# Inequality Between and Within Immigrant Groups in the United States

Hisham Foad*

January 2018

The increase in income inequality has been one of the defining economic trends of the past forty years. The increase in inequality in the United States has been attributed to skill-biased technological change, globalization, and a a changing institutional environment. To what extent can these factors explain rising inequality? I attempt to answer this question by looking at income inequality both within and between immigrant groups in the United States. There is tremendous variation in income inequality between these groups, with Gini coefficients ranging from 0.59 for immigrants from the MENA to 0.42 for immigrants from Mexico. There are also large differences in inequality between different enclaves of immigrants from the same source country. For example, MENA immigrants living in Michigan have an income Gini coefficient of 0.61 as compared to 0.55 for MENA immigrants living in New Jersey. To what extent are differences in inequality between immigrant groups driven by observable characteristics that differentiate these groups? What features of these immigrant enclaves drive differences in immigrant inequality? In this study, I exploit the variation in income inequality both between and within immigrant groups to estimate the micro level determinants of income inequality using a broad sample of 32 immigrant groups distributed across a wide range of ethnic enclaves derived from ACS data. I utilize a regression decomposition technique from Fields (2003) and find that most of the inequality between immigrant groups can be explained by observable characteristics like education, leaving little left over for unobservable "cultural factors." Within groups, there is some variation in the determinants of inequality. For groups like Iranian immigrants, inequality is driven by educational differences, suggesting a policy aimed at increasing educational opportunities. For Mexican immigrants, the largest determinant of inequality is gender, suggesting policies aimed at reducing gender disparities in income. Other groups like Vietnamese immigrants see inequality driven by the amount of time in the US, suggesting policies aimed at accelerating assimilation. That the sources of inequality differ across groups implies a more nuanced approach to crafting policies aimed at reducing income inequality.

*Department of Economics, San Diego State University. Please direct all correspondence to the author at hfoad@mail.sdsu.edu

# 1 Introduction

Rising income inequality has been one of the defining economic trends of the past forty years. Piketty and Saez (2003) show that the income share of the top 1% has steadily increased since the early 1970's, with the top 1% of earners holding over 20% of income in 2013.[1] The increase in inequality in the United States has been often been attributed to skill-biased technological change and globalization. Other explanations include the superstar hypothesis (Kaplan and Rauh, 2013) in which technology and a changing institutional environment have allowed gains to the very top earners to rise dramatically. Changes to the institutional environment include a less progressive tax system and a decline in union power. Another explanation related to this paper is that inequality has increased due to low skill migration. With low skill migration, wages paid to low-skill natives could theoretically decline (as in Borjas, 2003). However, Card (2009) and Ottaviano and Peri (2006) argue that low skill immigrants and low skill natives are imperfect substitutes. As such, immigrants tend to compete not with natives, but each other. Immigration has only a modest effect on inequality, with Card (2009) estimating that only 5% of the overall increase in inequality between 1980 and 2002 can be explained by immigration.

Concerns over income inequality have been magnified by a number of studies that have found that countries with greater inequality tend to have worse socio-economic outcomes than those with more even income distributions.[2] Given that the observed level of inequality in the United States may be economically inefficient, it is important that we understand the causes of inequality in order to craft effective policy addressing this issue.

The goals of my study are two-fold. First, I examine the determinants of income inequality by exploiting differences in inequality between and within immigrant groups in the United States. These immigrant groups may differ from natives in a number of ways such as language proficiency, citizenship, tenure in the US, education, and culture. Immigrants groups differ from each other along these same lines as well. To what extent are differences in inequality between immigrant groups due to observable characteristics? How much is left over for unobservable cultural characteristics? Within immigrant groups (i.e. holding culture constant), how much of the differences in inequality between immigrant communities can be explained by observable characteristics of these communities? Exploiting these variations can help to shed light on the various explanations for income inequality.

The second goal of my study is to generate policy recommendations from evaluating the differential causes of inequality across and within immigrant communities. For example, the education has the largest effect on inequality within Arab communities, while gender has the largest effect on inequality within Mexican communities. Thus, a policy targeting inequality in Arab communities should focus on increasing educational access, while one targeting inequality in Mexican communities should target gender inequality.

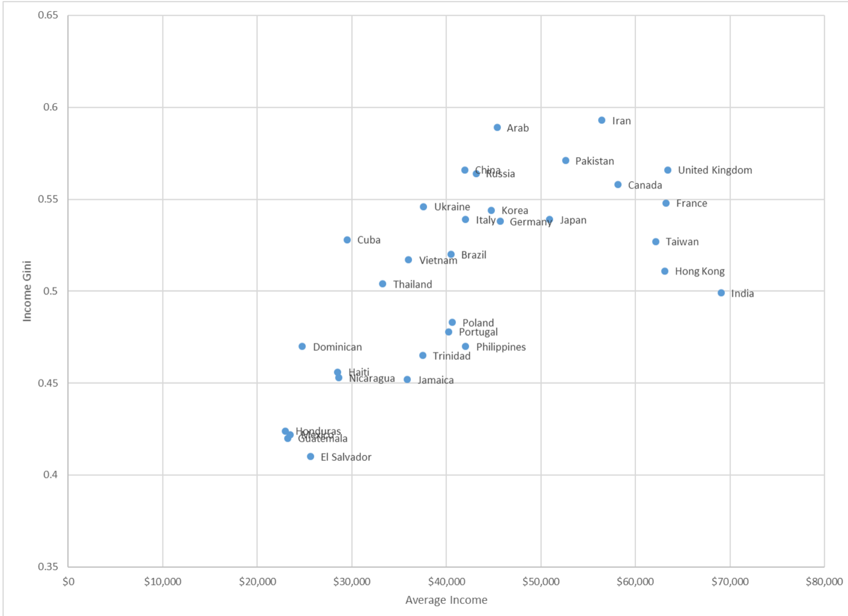
Of course this approach is only valid if there are differences in inequality both between and within immigrant communities. Figure 1 below indicates that there are indeed large differences in inequality across immigrant groups. For this figure, I estimate inequality using the Gini coefficient computed using total income data from the 2008-2014 American Community. The Gini coefficient is an index measuring inequality that ranges from 0 (perfect equality) to 1 (perfect inequality). As can be seen, there is considerable variation in inequality, with Iranian immigrants having a Gini coefficient of just under 0.6 and Salvadoran immigrants having a Gini coefficient just above 0.4.

Interestingly, there appears to be a strong and positive correlation between average income for each immigrant group and the degree of income inequality within each group. This would seem to suggest that much of the observed differences in inequality between these groups can be explained by factors that also influence differences in the level of income between these groups. Thus, the reasons why there is greater income inequality among Iranian immigrants than Salvadoran immigrants could be the same as why on average, Iranian migrants earn more than their Salvadoran counterparts.

[1] For recent inequality data, see inequality.org

[2] See for example Wilkinson and Pickett (2011)

Figure 1: Inequality and Income Across Immigrant Groups




Data from the 2008-2014 American Community Survey. Inequality is measured using the Gini coefficient computed from total personal income within each immigrant group. Average income is the average of this variable within each group.

