Social Networks, Ethnicity, and Entrepreneurship

William R. Kerr
Harvard University and NBER

Martin Mandorff*
Swedish Competition Authority

November 13, 2020

Abstract

We study the relationship between ethnicity, occupational choice, and entrepreneurship. Immigrant groups in the United States cluster in specific business sectors. For example, Koreans are 34 times more concentrated in self-employment for dry cleaning than other immigrant groups, and Gujarati-speaking Indians are 84 times more concentrated in managing motels. We quantify that smaller and more socially isolated ethnic groups display higher rates of entrepreneurial concentration. This is consistent with a model of social interactions where non-work relationships facilitate the acquisition of sector-specific skills and result in occupational stratification along ethnic lines via concentrated entrepreneurship.

Data Disclosure: Data and programs are available upon request.

Key words: entrepreneurship, self-employed, occupation, ethnicity, immigration, networks.

JEL codes: L26; D21, D22, D85, F22, J15, L14, M13.

*Comments are appreciated and can be sent to wkerr@hbs.edu. We thank Emek Basker, Gary Becker, Michel Beine, Ola Bengtsson, Gustaf Bruze, Dennis Carlton, Barry Chiswick, Rob Fairlie, Matthew Gentzkow, John Haltiwanger, Emil Iantchev, Svante Janson, Mini Kaur, Steven Lalley, Anne Le Brun, Ben Mathew, Trang Nguyen, Andriy Protsyk, Yona Rubinstein, Jesse Shapiro, Rachel Soloveichik, Chad Syverson, Catherine Thomas, Robert Topel, Nick Wormald, and seminar participants for very valuable comments. We thank Meir Brooks, Rahul Gupta, and Kendall Smith for excellent research support. The theory section of this paper draws heavily from Mandorff's Ph.D. Dissertation at the University of Chicago. Financial support from the Marcus Wallenberg Foundation, the Jan Wallander and Tom Hedelius Foundation, the Esther and T.W. Schultz Dissertation Fellowship, the Markovitz Dissertation Fellowship, the Kauffman Foundation, and Harvard Business School is gratefully acknowledged.

1 Introduction

Immigrants engage in self-employment and entrepreneurship more than natives. Fairlie and Lofstrom (2013) calculate that immigrants represent 25% of new US business owners but only 15% of the workforce. Moreover, immigrant business owners tend to specialize in a few industries, and these industries vary across ethnic groups. Prominent examples in the United States include Korean dry cleaners, Vietnamese nail care salons, Yemeni grocery stores, and Punjabi Indian convenience stores. Despite the importance of these patterns economically—for example, *The Economist* reported that one-third of all US motels in 2016 were owned by Gujarati Indians—few studies examine the origin or consequences of this ethnic specialization for self-employment.

We study how social interactions within isolated ethnic groups can generate entrepreneurial specialization without relying on inherent differences across groups. We develop a model that considers a small industry where self-employed entrepreneurs benefit from social interactions outside of work, such as family gatherings, religious and cultural functions, and meetings with friends. At these events, entrepreneurs can share industry knowledge and provide advice on topics such as: how to start up or take over a business; how to establish supplier, customer and employee relationships; how to handle licenses and taxes; how to navigate market trends; and how to adjust product offerings and set prices. The model shows how small ethnic minority groups can develop comparative advantages for self-employment in small industries in this way.

These model foundations are consistent with case examples of the origin and expansion of prominent ethnic clusters. The first Gujarati hotel came about when Kanji Manchhu Desai, along with two Gujarati farm workers, took over a 32-room hotel in Sacramento in 1942 after the hotel's Japanese-American owner was forced into a World

War II internment camp. Desai moved five years later to a San Francisco hotel and thereafter encouraged new Gujarati immigrants into the business: "If you are a Patel, lease a hotel" (Bhattacharjee, 2017). A sociologist described the subsequent spread (Dhingra, 2012; Virani, 2012): "...if a new Gujarati immigrant wanted to open up a florist, for instance, his relatives wouldn't know anything about it but if he wanted to open up a motel, he would have access to experienced investors and advice."[1]

The start of the Vietnamese nail care salon industry is even more serendipitous. In 1975, actress Tippi Hendren of Alfred Hitchcock's *The Birds* traveled to Hope Village, a Vietnamese refugee camp in California with the goal of helping the women identify a vocation. During the visit, the women became fascinated by Hendren's manicure, so Hendren subsequently brought her personal manicurist and additional support from a beauty school to the camp to teach 20 women the trade. Hendren further helped the women become properly licensed and find early employment in nail salons throughout Southern California (e.g., Moris, 2015; Hoang, 2015). The model spread, and Vietnamese today are by far the largest ethnic group working in nail care.

These and similar accounts suggest a general process towards entrepreneurial specialization with industry-specific skills being endogenously acquired. Millman writes in *The Other Americans* (1998), for example: "The Gujarati model for motels might be copied by Latinos in landscaping, West Indians in homecare or Asians in clerical services. By operating a turnkey franchise as a family business, immigrants will help an endless stream of service providers grow." Moreover, ethnic entrepreneurial specialization has deep historical roots and occurs in many countries. Examples of ethnic specializations are Jewish merchants in Medieval and Renaissance Europe, shopkeepers and traders among Armenians in the Ottoman Empire, Jains and Parsis in India,

[1] Chung and Kalnins (2006) show how Gujarati hotel owners use these networks to access resources.

Lebanese in West Africa, Indians in East Africa, Japanese in South America, and Chinese in Southeast Asia and the Caribbean, as well as the Chinese launderers in early twentieth century California.

We accordingly construct a general model that does not revolve around the traits of any single ethnic group or setting, and our empirical analysis includes as many immigrant groups in the United States as possible. Understanding the origin of group-level differences is important, as we know that the higher immigrant propensity towards entrepreneurship remains after controlling for the observable traits of individuals. Our model and subsequent empirical work emphasize how smaller group size and greater social isolation can lead to entrepreneurial specialization by an ethnic group to take advantage of the inherent social interactions among group members. These interactions yield a comparative advantage for ethnic self-employment in small industries.[2,3]

We analyze the model's predictions using Census Bureau data for the United States in 2000. The size of groups and their social isolation, which we measure using in-marriage rates among immigrants who arrived to the United States as children, strongly predict industrial concentration for immigrant self-employed entrepreneurs. A one

[2]In our setting, social interaction can increase the productivity of small minority groups, working in the opposite direction of market discrimination, often present at the same time. The latter, as analyzed by Becker (1957), acts as a tax on market interaction and tends to hurt the minority. An illustration of the dichotomy of social interaction and market interaction is found in Shakespeare's *The Merchant of Venice* (Act 1, Scene III). Following a negotiation over a large loan to a Christian man who has always scorned him, the Jewish moneylender Shylock comments: "I will buy with you, sell with you, talk with you, walk with you, and so following; but I will not eat with you, drink with you, nor pray with you."

[3]We do not explicitly model factors like access to finance, risk sharing, and sanctions for misbehavior that are frequently ascribed to ethnic networks. We likewise will not formally model behavioral factors prompting self-employment (e.g., Åstebro et al., 2014). Accounts like that of Gujarati hotel owners suggest these factors contribute to entrepreneurial specialization. For example, incumbent Gujarati owners were willing to provide new Gujarati immigrants access to funds to purchase hotel properties (Dhingra, 2012; Virani, 2012). As these incumbents would likely favor these hotel investments over investments in other sectors given their knowledge of the industry and ability to redeploy the property if the new arrival failed, this lending would serve to increase ethnic entrepreneurial specialization. But, ethnic bonds surely supported other lending as well, even if to a lesser degree.

standard-deviation decline in group size raises the group's industry concentration for self-employment by 0.6 standard deviations, and a one standard-deviation increase in group isolation boosts concentration by 0.3 standard deviations in our baseline model. Our work is robust to using a panel model covering 1980-2018, controlling for expected industry concentration based upon Monte Carlo simulations with each group's size, considering different measures of social isolation, exploiting variation in group size across metropolitan areas, and using instruments developed from a gravity model for migration to the United States and in-marriage rates present in the United Kingdom and Spain. Other extensions analyze income levels for immigrants and the industries chosen for entrepreneurial specialization, finding results consistent with our framework.

Our work connects to prior studies of immigrant entrepreneurship and self-employment (e.g., Fairlie and Lofstrom, 2013).[4] Classic accounts of entrepreneurship focus on factors like risk taking (Kihlstrom and Laffont, 1979), business acumen (Lucas, 1978) or skill mix (Lazear, 2005), with the connection of entrepreneurship to migration being frequently noted but unexplained. Fairlie and Robb (2007) find that more than half of business owners have close relatives who are self-employed, and a quarter of business owners have worked for these. The role of networks for entrepreneurs for giving and receiving advice has received extensive attention in the entrepreneurship literature.[5]

Building on these types of interactions, our model provides among the first joint explanations for immigrants engaging in entrepreneurship at greater rates and doing so in a pattern that emphasizes industry specialization by group. Our work relates

[4] See Chung and Kalnins (2006), Fairlie (2008), Fairlie et al. (2010), Hunt (2011), Patel and Vella (2013), Kerr and Kerr (2017, 2020a), and Kim and Morgan (2018). Fairlie and Lofstrom (2013) and Kerr (2013) provide reviews.

[5] For example, Birley (1985), Elfring and Hulsink (2003), Greve and Salaff (2003), Rosenthal and Strange (2012), Ghani et al. (2013), Leyden and Link (2015), Kerr and Kerr (2020b), and Bennet and Chatterji (2020).

to studies in sociology regarding entrepreneurial specialization and explanations like sojourner status, middleman minorities, discrimination in the labor market, social cohesion, social capital and networks, as well as cultural and/or religious traits in specific groups. See the online appendix for an overview.

We also relate to the recent literatures that have shown immigrants cluster in certain occupations (e.g., Patel and Vella, 2013) and the importance of ethnic networks for immigrants (e.g. Munshi, 2003; Beaman, 2012). Social interactions are important in job referrals, searching, and hiring (e.g., Granovetter, 1973; Bayer et al., 2008; Neumark, 2013), and the agglomeration literature describes how interactions can boost productivity (e.g., Arzaghi and Henderson, 2008; Glaeser and Gottlieb, 2009). Whereas group-level differences tend to decay over time in a basic referral model, e.g., due to random disturbances or skill heterogeneity, social interaction in our model yield increasing returns and stratification. Extensive literatures consider minority occupational specialization[6] and the importance of social interactions for economic behavior within or outside of the workplace.[7] Our paper builds on these literatures to provide unique insights to self-employment behavior that are traced out below.

2 A Model of Entrepreneurial Clustering

2.1 Model Set-Up

We construct a simple model to illustrate how social isolation and small group size can generate ethnic entrepreneurial clustering when social interactions and production

[6]Kuznets (1960) observes that "all minorities are characterized, at a given time, by an occupational structure distinctly narrower than that of the total population and the majority." Our theory is also related to the concept of ethnic capital (Borjas, 1992, 1995) and group assimilation (Lazear, 1999). Patel et al. (2013) provide a review.

