNardo, 2000) nor do they have much impact on industry mix (Lewis, 2004).

many of whom reside in the U.S. illegally. Their legal status may confine Mexicans to particular occupations, making it harder for them to compete head to head with natives. In the absence of a perfect way to identify immigrants who differ only in their language skills and not other factors (legal status in particular is not measured), the approach I take is to examine variation across different immigrant subgroups, such as by age at arrival. While variation across these different subgroups is likely to suffer from some of the same problems as variation across English skill categories measured directly, the hope is that the combination of evidence presents a consistent picture. Overall, the direction of bias is unclear, though in light of the evidence that relative demand is not leading to much bias in the estimates, the approach taken in this paper may be biased towards finding that language accounts for imperfect substitutability.

III.B. Data

Data for the regression analysis come from the 2000 Census of Population and 2007, 2008, and 2009 Community Surveys. The latter three (which are from Ruggles et al., 2010) are combined into what will be referred to as "2008" data. Information on hours worked and hourly wages were aggregated to the metropolitan area by year by broad education group, as described above. Included in the calculation of workers' hours were workers age 16 to 65, with positive potential experience (old enough to be out of school given a normal progression through school), living in one of the 136 metropolitan areas in the sample.[15] The wage sample is the subsample of these workers who are currently employed with positive wage and salary earnings and zero farm or

[15] Also, those living in group quarters were excluded from the analysis.

business earnings in the past year.[16] Metropolitan areas were defined consistently using "PUMAs" and 1990 metropolitan area boundaries.

Table 4 presents the (unweighted) means and standard deviations of these data. The relative log hours of immigrants in the average area is negative both overall and by education level, which says immigrants are on average a minority of workers; in the mean education-metro-year observation immigrants' hours represent about 14.4 percent ($=e^{-1.934}x100$) of natives' hours. There is a lot of variation across these metro areas in the relative hours of immigrants: the standard deviation is around 1, which will be useful for interpreting the magnitude of the regression estimates below. The table also shows the immigrant-native wage gap is around 12.5 percent in the average metropolitan area, and is much larger for low-English immigrants, even the ones matched to natives with similar education and experience (as these gaps are calculated to do). There is a lot of variation in the immigrant - native wage gap across metro areas to be explained, which I will now try to do.

IV. Results

IV.A. Basic U.S. Evidence

Table 5 presents estimates of (3). Panel A examines the relative wages of all immigrants. The -0.04 estimate in column (1) says that a one unit increase in immigrants' relative hours is associated with a 4 percent decline in the wages of immigrants relative to natives. Although a one unit increase in the independent variable means approximately tripling immigrants' labor supply, which sounds like a lot, it is actually a reasonable change to examine as it is both a standard deviation (Table 4)

[16] To be included in the wage sample, native-born workers also were required to report speaking English "only" or "very well." Hourly wages above $200 and below $2 in 1999 dollars were replaced with these thresholds.

and roughly equal to the increase in this variable since 1990 in the average metropolitan area.[17] In any case, the -0.04 response is similar in magnitude to previous estimates of this relationship, such as OP or Card (2009).

Panel B of Table 5 produces separate estimates for the wages of immigrants who report strong English language skills -- immigrants who say they speak English "only," "very well," or "well" -- and who report poor English language skills -- immigrants who say they speak English "not well" or "not at all." As expected, the relative wages of immigrants with poor English are more negatively associated the relative presence of immigrant labor: their wages decline 5.7 percent for a one unit increase in the independent variable, compared to 2.2 percent decline for immigrants with stronger English. The difference between the two coefficients is statistically significant at the one percent level, shown as the p-value less than 0.01 in the row beneath these estimates.

Columns (2) and (3) show estimates separately by broad education. Panel A indicates that educated immigrants' wages are more responsive than less educated immigrants', despite the fact that educated immigrants tend to report substantially better English language skills (Table 1). This may not be inconsistent with the importance of English, however; it is plausible that language skills are more important in jobs that require college education, a view for which there is some evidence (Berman, Lang, and Siviner, 2003). In addition, the difference in the wage responses of immigrants with poor and strong English are similar for more- and less-educated immigrants: in Panel B the gap in coefficients is 3.6 (=0.047-0.011) percentage points for less- and 3.3 (=0.078-0.045) percentage points for more-educated immigrants.

Panel C breaks out estimates by the more detailed English language skills shown in Tables 1 and 2. Going to this detail requires the examination of a smaller number of metro areas, 112, that are large enough to observe all immigrant subgroups. (In particular, wage earning college graduates who

[17] In the average metro area-education cell in the sample, this variable rose from -2.69 to -1.75 between 1990 and 2008.

report not speaking English at all are rare.) The wage response is monotonic in self-reported English language skills, both overall and separately by education, and the differences across English categories are statistically significant. Interestingly, there is a significant negative response even in the top English category, what I referred to as "fluent" above. While some of the workers in this category may not be truly fluent in English, a reasonable interpretation of this is that there are factors besides English skills that make immigrants imperfectly substitutable with natives.

Figure 2a examines how the wage response varies by age at arrival and time in the U.S. It plots coefficient estimates (and confidence intervals) from estimates of (3) -- that is, still using the same variation in immigrant relative labor by broad education, metropolitan area, and year -- separately for immigrants in the ten year age-at-arrival categories and five year years in U.S. categories shown on the figure's x-axis. Panel A shows that the response of wages by age-at-arrival follow the same "kinked" pattern that self-reported English language skills did in Figure 1a. While it is possible that other unobserved factors that affect immigrants' substitutability with natives follow this kinked pattern, this reinforces the direct evidence from English language skills in Table 5. On top of this, the kinked relationship in the response appears to be limited to the less-educated subsample where it was found (Figures 1b and 2b). There is also a monotonic relationship with time in the U.S. (Figure 2a, panel B).

IV.B. The Effects of Language Supply

All of the regressions presented so far have employed the same independent variable: the natural log of the ratio of immigrants' aggregate hours to natives' aggregate hours. In this section I instead employ direct measures of the supplies of language skills, including as the natural log of the aggregate hours of those who speak English not well or not at all (those with "poor English") to

those who speak English only, very well, or well (those with "strong English"). In addition, to try and understand what happens in markets with a large numbers of Spanish speakers, I will add controls for the relative hours of workers who speak Spanish at home.

The results are in Table 6. To keep things simple, the table examines only workers with a high school education or less. Column (1) repeats the estimate from Panel A of Table 5: it says the overall immigrant - native log wage gap declines 3.4 percent for a 1 unit increase in immigrants' relative hours. Column (2) replaces this independent variable with the one measuring the relative hours of those with poor English relative to strong English. The coefficient is the same to two decimal places. This supports the argument made in Section II that, as a practical matter, the relative supply of English language skills moves one-for-one with the relative supply of immigrants.

Column (3) repeats the estimates of column (2) for the subgroup of immigrants who report speaking Spanish at home. The estimated wage response is nearly identical for this subgroup. To investigate if the responses vary with the density of Spanish speakers in the market, column (4) adds controls for the aggregate hours of workers who speak Spanish at home relative to workers with strong English. This measure is entered separately for college and non-college workers, following the results in Table 3. Column (4) presents weak evidence that there is an additional depressing effect of having a large number of Spanish speaking non-college workers on the wages of non-college Spanish speakers, in addition to the impact having a large number of non-English speakers. However, this is offset by the positive wage impact of greater density of college-educated Spanish speaking immigrants.

A graphical version of the column (4) relationship is shown in Figure 3. It plots Spanish-speaking relative hours among workers with more than a high school education versus those with a high school education or less – the two variables in the lower half of Table 6 – in 2008 data, and dotted lines indicate the median of each variable. The dashed line is not the fitted line (although it is very

close); rather, it indicates the dividing line between areas with above- and below average wages for less-educated Spanish-speaking immigrants (relative to natives) according to the estimate in column (4) (ignoring the impact of the other variable, poor/strong English).[18] In areas above the dashed line, less educated Spanish-speaking immigrants tend to have relatively high wages: this includes two Cuban enclaves, Miami and Jersey City, other areas in Florida, as well as areas on the Texas border (not labeled on graph), and Riverside and Los Angeles. Interestingly, many "new" immigrant destinations (Singer, 2004) -- like Atlanta, Charlotte, Denver, Greensboro, Portland, and Raleigh -- are below the line, indicating the Mexican (or, really, any less-educated Spanish-speaking) immigrants there earn below average wages. This is a reversal of these areas' position not so long ago – a similar graph for 1990 would show most of these new destinations above the line. In fact, the model in column (4) applied to 1990 (Census of Population) data predicts that Mexican immigrants would have earned a relative wage seven percent higher in Singer's new destinations than in other large areas that were her historical immigrant gateways.[19] The relatively high wages may have been part of what attracted Mexicans to these areas starting in the 1990s, although other research indicates faster employment growth was the primary draw (Card and Lewis, 2007). Over the past 20 years, the wage advantage of these new areas for Mexicans has completely eroded because the Spanish inflows have been disproportionately less-skilled.[20]

[18] Specifically, it is a line with a slope of 0.7 = 0.045/0.065 – the ratio of the two lower coefficients in column 4 – that goes through the sample median of the two variables, which are shown with the dotted lines in the figure. The residual scatter plot, conditioning out poor/strong English, is shown in Appendix Figure A1 and has the same qualitative patterns.
[19] What I refer to as "new" immigrant destinations in the Singer (2004) typography include her three categories of "emerging," "re-emerging," and "pre-emerging." The predicted values are computed using all three variables in column (4). Another major factor in the wage determination of less educated workers is the ratio of college to non-college workers in a labor market (e.g., Card, 2009), but on this measure, new and old destinations were very similar in 1990. The actual average wage of Spanish speaking immigrants was 13 percent higher in new than old destinations in 1990.
[20] The patterns over time are similar for new relative to traditional Mexican destinations. I defined as "traditional" Mexican destinations the top 20 areas ranked on number of working age Mexicans in 1990. Drawing on Card and Lewis (2007), I defined 15 "new" Mexican destinations with a large residual inflow rate of Mexicans between 1990 and 2008 over what is predicted from a linear regression on 1990 Mexican (working-age population) share, and which are not traditional Mexican destinations. The large areas among the latter are all new destinations according to the Singer (2004) typography. Estimates in Table 6 imply that Mexicans would have earned 3.5 percent more in new than in traditional destinations in 1990, but by 2008 they would have earned one percent less.