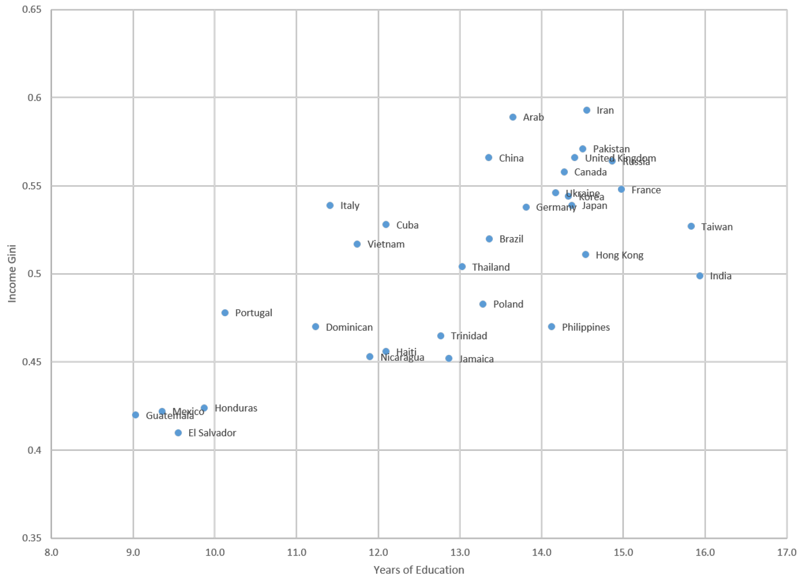
Evidence of this is given by Figure 2, which graphs the Gini coefficient against average years of education for the thirty-two immigrant groups in my sample. There is an even stronger positive correlation between inequality and years of education, with the most educated immigrant groups having the greatest amount of income inequality. This story fits with existing explanations that suggest that inequality has been driven less by reductions at the bottom and more by gains at the top. Members of better educated immigrant communities have more opportunities to become "earnings superstars," pushing up both average income and income inequality. Of course, we cannot perfectly explain inequality with education and average income, suggesting that other factors are important. For example, Arab and Filipino immigrants have around the same average years of education, but there is much more income inequality among Arab immigrants than among Filipino immigrants. What factors are leading to greater inequality among Arab immigrants? To what extent can these factors shed light on the root causes of income inequality?

The rest of this paper is organized as follows: Section 2 discusses the empirical methodology used to decompose the sources of income inequality both between and within immigrant groups. Section 3 describes the data used to conduct this analysis, justifying the inclusion of inequality determinants from the literature. Section 4 discusses the results of the analysis and section 5 concludes.

# 2 Methodology

The existing literature on the determinants of inequality can be split into two empirical methodologies. One attempts to decompose inequality by income sources following Fei et al (1978), Pyatt et al (1980), and Shorrocks (1982). The second decomposes inequality by population subgroup as in Blinder (1973), Oaxaca (1973), and Shorrocks (1984). Each methodology has its limitations. For example, decomposition by income source can help to identify how much inequality is due to, for example, investment income. However, it cannot assess how other important factors such as location, industry, education,

Figure 2: Inequality and Education Across Immigrant Groups




etc. affect inequality. Decomposition by subgroup can tell us how income varies between subgroups, but it does not tell us why there are inequality differences between these groups.

An attempt to bridge the gap between these methodologies was first proposed by Fields (2003) and Heltberg (2003). This technique combines income regression analysis with a Shorrocks-type income source decomposition. Essentially, the determinants of income are first estimated and then the contribution of each determinant to the variation in income across a particular group is computed. With this technique, we can go beyond decomposing inequality into simply income sources and differences between sub-groups. Rather, we can explain inequality by any factor that can be included in an income regression, including variables that could be policy relevant. As such, this methodology is well-suited for analyzing the causes of inequality as well as how differences in inequality between and within groups could be influenced by different variables.

The first step in the regression decomposition of inequality is to estimate the following equation:

$$y = \alpha + X\beta + u \tag{1}$$

Log income ($y$) is regressed on a vector of individual-level income determinants such as education, age, years in the US, English language ability, citizenship status, marital status, location, etc. Of course some of these income determinants may be endogenous. For example, education and income are both influenced by unobserved ability. For now, I am ignoring the endogeneity problem as I am less concerned with the determinants of income and more with the influence of these variables on income inequality. A future revision will pay closer attention to the endogeneity issue in this first stage regression.

The second step in this method is to use the estimates from the regression above to construct factor inequality weights for each variable in the regression. Following Shorrocks (1982) inequality decomposition by income source, the relative factor inequality weight of covariate $x_k$ is given by:

$$s_k = \frac{cov(\hat{\beta}_k x_k, y)}{var(y)} = \frac{\hat{\beta}_k cov(x_k, y)}{var(y)} \tag{2}$$

Looking at the expression in equation 2, we can see that this could be estimated by multiplying $\hat{\beta}_k$ by the OLS coefficient from a univariate regression of $x_k$ on $y$ (Ravallion and Chen, 1999). This relative factor inequality weight indicates the percentage change in income inequality due to $x_k$. We can further illustrate the inequality decomposition by considering $s_u = \frac{cov(u,y)}{\sigma_y^2}$, the proportion of inequality that is left unexplained by the variables in the model. Looking more closely, we can see that $s_u$ is simply the fraction of the variation in $y$ that is unexplained by the model. Therefore:

$$s_u = \frac{cov(u,y)}{var(y)} = 1 - R^2 \rightarrow \sum_{k=1}^{K} s_k = R^2 \tag{3}$$

To gauge the proportion of *explained* inequality by each variable in our analysis, we calculate the percentage contribution of each variable to the overall $R^2$ of the regression in equation 1:

$$p_k = \frac{s_k}{R^2} \tag{4}$$

A nice feature of this methodology is that it allows us to combine both continuous and categorical variables together as well as combine factor inequality weights by group. For example, suppose we have a subset of variables $g$ that represent aspects of assimilation such as years in the US, citizenship status, and English language ability. We can easily combine the group factor inequality weight as

$$s_g = \sum_{k \in g} s_k = \frac{cov(\sum_{k \in g} \hat{\beta}_k x_k, y)}{var(y)} \tag{5}$$

To evaluate inequality across subgroups, we can easily readjust equation 1 to:

$$y = \alpha + X\beta + D\Gamma + u \tag{6}$$

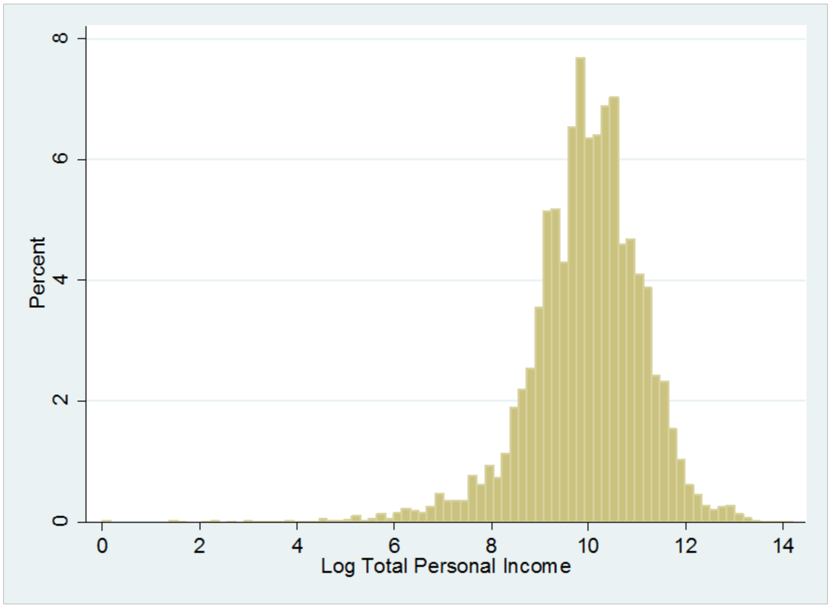
Where D represents dummy variables that identify different sub-groups. A key advantage over this approach is that our estimates of inequality decomposition will not depend on the choice of subgroup that is omitted (i.e. the control group with the dummy variables in $D$). Omitting a particular subgroup will not affect the $\hat{\beta}_k$ estimates or the overall explanatory power of the regression. It will affect $\alpha$ and the $\hat{\beta}_D$ estimates, but not the sum of all factor inequality weights for the set of all subgroup dummies.

