



University of  
Massachusetts  
Amherst

## Complexity, Diversity and Integration: Evidence from Recent US Immigration

Item Type	Working Paper
Authors	Wiener, Noé
DOI	<a href="https://doi.org/10.7275/k9yt-bk31">10.7275/k9yt-bk31</a>
Rights	UMass Amherst Open Access Policy
Download date	2026-03-14 05:52:45
Item License	<a href="http://creativecommons.org/licenses/by-nc-nd/3.0/">http://creativecommons.org/licenses/by-nc-nd/3.0/</a>
Link to Item	<a href="https://hdl.handle.net/20.500.14394/22314">https://hdl.handle.net/20.500.14394/22314</a>

# Complexity, Diversity and Integration: Evidence from Recent US Immigration

Noé M. Wiener

Department of Economics, University of Massachusetts Amherst,  
nwiener@umass.edu, ORCID-ID: 0000-0002-4362-0918

March 21, 2022

## Abstract

This article proposes alternative measures of immigrant integration founded in information theory. By considering differences in the heterogeneity of outcomes between immigrants and natives, the proposed measures provide robust and non-parametric estimates of the extent to which cohorts remain defined by their national origin. Integration is furthermore premised on equality in the association between economic characteristics and incomes, so that other factors can begin to shape outcomes for immigrants and natives alike. Results for successive immigrant cohorts in the post-war era are presented using Census income data for the United States. The speed by which the mark of migration on incomes is fading for different cohorts appears to decline significantly over the analyzed period. The paper also presents evidence on the disadvantageous relationship between immigrant status, education and incomes, as well as the integration of immigrants into the US racial and gender hierarchy. Integration appears to be increasingly reserved for college-educated and white men, whereas incomes for Hispanic and less educated immigrants remain strongly shaped by immigrant status.

**Keywords:** Immigration, integration, complexity, diversity, information theory

# 1 Introduction

Patterns of international migration have become increasingly variegated and diverse, not only in terms of geographic flows, but also the socioeconomic trajectories that migrants experience along the way (e.g. Portes, 1997; Crul, 2016). Assessing the prospects of immigrants for integration into the societies and economies of their destination has taken on a new urgency, driven by the prospect of ongoing mass migration and renewed efforts to manage such movements through deliberate policy. Quantitative empirical research on this question however has been hampered by a set of analytical tools that do not adequately capture the underlying notion of integration as a political- and socio-economic process (Borjas, 1991; Borjas, 1995; Chiswick, 1999; Grogger and Hanson, 2011). This paper presents a complementary framework for measuring migrants' socioeconomic integration that is more appropriate for a complex and diverse society.

A straight-line view of integration-as-assimilation has become increasingly unhelpful in assessing outcomes for immigrants. Yet the traditional modeling approaches are largely still tied to such a view, as they are attempting to explain trajectories in terms of individual characteristics and the time since immigration. The approach to integration explored in this paper instead foregrounds systemic, structural limits to immigrant outcomes. It matches the intuitive understanding of integration as a transition from outsider to insider status, from clearly separate to nearly indistinguishable. As long as there is a difference in the *range* of social and economic opportunities available to immigrants and natives, even though both may be expanding, we find migration leaving its mark on a population.

Integration will be defined here as a process whereby the immigrant status of a cohort loses its social significance with time in the destination country. As will be shown below, the observable implication of integration from this point of view is that migration status ceases to differentially homogenize outcomes for immigrants. This approach shifts the focus away from attempting to explain individual outcomes and towards the convergence in the outcome *distributions* of immigrants and natives. Integration in this sense is not simply the equalization of expected outcomes for immigrants and natives of a certain cohort, but the achievement of parity in outcomes throughout the distribution. Even though this paper presents results for integration in terms of incomes, the tools presented here are applicable to the analysis of a variety of social outcomes, including in the residential and occupational dimensions.<sup>1</sup>

The first part of this paper outlines the theoretical perspective of immigrant integration as social process. The section discusses the consequences of complexity for analyzing social systems, and the central notion of heterogeneity or entropy of outcomes. On this basis, a series of information-theoretical

---

<sup>1</sup>Incomes represent the condensate of a history of social interactions, which will be captured in reduced-form in the measures presented below. It is important to stress that incomes can be viewed as both “markers” and “means” of integration, among others (see e.g. Ager and Strang, 2008).

measures of integration are proposed that link the differences in heterogeneities between populations to the extent to which their national origin constrains their outcomes. The second part of the paper applies these measures to US census data on immigrant and native cohorts between 1970 and 2010. The section also provides evidence on the interactions among immigration status and various other characteristics in their influence on incomes, including educational attainment, race, ethnicity and gender. These measures are not tied to any particular theoretical presupposition beyond the choice of variables, and allow uncovering patterns for future research. The final section draws conclusions from these conceptual and empirical results for the broader literature on immigrant integration processes.

## 2 Visions of Integration and Issues of Measurement

Different visions of international migration have informed research on integration processes. These visions - some focusing on individual attributes, others on socio-political processes - provide different starting points for empirical analysis. I will argue that existing quantitative approaches to the measurement of integration, such as the ubiquitous linear regression framework, are strongly tied to an individualistic view.

The standard empirical framework for the measurement of integration focuses on the different characteristics that immigrants bring with them from their home country or acquire at their destination. Empirical studies using linear regressions have often followed the micro-econometric approach of modeling relative mean incomes conditional on various individual characteristics (e.g. Chiswick, 1978; Borjas, 1989; Lalonde and Topel, 1997; Bleakley and Chin, 2004; Borjas, 2015). From this point of view, observed integration must necessarily reflect the acquisition of location-specific human capital yielding returns equivalent to those of the native-born. Underlying the use of these empirical tools are rather strong assumptions regarding the process of income generation and allocation.

Although this framework readily admits that components of the endowment may be unobservable, and are subject to change with time spent at the destination, they remain attached to the individual immigrants. Rather than being shaped by social forces, the success or failure of integration becomes the reflection of largely individual efforts on the part of immigrants. In the following sections, I outline three core aspects of the immigrant experience that cannot adequately be captured using the standard regression framework.

### 2.1 Heterogeneity and Diversity

The recent history of immigration has led to renewed criticism of standard narratives of straight-line assimilation into a mainstream society. In the European

context, the increasing diversification of immigrant populations, and the emergence of majority-minority cities and towns, have led to descriptors such as “super-diversity” (Vertovec, 2007). While the explanatory benefit of the term is not beyond doubts (Grzymala-Kazłowska and Phillimore, 2018), it presents a useful reference for questioning the nature of the “mainstream” into which immigrants supposedly integrate.

Not only are the societies into which immigrants integrate ethnically and culturally increasingly diverse, they are also criss-crossed by divisions of class, gender, occupational status and a multitude of other elements of stratification. These divisions exist not only between, but also within ethnic groups, such that both natives and immigrants experience a variety of socioeconomic realities over time (Crul, 2016). At a general level, such divisions are to be expected in any system in which incomes are determined through a large number of uncoordinated interactions (Foley, 1994), and in which social identity categories play a major role in allocating individuals to differentiated social roles (e.g. Chris Tilly and Charles Tilly, 1998).

In the standard, neoclassically inspired framework, human capital “endowments” constitute the primitives of the explanation for divergent outcomes post-migration. Any residual variation in outcomes after accounting for observable differences in endowments are due to errors in measurement, or other other unobserved factors. As an epistemological stance, this separation of analyzed variation and residual “noise” does not seem adequate to understanding heterogeneity in diverse societies. A different perspective would set out from the *expectation* of heterogeneity, and seek to understand the degree of *organization* of outcomes, as well as differences in the extent of organization between populations.

## 2.2 Socio-political Complexity

The second important feature of integration that empirical measures ought to capture is its fundamentally social and political character. Rather than a purely individual project of human capital acquisition, integration can be understood as part of a larger, contested project of nation-building (Waldinger, 2009; Zolberg, 2009). From this point of view, the category of the “immigrant” is shaped and sustained by the system of laws, policies and practices that grant rights and impose restrictions on the economic opportunities of migrant populations.<sup>2</sup> Differences are also generated by constituencies in the destination country who seek to differentially exclude migrants from the full rights of citizenship (Bonacich, 1972), employ discriminatory hiring practices (Kenney and Wissoker, 1994), and so on.

The mechanisms involved in shaping socioeconomic outcomes are highly complex, largely unobservable and involve non-linearities and interactions of unknown form. Successful integration may not depend solely or primarily on

---

<sup>2</sup>A very explicit example of such systemic influences are visa categories that regulate legal durations of stay or access to the labor market (Anderson, 2010).

immigrants' individual characteristics and their effort in bringing them to bear in the host labor market, but instead on social forces operating at the meso- and macroscopic level. In economic terms, there are significant interdependencies beyond the interaction in perfectly competitive markets, both in the short-range correlation of decision-making (Radu, 2008), and in longer-range interactions due to indirect competition between workers in different labor market segments (dos Santos, 2017). Integration may be a gradual process of dissipation in all directions as barriers of various kinds are eroding over time, or it may happen in fits and spurts due to discrete changes in the political and legal environment.

Given this complexity, it is doubtful that the process of integration could be adequately specified using regression models that relate individual characteristics. Many of the factors involved are inherently unobservable, and certainly are not captured in the social surveys commonly used to track outcomes for immigrant populations. Furthermore, there are many ways that social categorization affects outcomes for different groups, beyond differences in group averages. Identifying integration with a conditional time-since-migration effect, as is attempted in the literature, requires the maintained hypothesis that the functional relationship relating years-since-migration and socioeconomic outcomes is correctly specified. Otherwise the resulting estimates are not valid measures of integration.

What is needed, then, is a measure of integration that can flexibly account for the extent to which outcomes are shaped by the categories of “immigrant” and “native”, irrespective of the mechanisms involved. If immigrant status ceases to be informative of outcomes, immigrants are no longer treated in a manner distinguishable from natives by the collection of social processes involved in allocating outcomes.