The estimates in Table 6 might additionally be interpreted as indicating that a sufficiently rich skill distribution of Spanish speakers allows a "Spanish-speaking" labor market to emerge within an area. To provide additional evidence for this, column (5) shows an estimate of the same relationship for non-Spanish speaking immigrants who, in this view, should not be affected by the density of Spanish speakers. The wage boost from college educated Spanish speakers is not present for them. Also in column (5) there is a positive coefficient on the less-educated Spanish/English hours ratio which nearly offsets the coefficient on poor English/English hours, indicating that an inflow of less-educated Spanish speakers has little overall effect on the wages of non-Spanish speaking immigrants. Put differently, Spanish and non-Spanish speakers' wages respond most strongly to skill ratios in their own language group. This is consistent with Spanish speakers having their own labor market.

IV.C. Puerto Rico

The U.S. contains one labor market, not examined in the previous analysis, in which nearly all immigrants and natives speak Spanish: Puerto Rico.[21] According to the view that language skills drive imperfect substitutability with natives, Puerto Rican immigrants should be perfect substitutes for Puerto Rican natives: their relative wages should not respond to immigrants' relative labor supply. The approach of examining a market where immigrants and natives share a language was previously taken by Castillo, Gilless and Raphael (2009) (CGR), who examined Costa Rica.

Since Puerto Rico is a single labor market, the analysis cannot exploit geographic variation in the size of the immigrant population. Instead, I follow the approach of CGR and OP and use variation

[21] In 1990, which is the last year the question was asked, 97 percent of those born outside and 99 percent of those born in Puerto Rico indicated that they spoke Spanish. Note that this is a broader language measure than was used in the previous analysis, which only indicated whether or not the respondent spoke Spanish "at home."

over time and across education-experience cells, to be detailed below, in the size relative size of the immigrant workforce.

That Puerto Rico has a non-trivial amount of immigration is perhaps surprising. Figure 4 shows the share of the working age population that was born outside Puerto Rico – what I define as "immigrants" for this analysis -- between 1970 and the present, broken out separately by the five education categories -- four years college, 1-3 years college, high school completion, 1-3 years high school, no high school -- used in the analysis. It is based on data from Puerto Rican population censuses, which are taken in parallel to the U.S. Census, and the 2007-9 Puerto Rican Community Surveys (Ruggles et al., 2010). Most of those born outside Puerto Rico are not truly immigrants but are other U.S. citizens, mostly of Puerto Rican heritage, who were born in the continental U.S. Figure 4 shows until recently, about 15 percent of college educated workers came from outside Puerto Rico. The less-educated "immigrant" share is lower but, until recently, had been rising.

The analysis will exploit variation across the cells in Figure 4, and, within these, across years of potential experience, grouped into five year cells up to 40. Usual weekly hours worked is not available in the 1970 Census, so hourly wages and aggregate hours worked will be replaced with weekly wages and weeks worked. In particular, the estimation equation is:

$$(4)\ \ln\left(w_{ikt}^{F}/w_{ikt}^{N}\right)=\alpha_{ik}+\gamma_{kt}+\delta_{it}+\beta\ln\left(WW_{ikt}^{F}/WW_{ikt}^{N}\right)+\varepsilon_{ikt}$$

where $\ln\left(WW_{ikt}^{F}/WW_{ikt}^{N}\right)$ represents the aggregate relative weeks worked last year by non-Puerto Rican born workers with education *i* and potential work experience *k*. Estimates of (4) control for an exhaustive set of education x experience, experience x year, and education x year dummies.[22]

[22] All estimates of (4) are unweighted, with standard errors computed to be robust to arbitrary error correlation within education x experience cells. In the Puerto Rican data, weeks worked is computed for those aged 16-65 who are not living in group quarters and who are old enough to be out of school. The Puerto Rican wage sample consists of those in the weeks sample with positive wage earnings, zero business and farm earnings, who are not currently enrolled in school and who report an occupation.

For comparison, (4) will also be estimated using data on the continental U.S., which will also serve as an additional check on the cross-city results presented above.[23]

Results are presented in Table 7, with Panel A for Puerto Rico and Panel B for the Continental U.S. Column (4) shows a negative and marginally significant coefficient of -0.033, for the U.S., which is surprisingly similar to the estimates in Table 5 given the large difference in methodology. The estimate for Puerto Rico is not zero, but positive and significant. A likely explanation for this that changes in the weeks worked partly reflects labor demand, which moves weeks worked and weekly wages in the same direction. To address this, I exploit the fact that much of the variation in the relative size of the immigrant workforce in Puerto Rico is driven by changes in the raw numbers of U.S.-born ethnic Puerto Ricans and Puerto Rican-born. As these reflect education and fertility decisions made long in the past, they are unlikely to be systematically related to present demand.[24]

The size of these two populations is computed using the combination of Puerto Rican and U.S. data, which is necessary because one third of the Puerto Rican born population lives in the U.S. When relative weeks worked is replaced by the relative population size, in column (2), the coefficient is indeed nonpositive and is close to zero. Column (3) shows that this variable moves almost exactly one-for-one with the weeks worked variable, a relationship that is highly significant.[25]

The relationship in column (2) is unfortunately very noisy, and Figure 5 shows why. It shows a scatter plot underlying the relationship in column (2), as well as the U.S. relationship in column (4). The Puerto Rican wage estimates are much noisier, as shown by the large mean squared error

[23] The U.S. samples are defined similarly to the Puerto Rican ones (see previous note), except for a few things. First, to be included in the wage sample, rather than reporting an occupation, workers need to report being currently employed. (This difference in methodology is due to the fact that employment rates are very low in Puerto Rico.) Also, weekly wages below 10 or above 10,000 in 1999 dollars were reset to these thresholds. Finally, the two lowest education groups are combined in U.S. estimates, owing to the very small number of non-high school educated natives in the U.S.

[24] Although demand conditions in Puerto Rico may influence the educational attainment of the Puerto-Rican born population, this influence may be limited since a substantial fraction (about one third) of Puerto Ricans end up in the continental U.S. labor market.

[25] This implies that an instrumental variables estimate of the effect of relative weeks works on relative weekly wages, using the population variable as an instrument, would have the same coefficient as appears in column (2).

("vertical" variation around the line) in the Puerto Rican figure compared to the U.S., owing to the much smaller number of observations in the Puerto Rican data.

The remaining columns of Table 7 examine separately the relative wage response (in the continental U.S.) of immigrants with poor and strong English, measured in the same way as in Panel B of Table 5.[26] This measure is not available until 1980, and so column (5) show estimates excluding 1970, which are smaller in magnitude and less precise. Broken out separately by English skills, in columns (6) and (7), the negative response is limited to immigrants with poor English. Though these estimates are not precise, they are similar in magnitude to the difference in response between high- and low-English immigrants that was found in Table 5, around 3 percentage points.

V. Implications for Poverty

What can we say about the effects of immigration on poverty in light of these estimates? A full answer to this question requires estimates of the effect of immigration on a broader set of wage outcomes than were studied in this chapter, but are explored in other research (Raphael and Smolensky, 2008, 2009). To simplify, the main thing the present estimates omit is the effect immigration has on wages by shifting the ratio of non-college to college labor. However, Card (2009) shows this is small (because the ratio is similar for immigrant and native workers). On top of these previously estimated effects, the estimates in this chapter imply that immigration has a larger impact on the wages of immigrants with poor English language skills, already a high poverty group (Table 2), and depending on immigrants' mix of language skills, on Spanish-speakers.