[7]Examples include Granovetter (1973), Glaeser et al. (1996), and Glaeser and Scheinkman (2002). Durlauf and Fafchamps (2006) and Durlauf and Ioannides (2010) provide reviews.

are complementary. To keep the model tractable and intuitive, we make several strong assumptions. Everyone has equal ability and is divided into two ethnic groups. Group A is the minority, with a continuum of individuals of mass N_A, and group B has mass $N_B > N_A$. Both groups have equal access to industries and there is no product market discrimination, but the groups are socially segregated and spend their leisure time separately. Social interactions are random within ethnic groups, such that each person interacts with a representative sample of individuals in their own group.

We analyze how these two ethnic groups sort across two industries. Industry 1 has a production structure where self-employed entrepreneurs obtain advantages through social interactions with other self-employed entrepreneurs in the same industry. When socializing during family gatherings and religious/cultural functions, entrepreneurs in this industry can mentor each other and share industry knowledge and professional advice. The more an entrepreneur socializes with other entrepreneurs, the more knowledge is exchanged. Industry 0, by contrast, exhibits constant returns to scale with worker productivity normalized to one. This industry can be equally comprised of individuals working in self-employment or in larger firms; the core assumption is that private social interactions do not have the same benefit in industry 0 as they do in industry 1.

More formally, define X_l for $l \in \{A, B\}$ as the fraction of the population in group l who are self-employed entrepreneurs in industry 1. Since social interaction is random within groups, a fraction X_l of the friends and family members of every individual in group l are also self-employed entrepreneurs in industry 1. For industry 1, denote individual entrepreneurial productivity in group l as $\theta(X_l)$. Our assumption that productivity increases when socializing with other entrepreneurs in industry 1 is formally stated as:

Assumption 1a *Entrepreneurial productivity in industry 1 increases in specialization:* $\theta' > 0$.

Denote aggregate output of industry 1 as Q_1, which is a function of the distribution (X_A, X_B):

$$Q_1(X_A, X_B) = X_A N_A \theta(X_A) + X_B N_B \theta(X_B). \tag{1}$$

Since social interaction plays no role for industry 0, its aggregate output is simply:

$$Q_0(X_A, X_B) = (1 - X_A) N_A + (1 - X_B) N_B. \tag{2}$$

Demands for the two industries need to be complementary enough to avoid the complications of multiple optima possibly generated by non-convexities. We simply assume them to be perfect complements via a Leontief utility function for consumers:

$$U(q_0, q_1) = \min\left(q_0, \frac{q_1}{v}\right), \tag{3}$$

where $v > 0$ is a preference parameter and q_0 and q_1 are individual consumption of each industry's output, respectively.

2.2 The Pareto Problem

We now describe the efficient outcome. Since the outputs of both industries have unitary income elasticities, distributional aspects can be ignored when characterizing the efficient outcome. The problem simplifies to choosing an industry distribution (X_A, X_B) that maximizes a representative utility function $U(Q_0(X_A, X_B), Q_1(X_A, X_B))$. A marginal analysis is inappropriate since this is a non-convex optimization problem. We consider instead the most specialized industry distributions, where as many individuals as possible from a single group A or B are self-employed entrepreneurs in industry 1.

Figure 1 depicts the production possibilities for the two specialized distributions. Define $V(X_A, X_B) \equiv Q_1/Q_0$ as the ratio of industry outputs under the distribution (X_A, X_B). Along the curve with the kink $V(1,0)$ in the figure, group A specializes as self-employed entrepreneurs in industry 1. Starting from a position on the far right where everyone works in industry 0, members of group A are added to the set of self-employed entrepreneurs in industry 1 as we move leftward along the x-axis. When the kink at $V(1,0)$ is reached, all members of group A are self-employed entrepreneurs in industry 1. Thereafter, continuing leftward, members of group B are also added to industry 1 until $Q_0 = 0$. Similarly, along the curve with the kink $V(0,1)$, group B first specializes as self-employed entrepreneurs in industry 1. Members of group B are added moving leftward along the x-axis until the kink at $V(0,1)$, where all Bs are working in industry 1. Thereafter members of group A are also added until $Q_0 = 0$.

The curve with minority specialization is above the curve with majority specialization, so long as the need for self-employed entrepreneurs in industry 1 is sufficiently small. A large fraction of As are self-employed entrepreneurs in industry 1 when the minority specializes, allowing minority entrepreneurs to socialize mostly with other entrepreneurs in industry 1, improving productivity. The same is not true for the majority, since even if a large fraction of self-employed entrepreneurs in industry 1 are Bs, most Bs are nevertheless employed in industry 0.

The argument can be generalized to show that minority specialization is Pareto efficient so long as industry 1 is small enough. Perfect complementarity simplifies the problem of solving for the optimal allocation, since any bundle where industrial outputs are in the exact ratio v of the Leontief preferences (3) is strictly preferable to all other bundles that do not include at least as much of each industry. The Pareto optimal distribution (X_A, X_B) must therefore satisfy $v = V(X_A, X_B)$. Define the total number of entrepreneurs in the population as $M \equiv X_A N_A + X_B N_B$. It follows that:

Proposition 1 *If $v \leq V(1,0)$, all self-employed entrepreneurs in industry 1 belong to minority group A.*

Proof: Take the distribution $(X_A, 0)$ where X_A is such that $v = V(X_A, 0)$. This is feasible since $v \leq V(1,0)$. Assume by contradiction that it is not the uniquely efficient distribution. Then there exists an alternative distribution (X'_A, X'_B) with $Q'_1 \geq Q_1$ and $Q'_0 \geq Q_0$. Given $Q'_0 \geq Q_0$, it follows that $M' \leq M$, or equivalently, $X'_A N_A + X'_B N_B \leq X_A N_A$, which implies $X'_A \leq X_A$ and $X'_B < X_A$, with $X'_A < X_A$ if $X'_B = 0$. Manipulating the expression for Q'_1:

$$\begin{aligned} Q'_1 &= (M' - X'_B N_B)\,\theta(X'_A) + X'_B N_B \theta(X'_B) \\ &< (M - X'_B N_B)\,\theta(X_A) + X'_B N_B \theta(X_A) = Q_1 \end{aligned}$$

This contradicts $Q'_1 \geq Q_1$.■

The efficient outcome requires that a single group specializes as self-employed entrepreneurs in industry 1, and importantly, which group specializes is not arbitrary. Minority specialization is more efficient since the minority's social isolation enables entrepreneurs in A to socialize mostly with other entrepreneurs in their small isolated group. For $v \leq V(1,0)$, the transformation curve and the curve with minority specialization in Figure 1 coincide.[8] Group A has absolute and comparative advantages as self-employed entrepreneurs in industry 1. If the demand for industry 1 is sufficiently great, however, then the minority is too small to satisfy demand by themselves. In the special case when $v = V(0,1)$, the demand for industry 1 is great enough for group B

[8] While our model does not depict competition or crowding-out among co-ethnic entrepreneurs, the size of the industry is governed by consumer tastes and the v parameter. Thus, a large ethnic group will not be able to specialize completely in a small sector.

to specialize completely. In this case minority involvement would dilute the majority's productivity advantage, and the Pareto efficient solution is for Bs to specialize in being self-employed entrepreneurs in industry 1.

Corollary *If* $v = V(0, 1)$*, all self-employed entrepreneurs in industry* 1 *belong to the majority,* B*.*

Thus, the relationship between group size and productivity is not monotonic, and the group with the absolute advantage is the group with a population size that most closely adheres to the size of industry 1. Other production possibilities generated by more unspecialized distributions, such as $X_A = X_B$, are not displayed in Figure 1. Our online theoretical appendix proves that a convex production function in social interactions ($\theta'' > 0$) is sufficient to ensure that at least one group specializes, in which case the efficient frontier is the outer envelope of the curves shown in Figure 1. Consequently, above a certain value of v, there is a discrete jump from minority specialization to majority specialization.

2.3 Model Discussion

This simple model provides a stark economic environment for considering how isolated social interactions impact the sorting of ethnic groups over industries. While our model considers only two industries, this simplification is not as limiting as it may first appear. The model captures a setting where a small industry of self-employed entrepreneurs can benefit through non-work interactions. Allowing the baseline industry 0 to be an aggregate of many constant-returns-to-scale industries would still lead to the efficient solution being for the small ethnic group to specialize in being the self-employed entrepreneurs if their group size matches the demand preferences for industry 1. In fact, framed this way, the baseline industry 0 would be expected to be quite large to any one industry, making it more likely that the minority group should specialize.

Another obvious simplification is that we only have two ethnic groups. Yet, a complex model allowing for several small industries and also several minority ethnic groups would lead to the same conclusions. For example, consider an economy with industries $1a$ and $1b$ that have equal demand and display the same productivity benefit for social interaction. Also allow there to be two minority groups of equal size. If the demands for industries $1a$ and $1b$ are sufficiently small, then the efficient outcome is for one minority group to specialize in being self-employed entrepreneurs in $1a$, and for the other minority group to specialize in $1b$. Which minority group specializes in which sector is arbitrary. In this multi-sector economy with sector-specific skills, otherwise-similar groups consequently specialize in different business sectors. Pushing further, if the economy has several small industries of varying sizes that benefit from these social interactions, and multiple minority ethnic groups, the efficient outcome will be characterized by minority groups specializing in specific self-employment industries as much as possible.

Our online theoretical appendix also provides several formal extensions to the model. We analyze competitive market outcomes and dynamics and show that initial conditions matter. Social interaction will reinforce early concentrations by attracting members of some groups and pushing out others. We also demonstrate that a small group size is inherently more likely to result in high initial concentrations in one or more industries that can then become reinforced and propagate. This reinforcing mechanism and the growing stratification over time are important features, as many referral models instead show decay over time due to imperfect transfer and a lack of a sustained earnings advantage.

An additional extension considers individual heterogeneity in ability and earnings and predicts that an ethnic group can achieve greater earnings at the group-level when specializing. The prediction becomes more complicated for entrepreneurs vs. wage workers within groups as it depends upon how high- vs. low-skilled members of the ethnic group are attracted by the gains from social interaction. The empirical work of Patel and Vella (2013) show a positive earning relationship for immigrant groups and common group occupational choices, and we note below some complementary evidence from our own data. This earnings premium provides evidence that the choice to engage in self-employment and specialize is due to more than just discrimination against minority groups (which could still nonetheless play a role), and it helps distinguish the theory from being just about referral networks for opportunities.[9]

A final extension looks at endogenous interactions. While our simple model takes social ties as given, in the extension we look at endogenous social interaction and show how a social network is formed through matching in a marriage market where social traits are diverse. We explore the potential for splinter groups to break out of the majority group in order to benefit from the increasing returns to social interaction in our model. Drawing on results from graph theory, we show that there are no such splinter groups in a first-best matching on social traits only. This demonstrates that there would be costs in terms of deteriorated matching quality if the majority were to duplicate the social structure of an (exogenously) isolated ethnic minority. Ethnicity consequently matters and can confer a productive advantage for self-employment even when interaction is endogenous.

[9] The favorable economic outcome does not necessarily carry over to utility, and we later discuss further the process of assimilation. Related work includes Chiswick (1978), Borjas (1987), Simon and Warner (1992), Rauch (2001), Mandorff (2007), Bayer et al. (2008), Beaman (2012), and Cadena et al. (2015).