### 2.3 Intersectionality

A third element that a measure of integration should incorporate is the ability to account for systematic differences in the diversity of trajectories that immigrants experience relative to natives. Beyond the separate roles that immigrant status, gender, race and ethnicity have in shaping outcomes, they also *interact* with each other to generate complex multitudes of relative advantage or disadvantage. The notion of “intersectionality” has also found its way into immigration scholarship (Crul, 2016). Yet quantitative intersectional work has continued to rely on the linear regression framework (e.g. Stewart and Dixon, 2010; Nawyn and Gjakaj, 2014). This means that such analysis continues to be susceptible to issues of misspecification, as well as a narrow view of the shape that intersectional relationships might take. A measure of integration ought to be able to flexibly account for the interaction among social identity characteristics in their influence on outcomes, without imposing strong extraneous assumptions about the particular causal channels involved.

### 3 Information-theoretic Measures of Integration

In information theory, we may find a set of tools that satisfies the three requirements outlined above, namely being appropriate for a diverse and heterogeneous society, accounting for complex and interdependent mechanisms of causation, and incorporating intersectional identities. The argument of this paper is that we may conceive of integration as a process whereby nativity loses its ability to differentially homogenize outcomes for a group of immigrants. I will adapt the informational association measures developed in dos Santos and Wiener (2019) and applied to the analysis of discrimination by social identity characteristics in dos Santos and Wiener (2020). The following discussion is fairly non-technical, and interested readers are referred to these references for mathematical details. The central notion in this argument is the entropy of a distribution, to which I turn next.

#### 3.1 Heterogeneity as Default

As argued above, disorder and heterogeneity should be the default expectation for the trajectories of both immigrants and natives over time. Rather than attempting to predict individual trajectories, we shift attention to the *distribution* of individuals over positions in the space of relevant social outcomes. The scientific process then involves increasingly detailed descriptions of such distributions in different contexts to uncover regularities and laws that act on the system as a whole.<sup>3</sup> Here I present one way of characterizing these distributions using entropic measures.

Let us consider the scenario of placing  $N$  individuals into  $s$  positions in the space of social outcomes. In our application below, these positions will be income brackets, but multi-dimensional outcomes could be envisaged. From the point-of-view of the system as a whole, the relevant result of allocating individuals in this manner is the relative occupancy of each outcome, as represented by a frequency distribution ( $f_k$ ). The precise identity of the individuals occupying various social states, beyond a set of social identity and economic characteristics to which we will return, is not relevant to the functioning of the system.

Clearly, there are many ways of rearranging individuals between positions in the outcome space that still result in the same frequency distribution. The number of permutations at the microscopic level that lead to the same macroscopic distribution is the multinomial coefficient  $W = \binom{N}{n_1, n_2, \dots, n_s} = \frac{N!}{n_1! \dots n_s!}$ , where  $n_k$  is the occupancy of the  $k$ -th position. This number, also referred to as the “multiplicity”, will vary for different outcome distributions. For a given number of outcomes  $s$ , the uniform (most heterogeneous) distribution has the highest multiplicity, whereas there is exactly one way to group all individuals into a single (homogeneous) social outcome category. It turns out that a measure from

---

<sup>3</sup>For foundational contributions to information-theoretic analysis, see Jaynes (2003) and Golan (2017) among others.

communication theory, Shannon’s entropy, is a very good approximation (for sufficiently large  $N$ ) of the logarithm of the multiplicity  $W$ :

$$H(f_k) = - \sum_i \frac{n_i}{N} \log_2 \left( \frac{n_i}{N} \right) \quad (1)$$

In informational terms, the entropy gives the expected number of bits of information that are gained by learning the outcome of any particular individual. Clearly, the expected information gained (or conversely, uncertainty removed) will be greater if there are many possible permutations of individuals that could have produced a particular outcome distribution. Thus, the multiplicity and the uncertainty associated with a distribution are intimately linked.

There are therefore at least two ways in which a measure of integration based on entropy incorporates heterogeneity as the default. First, more heterogeneous outcome distributions are combinatorially vastly more likely, given any restrictions that social institutions and mechanisms impose. Everything else equal, we should be less surprised to find relatively flat or uniform distributions of social outcomes than relatively peaked or otherwise strongly patterned ones. Second, more heterogeneous outcome distributions encode less information. A plausible heuristic in scientific inference is to favor the most non-committal, and hence maximally uncertain description that is still consistent with any prior knowledge we might have (Jaynes, 2003).

So far, I have proposed a measure of heterogeneity of outcomes, which also represents the degree of uncertainty we have about the outcome of a particular individual. I have also suggested the sense in which greater degrees of heterogeneity are less surprising, while less heterogeneity requires perhaps more urgent explanation. This idea comes into its own when we consider *differences* in heterogeneity between groups, as I will discuss next.

### 3.2 A Model-Free Measure of Integration

Over the course of their lives, people will change labor market status, occupation and income levels following changing social roles, career trajectories and a variety of other reasons. We expect the heterogeneity of outcomes to similarly evolve over time. However, time in the destination country has a different effect on immigrants which in turn ought to be reflected in a measure of integration. Here I propose an index of the influence of immigrant status on the degree of heterogeneity of outcomes.

The segregation of immigrants in particular geographic locations and the concomitant access or lack thereof to employment opportunities is well documented, (see e.g. Massey and Denton, 1985; Liu, 2009), as is the sorting of immigrants into particular industries or firms on the basis of social networks (Waldinger and Lichter, 2003; Andersson et al., 2014). From an observer’s perspective, such processes of segregation or sorting all serve to organize the diversity of possible trajectories of immigrant outcomes into more narrow and therefore predictable tracks. The empirical strategy proposed here focuses on

the manifestation of such differences in the degree of organization of social and economic outcomes.

As an illustration of these differences in the extent of income heterogeneity or organization, and their evolution over time, consider the group of immigrants that entered the US between 1960-69. Over time, their income distribution comes to resemble that of similarly aged native-born workers (Figure 1). This suggests that there was a greater degree of heterogeneity gained *relative* to the native population, allowing the distribution of incomes to approach one another. It appears, at least for the 1960s arrival cohort, as if most of the change occurred for men while the female wage distributions remained remarkably stable over time. We will return to these differences by social identity below.

[Figure 1 about here.]

Our interest is in the *differential* evolution of income heterogeneity between native and foreign-born workers of the same cohort. As discussed above, differences of heterogeneity or entropy are closely linked to the amount of information conveyed by the particular degree of freedom in question. Let  $Y$  be the social outcome measure under study, e.g. income, and let  $X$  be the immigration status. The entropy of  $Y$  conditional on  $X$  is the remaining uncertainty about  $Y$  once  $X$  is known:

$$H(Y|X) = - \sum_x f(x) \sum_y f(y|x) \log f(y|x) \quad (2)$$

The Mutual Information between  $X$  and  $Y$  is the uncertainty removed about  $Y$  once  $X$  is known, in other words the expected information conveyed about a social outcome by observation of immigration status.

$$I(X;Y) = H(Y) - H(Y|X) \quad (3)$$

As it turns out, the mutual information (3) is symmetrical. In order to produce an *account* of the uncertainty in social outcomes in terms of immigration status, we divide (3) by the original entropy of  $Y$  to obtain a normalized informational association coefficient  $A(Y|X)$ . This measure is bounded from below by 0, indicating that the residual uncertainty in  $Y$  after observing  $X$  is unchanged from the original uncertainty. In other words, immigration status conveys no information about the social outcome of interest. On the other hand, a value of 1 indicates that knowledge of immigration status is equivalent to knowledge of the social outcome.

Let us assume that the immigration status variable is binary, distinguishing foreign- and native-born. In this case, the integration measure  $A(Y|X)$  is the average of the informational association of foreign-born status and income,  $a(Y|foreign)$ , and the association of native-born status and income,  $a(Y|native)$ , each weighted by the respective population share. Note that while knowledge of immigration status overall can only reduce uncertainty about incomes (it is bounded by 0 from below), the measures for particular sub-populations may be negative. While positive values indicate that knowledge

of the particular sub-population reduces our uncertainty (conveys information) about their social position, negative values means it increases our uncertainty.

We can now draw on the connection between uncertainty and heterogeneity. If a particular immigration status is informative of social outcomes, it must reduce the heterogeneity of outcomes relative to the population as a whole. There is less room for factors other than this particular immigration status to shape social outcomes. On the other hand, an immigration status that increases our uncertainty must increase heterogeneity, relative to the population as a whole. Factors other than immigration status can play a bigger role in shaping outcomes.

Note that the reference category for these informational integration measures is the population as a whole (where immigration status has been “marginalized out”). These measures therefore help clarify the question of the mainstream into which integration occurs. In most contexts, the label “native-born” will not carry much information given the predominance of natives in the population as a whole. However, in majority-minority settings there might be interesting reversals as some immigrant groups come to dominate the reference population. In these circumstances, being native-born might grant substantial increases in heterogeneity, while immigrant status is nearly uninformative.

In this section, I argued that the information-theoretical approach conceives of integration as the loss of mutual information between immigration status and social outcomes. This loss of information goes together with relatively greater heterogeneity, and thus the possibility for other factors to shape outcomes. Note that this measure is entirely independent of any particular causal mechanisms through which such influence occurs. Unlike the standard approach of micro-econometric regression approach, the associations between immigration status and outcomes are not conceived of as being located at the level of individuals. This is of particular relevance as we turn to intersectional associations with other markers of social identity.

### **3.3 Interaction Information and the Intersectional Mark of Migration**

Immigration status interacts in complex ways with other elements of stratification. Intersections with economic and social identity characteristics, such as education, gender and race, raise at least two sets of questions in investigations into the extent of immigrant integration. First, do these characteristics constitute channels through which immigration status shapes social outcomes? From an informational perspective, this would be equivalent to asking whether immigrant status is no longer informative of social outcomes once the particular characteristic is known. Second, is the effect of immigrant status uniform in all sectors and among all groups, or are there differences? This question concerns the informational association of immigration status and outcomes within groups, relative to the population as a whole. As it turns out, the question is equivalent to asking whether the effect of a particular characteristic is different for immigrant and native groups.