Furthermore, the wage responses of these groups can be roughly translated into an effect on poverty rates, assuming all immigrants' wages are shifted down by the amount implied by the

[26] These columns exclude native-born workers who do not report speaking English "only" or "very well."

estimates. The estimates in both Table 7 and Table 5 are consistent with a one unit increase in the relative supply of immigrant labor -- again, roughly equal to the increase since 1990 -- lowering the wages of immigrant workers with poor English by 3 percentage points more than immigrants with strong English. The resulting three percent decline in wages, assuming all immigrants' income is from wages (which is close to true) would, without changing hours worked, be expected to drop the low-English immigrants between 100 and 103 percent of the poverty line into poverty. In 2008, this represented about one percent of low-English less-educated immigrants. If you also consider the fact that the decline in wages might induce some immigrants to work less, a reasonable approximation is to double this estimate.[27] In short, the rise in immigration since 1990 might have added one to two percentage points to the poverty rate of immigrants with poor English (compared to immigrants with strong English). Using the estimates in column (4) of Table 6, the change in the supply of Spanish, English, and other hours between 1990 and 2008 is expected to have lowered the wages of Spanish speaking immigrants by more, roughly 6 percent (relative to natives), which translates similarly into an increase in their poverty rates of Spanish-speaking immigrants of roughly 2 to 4 percentage points.[28] In practice, poverty rates among less-educated immigrants declined since 1990: from 35 to 30 percent among non-English speakers and from 25 to 22 percent among Spanish speakers. The estimates in this paper are consistent with the large immigrant influx since 1990 attenuating this decline.

[27] This comes from assuming a labor supply elasticity of one (following Raphael and Smolensky, 2009); that is to say, that a three percent decline in hourly wages is associated with a three percent decline in hours worked, for a total of a six percent reduction in annual labor income.

[28] Between 1990 and 2008, ln(poor/strong English hours) rose 1.39, and ln(Spanish/English hours) among non-college rose 1.16 and among college educated workers rose 0.58, for a -0.061 = -0.033*1.39 -0.045*1.16 + 0.065*0.57 change in the relative ln wage, using the coefficients from Table 6.

VI. Conclusions

On balance, the estimates in this paper suggest that while immigrants' imperfect English language skills may not be the only reason that they are imperfect substitutes for native-born workers in the U.S., they are a major reason for this. Using several different approaches and samples I have found that the wages of immigrants with poor English language skills tend to respond more negatively to a greater presence of immigrants than do the wages of immigrants with strong English language skills. A bottom line is that, at least among less-educated immigrants, immigration since 1990 might have pushed up the poverty rates of immigrants with poor English by one to two percentage points (relative to immigrants with strong English) and of Spanish-speaking immigrant by two to four percentage points (relative to natives). These impacts are overwhelmed by a downward trend in immigrant poverty rates, and are smaller for Spanish-speaking immigrants in markets where Spanish is prevalent in the college-educated workforce. These impacts may also help account for the new spread of Mexican immigrants to "new" immigrant destinations starting in the 1990s, as the lack of labor market competition from fellow Spanish speakers in those markets, compared to traditional Mexican destinations, would have made them a relatively attractive place to settle.

References

Berman, Eli, Kevin Lang, and Erez Siniver. "Language-Skill Complementarity: Returns to Immigrant Language Acquisition." *Labour Economics* 10(3): June 2003, pp. 265-290.

Bleakley, and Aimee Chin. "Language Skills and Earnings: Evidence from Childhood Immigrants." *The Review of Economics and Statistics* 86(2): May 2004, p. 481–496.

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

----. "Labor Outflows and Labor Inflows in Puerto Rico." *Journal of Human Capital* 2(1): Spring 2008, pp. 32-68.

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

----, "Immigration and Inequality," *The American Economic Review* 99(2), May 2009, pp. 1-21.

Card, David, and John DiNardo, "Do Immigrant Inflows Lead to Native Outflows?" *The American Economic Review* 90(2), May 2000, pp. 360-367.

Card, David, 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*. Chicago: University of Chicago Press, 2007.

Carliner, Geoffrey. "The Wages and Language Skills of U.S. Immigrants." *NBER Working Paper* 5763, September 1996.

Castillo, Federico, J. Keith Gilless, and Steven Raphael. "Comparing the Domestic Labor Market Impacts of a South-North and South-South Migration: The Cases of Costa Rica and the United States." Mimeo, UC Berkeley, September 2009.

Chiswick, Barry R. and Paul W. Miller. "The Endogeneity between Language and Earnings: International Analyses." *Journal of Labor Economics* 13(2): April 1995, pp. 246-288.

----. "Occupational Language Requirements and the Value of English in the U.S. Labor Market." IZA DP No. 2664, March 2007.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. "When Are Ghettos Bad? Lessons from Immigrant Segregation in the United States." *Journal of Urban Economics* 63(3): 759-74.

Dustmann, Christian and Arthur van Soest. "Language Fluency and Earnings: Estimation with Misclassified Language Indicators." *Review of Economics and Statistics* 83: 2001, 663-674.

----. "Language and the Earnings of Immigrants." *Industrial and Labor Relations Review* 55(3): 2002, pp. 473-492.

Dustmann Christian, and Francesca Fabbri. "Language Proficiency and Labour market Performance of Immigrants in the UK." *The Economic Journal* 113(489): July 2003, pp. 695-717.

Ferrer, Ana, David A. Green and Craig W. Riddell. "The Effect of Literacy on Immigrant Earnings." *The Journal of Human Resources* 41: 2006, 380-410.

Friedberg, Rachel and Jennifer Hunt. "The Impact of Immigrants on Host Country Wages, Employment and Growth." *Journal of Economic Perspectives* 9(2): Spring 1995, p. 23-44.

Goldin, Claudia and and Lawrence Katz. *The Race Between Education and Technology*, Cambridge, MA: Harvard University Press, 2008

Lang, Kevin. "A Language Theory of Discrimination." *The Quarterly Journal of Economics* 101(2): May 1986, pp. 363-382.

Lewis, Ethan, "How Do Local Labor Markets in the U.S. Adjust to Immigration?" Federal Reserve Bank of Philadelphia, Mimeo, November 2004.

Ottaviano, Gianmarco I.P., and Giovanni Peri, "Rethinking the Effect of Immigration on Wages," Mimeo, UC Davis, March 2010. Also forthcoming JEEA.

Peri, Giovanni, and Chad Sparber, "Task Specialization, Comparative Advantages, and the Effects of Immigration on Wages," *American Economic Journal: Applied Economics* 1(3), July 2009, pp. 135-69.

Raphael, Steven and Eugene Smolensky. "Immigration and Poverty in the United States." Institute for Research on Poverty Discussion Paper no 1347-08, University of Wisconsin-Madison, September 2008.

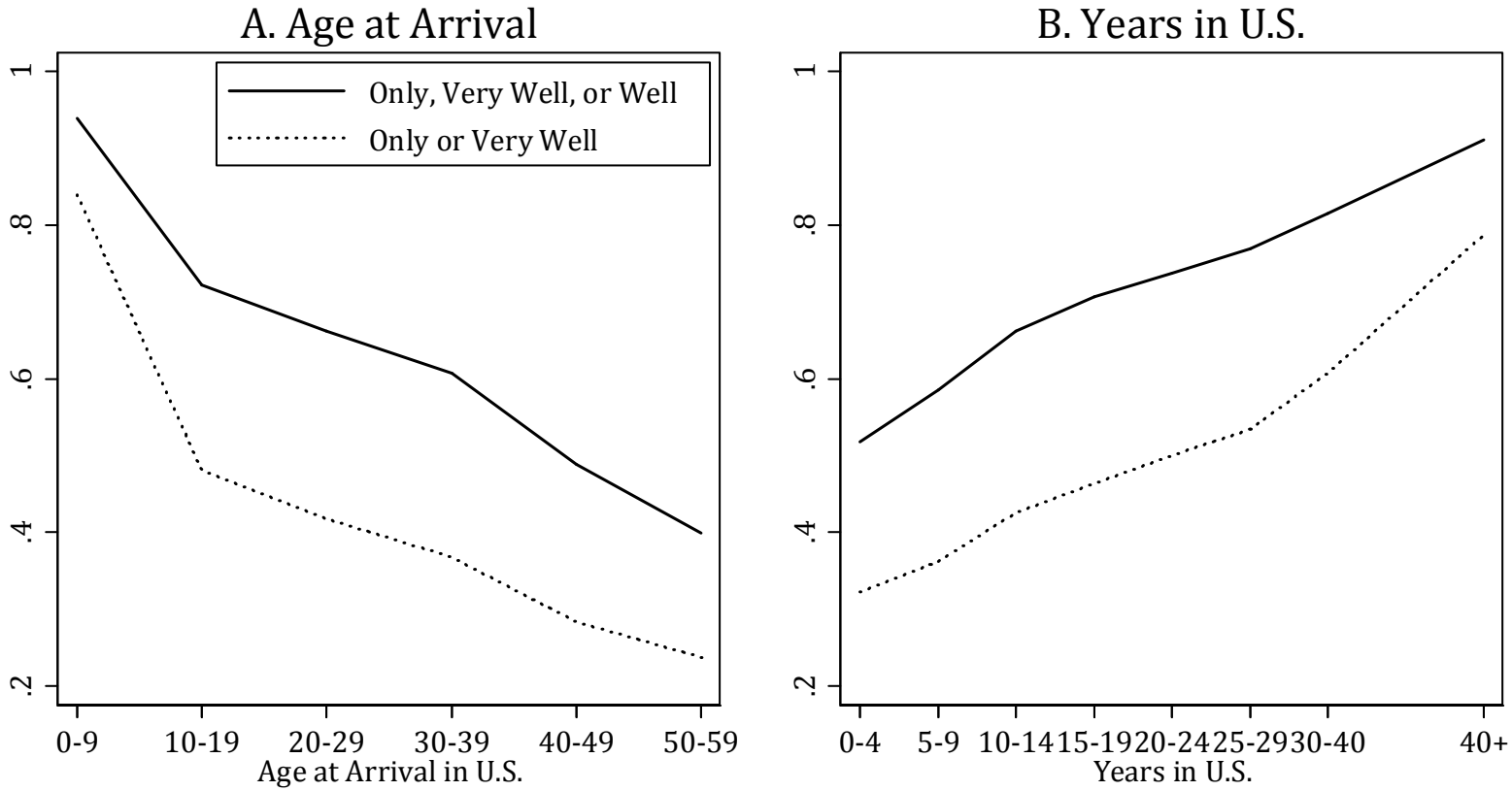
----. "Immigration and Poverty in the United States." *The American Economic Review* 99(2), May 2009, pp. 41-44.