The ability to decompose inequality by both source and group is a key reason why I chose this methodology. For example, decomposing the Gini coefficient would be problematic for my analysis as the Gini coefficient is not guaranteed to decompose perfectly into variations within and between groups. This is because there will invariably be some overlap between groups, leading to a lack of clear identification as to the true source of income inequality.

# 3 Data

The data used in this study comes from the 2008-2014 American Community Survey. Since the focus of this study is income inequality among immigrant groups, I restricted my sample to foreign born persons over the age of 17 and under 65 with positive income as classified by the total personal income variable that captures income from a variety of sources. Figure 3 displays a histogram of log income for my sample. The average log income in my sample is around 10, translating to a total personal income of around $22,000. However, the distribution has a negative skewness (-1.011), suggesting that the left tail (low income earners) is longer, but the distribution has more mass to the right of the mean. This is indicative of a high degree of income inequality.

Figure 3: Distribution of Income




To analyze differences between and within immigrant groups, I first need to define these groups. There is a tradeoff between the number of groups I consider and the number of individuals within each group. If I include every possible immigrant group, then some of these groups will have very few individuals and therefore weak identification of inequality within that group. As such, I choose to include only those immigrant groups (defined by country of birth) that have at least 10,000 individuals. Where possible, I combine countries of origin that have similar characteristics such as culture and language. For example, the Arab group includes immigrants born in North Africa, the Levant, the Arabian Peninsula, and Iraq. This group excludes, however, those born in Iran and Turkey, countries with different language. With this filter in place, I am able to identify 32 distinct immigrant groups in my sample.

Table 1 gives some summary statistics for the groups in my sample. By far the largest group in my sample is immigrants born in Mexico, comprising nearly a quarter of the sample. The three next largest groups after that are Filipinos (5.5%), Indians (4.4%), and Chinese (4.1%). There is considerable variation in income, ranging from a low of under $23,000 average income for Honduran immigrants to a high of over $69,000 for Indian immigrants. There is also considerable variation in years of education, with Guatamalans and Mexicans on average having around 9 years of education as compared to Taiwanese and Indians with nearly 16 years of education on average.

There are large differences in inequality, as previously illustrated in Figures 1 and 2. The lowest levels of income inequality appear for the least educated and lowest income groups (primarily from Mexico and Central America). Inequality is highest for the some of the richest and best educated groups (Pakistanis, Arabs, and Iranians). Interestingly, although average income and average education are positively correlated with income inequality, they are negatively correlated with education inequality. There is a simple correlation of -0.90 between the education Gini coefficient and average years of education, suggesting that the most educated groups also have the least inequality in education. That income inequality is higher for these groups supports the "superstar" hypothesis that inequality is driven by the highest performing individuals at the top. Highly educated groups offer more candidates to become these superstars, thus enabling greater income inequality. Stated differently, the most educated groups have higher income inequality because there are more opportunities for astronomical earnings than in the least educated groups.

**Table 1: Summary Statistics Across Immigrant Groups**

| Birthplace | Share | Avg. Income | Income Gini | Years Education | Education Gini |
|---|---|---|---|---|---|
| Arab Countries | 1.94% | $45,349 | 0.589 | 13.6 | 0.155 |
| Brazil | 0.79% | $40,460 | 0.520 | 13.3 | 0.133 |
| Canada | 2.96% | $58,135 | 0.558 | 14.3 | 0.118 |
| China | 4.11% | $41,924 | 0.566 | 13.4 | 0.209 |
| Cuba | 2.86% | $29,512 | 0.528 | 12.1 | 0.167 |
| Dominican Rep. | 1.82% | $24,744 | 0.470 | 11.2 | 0.183 |
| El Salvador | 2.52% | $25,652 | 0.410 | 9.5 | 0.246 |
| France | 0.64% | $63,206 | 0.548 | 15.0 | 0.117 |
| Germany | 3.65% | $45,665 | 0.538 | 13.8 | 0.108 |
| Guatamala | 1.62% | $23,208 | 0.420 | 9.0 | 0.281 |
| Guyana/Trinidad | 1.27% | $37,498 | 0.465 | 12.8 | 0.129 |
| Haiti | 1.25% | $28,497 | 0.456 | 12.1 | 0.147 |
| Honduras | 0.95% | $22,967 | 0.424 | 9.9 | 0.226 |
| Hong Kong | 0.68% | $63,078 | 0.511 | 14.5 | 0.132 |
| India | 4.41% | $69,048 | 0.499 | 15.9 | 0.103 |
| Iran | 0.96% | $56,428 | 0.593 | 14.5 | 0.144 |
| Italy | 1.41% | $42,030 | 0.539 | 11.4 | 0.218 |
| Jamaica | 1.71% | $35,864 | 0.452 | 12.9 | 0.124 |
| Japan | 1.32% | $50,908 | 0.539 | 14.4 | 0.109 |
| Korea | 2.55% | $44,756 | 0.544 | 14.3 | 0.124 |
| Mexico | 23.86% | $23,502 | 0.422 | 9.4 | 0.243 |
| Nicaragua | 0.60% | $28,616 | 0.453 | 11.9 | 0.163 |
| Pakistan | 0.64% | $52,597 | 0.571 | 14.5 | 0.136 |
| Philippines | 5.47% | $42,038 | 0.470 | 14.1 | 0.108 |
| Poland | 1.26% | $40,616 | 0.483 | 13.3 | 0.132 |
| Portugal | 0.55% | $40,224 | 0.478 | 10.1 | 0.255 |
| Russia | 0.92% | $43,124 | 0.564 | 14.9 | 0.124 |
| Taiwan | 1.08% | $62,164 | 0.527 | 15.8 | 0.106 |
| Thailand | 0.54% | $33,264 | 0.504 | 13.0 | 0.161 |
| Ukraine | 0.87% | $37,583 | 0.546 | 14.2 | 0.131 |
| United Kingdom | 2.41% | $63,388 | 0.566 | 14.4 | 0.111 |
| Vietnam | 3.41% | $35,987 | 0.517 | 11.7 | 0.209 |

Having defined immigrant groups, the next step is to identify the covariates that may influence income inequality both within and between immigrant groups. From the table above, it is clear that years of education should be included. To what extent is income inequality driven by differences in education? Hershbein, Kearny, and Summers (2015) estimate what would happen to inequality if 10% of non-college educated working men were to obtain a bachelor's degree. They find that although average earnings would rise for this group there would be a much smaller decline in inequality of less than 10%. Thus, education matters for inequality, but it is clearly not the only factor.

In addition to education, I consider several other factors commonly used in earnings regressions: age, gender, and marital status. Older workers tend to have more experience and should earn more. In addition, older workers have had a chance to accumulate more capital, which is significant since my income measure includes earnings from all sources. Gender may be significant given the observed wage gap between men and women. Furthermore, income inequality tends to be higher amongst men than women, mostly because of greater earnings at the top of the income distribution for men than at the top for women. Finally, marital status could have an effect on inequality through its effects on earnings. Numerous studies have shown that being married is associated with higher earnings (though here is certainly reverse causality). With higher earnings comes the opportunity for greater income inequality.