3 Analysis of US Entrepreneurial Stratification

3.1 US Census of Populations Data

We analyze the 2000 Census of Populations using the Integrated Public Use Microdata Series (IPUMS). We focus on the 5% sample, and we use person weights to create population-level estimates. In a panel exercise, we also use the 5% samples from 1980 and 1990 and the five-year American Community Survey (ACS) samples for 2006-2010 and 2014-2018. We will refer short-hand to the latter two datasets as the 2010 ACS and 2018 ACS, respectively. In addition, we build instruments from 1991 information on the United Kingdom and 2011 information on Spain from IPUMS-International.

We define ethnic groups using birthplace locations and, in a few cases, language spoken. We merge some related birthplace locations (e.g., combining England, Scotland, Wales, and non-specific UK designations into a single group). We also utilize the detailed language variable to separate Gujarati and Punjabi Indians and to identify Armenians and Chaldeans given their prominence. Our preparation yields 131 ethnic groups from 198 initial birthplace locations. Appendix Table 1a and 1b lists all ethnic groups and provides descriptive statistics on them.[10]

We assign industry classification and self-employment status through the industry and class-of-work variables. IPUMS uses a three-digit industry classification to categorize work setting and economic sector of employment. Industry is distinct from an individual's technical function or occupation, and those operating in multiple industries are assigned to the industry of greatest income or amount of time spent. The class-of-work variable identifies self-employed and wage workers,[11] and we examine

[10] A few ethnic groups represent categories not specified or elsewhere classified (e.g., "South America, ns"). We retain these for completeness, and our results are robust to excluding them.

[11] In the IPUMS data, self-employment is assigned when it is the main activity of an individual (e.g., not capturing academics who consulting part-time). The definition includes both owners of employer

{industry, class of work} pairings. For example, a self-employed hotelier is classified differently than a wage earner in the hotel industry. The sample excludes those whose self-employment status is unknown and industries without self-employment.[12]

Our core sample focuses on males who are 22-70 years old and not living in group quarters; for immigrants, we require that they have migrated to the United States at age 16 or older. Our final sample for 2000 contains 2.9 million observations, representing 59 million people. Of these, 0.26 million observations, representing 5.7 million people, are immigrants.

3.2 Clustering in Entrepreneurial Activities

We design "overage" ratios to quantify for an ethnic group the heightened rate of self-employment it displays for a particular industry and also across the full range of industries. Our primary metrics focus on the specialization evident among self-employed individuals only, while robustness checks build samples combining wage earners and self-employed.[13]

We first define $OVER_{lk}$ as the ratio of an ethnic group l's concentration in an

firms and sole proprietors.

[12]We utilize the 1990 IPUMS industry delineations for temporal consistency. Examples of excluded industries include the military, postal service, labor unions, religious and membership organizations, and public administration. Our final sample includes 126 industries, where we have aggregated some very small industries (principally in manufacturing) to ensure consistency over the 1980-2018 period. We are cautious to not rely on very aggressive definitions of industry boundaries, even if this leads us to underestimate some concentration. For example, Greek restaurateurs will sort into Greek restaurants and Chinese restaurateurs into Chinese restaurants, independent of social relationships, but we consider the restaurant industry as a whole to avoid taste-based factors or ethnic-specific skills. Similarly, we mostly look at industries on a national basis, even though additional clustering happens locally for some industries (e.g., taxi cabs). We use this uniform approach to be consistent over industries, vs. for example defining the motel industry in a different way from taxi cabs, and because ethnic connections can provide long-distance knowledge access (e.g., Rauch, 2001; Agrawal et al., 2008). An extension later in the paper considers variation over metropolitan areas.

[13]It may seem appealing to use wage earners instead as a counterfactual to self-employed workers. This approach is not useful, however, as ethnic entrepreneurs show a greater tendency to hire members of their own ethnic groups into their firms (e.g., Andersson et al., 2014a,b; Åslund et al., 2014; Kerr et al., 2015).

industry k to the industry k's national employment share. Thus, if an ethnic group l has N_l total workers and N_l^k workers in industry k, then $X_l^k = N_l^k/N_l$ and $OVER_{lk} = X_l^k/X^k$. This baseline metric measures the over- or under-representation of the ethnic group for a specific industry, and by definition both cases exist for an ethnic group across the full range of industries.

To aggregate these industry-level values into an overall measure of industry concentration for an ethnic group, our primary metric takes a weighted average using the share of the group's self-employment by industry as the weight:

$$OVER1_l = \sum_{k=1}^{K} OVER_{lk} X_l^k. \tag{4}$$

Intuitively, the metric is similar to a Herfindahl-Hirschman Index with an underlying adjustment for different industry sizes. Our estimations ultimately transform $OVER1$ to have unit standard deviation for interpretation. We also test the following variants:

1. Weighted average over the three largest industries for ethnic group l: $OVER2_l = \sum_{k'=1}^{3} OVER_{lk'} X_l^{k'} / \sum_{k'=1}^{3} X_l^{k'}$, where $k' = k$ such that $\sum_{k'=1}^{3} N_l^{k'}$ is maximized.

2. Weighted average over the three largest industry-level overages for ethnic group l: $OVER3_l = \sum_{k'=1}^{3} OVER_{lk'} X_l^{k'} / \sum_{k'=1}^{3} X_l^{k'}$, where $k' = k$ such that $\sum_{k'=1}^{3} OVER_{lk'}$ is maximized.

3. Maximum overage: $OVER4_l = \max_l [OVER_{lk}]$.

We investigate our entrepreneurial concentration hypotheses over the 131 ethnic groups using the metrics. $OVER1_l$ takes the weighted sum across industries, while $OVER2_l$ considers the three largest industries for an ethnic group. In most cases, $OVER2_l$ is bigger than $OVER1_l$ as concentration is often linked to substantial numerical representation; some exceptions happen when an ethnic group is focused on bigger industries. These calculations measure extreme values, and we need to be careful about small sample size, especially for $OVER3_l$ and $OVER4_l$ given their emphasis on outliers. We will thus focus mostly on $OVER1_l$ and also conduct Monte Carlo simulations of expected overage described later. We will also show the results are robust dropping ethnic-industry pairs with very limited observations.[14]

Figure 2 displays ethnic groups with the highest and lowest $OVER1_l$ metrics. There is substantial entrepreneurial clustering, with immigrants from Nepal (40.7), Senegal (37.0), Zimbabwe (36.5), and Yemen (36.3) displaying the overall highest industrial concentration for entrepreneurship. The national average for ethnic groups is 8.4, and lowest concentration rates are for immigrants from Poland (1.6), Germany (1.6), Canada (1.6), and Cuba (1.4). Appendix Tables 1a and 1b give a detailed list of overage ratios for each ethnic group and the industries with the largest overage ratio. In most cases, the industry where the ethnic group displays the highest concentration for self-employment is the same as the industry where the ethnic group shows the highest concentration for total employment. Appendix Tables 2a and 2b document the strong correlations among the overage metrics.

3.3 Social Isolation and In-Marriage Rates

We measure social isolation and concentrated group interactions through within-group marriage rates for child arrivals to the United States evident among ethnicities. This metric is a strong proxy if sorting in the marriage market is similar to sorting in other social relationships.[15] High marriage rates within an ethnic group, also termed

[14] Our NBER working paper focuses on 77 groups that have a minimum of 10 observations in at least one industry.

[15] Using the General Social Survey, Mandorff (2007) shows that in-marriage among religious groups within the United States (e.g., Catholic, Jewish, etc.) is tightly connected with high shares of close friendships being of the same religious group as the respondent.

in-marriage, suggest greater social isolation and stratification. Significant levels of in-marriage are often present in minority groups and along religious lines, with members of the ethnic group devoting more energy towards interacting with coethnics and ultimately transmitting the group's traits to future generations (e.g., Bisin and Verdier, 2000; Bisin, Topa, and Verdier, 2004). Such choices can come at the expense of better access to the formal labor market that can come through inter-marriage with natives (Furtado, 2010).[16]

We calculate in-marriage rates for ethnicities using a second dataset developed from IPUMS. We focus on women and men who immigrated to the United States when 0-15 years old and who are 22-70 years old at the time of the Census. Importantly, this sample is mutually exclusive from the earlier sample used to calculate our overage metrics, where we consider men who migrated at age 16 or older. By focusing on children at the time of migration, we also circumvent the joint migration of married couples to the United States.

Most immigrant groups are socially segregated with respect to marriage, some very strongly so. With random matching for marriage and equal male and female migration, in-marriage rates would roughly equal a group's fraction of the overall population. Group in-marriage rates (also shown in Appendix Table 1a) average 48% and often exceed 80%. Pairwise correlations of 0.31 and 0.45 exist for in-marriage rates and the $OVER1_l$ and $OVER2_l$ metrics, respectively. We later introduce some alternative metrics for social isolation.

[16] Classics include Kennedy (1944) and Herberg (1955), and Furtado and Trejo (2013) provide an extended review. Furtado and Theodoropoulos (2011) consider shifts in likelihood of inter-marriage by when someone migrates to the United States.

3.4 OLS Empirical Results

To quantify whether smaller and more-socially isolated ethnic groups have greater industrial concentration for entrepreneurship, we use the following regression approach:

$$OVER1_l = \alpha + \beta_1 SIZE_l + \beta_2 ISOL_l + \varepsilon_l, \tag{5}$$

where $SIZE_l$ is the negative of the log value of group size and $ISOL_l$ is the in-marriage rate of the group. We take the negative of group size so that our theoretical prediction is that β_1 and β_2 are positive. We report all coefficients in unit standard deviation terms for ease of interpretation. Our baseline regressions winsorize variables at their 1% and 99% levels to guard against outliers, weight estimations by log ethnic employment for each group, and report robust standard errors. +++, ++, and + indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Column 1 of Panel A in Table 1 measures that a one standard-deviation decrease in group size is correlated with a 0.58 standard-deviation increase in average entrepreneurial concentration across all industries. Similarly, a one standard-deviation increase in the in-marriage rate is correlated with a 0.33 standard-deviation increase in overage. Panel B introduces controls for the traits of the ethnic group in 2000: share who are 36-55 years old, share who are 55-70 years old (reference group is aged 22-35), share who are married, share who speak English well, share who have some college education, and share who have a college degree or higher (reference group is high school or less). The coefficients are more equal at 0.47 and 0.45, respectively, in the presence of these controls.

The next columns consider robustness checks on our metric design. Column 2 considers the metric that uses all employed workers for the ethnic group. Column 3 compares industry-level overages only to rates of other immigrant groups by excluding natives from the calculations of industry sizes. Column 4 drops ethnic-industry settings where fewer than three observation counts exist. Column 5 excludes new arrivals to America during the prior five years as some forms of employer-based migration are tied to specific jobs. Column 6 excludes the taxi cab industry, which is a frequent industry of maximum overage. The coefficients are stable across these variations.