We consider below the informational association between the joint occurrence of the (possibly multidimensional) set of characteristics  $Z$  and immigration status  $X$ , and social outcome  $Y$ . Let upper-case letters refer to variables, and lower-case letters to specific values of those variables. As shown in dos Santos and Wiener (2019), two related measures of informational association can be derived that provide answers to the questions raised above. The incremental informational association is the additional information conveyed by observation of  $x$  once  $z$  is known, normalized by the total uncertainty in  $Y$ :

$$\mathbb{I}_{x|z} = \frac{H(Y|z) - H(Y|(z, x))}{H(Y)} \quad (4)$$

Thus,  $\mathbb{I}_{x|z}$  measures the amount of information conveyed by immigration status that is independent of any influence exerted by characteristic  $z$ .

The second measure accounts for any information contained in the *interaction* among the variables  $(X, Y)$  in their influence on  $Y$ . This measure  $m(Y, x, z)$  indicates the extent to which the mutual information between immigration status and  $Y$  is different from that same mutual information among members of the group defined by  $z$ , and vice-versa:

$$m(Y, x, z) = a(Y|x) - \mathbb{I}_{x|z} = a(Y|z) - \mathbb{I}_{z|x} \quad (5)$$

Negative values of  $m(Y, x, z)$  indicate that there is information contained in the joint occurrence of  $(x, z)$  that is not contained in either set of variables by themselves. Immigration status  $x$  is more informative of outcomes for groups defined by  $z$ , and  $z$  is more informative of outcomes for groups defined by immigration status  $x$ , than for the population as a whole. Positive values on the other hand indicate that there is some redundant information in the joint observation of the two sets of variables about the outcome  $Y$ . Among all those with characteristic  $z$ , outcomes are more heterogeneous for those with immigration status  $x$  compared to the overall population defined by  $z$ .

The indices of incremental informational association, and the interaction information, provide additional non-parametric, model-free measurement tools for assessing causal channels and intersectional differences in the integration trajectories of immigrant cohorts.

## 4 Data and Research Design

The sample size of currently available longitudinal datasets is generally not sufficient to analyze outcomes for sub-populations such as immigrants in any detail. This study therefore uses repeated cross-sectional wage and salary data from the US census and the American Community Survey (ACS) to construct cohorts for the period between 1970 and 2010<sup>4</sup>. I restrict attention to immigrants arriving between 20-38 years of age. While we cannot identify individuals

---

<sup>4</sup>Details on sample selection and variable construction can be found in the online appendix.

across census waves, we can follow groups born around the same time. My analysis is therefore located at the level of community groups and cohorts, rather than the individual worker. The immigrant cohorts are then compared to native workers born in the same period. These comparisons allow the measurement of integration over time.

## 5 Immigrant Integration in the US from 1970-2010

The recent history of US immigration is one of significant changes, both in terms of the number and composition of the arriving population and their experience in US society. A notable feature of recent immigration is the apparent diversification of regions of origin, from a predominantly European to a Central American (primarily Mexican) and Asian immigrant population. This increased diversity of (non-European) origins has been documented across other countries with a settler-colonial history (Czaika and de Haas, 2014). The increasing numbers and diversity of the immigrant population in the US might lead us to expect that integration trajectories are becoming more heterogeneous, and perhaps less distinguishable from the experience of natives. On the whole, this is not what we find. Figure 2 shows how the mutual information measure of integration,  $a(Y|immig)$ , evolved for different cohorts of immigrants arriving in the US during the second half of the 20th century.

[Figure 2 about here.]

Figure 2 suggest that successive cohorts tend to start from a lower level of integration (represented by larger values in the figure), with stark differences between the genders. If we distinguish between 25 different income levels, we find that immigrant status removes around 2.5-5% of our uncertainty about incomes of recent immigrant cohorts in their first census after immigration. For earlier cohorts, knowledge of immigrant status could actually increase our uncertainty by 1-3% after 30-40 years since migration. The mutual information measure indicates integrative convergence for women in the 1950-1980 arrival cohorts, but some divergence for the 1990-2000 cohort. While initially seeing wages more tightly organized, wages of female immigrants in the 1970s and 1980s cohorts eventually become more heterogeneous than those of their native female counterparts. For men, the measure indicates integration for the 1950-1960 cohorts, but clear signs of divergence for the 1970-2000 cohorts.

So far, the informational association measures have documented how immigrant arrivals start out with successively lower measures of integration. Over time, immigrant women tend to integrate *towards native women* in most cases, while there is very little convergence after the 1960s for immigrant men.

## 5.1 Educational Integration and Downgrading

The distribution of educational achievements has shifted for natives relative to that of immigrants (Figure 3). In the 1950-1970 arrival cohorts, we find some educational upgrading for both natives and immigrants over time, with a diminishing share of high school dropouts and a non-negligible increase in the fraction of college graduates over time. The much more important change however is at the generational level, particularly with those born in 1940 and after.<sup>5</sup> Considering the more recent arrival cohorts, we find rates of college achievement for immigrants on par or even slightly greater than for the native population. On the other hand, a substantial fraction of immigrants up to the most recent cohorts have not graduated high school, while for natives this is an increasingly small share. Thus, the heterogeneity in educational achievement has decreased for the native population but remained elevated for immigrants.

[Figure 3 about here.]

While the distribution of immigrant and native educational achievements is interesting by itself, the question also arises whether immigrants and natives have the same ability to translate their education into incomes. Dos Santos and Wiener (2020) have documented how social identity significantly mediates the link between economic characteristics, such as education and experience, and income. Not only are educational systems and labor markets operating in ways that limit the ability of subaltern groups to acquire higher levels of education and experience. But the extent to which economic characteristics can influence incomes also depends on the social group under study. For example, dos Santos and Wiener (2020) show that low levels of educational attainment are *more* informative of incomes for women of all race and ethnic groups, compared to the population overall, while a college degree is more informative of incomes for white men (and recently, white women).

The translation of educational attainment into income possibilities is likely more difficult for immigrant populations. Dustmann, Schönberg, and Stuhler (2016) review evidence from multiple countries on the income and occupational “downgrading” of recent immigrants compared to natives with equivalent levels of education and experience. One hypothesis put forward in the literature to explain this pattern concerns the likely difficulty employers face in assessing skills and qualifications obtained in other countries, which would lead them to offer only jobs at the low end of the skill distribution to the first wave of immigrants from a particular region (Stark, 1993). A second possibility, for which some experimental evidence exists (Shinnaoui and Narchal, 2010; Oreopoulos, 2011), is that immigrant skills are actively discounted in a way suggestive of discrimination.

In order to investigate the link between immigrant qualifications and incomes, we apply the interaction information index presented in subsection 3.3

---

<sup>5</sup>After 1980, the increase is driven by women, while male college achievement rates stagnated.

above. While immigrant status appears to have remained or become increasingly informative of incomes for recent cohorts, some of this effect may be a reflection of the changing educational achievement of natives relative to immigrants. Once we account for broad education categories, the result appears somewhat more favorable. The incremental informational association of immigrant status conditional on education indicates a more consistently convergent trend for women with less than high school education (towards native-born women with the same level of education) and college graduates of both genders (Figure 4). For male college graduates, this is a convergence towards lesser heterogeneity. Stagnation or even divergence remains a feature for more recent male immigrants without a high school education, as well as for those with a high school education or some college.

[Figure 4 about here.]

Figure 5 combines information about the relative average incomes of three broad educational achievement groups, as well as the interaction information between immigrant status, educational achievement and incomes. Immigrants in the early arrival cohorts see very little difference to natives of all education levels. Starting with the 1970s arrival cohort, the pattern changes significantly. While higher levels of education are associated with higher relative incomes, this educational achievement is now generally *less* informative of incomes for immigrants than for the population as a whole. Low levels of education on the other hand, associated with lower levels of average incomes, are more informative for immigrants. For example, negative values for immigrants below grade 12 indicate that the unconditional informational association between this level of education and income is weaker than the same association conditional on immigrant status. Upon learning that an individual from the 1990s arrival cohort is an immigrant with less than 12 years of education, our uncertainty is reduced by around 10% in the first ten years after migration, compared to how much uncertainty is reduced by knowing about the educational achievement alone. This is particularly significant since, as was documented in Figure 3, immigrants are much more likely to be in the lowest educational category than the native-born.

[Figure 5 about here.]

With time in the US, there also appears to be a certain extent of convergence towards parity in the association of educational achievement and income towards the pattern of natives. The homogenizing effect of lacking a high school degree appears to be shrinking somewhat for immigrants towards the level for natives within cohorts. The results are much more ambiguous for intermediate levels of education, and for those with four years or more of college education. After 30-40 years since migration, a college degree is about as informative for immigrants of early arrival cohorts as it is for natives, but that trend is interrupted for those arriving in the 1970s and thereafter. This points to a concerning shift in the extent to which foreign-born workers can see their education valorized.

In summary, the findings suggests that immigrant status weakens the association between high levels of education and income but strengthens it for low levels. There is a very strong association of immigrant status and income among workers without a college degree. Immigrant status is less predictive of incomes among the highly educated, for whom there is even a heterogeneity premium compared to natives. This pattern is particularly evident for recent immigrant cohorts, who are increasingly distant from parity with their native counterparts.

## 5.2 Immigration and the Intersectional Inequality of Opportunity

The previous sections documented the extent to which foreign-born cohorts cease to find their incomes shaped by processes effectively different from those influencing their native-born peers. The speed by which the “mark of migration” is fading for different cohorts appeared to decline significantly over the analyzed period, a shift that coincided with a reconfiguration of the immigrant flow into the US. This section turns to the ways in which the influence of immigration status on incomes is shaped by gender, race and ethnicity. Do immigrants integrate into the racial and gender hierarchy of the US, or does immigrant status remain a distinctive marker?

As cohorts advance through their life course, their outcomes will tend to diverge. Above we reviewed some of the mechanisms generating heterogeneity for the population as a whole. There is ample evidence however for the pervasive influence of social identity, including gender, race and ethnicity, in shaping economic opportunities throughout the life course (e.g. Thomas, Herring, and Horton, 1994; Bertrand and Mullainathan, 2004; Goldsmith, Hamilton, and Darity, 2007). From the informational perspective of this paper, the question arises whether different social identity groups are afforded similar gains in income heterogeneity at every stage of their careers.