As my study focuses on immigrants, I also need to control for several factors that may influence inequality that are unique to the immigrant experience. First, tenure in the US could influence inequality as those who have been in the US for a longer period of time have had more opportunities to accumulate the social capital necessary to ascend to the highest earning potential. English language ability is also a key variable as a lack of proficiency in English will cap the earnings potential of an immigrant. Citizenship status could also matter. Bratsberg, Ragan, and Nasir (2002) argue that becoming a naturalized citizen increases earnings potential through two channels: improved job access and the acquisition of "US-specific human capital" which is incentivized by a decision to remain in the US permanently.

In addition to the variables mentioned above, I also include two measures of immigrant clustering. The first is whether or not an immigrant lives within an ethnic enclave. An ethnic enclave is a geographic area in which a particular ethnic group are overly represented (e.g. Chinatown, Little Italy, etc.) Several studies (Zhou and Logan, 1989; Gang et al, 2000; Edin et al 2003) find a positive effect of living in an enclave on earnings due to lower costs of migration/assimilation, less labor market discrimination, and higher returns to pre-immigration human capital. Toussaint-Comeau (2005) finds that there are higher rates of self-employment within enclaves. Waldfogel (2003) argues that enclaves provide the critical mass necessary for group-specific enterprises and immigrant institutions to flourish. On the other hand, numerous studies (Sanders and Nee, 1987; Borjas, 2000; Chiswick and Miller, 2001; Gronqvist, 2006) argue that living within an enclave limits earnings potential since the lower need to assimilate limits investments in human capital and skill sets in enclaves may not be transferable outside enclaves, leading to potential discrimination by employers within the enclave. Foad (2016) finds that the type of enclave matters, with better educated enclaves increasing earning potential and less educated enclaves reducing earning potential, mostly due to neighborhood effects (i.e. being surrounded by successful people increases the chances of your own success).

I measure the strength of an ethnic enclave by looking at the fraction of a particular immigrant group living in a location relative to that location's share of the national population:

$$Enclave_{g,l} = \frac{N_{g,l}/N_g}{N_l/N} \tag{7}$$

The size of an enclave for group $g$ in location $l$ is defined as the share of immigrants from group $g$ that live in location $l$ divided by location $l$'s share of the national population. This number essentially states how much more or less likely an immigrant from group $g$ is to live in location $l$. An Enclave score of 1 suggests that immigrants are just as likely to live in that location while a score of 3 would suggest that immigrants are three times as likely to live there. Thus, the larger the Enclave score, the more "over-represented" a particular immigrant group is in that location. The level of geographic disaggregation I use is the metropolitan statistical area (MSA).

Table 2 gives a sampling for the largest enclaves across several immigrant groups in my sample. There is considerable variation in enclaves, with different immigrant groups tending to be more prevalent in different locations. This is important as having a diversity of enclaves increases the likelihood that any estimated effect of enclaves on inequality will be due to living in an enclave rather than some unobserved features of cities that are common enclaves to all immigrant groups.

The enclaves variable captures clustering of immigrants in terms of where they live, but what about clustering of immigrants in specific industries? If a particular immigrant group is over-represented in an industry, then inequality between immigrant groups could simply be driven by the fact that inequality tends to be higher in certain industries. To control for this, I estimate industrial clustering using a similar method as computing enclave strength:

$$Industry_{g,i,l} = \frac{N_{g,i,l}/N_{g,l}}{N_{i,l}/N_l} \tag{8}$$

**Table 2: Largest Enclaves for Selected Immigrant Groups**

| *Arab Countries* | | *China* | |
|---|---|---|---|
| MSA | Enclave Score | MSA | Enclave Score |
| Detroit-Warren-Dearborn, MI | 6.7 | San Francisco-Oakland-Hayward, CA | 8.6 |
| Athens-Clarke County, GA | 4.0 | San Jose-Sunnyvale-Santa Clara, CA | 6.3 |
| Modesto, CA | 3.4 | Durham-Chapel Hill, NC | 5.0 |
| *India* | | *Mexico* | |
| MSA | Enclave Score | MSA | Enclave Score |
| San Jose-Sunnyvale-Santa Clara, CA | 8.4 | Kennewick-Richland, WA | 13.0 |
| Yuba City, CA | 6.4 | El Centro, CA | 9.4 |
| Trenton, NJ | 4.8 | Laredo, TX | 9.0 |
| *Philippines* | | *Vietnam* | |
| MSA | Enclave Score | MSA | Enclave Score |
| Honolulu, HI | 14.0 | San Jose-Sunnyvale-Santa Clara, CA | 13.6 |
| Vallejo-Fairfield, CA | 12.0 | Kennewick-Richland, WA | 6.0 |
| San Francisco-Oakland-Hayward, CA | 6.2 | Los Angeles-Long Beach-Anaheim, CA | 4.7 |

The Industry score for immigrant group $g$ working in industry $i$ in location $l$ is defined as the fraction of immigrants from group $g$ in location $l$ that work in that industry divided by the share of workers in location $l$ that work in industry $i$. Again, this number tells us how much more or less prevalent immigrants are in an industry in a particular location. For this draft of the paper, geographic disaggregation is defined at the state level, though I intend to refine this to the MSA level in a future draft. I classify 19 different industries based on definitions provided by the American Community Survey.[3]

Table 3 gives selected industry scores for immigrant groups and locations.There is a decent amount of variation in the top industries across groups and locations. For example, Arab immigrants are most prevalent in Michigan, where they are twice as likely to work in retail trade than the average Michigan resident. By contrast, Mexican migrants are most prevalent in California, where they are 4.3 times more likely to work in agriculture than the typical Californian, but only 30% as likely to work in IT. Compare this to Iranian immigrants in California, who are nearly twice as likely to work in household services (e.g. hair salons, dry-cleaning, etc.), but virtually non-existent in agriculture. Given this diversity, it may be important to control for industrial clustering when evaluating the determinants of inequality.

In summary, the regression decomposition method discussed in Section 2 will first be estimated by regressing log income on years of education, age, gender, marital status, years in the United States, English language ability, citizenship status, an individual's Enclave score, and an individual's Industry score. To these observable variables, I add country of origin dummy variables to capture any remaining between group differences that can be thought of as representing "cultural factors." The coefficients from this first stage will then be multiplied by the slope coefficients from a univariate regression of each covariate on log income to compute factor inequality weights. I then compute the proportion of explained inequality for each variable by dividing the factor inequality weight by the $R^2$ from the 1st stage regression.

[3]The industries are Agriculture, Arts/Entertainment, Civic/Religious Services, Construction, Education, Finance, Food/Accommodations, Healthcare, Household Services, IT, Manufacturing, Mining, Professional Services, Public Administration, Repair/Maintenance, Retail Trade, Transportation, Utilities and Wholesale Trade

**Table 3: Industrial Clustering for Selected Immigrant Groups and Locations**

| Group | State | Largest Industry | Smallest Industry |
|---|---|---|---|
| Arab Countries | Michigan (2.83) | Retail Trade (2.2) | Mining (0.2) |
| Brazil | Massachusetts (9.25) | Civic/Religious Svcs (5.2) | Utilties (0.1) |
| China | New York (3.36) | Mining (1.9) | Transportation (0.4) |
| Cuba | Florida (12.5) | Food/Accomodation (4.1) | Agriculture (0.1) |
| Germany | Colorado (1.93) | Civic/Religious Svcs (1.3) | Utilties (0.3) |
| Haiti | Florida (7.51) | Food/Accomodation (2.2) | Mining (0.2) |
| India | New Jersey (3.89) | Professional Svcs (2.2) | Agriculture (0.1) |
| Iran | California (4.64) | Household Services (1.8) | Agriculture (0) |
| Japan | Hawaii (13.52) | Household Services (2.9) | Construction (0.3) |
| Mexico | California (3.11) | Agriculture (4.3) | IT (0.3) |
| Pakistan | New York (2.65) | Transportation (4.1) | Arts/Entertainment (0.3) |
| Philippines | Hawaii (14.25) | Agriculture (2.3) | Education (0.4) |
| Poland | Illinois (7.49) | Construction (2.9) | Agriculture (0.2) |
| Portugal | Rhode Island (25) | Manufacturing (2.4) | Arts/Entertainment (0.2) |
| United Kingdom | District of Columbia (2.11) | IT (2.6) | Construction (0.3) |
| Vietnam | California (3.23) | Household Services (6.8) | Agriculture (0.1) |

The score in parentheses after State is the Enclave score computed as in equation 7, but using a stte level of geographic disaggregation. The scores in parentheses after the largest and smallest industries are the Industry scores computed as in equation 8.