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]. Minneapolis: University of Minnesota, 2010.

Singer, Audrey. "The Rise of New Immigrant Gateways." Brookings Institution Center on Urban and Metropolitan Policy, February 2004.

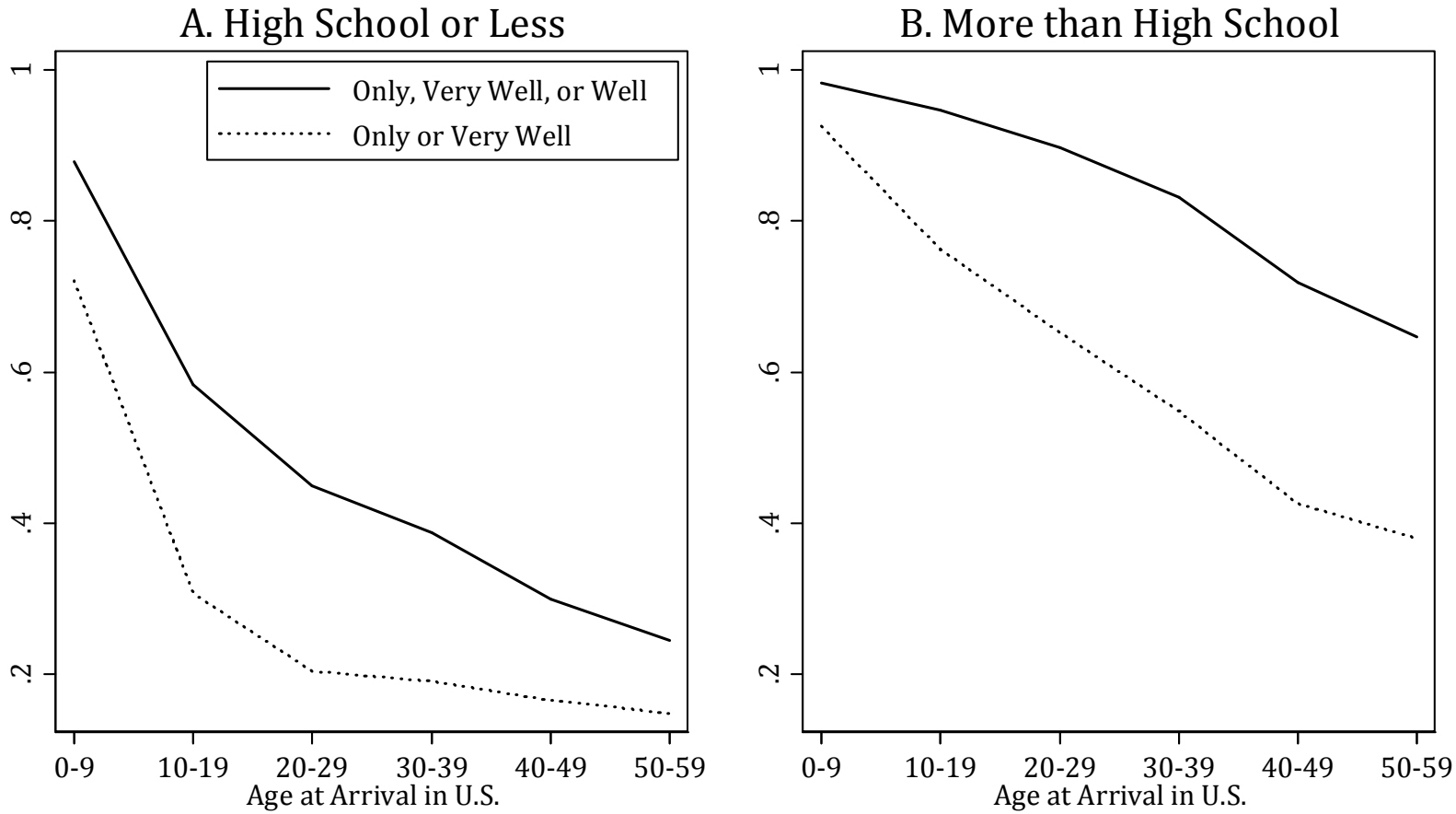
Figure 1a. Share of Immigrants Who Speak English, by Age At Arrival and Years in U.S., 2007-9

All Education Levels



Data Source: Combined 2007, 2008, and 2009 American Community Surveys (Ruggles et al., 2010). Sample limited to working age foreign-born living in 136 large metropolitan areas and not in group quarters.

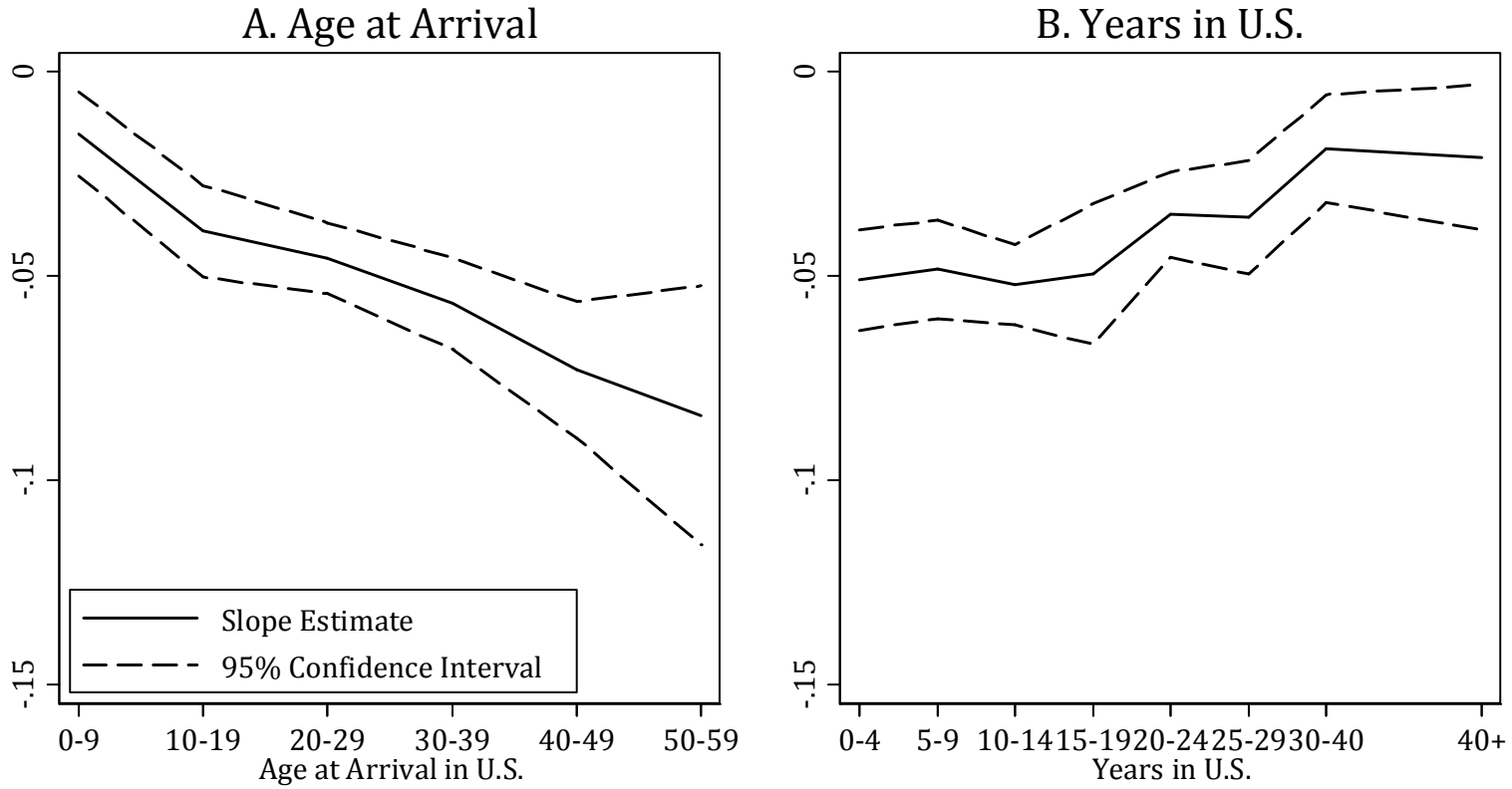
Figure 1b. Share of Immigrants Who Speak English, by Age At Arrival and Broad Education, 2007-9



Data Source: Combined 2007, 2008, and 2009 American Community Surveys (Ruggles et al., 2010). Sample limited to working age foreign-born living in 136 large metropolitan areas and not in group quarters.