Figure 6 shows how informational association of age and social identity with income declines with age. Rather strikingly however, there is a much higher floor for some groups, which keeps their incomes more organized throughout their observed economic life. While white and Asian men (and recently Asian women) with age come to achieve income distributions that are *more* heterogeneous than for the population as a whole, other groups continue experiencing a heterogeneity penalty. The relative influence of social identity decreases for the former, while increasing for Black and Hispanic women in particular. dos Santos and Wiener (2020) relate the extent to which social identity remains informative of incomes to the continuing effects of discrimination.

[Figure 6 about here.]

The interaction information among income, gender/race and immigration status sheds light on the integration trajectories of different social identity groups, and the differential impact of these identity markers on immigrants over the course of their working lives (Figure 7). Recall that the mutual information

represents both the extent to which immigration status is comparatively less informative for a particular gender/race compared to the population as a whole,  $A(Y|immig) - \mathbb{I}_{immig|ident}$ , and the corresponding difference in the informativeness of gender/race between immigrants and natives,  $A(Y|ident) - \mathbb{I}_{ident|immig}$ . The two expressions differ only in the order in which additional uncertainty about income is removed. This presentation not only facilitates relative assessments between groups, but also measures the mutual dependencies between gender/race and immigration status in the determination of incomes. The values of  $m(Y, ident, immig)$  indicate whether the information contained in the joint observation of  $(ident, immig)$  more nearly accounts for incomes compared to the simple sum of information in gender/race and immigration status separately.

[Figure 7 about here.]

Across all cohorts, being an immigrant is most informative of incomes for Hispanics of either gender and least informative for white men, closely followed by white women (Figure 7). This strong synergy in the relationship between Hispanic ethnic identity and immigration status in their influence on incomes suggests that there is information in the joint occurrence of these two features that is not present in the “Hispanic” or “immigrant” labels by themselves. Hispanic immigrants are singled out among all other groups in terms of the impact of their immigrant status on incomes. The 1970s immigrant cohort saw a particularly striking pattern of increasing distinction compared to their native-born Hispanic counterparts. Hispanic identity was initially around 3 percentage points more informative for immigrants in this cohort compared to natives, while in the third census after arrival this number increased to 10 percentage points. Every subsequent Hispanic arrival cohort saw further drops in their relative income heterogeneity compared to natives. White and Asian immigrants tend to enjoy greater degrees of heterogeneity than Black and Hispanic immigrants, when compared to the native-born population of the same race and ethnicity. The joint observation of a white racial identification and an immigrant status does very little to inform us about incomes, compared to the sum of the separate informational associations of immigrant status and race.

A comparison between immigrants and natives within each racial and ethnic category also allows us to see whether the informational impact of membership in subaltern or more privileged groups changes as immigrant cohorts spend time in the US. In particular, is there convergence to the native gendered, racial and ethnic structure? As Figure 7 reveals, knowledge that a person is white is less helpful in predicting incomes for immigrants than it is for natives. Once we are told that an immigrant is white, we know less about their likely incomes than we would if they were natives. For men, this effect appears to attenuate with time spent in the US, for white women the differential to natives remains. For Hispanics, the result is reversed. It is substantially more informative of incomes to learn about the Hispanic ethnicity of an immigrant than a native worker, even though native Hispanic workers already face a reduction in income heterogeneity in light of their identity. This difference appears to have been larger in each successive arrival cohort, and furthermore increased with years since migration. In

particular the 1970s arrival cohort saw large increases in the informativeness of their immigrant status relative to the native-born Hispanic cohort. Asian immigrants have seen somewhat greater heterogeneity gains from their racial identity than their native-born counterparts (the only exception being the 1950s arrival cohort, which appears with an initial heterogeneity penalty). Black workers show an interesting cross-over pattern for all cohorts (Figure 7). While “black” is more informative of incomes for immigrants than natives in the first census after arrival, the ordering switches for some of the recent cohorts and being black becomes less informative for immigrants with time spent in the US. While more detailed analysis is required to confirm this pattern, the result appears consistent with findings in the literature that Afro Caribbeans have better labor market outcomes than African Americans (Ifatunji, 2017).

The informational associations estimated in this section have shown that gender, race and ethnicity structure the immigrant experience for the entire life-course after arrival. We found that immigrant status has a smaller association with incomes for white men (and women) than other immigrant groups. White immigrant men in particular are on track to assimilate to white native men, while white immigrant women retain a heterogeneity premium over their native-born counterparts. Immigrant status is substantially more informative of incomes for Hispanics of either gender. Being Hispanic has an even greater association with incomes for immigrants than natives, and there is no convergence towards the native gendered and racial hierarchy.

## 6 Discussion

The informational account of incomes across immigrant cohorts and the corresponding native age group has revealed important differences in the observable significance of immigrant status over time. Similar to others in the literature, this study finds that integration appears to have slowed down for more recent entrants into the US. Other empirical studies have emphasized either the changing composition of the immigrant population (e.g. Borjas, 2014), or the changes in the wage and occupational structure of the US labor market (Portes, 1997; Butcher and Dinardo, 2002, e.g.). Our approach, while investigating the broad economic characteristics of immigrant populations relative to natives, problematizes the valuation of these characteristics in the host country. It also conceives of integration as an inherently *social* process that cannot be properly understood, and should not be empirically studied, using individualistic conceptions of income allocation.

Micro-econometric linear regression models claim to identify a time-since-migration “effect” that represents the speed and extent of integration. Such identification requires eliminating sources of endogeneity from omitted variables and other mis-specification. As argued above, fundamental limitations of our knowledge as well as available data cast doubt on this endeavor. The informational measures of association employed in this study are not susceptible to the problem of mis-specification, as they capture any dependence between the variables

in question. They are however also affected by the issue of self-selection on the basis of often unobservable features. Self-selection already occurs in the country of origin in systematic differences between movers and stayers (e.g. Chiswick, 1999), and in the non-random selection of return migrants (e.g. Constant and Massey, 2003). Both forms of selection arguably are aspects of “integration” as a community-level process, and I have not attempted to separate out their effects.

Whether particular communities and immigrant cohorts achieve improvements in their social outcomes reflects the context of incorporation each group faces, including political-legal and socioeconomic conditions. The period of analysis of this paper covers a number of significant changes in the stance of US immigration policy that contributed to reshaping migration flows (Zolberg, 2009; Massey, 2015, e.g.). The tools developed in this paper, particularly the interaction of racial and ethnic identity with immigration status, allows detailed analysis of the social impact of these contextual changes on integration trajectories. The relative degree to which immigrant status homogenizes income possibilities, and informs the valorization of socioeconomic characteristics, gives a reduced-form measure of the effect that the structure of opportunities and constraints in the host country has on immigrant outcomes.

As already mentioned, our analysis does not view the native-born population as a homogeneous group, but as further differentiated along lines of social identity. Immigrant status intersects with dimensions such as gender and race that themselves exert tremendous influence. Which social identity characteristics become socially salient is the result of complex and contested efforts of self- and other-classification. These classificatory struggles are themselves reflected in the terminology of “race” and “ethnicity” as social categories, and their respective administrative, political and disciplinary histories (Cornell and Hartmann, 2004). The informational association measures employed in this study have the unique advantage of identifying empirically the extent to which a certain identifier shapes social outcomes, without imposing *any* ex-ante assumption regarding the particular mechanism through which it operates. The salience of social identities, in other words, emerges from the analysis of the data.

The informational definition of integration may also help clarify the barriers and possibilities of solidarity between social groups. While integrated immigrant cohorts still experience various forms of inequality based on social class and other factors, incomes for such groups are no longer predictable from immigrant status. This may explain some of the difficulties earlier immigrants (and natives) have with empathizing with the experience of more recent immigrant cohorts, whose incomes continue to be shaped by immigrant status, as well as efforts to distance themselves from the immigrant identity.

Comparative studies across different institutional and political-economic contexts would allow for a more detailed assessment of such processes of group formation and dissolution, and may be fruitfully approached using information-theoretic integration measures. For example, one might be interested in studying conditions under which particular economic or identity characteristics become more or less informative for immigrants as compared to natives (see also the

literature on boundary-making practices, e.g. Barth, 1998; Wimmer, 2009). We have already discussed some of the institutional features that might render educational achievements more or less relevant in the distinction of immigrant and native workers.

## 7 Conclusion

Immigrants to the US have not been advancing through the social ladder in unison like a regiment of soldiers. Our results have documented a fracturing of the integration process for immigrants arriving after 1970 along lines of gender, race and ethnicity. That there should be significant heterogeneity in economic success over time is not surprising in light of the diversity of the immigrant population. This paper has advanced a definition of integration based on interpreting differences in the extent of heterogeneity between groups. If incomes are relatively more homogeneous for immigrant cohorts, this suggests that the collection of social forces allocating incomes in the US continue to treat immigrants differently than their native counterparts. Similarly, if immigration status is informative of outcomes it means that immigrants effectively have a smaller range of economic opportunities and remain marked by their history of migration.

The information-theoretic measures proposed here have the potential of being widely applicable to the study of immigrant integration. Future work might study the outcomes of immigrants in the second or third generation, or contrast the findings for international immigrants with those of domestic migrants. The interaction with additional degrees of freedom related to employment could shed light on the role of occupational and workplace segregation in preventing or facilitating integration. Preliminary work suggests that recent immigrant cohorts are concentrated in industries where they receive a heterogeneity penalty (such as retail and construction), which may serve as an index of workplace segmentation. A further possible extension of this work would analyze integration in dimensions other than income. This would be particularly relevant in comparing immigrant outcomes in the US to the European context of reception, where differential access to employment for example may play a larger role (e.g. Li and Heath, 2020).

### Acknowledgments

I would like to thank Paulo dos Santos, Duncan K. Foley, Katherine Moos, Ina Ganguli and Ying Chen for helpful discussions and advice.