Table 2 continues with additional robustness checks on the $OVER1_l$ outcomes. Columns 2 and 3 drop sample weights and winsorization steps, respectively, Column 4 introduces fixed effects for each origin continent, Column 5 uses a median regression format, and Column 6 bootstraps standard errors. Columns 5 and 6 should be compared to Column 2 given their unweighted nature. Column 7 adds an additional control to capture any mechanical relationship between ethnic group size and entrepreneurial overage. We conduct for each ethnic group 100 Monte Carlo simulations using the same count of self-employed as observed for the group but randomly picking the industry in accordance with the aggregate US distribution for self-employment. From these simulations, we calculate for each ethnic group the average expected overage. Introducing these controls does not significantly impact our estimations except that the size relationship diminishes modestly.[17]

Table 3 shows our other forms of the overage metric. Column 2 shows that a focus on the three largest industries for an ethnic group (i.e., $OVER2_l$ discussed above) increases the relative importance of social isolation for predicting overages. Columns 3 and 4 examine extreme values using the $OVER3_l$ and $OVER4_l$ metrics defined above. The estimates remain statistically significant and now show a smaller connection to group isolation relative to group size.[18]

[17] Considered as a distribution, 90.1% of ethnic groups have a realized overage that exceeds the median value of their simulations, and 34.4% have a realized value greater than the 95th percentile.

[18] We obtain similar results when modifying of our overage measures with industry-level propensities

Appendix Tables 3a and 3b further test the relationships of relative size and isolation on entrepreneurial clustering by using non-parametric regressions. We partition our size and isolation variables into terciles and create indicator variables for each combination of {smallest size, medium, largest size} and {most isolated, medium, least isolated}, and assign ethnic groups that fall into [largest size, least isolated] as the reference category.

The results continue to support the theory, as depicted in Figure 3. The [smallest size, most isolated] groups have entrepreneurial concentrations that are 1.8 standard deviations greater than the [largest size, least isolated] groups. Equally important, the pattern of coefficients across the other indicator variables shows the relationships are quite regular and not due to a few outliers. For example, holding the ethnic group size constant, higher levels of social isolation strongly and significantly correspond to larger overages. Flipping it around and holding social isolation constant, smaller group sizes mostly promote greater concentration within each isolation category.

3.5 Panel Data Models and Assimilation

We next consider panel estimations to remove time-invariant features of the data. Some ethnic groups may face persistent discrimination that contributes to both social isolation and entrepreneurial specialization. This could be particularly true for non-white immigrants, who feature prominently in Figure 2. Our cross-sectional results could also be overly dependent on a single wave of migration to the United States, possibly to fill short-term needs around the year 2000, and thus be incomplete for the longer-term dynamics we hope to capture. Showing similar results with a different source of identifying variation provides greater confidence in our estimations, and we

for being an employer firm vis-à-vis sole proprietors using data from the Survey of Business Owners.

can use panel models to also study the process of immigrant assimilation and the persistence of entrepreneurial specialization.

Table 4 extends our work to a panel model covering 107 ethnic groups over the five time periods of 1980, 1990, 2000, 2010, and 2018. The 24 excluded groups lack information for one or more years because of changes in the birthplaces recorded in IPUMS. Preparation steps are consistent across the time periods, and the controls for ethnic groups' traits are time varying as well in Panel B. We cluster standard errors by ethnic group.

Column 1 finds a longitudinal size relationship that is much stronger than that observed with the 2000 cross-section, while the group isolation is comparable in economic magnitude. Column 2 adds the control for expected overage based upon Monte Carlo simulations with ethnic observation counts in each year. With this control, the results look even more like those measured in the cross-section. Column 3 adds a linear time trend interacted with the 1980 level of overage as an alternative control strategy. Overall, the panel data model is quite consistent with the results present in the 2000 Census.

The process of assimilation of new arrivals receives great attention in the immigrant literature. Our model of entrepreneurial specialization does not undertake a detailed treatment of the issue and how later generations can be affected. It would be feasible, for example, for entrepreneurial specialization to weaken assimilation, being statically efficient and dynamically inefficient by creating "cul-de-sacs" of entrepreneurial specialization that limit further assimilation (e.g., Andersson Joona and Wadensjö, 2009). Furtado and Song (2005) also speak to the growing wage premiums connected to marrying a US native since 1980. On the other hand, greater earnings with entrepreneurial specialization can be a route for new immigrants to afford better educations and future career opportunities for their children.

The results in Table 4 shed some light on this issue. First, the panel coefficient for social isolation is very similar to the cross-section. This suggests that continued assimilation of an ethnic group into the United States as measured by reduced in-marriage rates would be connected to continued declines in entrepreneurial clustering. That said, the data suggest that this is not happening for many ethnic groups. From the 1980 and 1990 Censuses to the 2010 and 2018 ACS, the measured in-marriage rates among child arrivals to the country increased on average by eight percent points. Indeed, it may be difficult to find same-origin partners in small groups, leading to in-marriage rates increasing as the group grows in size.

Additionally, the unreported age controls for the group in Panel B capture the aging of the migrants in the United States. Conditional on in-marriage rate adjustments, aging as captured by these controls does not connect very strongly to lower entrepreneurial clustering. This is similarly true when considering changes over decades in the share of the ethnic group that has been in the United States for longer than 15 years. Future research with data that combine the records of parents and children can further investigate the assimilation outcomes and long-term consequence of entrepreneurial clustering by first-generation immigrants.

We next consider two complements to the panel model. We have established a tight empirical relationship of the in-marriage rate to ethnic entrepreneurial specialization, but we should consider other measures of social isolation. We undertake this comparison next to better ground the use of the in-marriage rate and learn more about other types of social distance between groups. We then test for reverse causality concerns: for example, that growing entrepreneurial specialization leads to more in-marriage among the ethnic group. For this, we use IV models that exploit sources of variation outside of the United States.

3.6 Additional Measures of Social Isolation

Table 5 considers additional measures of social isolation. We first measure the residential segregation of the ethnic group. Ethnic enclaves can be important early homes for new arrivals, with links to social isolation like those we measured via in-marriage rates. While residential segregation could generate self-employment activity to satisfy local consumer demand of the ethnic group, extensive specialization of entrepreneurial activity would require serving customers from other ethnic groups. Many common industries of entrepreneurial specialization, such as taxi drivers, construction and building trades, and landscape services, could be well aligned with self-employed members traveling to other local areas to serve customers.

Our data here are limited to exploiting the Public Use Micro Areas (PUMA) of residence within metropolitan areas captured by the 2000 Census. We only consider metro areas with more than one PUMA, and we calculate residential segregation for an ethnic group relative to 100 randomized counterfactuals that considered if an equivalent number of Census observations were drawn at random in proportion to local population from PUMAs in the same metropolitan areas where the ethnic group resides. Transformed to have unit standard deviation for comparability, residential segregation is also a strong predictor for entrepreneurial clustering in Column 2 and with an economic magnitude comparable to the in-marriage rate.

Columns 3-5 alternatively take data from Spolaore and Wacziarg (2016) on the genetic, linguistic, and religious distance of countries to each other. We applied these country-based distances to our setting by measuring a weighted average for an ethnic group from the ethnic composition of the United States as measured by country of birth for US residents. Metrics are again expressed in unit standard deviations. Regressions cluster standard errors by 120 unique observations from Spolaore and Wacziarg (2016) that we map to our sample. While we can map measures of genetic distance for our full sample, linguistic and religious distances are only available for 113 groups (112 overlapping).

Without controls for ethnic group traits, genetic and religious distance most closely connect to entrepreneurial clustering, while linguistic and religious distance are strongest in the presence of the controls. When combining all of our measures together in Column 6, in-marriage rates stand out, with genetic distances also being important in Panel A. These results, in combination with their longitudinal consistency in Table 4, suggest that our measure of social isolation via in-marriage rates captures a salient part of the group's social dynamics that is not just due to residential segregation, linguistic isolation, or an even more fixed component like genetic distance.

3.7 IV Empirical Tests

We next consider IV specifications to test against reverse causality concerns (e.g., where isolated business ownerships lead to greater social isolation or lower group sizes) or omitted variables. Some omitted factors could center on sector-specific skills gained by ethnic groups abroad that are then ported to the United States with migration (especially if booming local demand for an ethnic group's services leads them to draw more migrants with similar skills from their home country over). Others could be due to local traits, such as state-level adoption of stringent employment verification procedures (e.g., Amuedo-Dorantes and Bansak, 2012; Orrenius and Zavodny, 2016) leading to more social and workplace isolation.

Our primary IV approach uses as instruments the predicted ethnic group size from a gravity model and in-marriage rates from the United Kingdom in 1991. To instrument for ethnic group size, we use a gravity model to quantify predicted ethnic size based upon worldwide migration rates to the United States. The original application of gravity models was to trade flows, where studies showed that countries closer to each other and with larger size tended to show greater trade flows, similar to the forces of planetary pull. This concept has also been applied to the migration literature, and we similarly model

$$SIZE_l = \alpha + \beta_1 DIST_l + \beta_2 POP_l + \varepsilon_l, \tag{6}$$

where $DIST_l$ is the log distance to the United States from the origin country and POP_l is the log population of the origin country. For this purpose, we estimate log ethnic group size in the United States as the dependent variable (without a negative value being taken as in earlier estimations). Unsurprisingly, lower distance ($\beta_1 = -1.43$ (s.e.=0.24)) and greater population ($\beta_2 = 0.42$ (s.e.=0.05)) are strong predictors of ethnic group size in the United States. We take the predicted values from this regression for each ethnic group as our first instrument.

For our instrument of in-marriage rates in the United States, we calculate the in-marriage rates in the 1991 UK Census of Populations. This approach is attractive as the social isolation evident in the United Kingdom a decade before our study is most likely to be predictive of US self-employment rates to the extent that the British isolation captures a persistent trait of the ethnic group. The instrument is not completely foolproof (e.g., a third factor like specialized ethnic-specific skills could be present in the diaspora in both countries and lead to similar outcomes), but the instrument does provide assurance against some of the most worrisome endogeneity arising in local areas. A limitation of this instrument is that we are only able to calculate it for 34 broader ethnic divisions. We map our observations to these groups and cluster the standard errors at the UK group level.

The first-stage results with this instrument set are quite strong. The first two columns of Table 6 show that these instruments have very strong individual predictive power with and without the ethnic group controls. The second-stage results in Column 3 are similar to the OLS findings. The IV specifications in Panel A suggest that a one standard-deviation decrease in ethnic group size increases overage by 0.46 standard deviations. A one standard-deviation increase in isolation leads to a 0.32 standard-deviation increase in entrepreneurial concentration. These results are well-measured and economically important. The results are close enough to the OLS findings that we cannot reject the null hypothesis in Wu-Hausman tests that the instrumented regressors are exogenous. These IV results strengthen the predictions of our theory that smaller, more isolated groups are more conducive to entrepreneurial clustering.

Ideally, we would be able to build a broader instrument that used in-marriage rates from many countries for an ethnic group. This would help counteract any persistent bias due to similarities for immigrant experiences in the UK and US economies, and it would overcome measurement error in the instruments. Unfortunately, the data requirements for our in-marriage rate calculation are steep, especially for knowing detailed countries of birth of spouses within a household, and the only additional source we could identify from IPUMS International is Spain 2011. These data have 60 ethnic origin groups that we can map to the US data.