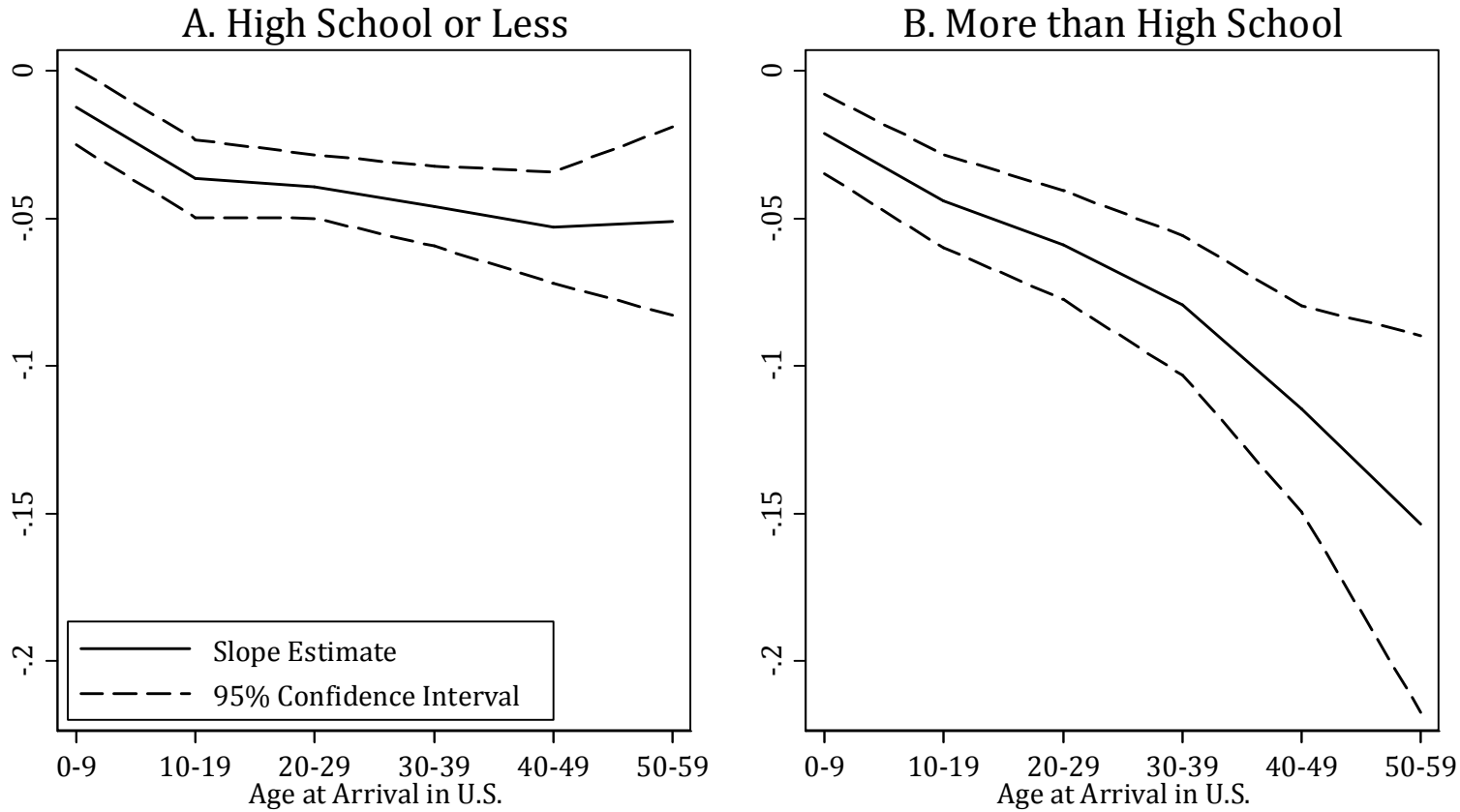
Figure 2a. Response of Immigrant Relative Wages to Relative Supply, by Age At Arrival and Years in U.S., 2000 and 2008

All Education Levels



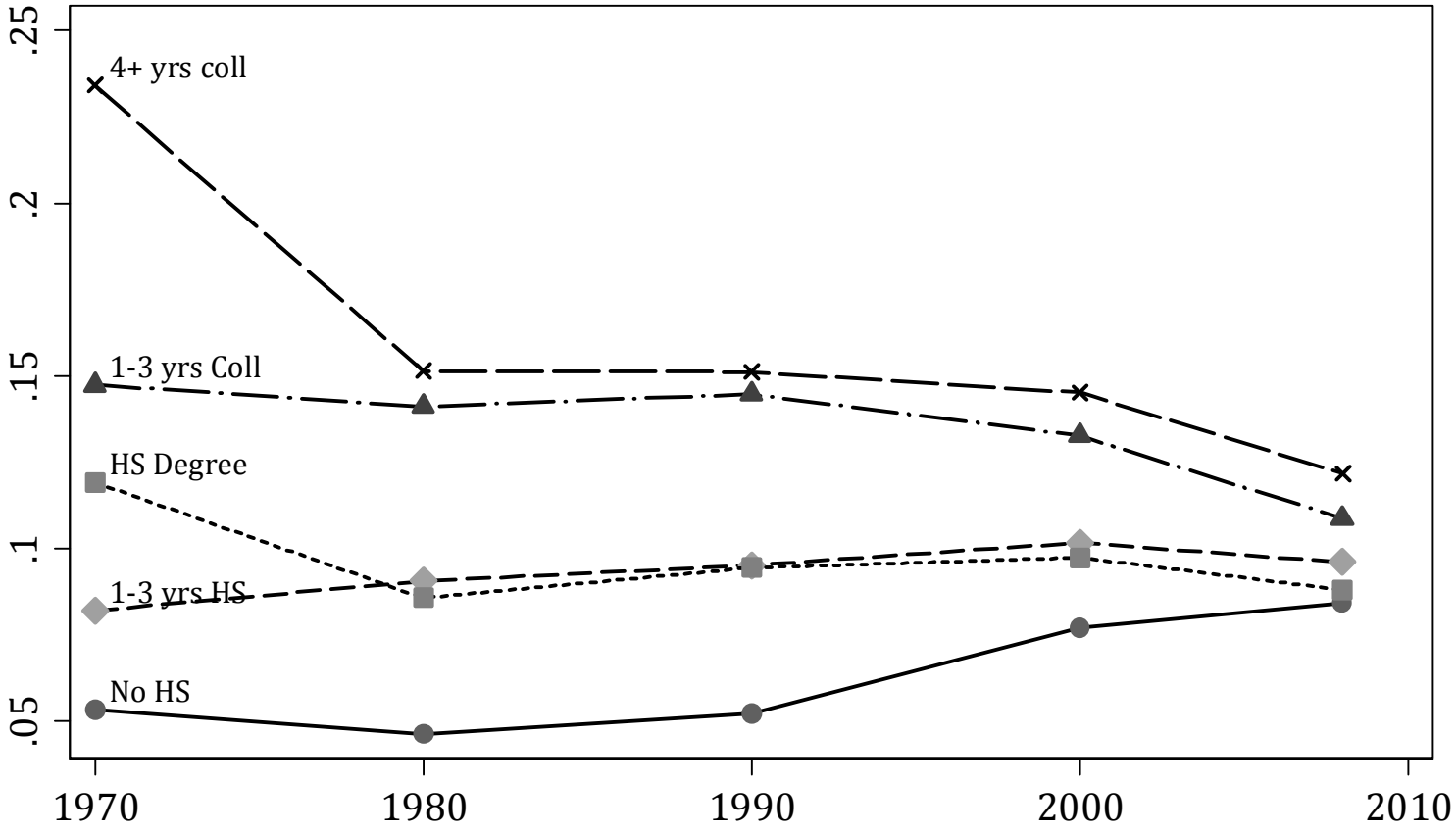
Data Source: 2007, 2008, and 2009 American Community Surveys (Ruggles et al., 2010) and 2000 Census of Population. See notes to Table 5 for details of data construction and estimation.