### Declaration of Interest

The author declares no conflict of interest.

## References

- Ager, Alastair and Alison Strang (2008). "Understanding Integration: A Conceptual Framework". en. In: *Journal of Refugee Studies* 21.2, pp. 166–191.
- Anderson, Bridget (2010). "Migration, Immigration Controls and the Fashioning of Precarious Workers". English. In: *Work, Employment and Society* 24.2, pp. 300–317.
- Andersson, Fredrik et al. (2014). "Workplace Concentration of Immigrants". en. In: *Demography* 51.6, pp. 2281–2306.
- Barth, Fredrik (1998). "Introduction". English. In: *Ethnic Groups and Boundaries: The Social Organization of Culture Difference*. Ed. by Fredrik Barth. Reissued. OCLC: 247609912. Long Grove, Ill: Waveland Press, pp. 9–38.
- Bertrand, Marianne and Sendhil Mullainathan (2004). "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination". English. In: *American Economic Review* 94.4, pp. 991–1013.
- Bleakley, Hoyt and Aimee Chin (2004). "Language Skills and Earnings: Evidence from Childhood Immigrants\*?". English. In: *Review of Economics and Statistics* 86.2, pp. 481–496.
- Bonacich, Edna (1972). "A Theory of Ethnic Antagonism: The Split Labor Market". In: *American Sociological Review* 37.5, p. 547.
- Borjas, George J. (1989). "Immigrant and Emigrant Earnings: A Longitudinal Study". In: *Economic inquiry* 27.1, pp. 21–37.
- (1991). "Immigration and Self-Selection". In: *Immigration, Trade, and the Labor Market*. Ed. by John M. Abowd and Richard B. Freeman. A National Bureau of Economic Research Project Report. tex.lccn: HD8081.A5 I53 1991. Chicago: University of Chicago Press, pp. 29–76.
- (1995). "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?" English. In: *Journal of Labor Economics* 13.2, pp. 201–245.
- (2014). *Immigration Economics*. English. Harvard University Press.
- (2015). "The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again". In: *Journal of Human Capital* 9.4, pp. 483–517.
- Butcher, Kristin F. and John Dinardo (2002). "The Immigrant and Native-Born Wage Distributions: Evidence from United States Censuses". In: *ILR Review* 56.1, pp. 97–121.
- Chiswick, Barry R. (1978). "The Effect of Americanization on the Earnings of Foreign-Born Men". In: *Journal of Political Economy* 86.5, pp. 897–921.
- (1999). "Are Immigrants Favorably Self-Selected?" In: *The American Economic Review* 89.2, pp. 181–185.
- Constant, Amelie and Douglas S. Massey (2003). "Self-Selection, Earnings, and out-Migration: A Longitudinal Study of Immigrants to Germany". English. In: *Journal of Population Economics* 16.4, pp. 631–653.
- Cornell, Stephen and Douglas Hartmann (2004). "Conceptual Confusions and Divides: Race, Ethnicity, and the Study of Immigration". English. In: *Not*

- Just Black and White: Historical and Contemporary Perspectives on Immigration, Race, and Ethnicity in the United States*. Ed. by Nancy Foner and George M. Fredrickson. Russell Sage Foundation, pp. 23–41.
- Crul, Maurice (2016). “Super-Diversity vs. Assimilation: How Complex Diversity in Majority–Minority Cities Challenges the Assumptions of Assimilation”. English. In: *Journal of Ethnic and Migration Studies* 42.1, pp. 54–68.
- Czaika, Mathias and Hein de Haas (2014). “The Globalization of Migration: Has the World Become More Migratory?” English. In: *International Migration Review* 48.2, pp. 283–323.
- Dos Santos, Paulo L. (2017). “The Principle of Social Scaling”. English. In: *Complexity* 2017, pp. 1–9.
- Dos Santos, Paulo L. and Noé Wiener (2019). “Indices of Informational Association and Analysis of Complex Socio-Economic Systems”. en. In: *Entropy* 21.4. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 367.
- (2020). “By the Content of Their Character? Comparative Group Income Heterogeneity and Socio-Economic Discrimination”. In: *Journal of Mathematical Sociology* 44.1, pp. 12–41.
- Dustmann, Christian, Uta Schönberg, and Jan Stuhler (2016). “The Impact of Immigration: Why Do Studies Reach Such Different Results?” English. In: *Journal of Economic Perspectives* 30.4, pp. 31–56.
- Foley, Duncan K. (1994). “A Statistical Equilibrium Theory of Markets”. In: *Journal of Economic Theory* 62.2, pp. 321–345.
- Golan, Amos (2017). *Foundations of Info-Metrics: Modeling and Inference with Imperfect Information*. tex.lccn: T50 .G64 2017. New York, NY: Oxford University Press.
- Goldsmith, Arthur H., Darrick Hamilton, and William Darity (2007). “From Dark to Light: Skin Color and Wages among African-Americans”. In: *The Journal of Human Resources* 42.4, pp. 701–738.
- Grogger, Jeffrey and Gordon H Hanson (2011). “Income Maximization and the Selection and Sorting of International Migrants”. In: *Journal of Development Economics* 95.1, pp. 42–57.
- Grzymala-Kazłowska, Aleksandra and Jenny Phillimore (2018). “Introduction: rethinking integration. New perspectives on adaptation and settlement in the era of super-diversity”. en. In: *Journal of Ethnic and Migration Studies* 44.2, pp. 179–196.
- Ifatunji, Mosi Adesina (2017). “Labor Market Disparities Between African Americans and Afro Caribbeans: Reexamining the Role of Immigrant Selectivity”. en. In: *Sociological Forum* 32.3, pp. 522–543.
- Jaynes, Edwin T. (2003). *Probability Theory: The Logic of Science*. Ed. by G. Larry Bretthorst. tex.lccn: QA273 .J36 2003. Cambridge, UK ; New York, NY: Cambridge University Press.
- Kenney, G. and D. Wissoker (1994). “An Analysis of the Correlates of Discrimination Facing Young Hispanic Job-Seekers”. In: *American Economic Review* 84.674-683.

- Lalonde, Robert J. and Robert H. Topel (1997). “Economic Impact of International Migration and the Economic Performance of Migrants”. English. In: *Handbook of Population and Family Economics*. Ed. by Mark Rosenzweig and Oded Stark. Vol. 1. Elsevier, pp. 799–850.
- Li, Yaojun and Anthony Heath (2020). “Persisting disadvantages: a study of labour market dynamics of ethnic unemployment and earnings in the UK (2009–2015)”. en. In: *Journal of Ethnic and Migration Studies* 46.5, pp. 857–878.
- Liu, Cathy Yang (2009). “Ethnic enclave residence, employment, and commuting of Latino workers”. en. In: *Journal of Policy Analysis and Management* 28.4, pp. 600–625.
- Massey, Douglas S. (2015). “A Missing Element in Migration Theories”. In: *Migration letters : an international journal of migration studies* 12.3. tex.pmcid: PMC4933523 tex.pmid: 27398085, pp. 279–299.
- Massey, Douglas S. and Nancy A. Denton (1985). “Spatial Assimilation as a Socioeconomic Outcome”. en. In: *American Sociological Review* 50.1, p. 94.
- Nawyn, Stephanie J. and Linda Gjakaj (2014). “The Magnifying Effect of Privilege: Earnings Inequalities at the Intersection of Gender, Race, and Nativity”. en. In: *Feminist Formations* 26.2, pp. 85–106.
- Oreopoulos, Philip (2011). “Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes”. English. In: *American Economic Journal: Economic Policy* 3.4, pp. 148–171.
- Portes, Alejandro (1997). “Immigration Theory for a New Century: Some Problems and Opportunities”. In: *The International Migration Review* 31.4, pp. 799–825.
- Radu, Dragos (2008). “Social Interactions in Economic Models of Migration: A Review and Appraisal”. en. In: *Journal of Ethnic and Migration Studies* 34.4, pp. 531–548.
- Ruggles, Steven et al. (2015). *Integrated Public Use Microdata Series: Version 6.0 [Machine-Readable Database]*. tex.howpublished: University of Minnesota. Minneapolis, MN.
- Shinnaoui, Doha and Renu Narchal (2010). “Brain Gain to Brain Waste: Individual Biases, Prejudice, and Discounting of Migrant Skills”. English. In: *Journal of International Migration and Integration / Revue de l'integration et de la migration internationale* 11.4, pp. 423–437.
- Stark, Oded (1993). *Migration of Labor*. English. Wiley.
- Stewart, Quincy Thomas and Jeffrey C. Dixon (2010). “Is It Race, Immigrant Status, Or Both: An Analysis of Wage Disparities among Men in the United States”. English. In: *International Migration Review* 44.1, pp. 173–201.
- Thomas, Melvin E., Cedric Herring, and Hayward Derrick Horton (1994). “Discrimination over the Life Course: A Synthetic Cohort Analysis of Earnings Differences between Black and White Males, 1940-1990”. In: *Social Problems* 41.4, pp. 608–628.
- Tilly, Chris and Charles Tilly (1998). *Work Under Capitalism*. New perspectives in sociology. Boulder, Colo: Westview Press.

- Vertovec, Steven (2007). “Super-diversity and its implications”. en. In: *Ethnic and Racial Studies* 30.6, pp. 1024–1054.
- Waldinger, Roger D (2009). “Transforming Foreigners into Americans”. English. In: *New Americans: A Guide to Immigration Since 1965*. Ed. by Mary C Waters and Reed Ueda. OCLC: 923108648. Cambridge: Harvard University Press, pp. 137–148.
- Waldinger, Roger D and Michael Ira Lichter (2003). *How the Other Half Works: Immigration and the Social Organization of Labor*. English. OCLC: 896924569. Berkeley, Calif.: University of California Press.
- Wimmer, Andreas (2009). “Herder’s Heritage and the Boundary-Making Approach: Studying Ethnicity in Immigrant Societies”. English. In: *Sociological Theory* 27.3, pp. 244–270.
- Zolberg, Aristide R (2009). *A Nation by Design: Immigration Policy in the Fashioning of America*. Harvard University Press.