In Columns 4-6, we use average in-marriage rate for an ethnic group from the UK 1991 and Spain 2011 as instruments for US 2000 in-marriage rates. As anticipated, the results are a bit sharper and, due to the growth of the isolation coefficient in the second stage, we are now more likely to reject that the instrumented regressors are exogenous. We remain cautious of the Spain instrument but take comfort in the overall stability evident in this modification.[19]

Appendix Tables 5a-8b show robustness checks to the instruments. Results are very similar with simple adjustments like excluding sample weights and dropping winsorization. Some results for the social isolation metric have larger standard errors when bootstrapping and including ethnic group controls, which is not too surprising given the smaller number of underlying UK groups. Another weak spot is that the expected overage controls from simulations can crowd out the size instrument in a dual IV as the instrument and predicted overage are being built upon the same data, making it hard to separate them. Beyond these caveats, however, the IV is quite robust overall. We also find very similar results when expanding the gravity equation to have a squared distance term or an indicator for Canada and Mexico as bordering countries or when using underlying components of the gravity equation as direct instruments.[20]

3.8 Extension: Earnings

Our model predicts that members of an ethnic group can achieve greater earnings when entering a common entrepreneurial setting. In our framework, social complementarities produce a positive relationship between earnings and entrepreneurship at the group

[19] Appendix Table 4 shows first- and second-stage outcomes from using the in-marriage rates in Spain as their own instrument. The isolated Spain instrument is weak, especially in the presence of ethnic group controls. This appendix table also shows similar results to those reported in Table 6 when we model the UK and Spain instruments individually in same specification.

[20] Diagnostics that compare the US, UK and Spanish industry distributions for entrepreneurial specialization support the instrument. While in-marriage rates for ethnic groups in both European countries exhibit a strong correlation to those in the United States, their industry distributions show less commonality. When comparing the industries across countries that contain the most self-employed for an ethnic group, the overlap with the United States is 37% and 25% for the United Kingdom and Spain, respectively. This calculation is done with cases where the ethnicity is precisely identified in both data sets, and the overlap is even less when including ethnicities where data require less-precise mappings (e.g., "New Zealand" in the United States data to "Oceania" in the Spanish data). Very rarely is the industry of maximum entrepreneurial specialization the same across countries for an ethnic group. While encouraging, we treat these comparisons cautiously given the many challenges in aligning Census data across countries that were developed with different industry classifications.

level. Evidence for this prediction helps show discrimination is not solely responsible for our findings, and this also helps differentiate our work from job search networks. To the extent that our person-level controls on education and language fluency capture skill levels, we may also anticipate that self-employed individuals earn more.[21] This net relationship must be empirically investigated in the data, and an earnings premium for self-employed workers would provide evidence against the entrepreneurial clustering being due to herding behavior or other forms of inefficient entry.

Patel and Vella (2013) comprehensively show a positive earning relationship for immigrant groups and common group occupational choices using the 1980-2000 Census of Populations data. Table 7 provides complementary pieces of evidence that look at variation within MSA-industry cells and within ethnic groups. As in our prior estimations, the sample includes immigrant males who arrived into the United States after age 16 and are aged 22-70 in 2000. The outcome variable is log annual income.[22] Estimations include fixed effects for the following person-level traits (category counts in parentheses): age (5), age at immigration (5), education (4), and English language fluency (2). Regressions use person weights and cluster standard errors by ethnic group. Explanatory variables are transformed to have unit standard deviation for easy comparison and interpretation.

[21] Without conditioning on skill, our model does not make universal predictions about whether the self-employed or wage earners of an ethnic group earn more overall, as the online theoretical appendix shows this depends upon the skill distribution for an ethnic group. The prediction emerges if one can control for skill levels. Many articles have noted the challenges of measuring skills for immigrants via common metrics like education, as foreign degrees may be under-recognized for example, and so we approach this prediction cautiously.

[22] Evaluation of entrepreneurial earnings is challenging, due to issues like greater income volatility, under-reporting or tax avoidance schemes, and the experimentation value of trying out new ideas (e.g., Manso, 2016; Dillon and Stanton, 2018). We have some instances where the data show zero or negative earnings for self-employed, as well as very low values for wage earners. We bottom code annual earnings at \$1000 before taking the log transformation. We achieve very similar patterns with other earnings floors or simply dropping zero and negative values.

Panel A considers self-employed individuals, and Panel B considers wage workers. The first column simply considers the share of an individual's ethnic group who are self-employed in the industry of the focal worker. There is a positive relationship for both worker types, even conditional on MSA-industry fixed effects. For the self-employed, a one-standard deviation increase in the concentration of the ethnic group for self-employment in the industry is associated with about a seven percent increase in annual earnings. For wage workers, the relationship is measured to be four percent.

Column 2 adds into the estimation the overall share of the ethnic group who are self-employed—which is very predictive of group earnings, per the model—and Column 3 further adds the total ethnic group employment in the focal industry. Columns 4 and 5 add ethnic group fixed effects, which absorb the group's overall rate of self-employment and focus on variation across industries within each ethnic group. Looking across these estimations, there is strong confirmation of the model's prediction that members of an ethnic group can achieve greater earnings through entrepreneurial clustering. The whole group earns more when entrepreneurial activity is higher, and the earnings of the self-employed in an industry show a tight relationship to other members of the ethnic group being self-employed in the same industry space.

Table 7's split-sample approach does not quantify whether self-employed earn more than immigrant wage workers in the same setting as the fixed effects and controls can change values. Appendix Table 9 shows a combined analysis with self-employment interactions and groups traits, thus requiring control variables to have the same values. These estimations confirm that within the same MSA-industry cell and conditional on covariates, self-employed do earn more. Given the challenges for measuring entrepreneurial income noted earlier, this differential is likely also an underestimate.

These results support the model's structure and are consistent with a potential positive benefit from immigrant entrepreneurial concentration. It is important for future theoretical and empirical work to consider both owners and employees of firms. Empirical work can particularly target employer-employee datasets to observe more detailed hiring and wage patterns; such work can also evaluate job transitions during the assimilation of new members of ethnic groups, perhaps ultimately leading to starting their own business.

3.9 Extension: Industry Variation

We conclude our analysis with two extensions that consider industry and metropolitan variations. The Pareto version of our model, presented in Section 2, makes the compelling prediction that ethnic groups should match in terms of size with the industry of self-employment; that is, smaller ethnic groups are a better fit for small self-employment industries, while larger groups should be in larger sectors.[23]

Figure 4 shows descriptive evidence in this regard. We plot for five aggregated groups the cumulative distribution in self-employment as we move from the smallest industries for self-employment, starting with Petroleum and coal products (left-hand side, #1), to the largest industry of Construction (right-hand side, #126). The solid line captures the self-employment distribution of US natives. We parse immigrant ethnicities into four equal-sized groups based upon whether they are above/below the median social isolation and group size.

The figure visually aligns with the model's prediction. All immigrant groups are shifted to the left of the cumulative distribution of US natives, indicating a greater

[23]The assortative size matching prediction is very stark in Section 2's Pareto efficient problem, and a competitive dynamic model yields the generalized prediction that small industries will be matched with small ethnic groups. The strict ordering may not necessarily hold in a competitive dynamic version of the model. For example, an early saturation of the self-employment opportunities in a given industry by an ethnic group may foreclose future entry by a new ethnicity under some forms of the model.

share of self-employment work in smaller sectors. The smallest and most isolated ethnic groups are the most concentrated in smaller industries, followed by the smallest and least isolated ethnic groups. The figure highlights some of the industries (e.g., taxis, grocery stores, physicians, eating and drinking places) where concentration emerges.[24]

Table 8 confirms these patterns with regressions, including adding controls in Panel B for ethnic group traits. Columns 1 and 2 show that smaller and more isolated groups have their self-employment activity concentrated in industries with smaller sizes as measured in terms of self-employed workers only or all workers, respectively. Columns 3 and 4 find similar results when isolating the largest industry of self-employed workers for an ethnicity. Columns 5-8 show these results are not present for wage workers. The wage worker results are an interesting extension beyond our model as they suggest the co-ethnic hiring of immigrants, which has been frequently observed, is not so extensive as to replicate the industry concentration pattern that is experienced for self-employment.

At an aggregated level, we can also use the industries in Figure 4 to provide some calculations broadly consistent with the model's mechanism of interactions. The 2016 Annual Survey of Entrepreneurs (ASE) asked entrepreneurs and small business owners their sources of advice for business. The publicly available ASE data are only available at the two-digit NAICS level, so we compare Accommodation and food services (NAICS 72) to Construction (NAICS 23) and a composite of other industries. Table 9 shows that entrepreneurs and small business owners in Accommodation and food services report the greatest likelihood of collecting advice from customers, family, and friends.

[24] Additional analyses merged O*Net data from Deming (2017) into the occupational structure for industries. As suggested by Figure 4, ethnic self-employment is strongest in settings and roles that have required social and customer connections; it is not connected to settings and roles with routine tasks or those heavy in numbers and reasoning.

Construction has a higher reported reliance on colleagues, while legal and professional advisors feature more strongly among other two-digit NAICS sectors.[25]

3.10 Extension: Metropolitan Variation

We close our study by examining variation in ethnic group size across metropolitan areas. Our theoretical framework is built around a single economy and does not include spatial variation. While some industries of self-employment concentration are spatially distributed by nature (e.g., the concentration in motels by Gujarati Indians), many industries like taxis and landscape services are oriented towards local markets. This localization of service does not necessarily prevent a group from consistently specializing in an industry, as there can be sharing over communities and regional gatherings. Also, Basso and Peri (2020) show that the most recent immigrant arrivals have the highest rates of internal migration across locations within America. Such migration can transport a local specialization to new locations, such as the spreading of Vietnamese nail care salons and Gujarati motels from their points of origin in California in 1975 and 1942, respectively.[26]

To examine whether local group size connects with local entrepreneurial clustering,

[25] While the literature has emphasized this networking dimension, we are not aware of a study that specifically tabulates the differential for entrepreneurs compared to employees (vs. measuring variation among entrepreneurs). Kerr and Kerr (2020b) surveyed 1,334 entrepreneurs and employees working in four co-working centers owned and operated by CIC. Across six surveyed factors (e.g., business operations, venture financing, technology, suppliers, people to recruit, and customers), entrepreneurs averaged a 25% higher likelihood of giving or receiving advice. The positive differential for entrepreneurs remained and was statistically signficant when including fixed effects for firms. While the difference to employees was present in all categories, it was strongest for venture financing, suppliers, and customers.

[26] When examining MSAs in IPUMS where adult-arrival migrants of an ethnic group appear in one Census after 1980 and are not present in the prior decade for the MSA, about 45% of the adult-arrivals have migrated to the United States over the prior ten years. In cases where 10+ adult-arrivals are present for the first time in an MSA, this share is 57%. While caution should be exercised given the population sampling in IPUMS, these statistics suggest that an important share of MSA entry comes from internal migration within the United States of an ethnic group.

Table 10 presents regressions with group size measured at the metropolitan level.[27] We include metropolitan fixed effects to control for the overall scale of local activity, and we control for the in-marriage rate measured nationally.[28] Column 1 provides the estimation with size by itself, while Columns 2 and 3 add the expected overage based upon Monte Carlo simulations for the local ethnic group observation count. Column 3 further adds ethnic group fixed effects. Across these specifications, there is again very consistent evidence that smaller ethnic group size is connected to greater entrepreneurial clustering. We hope that future research can develop frameworks to jointly quantify industry and geographic spans for entrepreneurial concentration of ethnic groups and their dynamics.