# 4 Results

## 4.1 Inequality Between Groups

Table 4 gives the regression inequality decomposition estimates when all immigrant groups are pooled together as well as the same analysis for native-born residents of the United States. By far, the largest determinant of inequality across all immigrants is education. Years of education is responsible for nearly 50% of the explained variation in income. In other words, the main reason why we observe income inequality across immigrants is that some immigrants have more education than others. This suggests that increased access to education could be an effective policy at reducing inequality. Gender also has a significant effect, with over 18% of explained inequality driven by this variable. Thus, one of the reasons why there is inequality is that men tend to earn more than women. Policies targeting increased opportunities for women could be an effective policy response to this result. Marital status can explain just over 10% of inequality, supporting the result that there is a marriage premium that raises wages.

Variables unique to the immigrant experience do not have nearly the same effect on inequality as those that are common to both natives and immigrants. Collectively, the unique immigrant variables (tenure in the US, English language ability, citizenship, enclave score, and industry score) account for only 15.5% of the explained variation in inequality. English language ability can explain 6.6%, years in the US explains 5.5%, and citizenship status explains 3.1%. Interestingly, neither living in an ethnic enclave or working in an industry dominated by a particular ethnic group appears to have any effect on income inequality. This is somewhat surprising given the debate over the effects of ethnic enclaves and the potential for industrial clustering to influence income distributions. However, it may be the case that I need to refine the definitions of enclaves to account for specific enclave characteristics and the industry definitions themselves may be too broad.

Another notable observation is that the listed covariates account for over 95% of the explained variation in income inequality across groups. As such, there is very little left over for the individual country dummy variables. This suggests that the unobservable cultural features associated with the immigrant dummies are not very important in influencing income inequality between immigrants. By far the largest coefficient on an immigrant dummy is for India, but even then the proportional inequality share is only 3%. Thus, even the high social capital and tight-knit communities of Indian migrants do not appear to matter much for income inequality once we control for factors like education and measures of immigrant assimilation.

Finally, comparing the determinants of inequality between immigrants and natives, there are a lot of similarities. As with immigrants, education is the key factor influencing inequality for natives. The effect of gender is slightly smaller, but there are greater effects of both age and marital status. This suggests that there may be a larger experience for natives, perhaps reflecting greater advancement opportunities for natives than immigrants. As for marital status, natives may engage in a greater degree of "marrying up," where people from lower socio-economic backgrounds marry a partner from a higher background. As such, marriage could increase income more for lower earning individuals, creating a larger income gap between unmarried and married people.

**Table 4: Regression Inequality Decomposition Between Immigrant Groups**

| | *Immigrants* | | *Natives* | |
|---|---|---|---|---|
| Variable | Factor Inequality Weight | Proportional Share | Factor Inequality Weight | Proportional Share |
| Yrs Educ | 9.1% | 47.8% | 12.0% | 53.3% |
| Age | 0.6% | 3.3% | 3.3% | 14.6% |
| Male | 3.4% | 18.1% | 3.2% | 14.2% |
| Married | 2.0% | 10.6% | 4.1% | 18.4% |
| Yrs USA | 1.0% | 5.5% | . | . |
| English | 1.3% | 6.6% | . | . |
| Citizenship | 0.6% | 3.1% | . | . |
| Enclave | 0.0% | 0.0% | . | . |
| Industry | 0.1% | 0.3% | . | . |
| *Sum* | *18.1%* | *95.3%* | *22.7%* | *100%* |

See Section 2 for a detailed description of the empirical methodology. Summing up the factor inequality weights yields the $R^2$ for the first stage regression The proportional inequality share are the fraction of the explained variation in inequality that is due to each variable. The sum of the proportional inequality share is 1 (100% of the explained variation in inequality must be attributed to all of the variables in the model.) Not pictured are coefficients on 31 immigrant group dummy variables. Mexico is the excluded (reference) group.

## 4.2 Inequality Within Groups

The preceding results suggest that once we control for observable factors, there are not significant differences in inequality between immigrant groups. Another dimension through which inequality can be examined is differences in inequality within immigrant groups. Holding birthplace constant, how do observable factors like education, tenure in the US, enclave score, etc. affect inequality? Do these within-group determinants of inequality vary across migrant group?

Table 5 gives the regression inequality decomposition estimates by immigrant group. These same numbers are also presented in a more visually intuitive way in Figure 4. The within-group determinants of inequality differ substantially across the immigrant groups in my sample. For example, years of education can only explain 15.6% of inequality among Mexican immigrants as compared to over 65% of inequality for Italian and Jamaican immigrants. Rather, inequality among Mexican immigrants is driven much more strongly by gender, which explains 46% of income inequality. Thus, a policy targeting inequality among Jamaican immigrants should focus more on increasing educational opportunities, while one targeting inequality among Mexican immigrants should focus more on gender inequities.