## A Data Source and Variable Construction

The sample was extracted from (Ruggles et al., 2015), using the 1% metropolitan samples from the 1970 census, the 5% state samples for 1980 and 1990, the 5% sample for 2000, as well as the pooled 2007-2011 American Community Survey (ACS) data, corresponding to a 5% sample. For simplicity, we refer to the 5-year ACS as the 2010 census. We restrict the sample to respondents living in households according to the 1970 definition only, which excludes those living in group quarters and households with five or more persons unrelated to the head of household. We also exclude minors under 18 and those aged 65 and above in the census year.

The economic outcome of interest in this study are weekly wages and salaries, adjusted for inflation using the 1999 Consumer Price Index. We therefore further restrict the sample to those reporting at least one week of paid work during the previous year and a strictly positive weekly wage. Within each census year, we remove observations above the 98th percentile of the sample-weighted weekly wage distribution, as well as those with top-coded annual wages. We also remove observations with imputed wage information from our sample, including only logical edits as the most reliable form of imputation.

An important limitation of the present study relates to the particular choice of racial and ethnic identifiers. The desire to obtain consistent groupings of sufficient size over the period of analysis led to categorizations that are highly aggregated. Clearly the “Asian” category, in particular, encompasses groups with very different trajectories and histories in the US. Vastly more heterogeneous aggregations mask the very distinct ways that particular national origin groups experience their immigration and settlement. As estimation methods for the measures in this study improve and available sample sizes increase, a more detailed analysis will become feasible. Until such time, we make use of the information we do have to learn about the functioning of the economic system that generated the macroscopic records we observe in the wage distributions.

## A.1 Construction of Cohorts

In constructing cohorts for the study of integration trajectories, the dimension of age-at-immigration needs to be considered as well as the birth year. The likely trajectories of an immigrant arriving as a child will be very different from one immigrating close to retirement. Immigrants arriving as children or young adults may complete some or all of their schooling in the US, and have labor market experiences similar to the second generation of children born in the US to immigrant parents.

The year-of-immigration variable in the census data is very coarsely measured in earlier census waves, which creates two sets of challenges. First, age at migration as well as years-since-migration have to be constructed on the basis of the coarse year-of-immigration variable. Second, due to changes in the year-of-immigration variable between the 1970, 1980 and 1990 censuses, it is not possible to create entirely consistent arrival cohorts over time. The 1960s arrival cohort includes respondents first entering the US to live/stay between 1960-1970 in the 1970 census, but between 1960-69 in the remaining census waves. Similarly, the 1970s cohort includes arrivals between 1970-1980 for the 1980 census, but 1970-79 for the 1990/2000 censuses and the ACS.

Due to these limitations of our data, we are faced with a trade-off between using all available observations and obtaining narrow age-at-migration windows. By constructing year-of-immigration windows by decade and combining them with birth cohorts that span 10 years, we obtain a broad age-at-migration window of ages 20-38. This cohort definition focuses on immigrants who have the potential to stay in the US for a substantial part of their working lives, although it may include some who have immigrated for the purpose of obtaining higher education.<sup>6</sup>

## B Coarse-Graining Procedure

Coarse-graining refers to the (informationally) compressed representation of the state of a system under study, arguably the central task of any scientific theory. In this paper, we retained information on a small set of social and economic characteristics, such as national origin and income levels, but glossed over remaining differences, such as the specific identities of the individuals involved. The uncertainty about the particular permutation of individuals that produced the observed distributions of income is the basis for our accounting exercises.

One element of the coarse-graining proposed here is to consider small income differences to be equivalent. The proposed approach involves defining a constant number of equal-sized histogram bins, generally 25 in this study. We use a lower number of 15 wage bins when studying groups with relatively small sample sizes, such as arise in the interaction of immigrant status and race/ethnicity.

---

<sup>6</sup>We have repeated the analysis with a more narrow range of age-at-migration, 25-39 years, which leads us to use only respondents born in the first half of their birth decade. Results for this alternative age-at-migration window are qualitatively similar.

Supplementary analyses conducted by the author (not reported here) suggest that the qualitative results of this paper are very robust to alternative coarse-graining procedures, including a range of different bin specifications.

## C Sample sizes

Despite the substantial sample sizes achieved by the use of census data, a small number of sub-populations are only poorly captured. These cases arise when immigration cohorts are interacted with social identity or educational achievement variables in earlier census years. The approach taken in this paper is to drop categories with fewer than 1000 observations from our analysis (highlighted in the tables). Future research in this area might benefit from more advanced estimation techniques for the entropic measures used in this study.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

## List of Figures

1	Wage distributions for the 1960s arrival cohort compared to their native counterparts, by gender. . . . .	25
2	Income integration by arrival cohort and sex. 1970-2010 Census data. . . . .	26
3	Educational distribution by nativity and arrival cohort. 1970-2010 Census data. . . . .	27
4	Income integration conditional on education, by arrival cohort and sex. 1970-2010 Census data. . . . .	28
5	Mutual information between income, education, and immigration status, and relative average income of education level. Faint gray lines show results for native cohorts. 1970-2010 Census data. . . . .	29
6	Incremental informational association of age and social identity with income. The measure shows how much information is gained for different cohorts when compared to the overall population. 1970-2010 Census data. . . . .	30
7	Interaction information of nativity, social identity and income, by arrival cohort. 1970-2010 Census data. . . . .	31

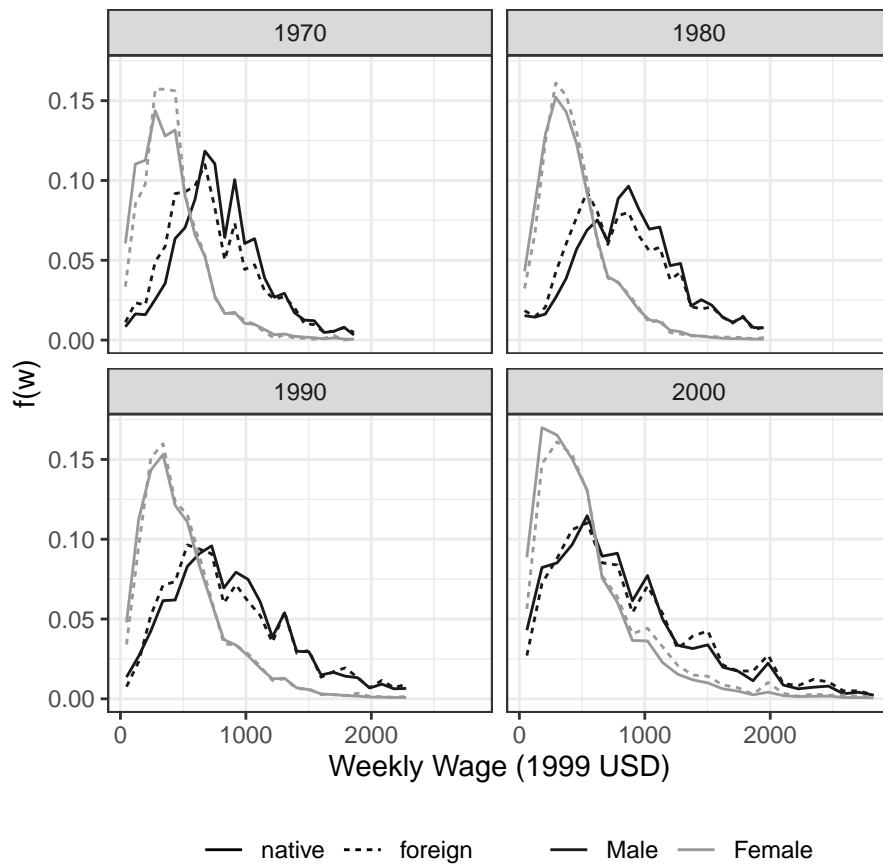


Figure 1: Wage distributions for the 1960s arrival cohort compared to their native counterparts, by gender.

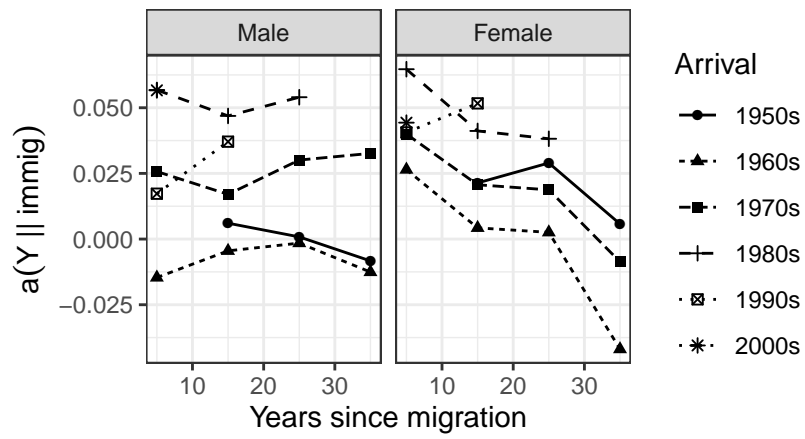


Figure 2: Income integration by arrival cohort and sex. 1970-2010 Census data.