4 Conclusions

A striking feature of entrepreneurship is the degree to which immigrants of different ethnic backgrounds cluster into self-employment in different industries. These concentrations are sufficiently visible to be captured in popular culture (e.g., the Indian immigrant entrepreneur Apu who runs the convenience store in *The Simpsons*), and the cumulative magnitudes can be shocking: the Asian American Hotel Owners Association claims to be the largest hotel owners association in the world and represent half of the hotels in the United States. Yet, while noticeable, the economic implications of these tendencies are underexplored.

Our model outlines how the social interactions of small, socially isolated groups can give rise to this self-employment pattern by reducing the cost of acquiring sector-

[27]We drop rural areas from this analysis. Faggio and Silva (2014) analyze differences in self-employment alignment to entrepreneurship in urban and rural areas.

[28]We do not measure in-marriage locally because many ethnic groups have events (e.g., national camps, regional balls) that are intended to encourage in-marriage. At a more mundane level, we also do not observe where a couple was married.

specific skills. Our online appendix explores several extensions to the basic framework, and many other avenues for future research exist. A fruitful path would be to model the intergenerational transmission of skills and to follow occupational structure and entrepreneurial persistence across generations. This interaction mechanism can also be applied to the study of the transmission of other types of skills beyond entrepreneurship.

Empirically, the Census data confirm small and socially isolated immigrant groups in the United States display heightened entrepreneurial clustering. Further quantifying these forces in employer-employee data and firm operating data are important to understand hiring patterns, career trajectories, and market power. The recent US patterns resemble many earlier observations of the economic success and social isolation of specialized minority groups throughout history. We hope this study can be replicated in settings outside of the United States given its general nature (Fairlie et al., 2010).

References

[1] Agrawal, Ajay, Devesh Kapur, and John McHale. 2008. How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics* 64: 258-269.

[2] Amuedo-Dorantes, Catalina, and Cynthia Bansak. 2012. The labor market impacts of mandated employment verification systems. *American Economic Review* 102(3): 543-548.

[3] Andersson, Fredrik, Monica Garcia-Perez, John Haltiwanger, Kristin McCue, and Seth Sanders. 2014. Workplace concentration of immigrants. *Demography* 51(6): 2281-2306.

[4] Andersson, Fredrik, Simon Burgess, and Julia Lane. 2014. Do as the neighbors do: The impact of social networks on immigrant employment. *Journal of Labor Research*.

[5] Andersson Joona, Pernilla, and Eskil Wadensjö. 2009. Being employed by a co-national: A cul-de-sac or a short cut to the main road of the labour market? *Journal of International Migration and Integration* 13(1): 99-120.

[6] Arzaghi, Mohammad, and J. Vernon Henderson. 2008. Networking off Madison Avenue. *Review of Economic Studies* 75(4): 1011-1038.

[7] Åslund, Olof, Lena Hensvik, and Oskar Skans. 2014. Seeking similarity: How immigrants and natives manage in the labor market. *Journal of Labor Economics* 32(3): 405-442.

[8] Åstebro, Thomas, Holger Herz, Ramana Nanda, and Roberto Weber. 2014. Seeking the roots of entrepreneurship: Insights from behavioral economics. *Journal of Economic Perspectives* 28: 49-70.

[9] Basso, Gaetano, and Giovanni Peri. 2020. Internal mobility: The greater responsiveness of foreign-born to economic conditions. *Journal of Economic Perspectives* 34(3): 77-98.

[10] Bayer, Patrick, Stephen Ross, and Giorgio Topa. 2008. Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy* 116: 1150-1180.

[11] Beaman, Lori. 2012. Social networks and the dynamics of labor market outcomes: Evidence from refugees resettled in the US. *Review of Economic Studies* 79: 128-161.

[12] Becker, Gary. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.

[13] Bennett, Victor, and Aaron Chatterji. 2020. The entrepreneurial process: Evidence from a nationally representative survey. *Strategic Management Journal*, forthcoming.

[14] Bhattacharjee, Yudhijit. 2018. https://www.nationalgeographic.com/culture/2018/09/south-asia-america-motels-immigration/.

[15] Birley, Sue. 1985. The role of networks in the entrepreneurial process. *Journal of Business Venturing* 1(1): 107–117.

[16] Bisin, Alberto, and Thierry Verdier. 2000. Beyond the melting pot: Cultural transmission, marriage, and the evolution of ethnic and religious traits. *Quarterly Journal of Economics* 115: 955-988.

[17] Bisin, Alberto, Giorgio Topa, and Thierry Verdier. 2004. Religious intermarriage and socialization in the United States. *Journal of Political Economy* 112: 615-664.

[18] Borjas, George. 1987. Self-selection and the earnings of immigrants. *American Economic Review* 80: 531-553.

[19] Borjas, George. 1992. Ethnic capital and intergenerational mobility. *Quarterly Journal of Economics* 107: 123-150.

[20] Borjas, George. 1995. Ethnicity, neighborhoods and human capital externalities. *American Economic Review* 85: 365-390.

[21] Cadena, Brian, Brian Duncan, and Stephen Trejo. 2015. The labor market integration and impacts of U.S. immigrants. In *Handbook of the Economics of International Migration*, edited by Barry Chiswick and Paul Miller. Amsterdam: North Holland. 1197-1259.

[22] Chiswick, Barry. 1978. The effect of Americanization on the earnings of foreign-born men. *Journal of Political Economy* 86: 897-921.

[23] Chung, Wilbur, and Arturs Kalnins. 2006. Social capital, geography, and survival: Gujarati immigrant entrepreneurs in the U.S. lodging industry. *Management Science* 52(2): 233-247.

[24] Deming, David. 2017. The growing importance of social skills in the labor market. *Quarterly Journal of Economics* 132(4): 1593-1640.

[25] Dhingra, Pawan. 2012. *Life Behind The Lobby: Indian American Motel Owners and the American Dream*. Palo Alto, CA: Stanford University Press, 2012.

[26] Dillon, Eleanor, and Christopher Stanton. 2018. Self-employment dynamics and the returns to entrepreneurship. HBS Working Paper 17-022.

[27] Durlauf, Steven, and Marcel Fafchamps. 2006. Social capital. In *Handbook of Economic Growth*, edited by Philippe Aghion and Steven Durlauf. Amsterdam: North Holland.

[28] Durlauf, Steven, and Yannis Ioannides, 2010. Social interactions. *Annual Review of Economics* 2: 451-478.

[29] Elfring, Tom, and Willem Hulsink. 2003. Networks in entrepreneurship: The case of high-technology firms. *Small Business Economics* 21: 409-422.

[30] Faggio, Giulia, and Olmo Silva. 2014. Self-employment and entrepreneurship in urban and rural labour markets. *Journal of Urban Economics* 83(1): 67-85.

[31] Fairlie, Robert. 2008. *Estimating the Contribution of Immigrant Business Owners to the U.S. Economy*. Small Business Administration, Office of Advocacy Report.

[32] Fairlie, Robert, Harry Krashinsky, and Julie Zissimopoulos. 2010. The international Asian business success story? A comparison of Chinese, Indian and other Asian businesses in the United States, Canada and United Kingdom. In *International Differences in Entrepreneurship*, edited by Josh Lerner and Antoinette Schoar. Chicago, IL: University of Chicago Press.

[33] Fairlie, Robert, and Magnus Lofstrom. 2013. Immigration and entrepreneurship. In *The Handbook on the Economics of International Migration*, edited by Barry Chiswick and Paul Miller. Amsterdam: North-Holland Publishing.

[34] Fairlie, Robert, and Alicia Robb. 2007. Families, human capital, and small business: Evidence from the Characteristics of Business Owners Survey. *Industrial and Labor Relations Review* 60: 225-245.

[35] Furtado, Delia. 2010. Why does intermarriage increase immigrant employment? The role of networks. *B.E. Journal of Economic Analysis & Policy* 10(1): Article 101.

[36] Furtado, Delia, and Nikolaos Theodoropoulos. 2011. Interethnic marriage: A choice between ethnic and educational similarities. *Journal of Population Economics* 24(4): 1257-1279.

[37] Furtado, Delia, and Tao Song. 2015. Intermarriage and socioeconomic integration: Trends in earnings premiums among U.S. immigrants who marry natives. *The ANNALS of the American Academy of Political and Social Science* 662(1): 207-222.

[38] Furtado, Delia, and Stephen Trejo. 2013. Interethnic marriages and their economic effects. In *International Handbook on the Economics of Migration*, edited by Amelie Constant and Klaus Zimmermann. Northampton, MA: Edward Elgar, 276-292.

[39] Ghani, Ejaz, William Kerr, and Stephen O'Connell. 2013. Local industrial structures and female entrepreneurship in India. *Journal of Economic Geography* 13(6): 929–964.

[40] Glaeser, Edward, and Joshua Gottlieb. 2009. The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature* 47(4): 983-1028.

[41] Glaeser, Edward, Bruce Sacerdote, and José Scheinkman. 1996. Crime and social interactions. *Quarterly Journal of Economics* 111: 507-548.

[42] Glaeser, Edward, and José Scheinkman. 2002. Non-market interaction. In *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress*, edited by Mathias Dewatripont, Lars Peter Hansen, and Stephen Turnovsky. Cambridge, UK: Cambridge University Press.

[43] Granovetter, Mark. 1973. The strength of weak ties. *American Journal of Sociology* 78: 1360-1380.

[44] Greve, Arent, and Janet Salaff. 2003. Social networks and entrepreneurship. *Entrepreneurship Theory and Practice* 28(1): 1–22.

[45] Herberg, Will. 1955. *Protestant, Catholic, Jew: An Essay in American Religious Sociology*. Garden City, NY: Doubleday.

[46] Hoang, Cleste. 2015. http://www.takepart.com/article/2015/05/05/tippi-hedren-vietnamese-refugees-nail-industry.

[47] Hunt, Jennifer. (2011). Which immigrants are most innovative and entrepreneurial? Distinctions by entry visa. *Journal of Labor Economics* 29(3): 417-457.

[48] Kennedy, Ruby. 1944. Single or triple melting-pot? Intermarriage trends in New Haven, 1870-1940. *The American Journal of Sociology* 49: 331-339.

[49] Kerr, Sari Pekkala, and William Kerr. 2017. Immigrant entrepreneurship. In *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, edited by John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar. Chicago, IL: University of Chicago Press.

[50] Kerr, Sari Pekkala, and William Kerr. 2020a. Immigrant entrepreneurship in America: Evidence from the Survey of Business Owners 2007 & 2012. *Research Policy* 49(3): 103918.

[51] Kerr, Sari Pekkala, and William Kerr. 2020b. Immigrant networking and collaboration: Survey evidence from CIC. In *The Roles of Immigrants and Foreign Students in U.S. Science, Innovation, and Entrepreneurship*, edited by Ina Ganguli, Shulamit Kahn, and Megan MacGarvie. Chicago: University of Chicago Press.