Figure 2b. Response of Immigrant Relative Wages to Relative Supply, by Age At Arrival and Broad Education, 2000 and 2008



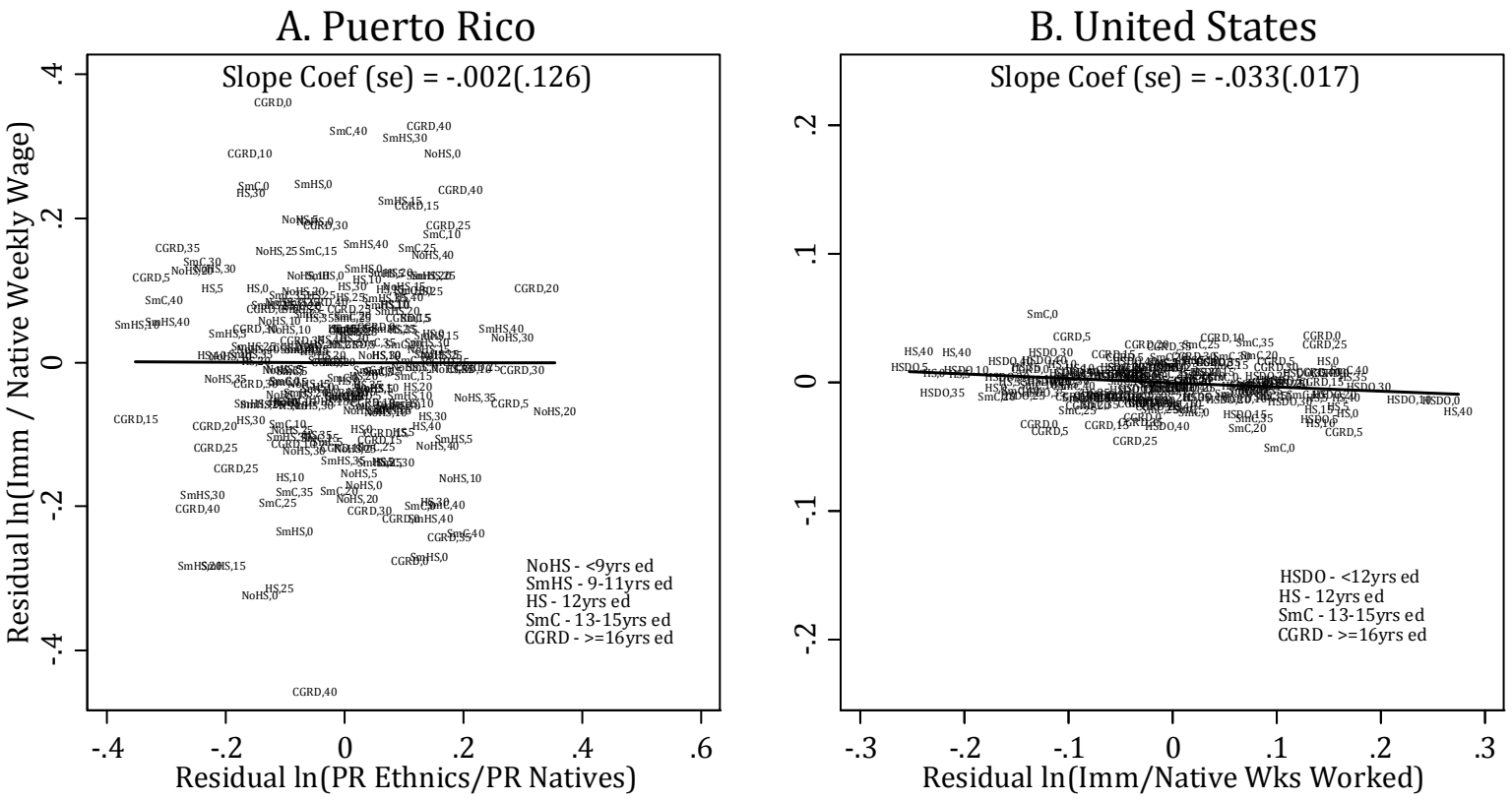
Data Source: 2007, 2008, and 2009 American Community Surveys (Ruggles et al., 2010) and 2000 Census of Population. See notes to Table 5 for details of data construction and estimation.

Figure 4. Non-Puerto Rican born share of Puerto Rican population, by education and year



Data source: 1970-2000 U.S. Censuses, and combined 2007-9 Puerto Rican Community Surveys (Ruggles et al., 2010). Sample limited to population aged 16-65, not living in group quarters and old enough to be out of in school with normal progression.

Figure 5. Response of immigrant relative weekly wage to immigrant relative supply: U.S. and Puerto Rico



Note: Scales of the two graphs are (roughly) proportional. Labels identify education x five year experiences cells. The x-variable in Panel A is the residual ln(ethnic Puerto Rican US-born/Puerto Rican-born population). (A one unit increase in this variable is associated with a one unit increase in the relative weeks worked of non-Puerto-Rican born workers -- see Table 7). All variables are residuals from a regression on an exhaustive set of education x year, experience x year, and education x experience dummies. Raw data sources are U.S. and Puerto Rican Population Censuses, and American and Puerto Rican Community Surveys (Ruggles et al., 2010). See text for details.

Table 1. Distribution Across English Speaking Categories, by Nativity, Education and Home Language

Subgroup:	A. All Working Age			B. Speaks Spanish At Home		
Education Level:	All Educ. Levels	High School or Less	More than High School	All Educ. Levels	High School or Less	More than High School
	(1)	(2)	(3)	(4)	(5)	(6)
Foreign-Born, Share Speaks English…						
Only/Very Well	0.464	0.278	0.679	0.261	0.177	0.516
Well	0.221	0.227	0.214	0.234	0.227	0.257
Not Well	0.211	0.318	0.088	0.317	0.365	0.173
Not At All	0.104	0.178	0.019	0.188	0.232	0.054
Native-Born, Share Speaks English…						
Only/Very Well	0.983	0.971	0.990	0.820	0.764	0.879
Well	0.011	0.017	0.007	0.113	0.139	0.086
Not Well	0.005	0.009	0.003	0.054	0.074	0.033
Not At All	0.001	0.003	0.000	0.013	0.023	0.002

Data Source: Combined 2007, 2008, and 2009 American Community Surveys (Ruggles et al. 2010). Sample limited to working-age population (age 16-65 with positive years of potential work experience) residing in 136 large metropolitan areas and not in group quarters. Panel B is further restricted to those who report speaking Spanish at home. Sample weights used to construct shares.

Table 2. Mean ln Wages and Poverty Rates By Nativity, Education, and Language Skills, 2008

Outcome:	A. Mean ln(hourly Wage), 1999$			B. Share of Group in Poverty		
Education Level:	All Educ. Levels	High School or Less	More than High School	All Educ. Levels	High School or Less	More than High School
	(1)	(2)	(3)	(4)	(5)	(6)
			Native Born			
All Native-Born	**2.25**	**1.94**	**2.39**	**0.10**	**0.16**	**0.06**
			Foreign-Born			
All Foreign-Born	**2.05**	**1.76**	**2.35**	**0.15**	**0.20**	**0.08**
Foreign-Born Who Speak English:						
Only/Very Well	2.31	1.91	2.46	0.09	0.14	0.06
Well	1.99	1.83	2.19	0.13	0.16	0.10
Not Well	1.70	1.67	1.87	0.21	0.22	0.17
Not at All	1.55	1.54	1.71	0.29	0.30	0.23
			Foreign-Born who Speak Spanish At Home			
All FB Spanish Speakers	**1.80**	**1.70**	**2.07**	**0.19**	**0.22**	**0.11**
Speaks no English	1.54	1.53	1.66	0.30	0.30	0.24

Data Source: Combined 2007, 2008, and 2009 American Community Surveys (Ruggles et al. 2010). In Panel B, sample is limited to working-age population (age 16 - 65 and old enough to be out of school, given normal progression) residing in 136 large metropolitan areas and not in group quarters. In Panel A, sample is limited to respondents from the Panel B sample that are currently employed and that had positive hours worked, positive wage and salary earnings, and zero self-employment and farm earnings in the past year. Wages are adjusted to 1999 dollars using the consumer price index, and wages exceeding 200 dollars and less than 2 dollars in 1999 dollars are replaced with these thresholds.

Table 3: The Role of Language Skills in Immigrant-Native Wage Gaps.

Dependent Variable:	ln(Hourly Wage), Workers with HS Education or Less					
Subgroup:				Spanish Speakers Only		
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant	-0.186	-0.041	-0.021	0.000	-0.003	-0.004
	(0.009)	(0.007)	(0.007)	(0.013)	(0.013)	(0.013)
Speaks English:						
Only or Very Well		0.209	0.168	0.149	0.208	0.174
		(0.008)	(0.009)	(0.008)	(0.021)	(0.019)
Only or Very Well **x** Share of MSA who speak Spanish at home						
Among entire working-age pop					-0.212	
					(0.081)	
Among those with HS or less						0.069
						(0.081)
Among those with more than HS						-0.306
						(0.165)
Sample Size	724,737	724,737	724,737	173,590	173,590	173,590
R^2	0.019	0.028	0.189	0.016	0.018	0.018
Other Controls?*	No	No	Yes	No	No	No

Data Source: Combined 2007, 2008, and 2009 American Community Surveys (Ruggles et al., 2010). Sample limited to working-age respondents (age 16 - 65 with positive years of potential work experience) that have 12 or fewer years of education, that reside one of 136 large metropolitan areas and not in group quarters, that are currently employed, and that had positive hours worked, positive wage and salary earnings, and zero self-employment and farm earnings in the past year. Wages exceeding 200 dollars or less than 2 dollars in 1999 dollars are replaced with these thresholds. Standard errors, in parentheses, computed to be robust to arbitrary error correlation within metropolitan area. *Other controls are a quartic in potential work experience; years of education; years of interacted with education below 9 years, born after 1950, and both; and dummies for education less than 9 years, born after 1950, female, black, Hispanic, female*black, and female*hispanic.