**Table 5: Regression Inequality Decomposition Within Immigrant Groups**

| Group | Yrs Educ | Age | Male | Married | Yrs USA | English | Citizenship | Enclave | Industry |
|---|---|---|---|---|---|---|---|---|---|
| Arab Countries | 45.7% | 6.9% | 17.8% | 7.9% | 11.9% | 3.2% | 3.4% | 0.1% | 3.5% |
| Brazil | 39.0% | 19.1% | 27.5% | 7.1% | 2.6% | 1.5% | 2.3% | 0.5% | 1.5% |
| Canada | 62.3% | 0.9% | 27.1% | 8.0% | 1.2% | 0.4% | 0.0% | 0.2% | 1.8% |
| China | 59.9% | 3.0% | 5.4% | 13.7% | 5.2% | 14.4% | 2.1% | 0.5% | 1.6% |
| Cuba | 42.6% | 1.9% | 12.6% | 7.3% | 18.2% | 9.5% | 7.8% | 0.0% | 0.2% |
| Dominican Republic | 41.4% | 3.8% | 17.8% | 12.1% | 9.2% | 14.5% | 7.6% | 0.6% | 1.7% |
| El Salvador | 15.0% | 0.3% | 39.1% | 8.5% | 14.7% | 14.1% | 7.9% | 0.0% | 0.3% |
| France | 60.7% | 1.1% | 25.9% | 9.6% | 0.5% | 0.2% | 0.4% | 3.9% | 0.6% |
| Germany | 57.5% | 11.0% | 19.1% | 9.2% | 1.8% | 0.0% | 0.1% | 0.2% | 1.2% |
| Guatemala | 17.6% | 8.2% | 29.7% | 7.2% | 15.0% | 11.4% | 7.0% | 0.0% | 3.9% |
| Guyana/Trinidad | 61.8% | 0.8% | 7.8% | 12.6% | 8.3% | 0.2% | 10.8% | 0.6% | 0.3% |
| Haiti | 42.9% | 7.3% | 3.7% | 11.8% | 12.1% | 6.9% | 10.8% | 3.6% | 0.8% |
| Honduras | 18.0% | 10.3% | 26.8% | 11.0% | 19.6% | 7.7% | 6.1% | 0.0% | 0.5% |
| Hong Kong | 61.7% | 1.1% | 7.5% | 7.7% | 8.4% | 8.4% | 1.6% | 2.3% | 1.3% |
| India | 48.3% | 0.3% | 17.3% | 14.6% | 3.4% | 3.1% | 0.0% | 2.1% | 11.5% |
| Iran | 54.0% | 0.8% | 12.4% | 5.0% | 16.4% | 9.1% | 4.6% | 0.8% | 0.0% |
| Italy | 65.3% | 16.3% | 39.9% | 6.5% | 0.7% | 3.2% | 0.0% | 0.0% | 2.0% |
| Jamaica | 65.2% | 2.0% | 1.6% | 11.2% | 9.4% | 0.1% | 9.9% | 0.0% | 0.6% |
| Japan | 52.4% | 6.8% | 32.1% | 4.8% | 1.3% | 0.9% | 0.0% | 0.0% | 1.6% |
| Korea | 54.8% | 1.0% | 11.7% | 13.7% | 11.2% | 3.8% | 1.0% | 1.6% | 1.2% |
| Mexico | 15.6% | 2.4% | 46% | 8.4% | 7.3% | 8.1% | 9.3% | 1.7% | 1.2% |
| Nicaragua | 34.2% | 2.2% | 20.9% | 8.7% | 9.3% | 17.2% | 9.1% | 0.8% | 1.9% |
| Pakistan | 47.8% | 7.1% | 20.2% | 13.7% | 10.6% | 1.0% | 0.6% | 0.1% | 0.4% |
| Philippines | 54.1% | 0.8% | 3.6% | 10.3% | 12.6% | 1.6% | 6.6% | 0.5% | 12.4% |
| Poland | 46.4% | 4.4% | 35.7% | 13.7% | 0.3% | 3.3% | 4.1% | 0.1% | 0.9% |
| Portugal | 44.4% | 10.0% | 41.2% | 11.0% | 1.1% | 7.7% | 1.6% | 3.6% | 1.6% |
| Russia | 58.8% | 2.8% | 8.5% | 12.3% | 3.5% | 10.0% | 0.0% | 3.9% | 0.3% |
| Taiwan | 49.6% | 0.5% | 13.5% | 5.0% | 16.9% | 6.2% | 4.9% | 2.0% | 1.3% |
| Thailand | 38.3% | 15.9% | 8.9% | 7.1% | 20.3% | 0.8% | 2.6% | 1.6% | 4.5% |
| Ukraine | 54.6% | 7.9% | 13.7% | 14.8% | 5.3% | 16.5% | 2.6% | 0.5% | 0.7% |
| United Kingdom | 55.0% | 0.7% | 29.9% | 7.0% | 1.1% | 0.0% | 0.0% | 2.3% | 5.4% |
| Vietnam | 37.5% | 1.6% | 8.7% | 9.3% | 25.7% | 11.6% | 3.9% | 0.4% | 4.5% |

The percentages are the proportional inequality shares as defined by equation 4 in section 2. These shares represent the fraction of explained income inequality that each variable can be allocated for each individual immigrant group.

Looking across these results, a few other notable estimates stand out. First, an immigrant's tenure in the US matters more for some groups than for others. For Vietnamese immigrants, tenure in the US can explain over a quarter of income inequality, while tenure matters much less for several European migrant groups (France, Germany, Italy, UK). That tenure matters more for groups like Thai and Vietnamese immigrants suggests that it may take longer for these groups to assimilate to the point where high levels of income are attainable. The effects of English language ability are more modest than would have been anticipated, ranging from explaining 17% of inequality for Nicaraguan immigrants to 0% of inequality for British immigrants. Clearly, the more variation in English language ability within a group, the larger effect. As most immigrants attain decent English language ability or live in areas where a lack of this ability is not a major obstacle, the impact of this variable on inequality is diminished.

As with the pooled results in Table 4, living in an ethnic enclave does not appear to have much of an effect on inequality, even within immigrant groups. This would indicate that enclaves are neither creating significant opportunities for inequality nor helping to redistribute income within a group. Of course, the enclave definition may need to be refined to include characteristics of the enclave such as average education. Some more meaningful results are found when looking at industrial clustering. On the high end, industrial clustering can explain 12.4% and 11.5% of income inequality for Filipino and Indian immigrants respectively. Across all locations, the most prevalent industry for Filipinos is healthcare and for Indians, information technology. Both of these industries tend to yield higher incomes at the top, leading to greater within industry inequality. Thus, the observed inequality for these groups can at least be partially attributable to the industries that many of their constituents work in.

Figure 4a: Proportional Inequality Shares for Years of Education and Gender

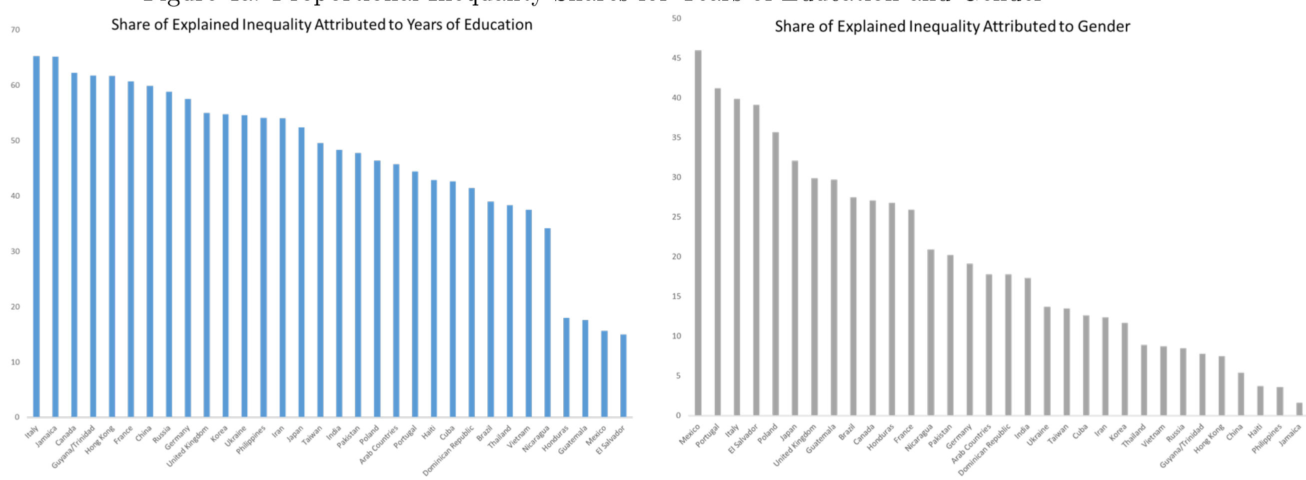


Figure 4b: Proportional Inequality Shares for Years in the USA and English Language Ability