[52] Kerr, Sari Pekkala, William Kerr, and William Lincoln. 2015. Skilled immigration and the employment structures of U.S. firms. *Journal of Labor Economics* 33(S1): S147-S186.

[53] Kerr, William. 2013. U.S. high-skilled immigration, innovation, and entrepreneurship: Empirical approaches and evidence. Working Paper no. 19377, NBER, Cambridge, MA.

[54] Kerr, William, and Martin Mandorff. 2018. Social networks, ethnicity, and entrepreneurship. Working Paper no. 21597 (revised), NBER, Cambridge, MA.

[55] Kihlstrom, R., and Jean-Jacques Laffont. 1979. A general equilibrium entrepreneurial theory of firm formation based on risk aversion. *Journal of Political Economy* 87: 719-748.

[56] Kim, Yoonha, and John Morgan. 2018. Hidden gems? Cultural barriers and the self-employment of high skilled U.S. immigrants. Working paper.

[57] Kuznets, Simon. 1960. Economic structure and life of the Jews. In *The Minority Members: History, Culture, and Religion*, edited by Louis Finkelstein. Philadelphia, PA: Jewish Publication Society of America.

[58] Lazear, Edward. 1999. Culture and language. *Journal of Political Economy* 107: 95-126.

[59] Lazear, Edward. 2005. Entrepreneurship. *Journal of Labor Economics* 23: 649-680.

[60] Leyden, Dennis, and Albert Link. 2015. Toward a theory of the entrepreneurial process. *Small Business Economics* 44(3): 475-484

[61] Lucas, Robert. 1978. On the size distribution of business firms. *Bell Journal of Economics* 9: 508-523.

[62] Mandorff, Martin. 2007. Social networks, ethnicity, and occupation. University of Chicago Ph.D. Dissertation.

[63] Manso, Gustavo. 2016. Experimentation and the returns to entrepreneurship. *Review of Financial Studies* 29(9): 2319-2340.

[64] Milman, Joel. 1998. *The Other Americans: How Immigrants Renew Our Country, Our Economy, and Our Values.* London: Penguin Books.

[65] Moris, Regan. 2015. How Tippi Hedren made Vietnamese refugees into nail salon magnates. https://www.bbc.com/news/magazine-32544343.

[66] Munshi, Kaivan. 2003. Networks in the modern economy: Mexican migrants in the U. S. labor market. *Quarterly Journal of Economics* 118(2): 549-599.

[67] Neumark, David. 2013. Ethnic hiring. In *International Handbook on the Economics of Migration.* Edward Elgar Publishing.

[68] Orrenius, Pia, and Madeline Zavodny. 2016. Do state work eligibility verification laws reduce unauthorized immigration? *IZA Journal of Migration* 5:5.

[69] Patel, Krishna, Yevgeniya Savchenko, and Francis Vella. 2013. Occupational sorting of ethnic groups. In *International Handbook on the Economics of Migration.* Edward Elgar Publishing.

[70] Patel, Krishna, and Francis Vella. 2013. Immigrant networks and their implications for occupational choice and wages. *Review of Economics and Statistics* 95(4): 1249-1277.

[71] Rauch, James. 2001. Business and social networks in international trade. *Journal of Economic Literature* 39: 1177-1203.

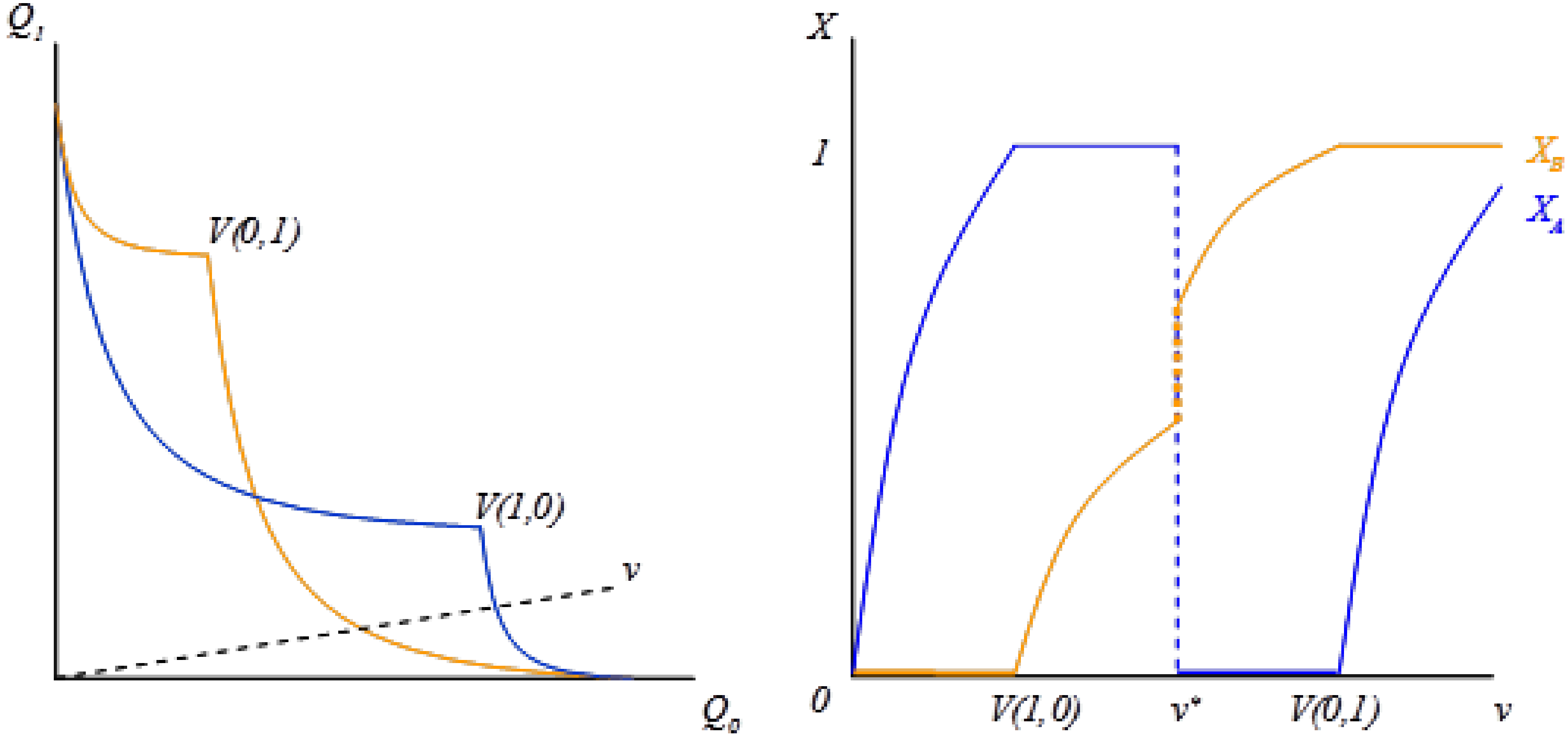
[72] Rosenthal, Stuart, and William Strange. 2012. Female entrepreneurship, agglomeration, and a new spatial mismatch. *Review of Economics and Statistics* 94(3): 764-788.

[73] Simon, Curtis, and John Warner. 1992. Matchmaker, matchmaker: The effect of old boy networks on job match quality, earnings and tenure. *Journal of Labor Economics* 10(3): 306-330.

[74] Spolaore, Enrico, and Romain Wacziarg. 2016. Ancestry, language and culture. In *Palgrave Handbook of Economics and Language*, edited by Victor Ginsburgh and Shlomo Weber. Palgrave McMillan: 174-211.

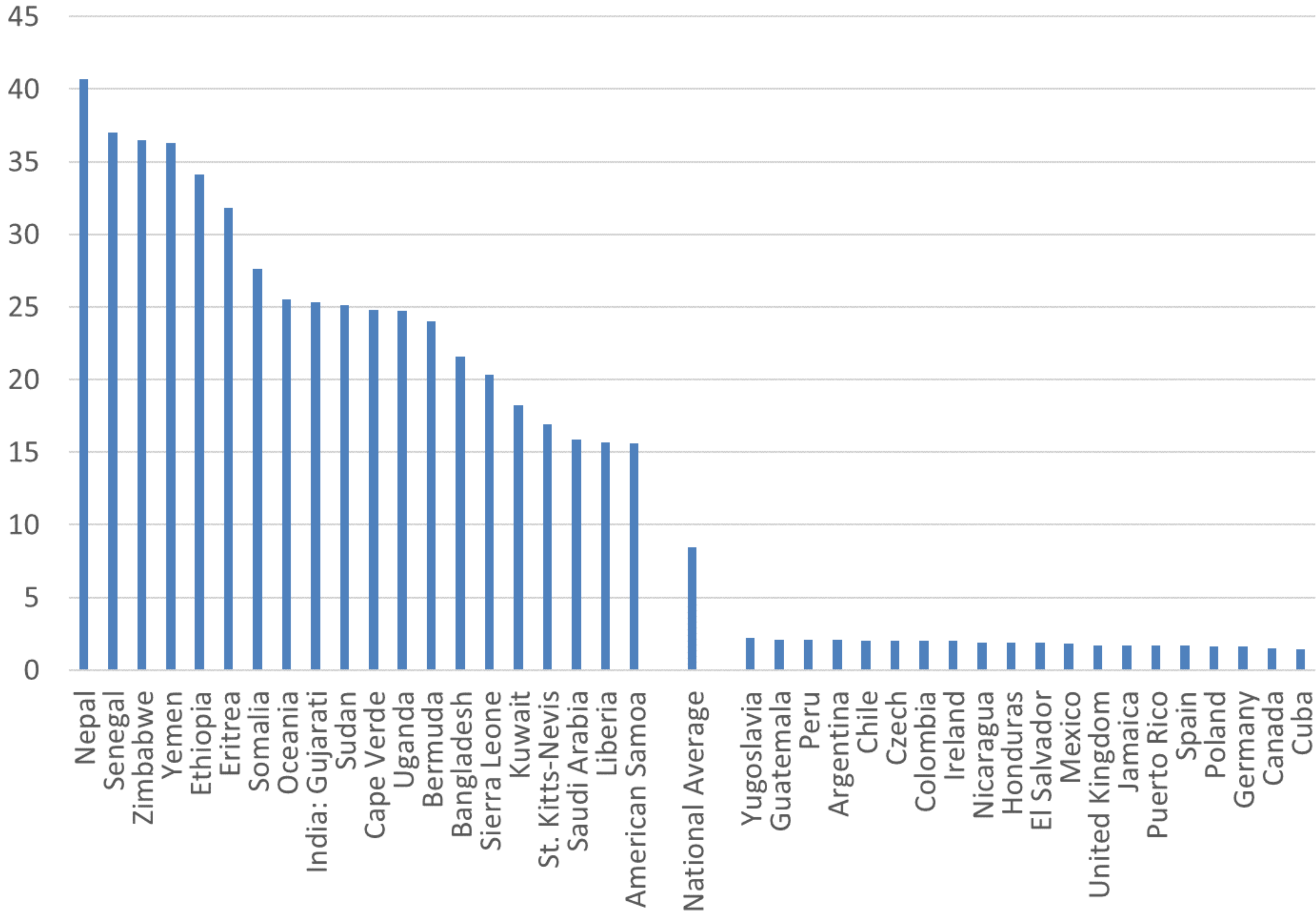
[75] Virani, Aarti. 2012. https://blogs.wsj.com/indiarealtime/2012/06/11/why-indian-americans-dominate-the-u-s-motel-industry/?mg=prod/accounts-wsj.