Table 4. Regression Data Descriptive Statistics

	All Education Levels		High School or Less		More than High School	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
	(1)	(2)	(3)	(4)	(5)	(6)
ln(imm hours/nat hours)	-1.934	1.019	-1.581	1.113	-2.288	0.769
Immigrant-Native Wage Gap						
All Immigrants	-0.125	0.079	-0.156	0.072	-0.093	0.073
High English Immigrants	-0.076	0.066	-0.088	0.064	-0.063	0.065
Low English immigrants	-0.360	0.201	-0.251	0.105	-0.469	0.215
Observations	544		272		272	

Data source: Combined 2007, 2008, and 2009 American Community Surveys (Ruggles et al., 2010) and 2000 Census of Population. Sample for constructing hours worked includes all those aged 16-65 who are old enough to be out of school (given normal progression) and residing in one of the 136 metropolitan areas in the sample and not in group quarters. To be in the wage sample requires being in the hours worked sample plus being employed, with positive weeks and hours worked last year, nonzero wage and salary earnings, and zero business and farm earnings; for natives it also requires speaking English "only" or "very well." Hourly wages above $200 and below $2 in 1999 dollars are reset to these thresholds. Data have been aggregated to 136 metropolitan areas x two education groups x two years (2000 and "2008," combining the three ACSs). Table shows unweighted means and standard deviations.

Table 5: Response of the Difference in the Mean ln(Hourly Wage) of Immigrants and Similar* Natives to Changes in Immigrant Relative Aggregate Hours Worked

Education Levels:	All	High School or Less	More than HS
	(1)	(2)	(3)
A. All Immigrants	-0.040	-0.034	-0.054
	(0.003)	(0.004)	(0.006)
B. By Broad English Language Skills:			
Speaks English Only,	-0.022	-0.011	-0.045
Very Well, or Well	(0.004)	(0.004)	(0.006)
Speaks English Not Well	-0.057	-0.047	-0.078
or Not At All	(0.006)	(0.006)	(0.016)
P-value, Equal Coefficients	0.000	0.000	0.030
C. By Detailed English Language Skills (112 MSAs):			
Speaks English Only or	-0.020	-0.012	-0.036
Very Well	(0.004)	(0.004)	(0.007)
Speaks English Well	-0.028	-0.018	-0.047
	(0.005)	(0.006)	(0.011)
Speaks English Not Well	-0.049	-0.037	-0.071
	(0.006)	(0.007)	(0.013)
Speaks English Not at All	-0.083	-0.050	-0.144
	(0.014)	(0.008)	(0.035)
P-value, Equal Coefficients	0.000	0.000	0.005

Data Source: Combined 2007, 2008, and 2009 American Community Surveys (Ruggles et al., 2010) and 2000 Census of Population. Table shows coefficient estimates from regressions of the wage gap between specified immigrants and "similar" natives (see below) on the natural log of the ratio of aggregate hours worked of immigrants and natives, using variation across metropolitan areas, year (2000 or "2008") and the two broad education of columns (2) and (3). All regressions control for year by education effects. Standard errors, in parentheses, computed to be robust to arbitrary error correlation within metropolitan area. Sample for constructing mean wages limited to working-age respondents (age 16 - 65 and old enough to be out of school given normal progression), that reside one of 136 large metropolitan areas (112 in Panel C) and not in group quarters, that are currently employed, and that had positive hours and weeks worked, positive wage and salary earnings, and zero business and farm earnings in the past year; for natives, sample is further limited to those who report speaking English "only" or "very well." Hourly wages above $200 and below $2 in 1999 dollars were reset to these thresholds. *The mean ln hourly wage of "similar" natives is computed by weighting natives to have the same distribution across potential experience (in five year bands) x education (four groups: high school dropouts, high school, some college, and at least four years college) cells as the specified group of immigrants in the metropolitan area and year.

Table 6: Immigrants' Relative Wages and Language Supplies, 2000 and 2008

Dependent Variable:	ln(Immigrant Wage/Native Wage), High School or Less				
Subgroup of Immigrants:	All		Spanish-Speaking		Other
	(1)	(2)	(3)	(4)	(5)
Ln (Immigrant Hours/	-0.034				
Native-Born Hours)	(0.004)				
Ln(Poor/Strong English-Speaking		-0.034	-0.036	-0.033	-0.060
Hours)		(0.004)	(0.007)	(0.028)	(0.018)
ln(Spanish-speaking Hours/Strong English Speaking Hours):					
Among workers with High School				-0.045	0.052
Education or Less				(0.035)	(0.024)
Among workers with More than				0.065	-0.015
High School Education				(0.017)	(0.017)
Metro x Year Observations	272	272	272	272	272
R^2	0.300	0.298	0.157	0.224	0.110

Data Source: Combined 2007, 2008, and 2009 American Community Surveys (Ruggles et al., 2010) and 2000 Census of Population. Wage sample limited to working-age respondents (age 16 - 65 with and old enough to be out of school, given normal progression) that have 12 or fewer years of education (or a GED), that reside one of 136 large metropolitan areas and not in group quarters, that are currently employed, and that had positive hours worked, positive wage and salary earnings, and zero self-employment earnings in the past year; for natives it also requires speaking English "only" or "very well." The dependent variable is the difference in the mean ln hourly wage between the specified group of immigrants and similar natives, where the mean ln hourly wage of "similar" natives is computed by weighting natives to have the same distribution across potential experience (in five year bands) x education (high school dropouts or completers) cells as the specified group of immigrants in the metropolitan area and year. Strong English-speaking hours worked is the sum of hours worked by those who report speaking English "Only" "Very Well," or "Well," while poor English-speaking hours are the sum of hours worked reported by those who speak English "Not Well" or "Not at All" among working-age respondents. "Spanish-speakers" are respondents who report speaking Spanish at home. All regressions are unweighted and control for year effects. Standard errors, in parentheses, computed to be robust to arbitrary error correlation within metropolitan area.

Table 7: Continental U.S. and Puerto Rico: Aggregate estimates, 1970-2000

	A. Puerto Rico			B. Continental United States			
	Y=ln(foreign/native weekly wage)		Y = ln(for/ nat weeks)	Subsample:	Excluding 1970		
						Strong English	Poor English
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln(Foreign-born weeks/	0.150			-0.033	-0.021	-0.002	-0.031
Native-born weeks)	(0.061)			(0.017)	(0.019)	(0.020)	(0.049)
ln(Cont'l U.S. born ethnic Puerto		-0.002	1.003				
Ricans/Puerto-Rican born)		(0.126)	(0.126)				
Observations	224	224	224	180	144	144	144
R-squared	0.489	0.457	0.919	0.935	0.942	0.913	0.976

Data Source: 1970, 1980, 1990, and 2000 Puerto Rican Censuses of Population and combined 2007, 2008, and 2009 Puerto Rican Community Surveys (Panel A) and 1970, 1980, 1990, and 2000 U.S. Censuses of Population and combined 2007, 2008, and 2009 American Community Surveys (Panel B), all from Ruggles et al. (2010). Sample for independent variable includes workers age 16-65 who are old enough to be out of school given normal progression through school and not living in group quarters. The sample used to compute the dependent variable, weekly wages, is this sample with the additional requirement of being currently employed (U.S.) or reporting an occupation (Puerto Rico), not enrolled in school, and without business or farm income. In U.S. data (columns 4-7), weekly wages exceeding $10,000 or below $10 in 1999 dollars were reset at these thresholds. In columns 5-7, native-born workers who did no report speaking English "only" or "very well" were excluded from the wage sample. Sample weights used to aggregate variables to the 5-year experience x education cells used in the analysis (see text). Standard errors are calculated to be robust to arbitrary error correlation within education x experience cells.