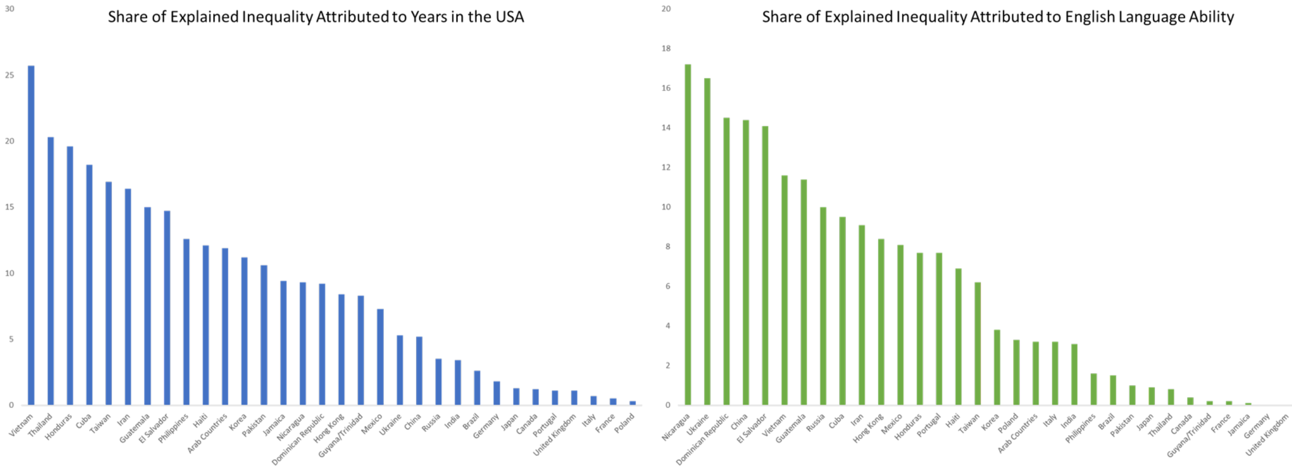
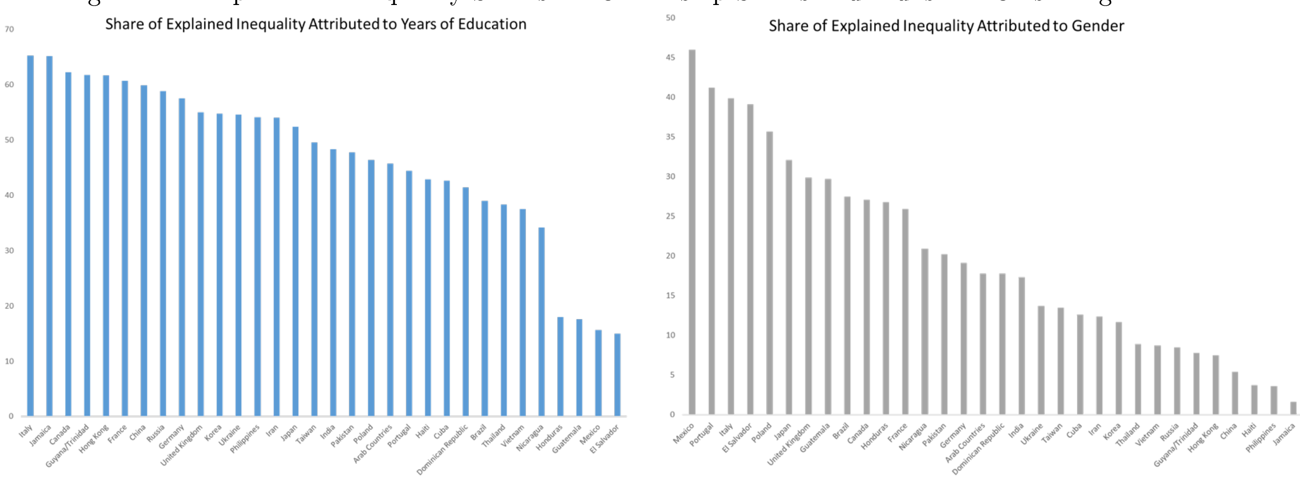



Figure 4c: Proportional Inequality Shares for Citizenship Status and Industrial Clustering




### 4.3 Determinants of Inequality by Group and Location

For each metropolitan area in the ACS sample, I compute the Gini coefficient across total personal income for each of the 32 immigrant groups defined previously. Any location with fewer than 50 observations for a particular immigrant group is dropped from the sample, leaving me with 2,121 immigrant-location observations. Table 6 provides summary statistics for this new sample.

There is considerable variation in Gini coefficients across immigrant-location groups, ranging from a low of 0.26 for Guatemalan immigrants in Chattanooga, TN to a high of 0.71 for Italian immigrants in Richmond, VA. To what extent are these differences in Gini coefficients explained by observable variables and how much is due to unobserved characteristics of immigrant groups of location? Some of these observable characteristics are listed in Table 6. There is considerable variation in average years of education and income across these immigrant communities. Industrial clustering also exhibits a wide range of values, with Indian immigrants in Atlantic City, NJ being less likely than natives to work in a particular industry up to Vietnamese immigrants in Milwaukee, WI being nearly 14 times more likely than natives to work in particular industry. Similarly, enclave strengths vary widely, with the Mexican community in Pittsburgh, PA being very small relative to the city population, while Portuguese immigrant community is 43 times more likely to live in Providence, RI than natives.

**Table 6: Summary Statistics for Immigrant-Location Sample**

| Variable | Mean | St.Dev. | Minimum | Maximum |
|---|---|---|---|---|
| Observations | 677 | 2,753 | 50 | 87,238 |
| Gini | 0.490 | 0.068 | 0.257 | 0.707 |
| YrsEd | 13.1 | 2.2 | 5.3 | 17.7 |
| Income | $41,757 | $16,228 | $11,196 | $152,808 |
| Industry | 1.6 | 1.0 | 0.8 | 13.8 |
| Enclave | 1.5 | 2.2 | 0.0 | 43.2 |
| YrsUSA | 24.5 | 8.8 | 5.6 | 53.4 |
| Male | 0.52 | 0.11 | 0.19 | 0.97 |
| English | 0.82 | 0.18 | 0.24 | 1.00 |
| Manager | 0.08 | 0.05 | 0.00 | 0.34 |

Summary statistics across immigrant-location groups. After dropping locations with fewer than 50 immigrants, I ended up with 2,121 immigrant-location pairs. Location is defined at the metropolitan area. Gini coefficient is computed from total personal income. In addition to the variables previously defined, Manager refers to the fraction immigrants in a particular location that hold managerial positions.

The age of an immigrant community varies widely, with newer communities like the 126 Chinese immigrants in Lansing, MI who have been in the US an average of 6.4 years as compared to 404 Italian immigrants in Pittsburgh, PA who have been in the US an average of 52 years. Some communities have gender imbalances and others have better English language skills than others. Finally, I have added a new variable "Manager" that measures the fraction of immigrants in a location with managerial responsibilities at work. This variable is intended to capture the effects of immigrant influence on hiring and pay in a particular location. For example, almost none of the 507 Mexican immigrants living in Fort Myers, FL hold managerial positions, while about 25% of British immigrants in San Jose, CA hold these positions. The expectation is that communities with more managerial positions will have less inequality as discrimination should be lower in these areas.