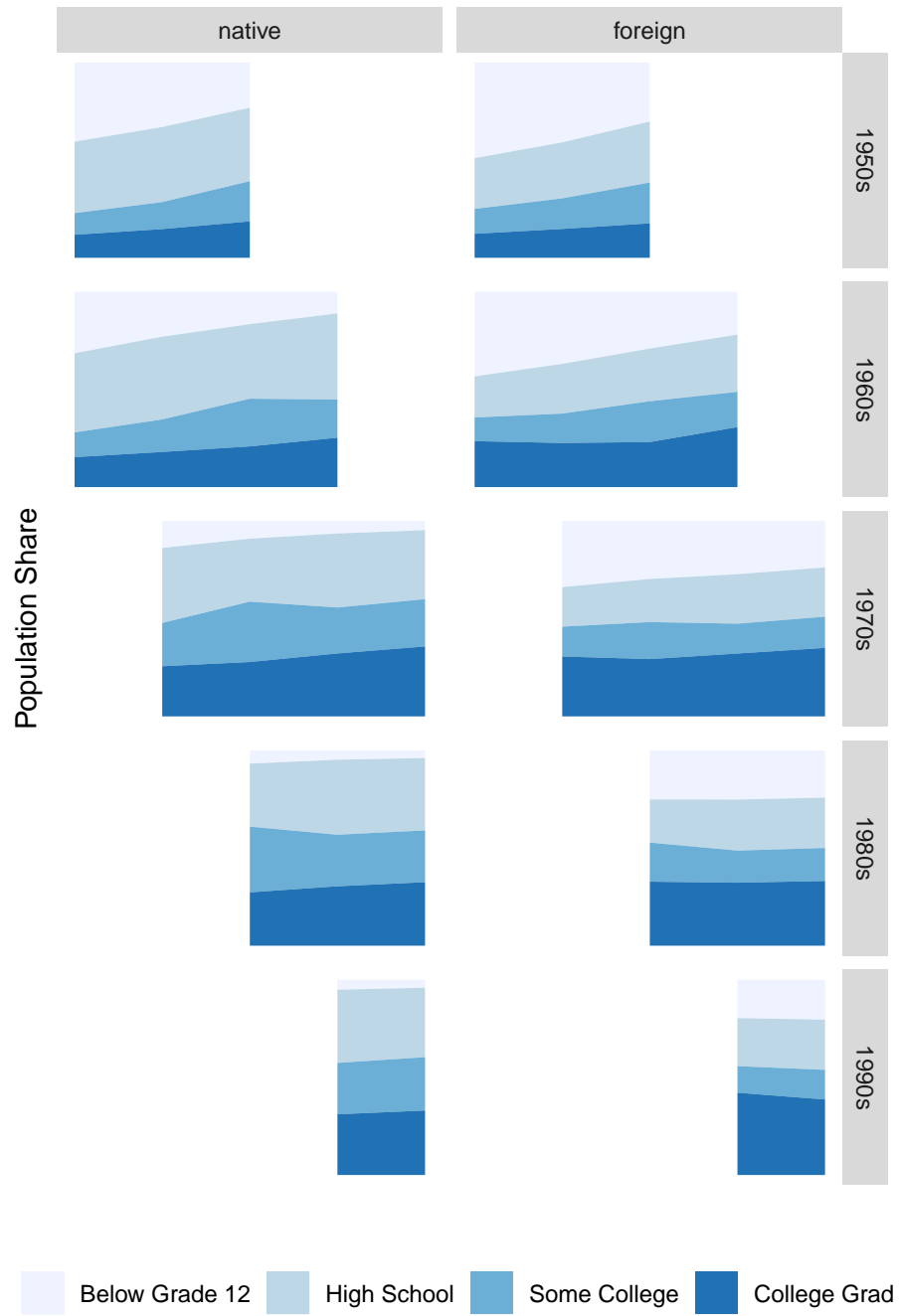


Figure 3: Educational distribution by nativity and arrival cohort. 1970-2010 Census data.

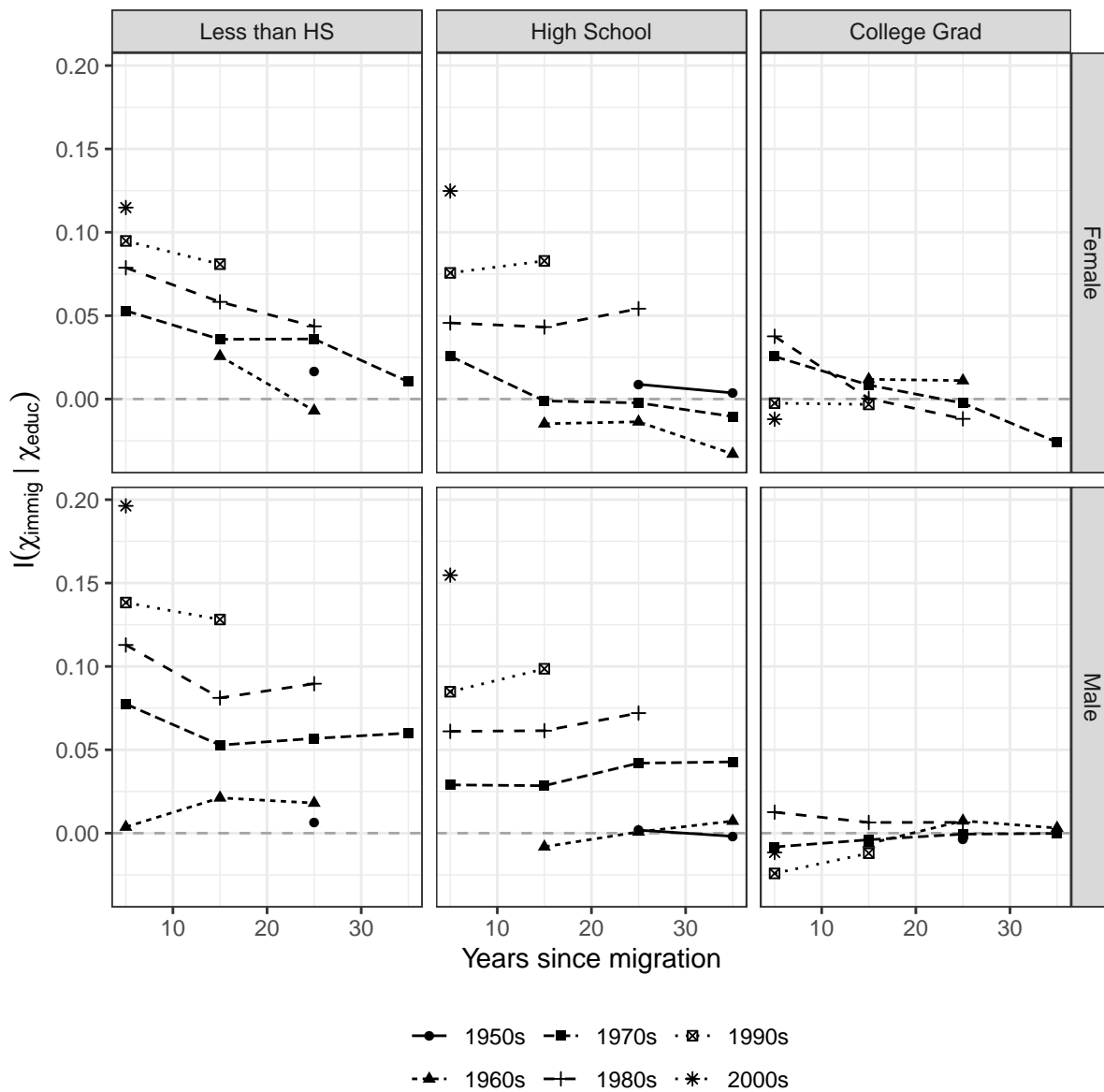


Figure 4: Income integration conditional on education, by arrival cohort and sex. 1970-2010 Census data.

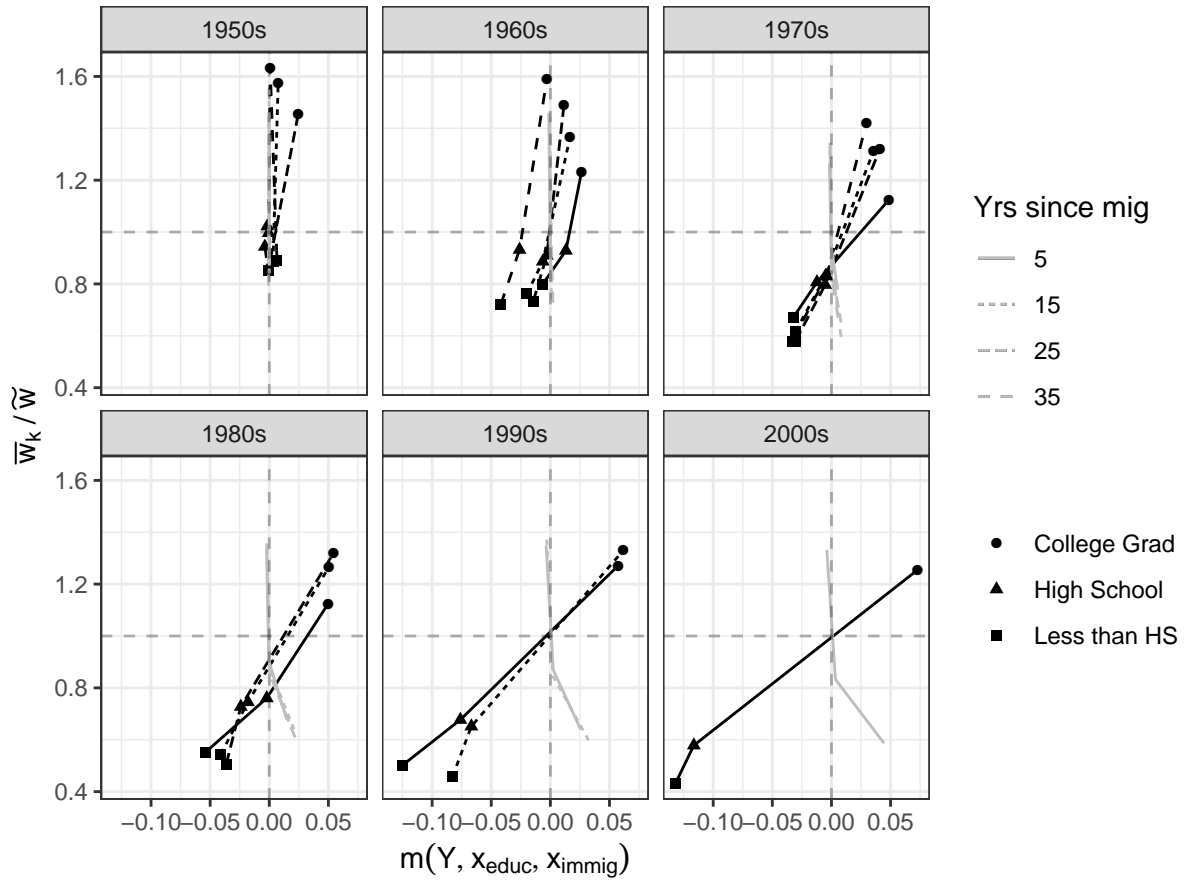


Figure 5: Mutual information between income, education, and immigration status, and relative average income of education level. Faint gray lines show results for native cohorts. 1970-2010 Census data.