Appendix A

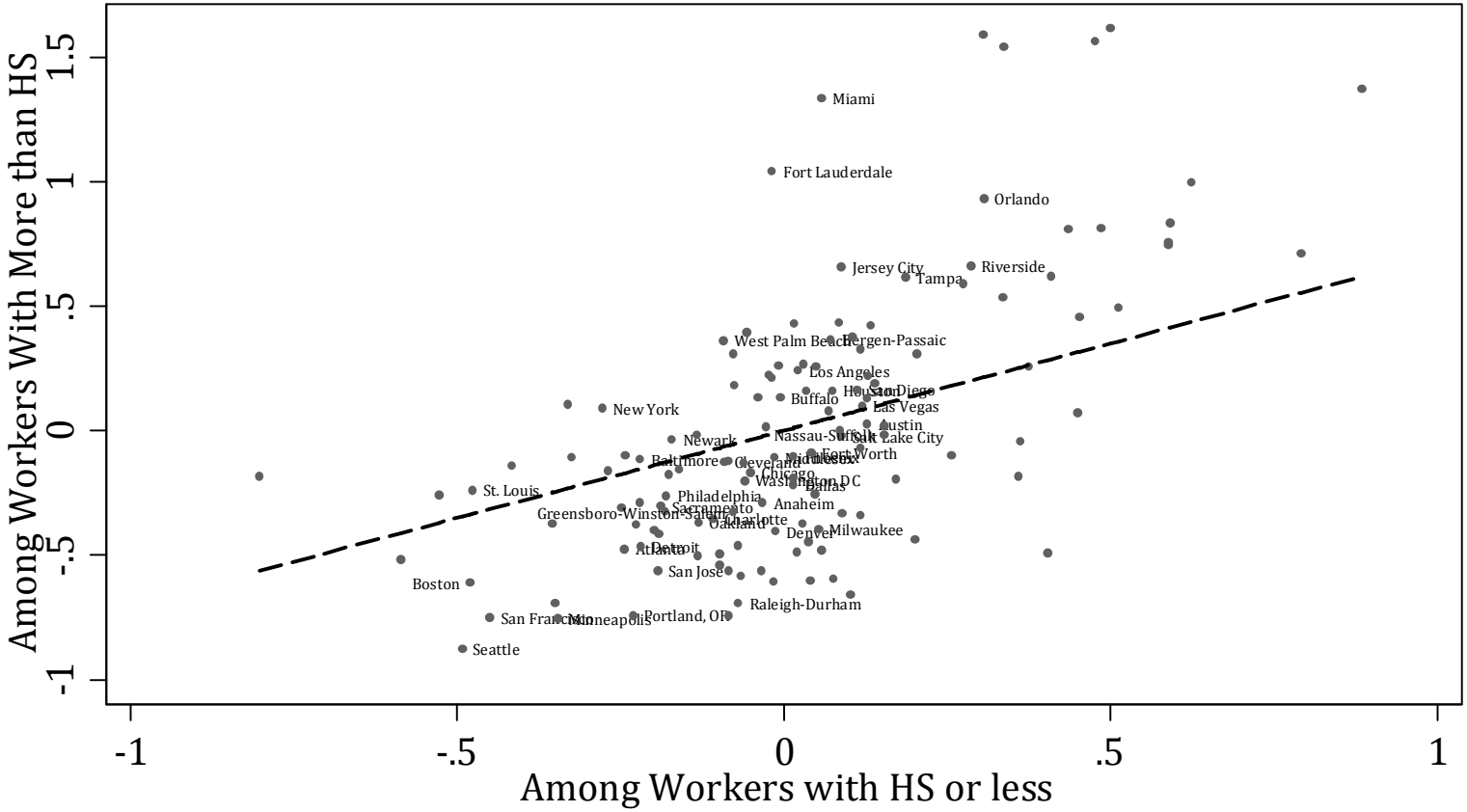
To address the possibility that a high presence of immigrants in a market reflects high wages in those markets -- which would tend to bias the coefficients in Tables 5 and 6 towards zero, I use an instrumental variable for $\ln(H^F/H^N)$ from Card (2009), which measures the immigrant inflow rates predicted by the area's lagged origin mix of immigrants. In particular, it is:

$$z_{jct} = \frac{\sum_o f_{oc,t-2} \Delta P^F_{ojt}}{P_{c,t-1}},$$

where $f_{oc,t-2}$ represents a two decade lag in the fraction of U.S. immigrants from region *o* living in metropolitan area *c*, which apportions ΔP^F_{ojt}, the number of immigrants (nationally) arriving from *o* in skill group *j* in the past decade. (The regions are listed in Table A2, along with ΔP^F_{ojt} figures.) The numerator is thus the predicted number of immigrant arrivals in cell *jc*. This is converted to a predicted arrival rate by dividing by the area's beginning of decade population, $P_{c,t-1}$. The assumption behind this instrument is immigrants persist in locating in certain areas because they value being near similar immigrants, and not because these areas have persistently stronger wage growth for that type of immigrant.

This instrument is a strong predictor of immigrants' relative hours. F-statistics on the instrument in the first stage are in the 50-100 range. In addition, instrumental variables estimates using this instrument are similar to the OLS estimates presented in Table 5. This is shown in Appendix Table A3, which is identical in structure to Table 5 but shows IV estimates. It shows the same pattern of coefficients, with greater (magnitude) wage responses for immigrants with worse English language skills. The differences in coefficients across English-speaking categories are similar in magnitude to the OLS estimates, and are, as in OLS, statistically significant.

Appendix Figure A1. Residual ln(Spanish/English Hours), 2008: more than high school vs. high school or less



Data Source: 2007-9 American Community Surveys. Points are residuals of a regression of ln(Spanish/English hours) on ln(poor/strong English hours), separately by education (high school or less on the x-axis, more than HS on the y-axis). Points above sloped line have above average wages for Spanish-speaking immigrants, relative to natives, (conditional on ln(poor/strong English hours) according to estimates in Table 6, column 4.

Table A1. Share Speaks Spanish at Home, by Broad Education, Selected Metropolitan Areas

Area	High School or Less	More than High School	Area	High School or Less	More than High School
Anaheim, CA	0.543	0.118	McAllen, TX	0.928	0.789
Aurora, IL	0.502	0.097	Miami, FL	0.684	0.612
Bakersfield, CA	0.507	0.198	Oxnard-Ventura, CA	0.528	0.147
Brownsville, TX	0.816	0.670	Riverside, CA	0.505	0.213
El Paso, TX	0.846	0.678	Salinas, CA	0.673	0.193
Jersey City, NJ	0.526	0.270	San Antonio, TX	0.540	0.292
Laredo, TX	0.890	0.818	Santa Barbara, CA	0.589	0.151
Los Angeles, CA	0.633	0.216	Santa Cruz, CA	0.515	0.093

Data source: stacked 2007, 2008 and 2009 American Community Surveys (ACSs) (Ruggles et al., 2010). Sample limited to working-age population (age 16-65 and old enough to be out of school, given a normal progression through school), and not living in group quarters. Computed using ACS sample weights.

Table A2. National Immigrant Arrivals, by Education, Origin, and Decade

Origin Group:	2000-2008 High School or Less	2000-2008 More than High School	1990-2000 High School or Less	1990-2000 More than High School
	(1)	(2)	(3)	(4)
Mexican	1,944,656	292,542	2,618,328	296,963
Central American	517,066	105,261	493,669	92,671
South American	290,534	302,840	333,430	275,063
Caribbean (ex Cuban)	202,625	131,153	331,827	148,237
SE Asian (ex Filipino)	139,257	104,244	288,013	173,567
Chinese	135,836	220,608	158,375	302,729
Russian or E European	133,065	286,665	283,883	385,346
Sub-Saharan African	129,245	178,315	129,346	173,449
South Asian	123,072	497,999	148,698	430,311
Cuban	89,306	56,648	109,769	56,659
Middle Eastern (ex Israeli)	88,988	165,310	94,684	137,885
Filipino	56,810	229,456	91,406	219,320
Commonwealth	51,432	189,733	74,478	264,485
Korean or Japanese	50,217	220,028	84,669	248,958
Southern European	27,374	46,875	34,168	48,243
Northern European*	9,521	63,872	55,668	169,033

* Includes Israelis. Data Source: Combined 2007, 2008, 2009 American Community Surveys (columns (1) and (2), from Ruggles et al., 2010) and 2000 Census of Population (columns (3) and (4)). Sample limited to working age population (age 16-65 and old enough to be out of school, given normal progression through school) and not living in group quarters.

Table A3: Instrumental Variables Estimates of the Response of the Difference in the Mean ln(Hourly Wage) of Immigrants and Similar* Natives to Changes in Immigrant Relative Aggregate Hours Worked

Education Levels:	All	High School or Less	More than HS
	(1)	(2)	(3)
A. All Immigrants	-0.035	-0.030	-0.052
	(0.005)	(0.006)	(0.007)
B. By Broad English Language Skills:			
Speaks English Only,	-0.018	-0.010	-0.044
Very Well, or Well	(0.005)	(0.005)	(0.006)
Speaks English Not Well	-0.043	-0.033	-0.074
or Not At All	(0.009)	(0.009)	(0.016)
P-value, Equal Coefficients	0.000	0.001	0.060
C. By Detailed English Language Skills (112 MSAs):			
Speaks English Only or	-0.022	-0.018	-0.036
Very Well	(0.005)	(0.005)	(0.007)
Speaks English Well	-0.015	-0.005	-0.046
	(0.007)	(0.008)	(0.011)
Speaks English Not Well	-0.030	-0.017	-0.069
	(0.010)	(0.011)	(0.015)
Speaks English Not at All	-0.057	-0.030	-0.137
	(0.014)	(0.013)	(0.029)
P-value, Equal Coefficients	0.000	0.009	0.001

Data Source: Combined 2007, 2008, and 2009 American Community Surveys (Ruggles et al., 2010) and 2000 Census of Population. Table shows coefficient estimates from regressions of the wage gap between specified immigrants and "similar" natives (see below) on the natural log of the ratio of aggregate hours worked of immigrants and natives, using variation across metropolitan areas, year (2000 or "2008") and the two broad education of columns (2) and (3). All regressions control for year by education effects and are estimated by instrumental variables using the lagged origin mix instrument described in the appendix. Standard errors, in parentheses, computed to be robust to arbitrary error correlation within metropolitan area. Sample for constructing mean wages limited to working-age respondents (age 16 - 65 and old enough to be out of school given normal progression), that reside one of 136 large metropolitan areas (112 in Panel C) and not living in group quarters, that are currently employed, and that had positive hours and weeks worked, positive wage and salary earnings, and zero business and farm earnings in the past year; for natives, sample is further limited to those who report speaking English "only" or "very well." Hourly wages above $200 and below $2 in 1999 dollars were reset to these thresholds. *The mean ln hourly wage of "similar" natives is computed by weighting natives to have the same distribution across potential experience (in five year bands) x education (four groups: high school dropouts, high school, some college, and at least four years college) cells as the specified group of immigrants in the metropolitan area and year.