In order to more fully assess the determinants of inequality between these communities, I estimate the following regression across immigrant group $i$ in location $j$. To control for unobservable group and location characteristics, I also include dummy variables for each immigrant group and each metropolitan area in the sample.

$$\begin{aligned} Gini_{i,j} &= \beta_1 YrsEd_{i,j} + \beta_2 LogY_{i,j} + \beta_3 Industry_{i,j} + \beta_4 Enclave_{i,j} + \beta_5 YrsUS_{i,j} \\ &\quad + \beta_6 Manager_{i,j} + \beta_7 Male_{i,j} + \beta_8 English_{i,j} + \sum_i \gamma_i Grp_i + \sum_j \delta_j Location_j + u_{i,j} \end{aligned} \tag{9}$$

The result from this regression are given in Table 7:

The strongest determinant of inequality is the average income in an immigrant community, with a one standard deviation increase in income leading to a 0.66 standard deviation increase in the Gini coefficient. This suggests that communities with the highest average income tend o also have the most unequal distribution of income. Holding income constant, better educated communities have less inequality, suggesting a potential path to reducing inequality. Industrial clustering appears to reduce inequality, though the effect is much smaller than that of income and education. Similarly, communities with stronger enclave strengths also have less inequality. Together, these results suggest that communities that have a larger presence in a particular location exhibit less inequality, perhaps due to greater immigrant influence in that location. This is supported by the estimate on Manager, with communities with a higher share of managers exhibiting less inequality. Older communities also

**Table 7: Income Inequality Determinants across Immigrant-Location Groups**

| Variable | OLS Estimate | Beta Coefficient |
|---|---|---|
| Years Education | -0.008 | -0.250 |
| | [0.000] | |
| Log Income | 0.118 | 0.661 |
| | [0.000] | |
| Industry | -0.006 | -0.091 |
| | [0.000] | |
| Enclave | -0.001 | -0.035 |
| | [0.049] | |
| Manager | -0.119 | -0.084 |
| | [0.000] | |
| Years USA | -0.001 | -0.090 |
| | [0.024] | |
| Male | -0.049 | -0.076 |
| | [0.006] | |
| English | -0.007 | -0.017 |
| | [0.747] | |
| Obs | 2,121 | |
| $R^2$ | 0.642 | |

OLS estimates of equation 9. P-values given in brackets. The beta coefficients refer to OLS estimates after standardizing all variables.

tend to observe less inequality, suggesting that immigrants assimilate, income gaps shrink. Gender also plays a role, with more male dominated communities exhibiting less inequality. This suggests that gender imbalances in pay are driving some of the observed income inequality. Finally, English language ability does not appear to influence inequality once we control for group and location effects.

# 5 Conclusion

The goal of this paper is to better understand the determinants of income inequality and to see how these determinants vary across immigrant groups. The results suggest that overall, access to education is the key determinant of income inequality in the United States. This holds whether we are looking at immigrants or at natives. As such, policies that increase educational access should be effective at reducing income inequality. Other variables that contribute to inequality include gender, marital status, tenure in the US, and English language ability. Thus, policies targeting unequal opportunities for women and barriers to marriage could be effective. As to the immigrant experience, policies that hasten and ease the assimilation process could help to reduce inequality. Interestingly, the observed differences in income inequality between immigrant groups disappear after we control for the observable factors described above.

Looking within immigrant groups, we see that the key determinants of inequality will depend on what immigrant group we are looking at. For some immigrant groups, education is still the main driver of inequality. For others inequality is primarily determined by gender, while for others time spent in the US is the most important. These differences suggest a nuanced approach to policies targeting inequality within these groups. Increasing educational opportunities will be more effective for some groups than for others. The focus for other groups should instead by things like gender equality and accelerating the process of assimilation and the spread of social capital.

Controlling for unobservable characteristics of immigrant communities (by group and by location), income inequality is affected most by income and education. High income communities tend to be more unequal, again supporting the notion that inequality is driven by variation at the top of the distribution. Education reduces inequality, again suggesting that increasing access to education could be an effective way to reduce inequality. Stronger immigrant communities (as measured by size relative to location, presence in a particular industry, representation at the managerial level, and age of the community) tend to exhibit less inequality. This suggests that some of the observed inequality among immigrant groups is driven by problems associated with assimilation such as discrimination.

# 6 References


Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8(4), 436-455

Borjas, G. (2000). The economic progress of immigrants. *Issues in the Economics of Immigration*, NBER.

Borjas, G. (2003). The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market. *Quarterly Journal of Economics*, 118(4), pp. 1335-1374

Bratsberg, B. Ragan, J.F., and Nasir, Z. (2002). The effect of naturalization on wage growth: a panel study of young male immigrants

Card, D. (2009). Immigration and inequality. *American Economic Review: Papers and Proceedings*, 99(2), 1-21

Chiswick, B., Miller, P. (2001). A Model of Destination-Language Acquisition: Application to Male Immigrants in Canada. *Demography, 38*, 391-409.

Edin, P-A., Fredriksson, P., Åslund, O. (2003). "Ethnic Enclaves and the Economic Success of Immigrants—Evidence from a Natural Experiment." *Quarterly Journal of Economics, 118*(1), 329-357.

Fei, J., Ranis, G. and Kuo, S. (1978) Growth and the family distribution of income by factor components. *Quarterly Journal of Economics*, 92(1), pp. 17–53.

Fields, G.S. (2003) Accounting for income inequality and its change: a new method with application to the distribution of earnings in the United States, in: S.W. Polachek (ed.) *Worker Well-being and Public Policy* (Oxford: Elsevier Science Ltd.), pp. 1–38.

Foad, H. (2016) Ethnic Enclaves: Help or Hindrance? Working paper available at www.foad.sdsu.edu

Gang, I.N., Zimmermann, K.F. (2000). "Is Child like Parent? Educational Attainment and Ethnic Origin." *Journal of Human Resources, 35*(3), 550-69.

Grönqvist, H. (2006). "Ethnic Enclaves and the Attainments of Immigrant Children." *European Sociological Review*, 22(4), 369-382

Heltberg, R. (2003) *Spatial Inequality in Vietnam: a Regression-based Decomposition* (Copenhagen: Institute of Economics, University of Copenhagen).

Hershbein, B., Kearney, M., and Summers, L. (2015). Increasing education: What it will and will not do for earnings and earnings inequality. *Brookings Up Front*. Accessible at www.brookings.edu/blog/up-front

Kaplan, S. and Rauh, J. (2013). It's the Market: The Broad-Based Rise in the Return to Top Talent. *Journal of Economic Perspectives*, 27(3), pp.35-56

Oaxaca, R.L. (1973) Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3), pp. 693–709.

Ottaviano, G. and Peri, G. (2006) Rethinking The Effect Of Immigration On Wages. *Journal of the European Economic Association*, 10(1), pp 152-197

Naschold, F. (2009) Microeconomic determinants of income inequality in rural Pakistan. *The Journal of Development Studies*, 45(5), pp. 746-768

Piketty, T. and Saez, E. (2003) Income inequality in the United States: 1913-1998. *Quarterly Journal of Economics*, 118(1), pp. 1-39

Pyatt, G., Chen, C.-N. and Fei, J. (1980) The distribution of income by factor components. *Quarterly Journal of Economics*, 95(3), pp. 451–473.

Ravallion, M. and Chen, S. (1999) When economic reform is faster than statistical reform: Measuring and explaining income inequality in rural China. *Oxford Bulletin of Economics and Statistics*, 61(1), pp. 33– 56.

Sanders, J.M. and Nee, V. (1987). "Limits of Ethnic Solidarity in the Enclave Economy." *American Sociological Review*, 52(6), 745-73

Shorrocks, T. (1982) Inequality decomposition by factor component. *Econometrica*, 50(1), pp. 193–211.

Shorrocks, T. (1984) Inequality decomposition by population subgroup. *Econometrica*, 52(6), pp. 1369– 1385.

Toussaint-Comeau, M. (2005). Do Enclaves Matter in Immigrants' Self-Employment Decision? Federal Reserve Bank of Chicago Working Paper, No. 2005-23

US Census Bureau (2017). American Community Survey. Accessed at www.ipums.org

Waldfogel, J. (2003). "Preference Externalities: An empirical study of who benefits whom in differentiated-product markets." *Rand Journal of Economics*, 34(3), 557-568

Wilkinson, R. and Pickett, K. (2009). *The Spirit Level: Why More Equal Societies Almost Always Do Better.* London: Allen Lane

Zhou, M. and Logan, J.R. (1989). "Returns on Human Capital in Ethnic Enclaves: New York City's Chinatown." *American Sociological Review*, 54(5), 809-20