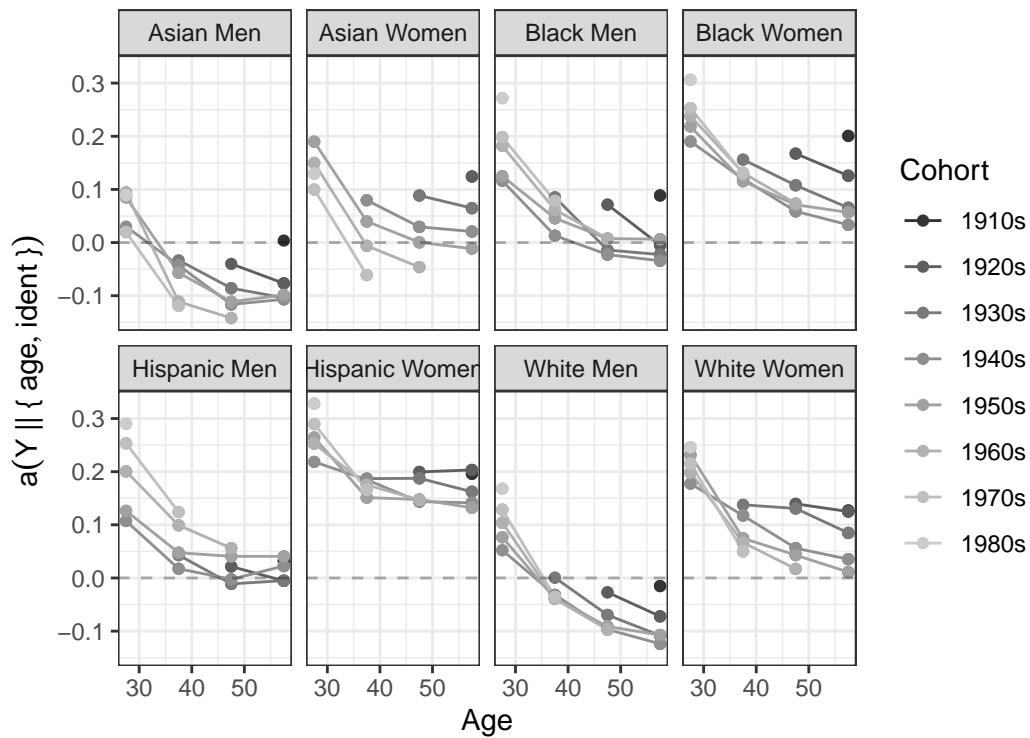


Figure 6: Incremental informational association of age and social identity with income. The measure shows how much information is gained for different cohorts when compared to the overall population. 1970-2010 Census data.

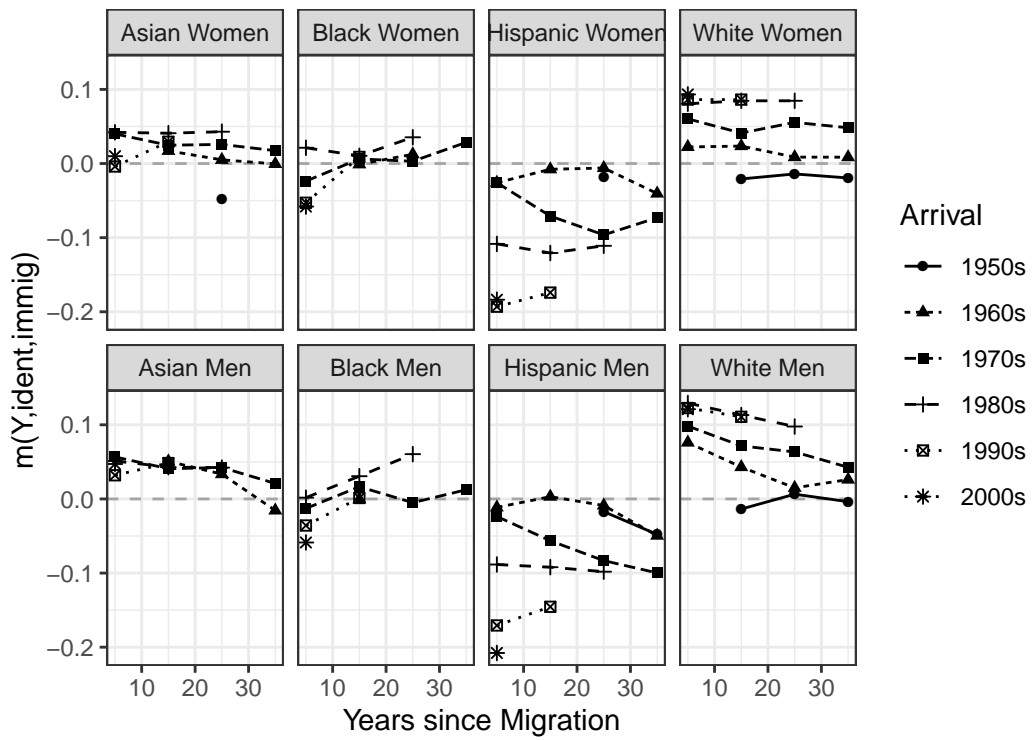


Figure 7: Interaction information of nativity, social identity and income, by arrival cohort. 1970-2010 Census data.

## List of Tables

1	Foreign-born sample size by census year, decade of immigration and educational achievement. Dropped categories highlighted. . .	33
2	Foreign-born sample size by census year, decade of immigration, sex and educational achievement. Dropped categories highlighted.	34
3	Foreign-born sample size by census year, decade of immigration and race/ethnicity. Dropped categories highlighted. . . . .	35

Table 1: Foreign-born sample size by census year, decade of immigration and educational achievement. Dropped categories highlighted.

Year	Arrival	Below Grade 12	High School	Some College	College Grad	Total
1970	1950s	1469	775	380	351	2975
1970	1960s	1933	933	540	1039	4445
1980	1950s	5671	3960	2157	1965	13753
1980	1960s	8254	5688	3347	4946	22235
1980	1970s	12615	7563	5700	11418	37296
1990	1950s	1610	1649	1113	920	5292
1990	1960s	5418	4936	3807	4049	18210
1990	1970s	9671	6837	5937	8996	31441
1990	1980s	13599	11478	10378	16733	52188
2000	1960s	1628	2068	1283	2120	7099
2000	1970s	7376	6429	3899	8013	25717
2000	1980s	14438	14212	8952	17203	54805
2000	1990s	11585	13972	7591	22922	56070
2010	1970s	3521	4031	2637	5913	16102
2010	1980s	12123	13757	9345	19300	54525
2010	1990s	11808	15968	10126	27732	65634
2010	2000s	9074	10730	6572	28280	54656

Table 2: Foreign-born sample size by census year, decade of immigration, sex and educational achievement. Dropped categories highlighted.

Year	Arrival	Sex	Less than HS	High School	College Grad	Total
1970	1950s	Male	844	577	249	1670
1970	1950s	Female	625	578	102	1305
1970	1960s	Male	1129	783	766	2678
1970	1960s	Female	804	690	273	1767
1980	1950s	Male	3251	2810	1265	7326
1980	1950s	Female	2420	3307	700	6427
1980	1960s	Male	4532	3871	3075	11478
1980	1960s	Female	3722	5164	1871	10757
1980	1970s	Male	7486	6976	7267	21729
1980	1970s	Female	5129	6287	4151	15567
1990	1950s	Male	934	1214	579	2727
1990	1950s	Female	676	1548	341	2565
1990	1960s	Male	3005	3764	2321	9090
1990	1960s	Female	2413	4979	1728	9120
1990	1970s	Male	5488	5790	4923	16201
1990	1970s	Female	4183	6984	4073	15240
1990	1980s	Male	8255	12234	10092	30581
1990	1980s	Female	5344	9622	6641	21607
2000	1960s	Male	945	1453	1253	3651
2000	1960s	Female	683	1898	867	3448
2000	1970s	Male	4374	4739	4329	13442
2000	1970s	Female	3002	5589	3684	12275
2000	1980s	Male	8367	12074	9371	29812
2000	1980s	Female	6071	11090	7832	24993
2000	1990s	Male	7325	11898	13715	32938
2000	1990s	Female	4260	9665	9207	23132
2010	1970s	Male	2042	2974	3039	8055
2010	1970s	Female	1479	3694	2874	8047
2010	1980s	Male	6957	11378	9820	28155
2010	1980s	Female	5166	11724	9480	26370
2010	1990s	Male	6311	12562	14349	33222
2010	1990s	Female	5497	13532	13383	32412
2010	2000s	Male	5883	9336	15277	30496
2010	2000s	Female	3191	7966	13003	24160

Table 3: Foreign-born sample size by census year, decade of immigration and race/ethnicity. Dropped categories highlighted.

Year	Arrival	asian	black	hispanic	white	Total
1970	1950s	147	57	542	2216	2962
1970	1960s	548	238	1516	2073	4375
1980	1950s	923	258	2570	9971	13722
1980	1960s	3380	1537	7895	9316	22128
1980	1970s	12489	2791	12341	9295	36916
1990	1950s	473	112	988	3714	5287
1990	1960s	3063	961	6548	7604	18176
1990	1970s	11016	2121	11168	7047	31352
1990	1980s	18461	4138	19213	10101	51913
2000	1960s	1354	406	2317	3002	7079
2000	1970s	9128	1995	8801	5706	25630
2000	1980s	18985	5249	20773	9478	54485
2000	1990s	16883	3997	20556	14341	55777
2010	1970s	6121	1372	4697	3857	16047
2010	1980s	20385	5294	18869	9652	54200
2010	1990s	21571	5404	23207	15038	65220
2010	2000s	19016	4348	19277	11569	54210