Figure 1: Model depiction of entrepreneurial specialization



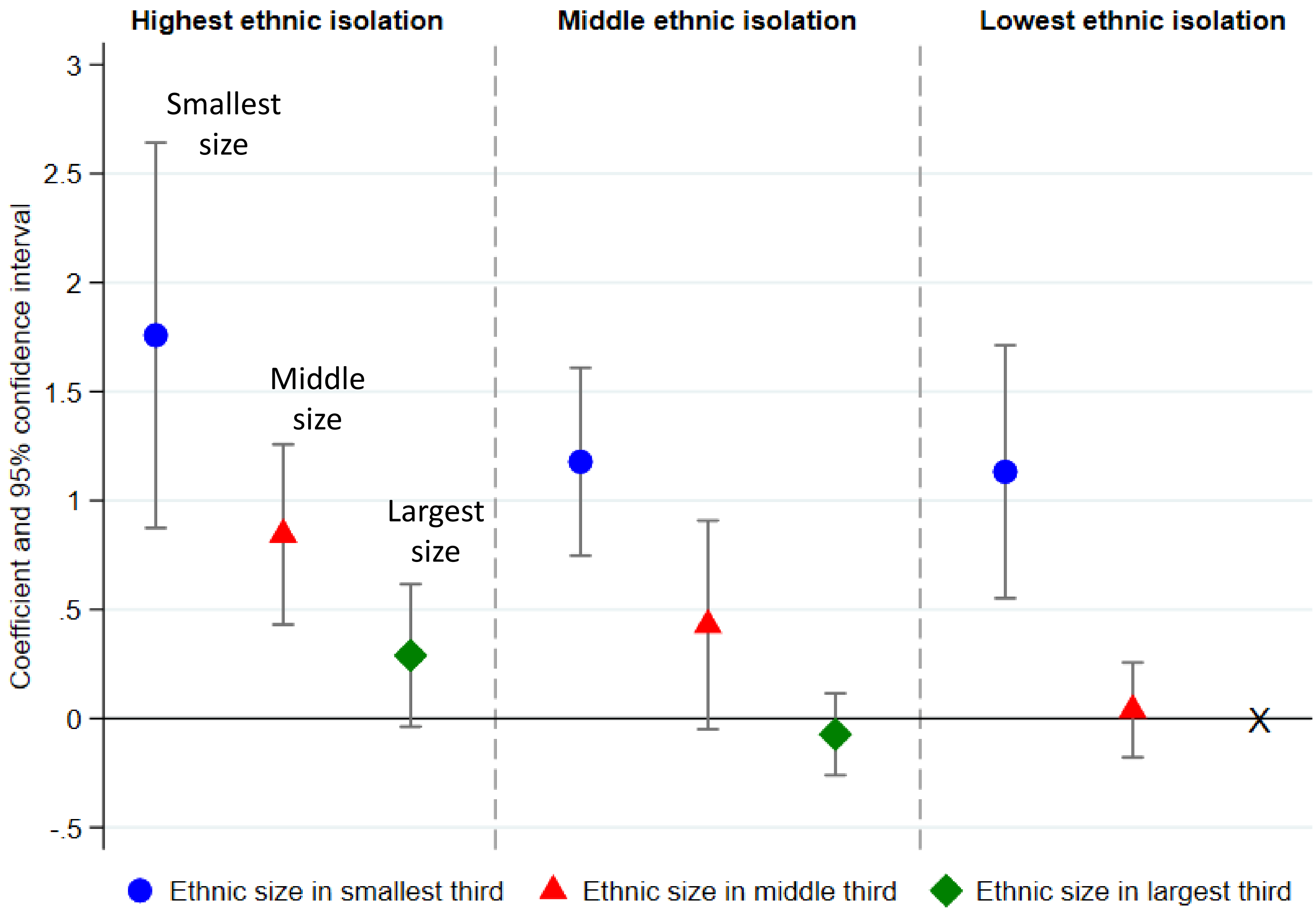
Notes: Left panel shows production possibilities with specialized occupational distributions. The ray *v* is the preference parameter over goods in the Leontief utility function. Along the curve with the kink *V*(1,0), all entrepreneurs belong to group *A* (below the kink) or all members of group *A* are entrepreneurs (above). Similarly, along the curve with the kink *V*(0,1), all entrepreneurs belong to group *B* (below) or all members of group *B* are entrepreneurs (above). The right panel shows the efficient occupational distribution for different values of *v* assuming productivity is convex in interactions. The minority group A specializes as entrepreneurs so long as the entrepreneurial sector is small enough.

Figure 2: Levels of entrepreneurial specialization by ethnic group



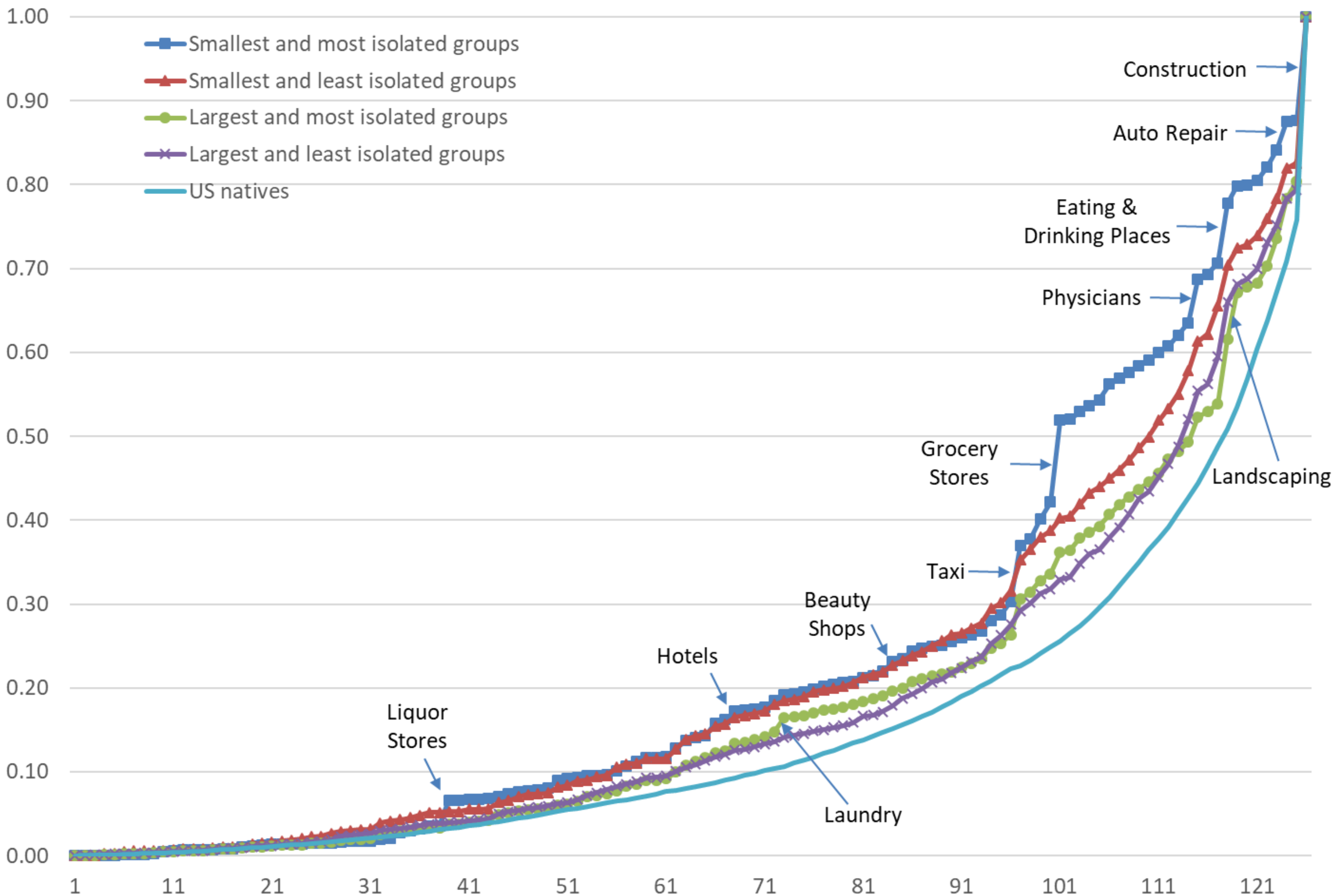
Notes: Figure shows the weighted average overage for the entrepreneurial concentration by ethnic immigrant group. The top 20 and bottom 20 values are shown, along with the national average.

Figure 3: Non-parametric estimations of entrepreneurial specialization



Notes: See Table 1. Ethnic groups are divided into equal-sized bins based upon group size and social isolation using terciles. Within each isolation triplet, groups are ordered smallest to largest as shown for the most isolated groups. Effects are measured relative to largest and least isolated ethnic groups. Coefficient estimates and 95% confidence bands are reported. Full results are provided in Appendix.

Figure 4: Industry distributions of self-employment



Notes: Table shows the cumulative distribution of self-employment for groups moving from the smallest [#1] to largest [#126] industries for self-employment. Immigrant ethnic groups are divided equally into four groups based upon being above or below median group size and social isolation.

Table 1: OLS estimations of weighted average overage across all industries

	Baseline estimation with OVER1	Using total worker sample	Excluding natives from denominator shares	Imposing min counts on ethnic industry presence	Excluding new arrivals over the prior five years	Excluding the taxi industry
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Inverse of log ethnic group size	0.582+++	0.440+++	0.615+++	0.588+++	0.475+++	0.472+++
(small groups have larger values)	(0.076)	(0.074)	(0.078)	(0.081)	(0.074)	(0.088)
Isolation of ethnic group	0.325+++	0.557+++	0.307+++	0.326+++	0.529+++	0.483+++
	(0.076)	(0.090)	(0.085)	(0.086)	(0.085)	(0.095)
Adjusted R-Squared value	0.373	0.373	0.385	0.363	0.378	0.337
	B. Including controls for ethnic group's traits					
Inverse of log ethnic group size	0.465+++	0.325+++	0.460+++	0.432+++	0.370+++	0.416+++
(small groups have larger values)	(0.079)	(0.072)	(0.082)	(0.082)	(0.071)	(0.087)
Isolation of ethnic group	0.447+++	0.674+++	0.428+++	0.491+++	0.672+++	0.561+++
	(0.094)	(0.103)	(0.113)	(0.111)	(0.104)	(0.125)
Adjusted R-Squared value	0.455	0.479	0.501	0.494	0.484	0.403

Notes: Estimations describe the OLS relationship between industry concentration for ethnic entrepreneurship and ethnic group size and in-marriage isolation in 2000. The outcome variable is the weighted average overage ratio across industries for each ethnic group, where the weights are levels of self-employment in each industry per group. Variables are winsorized at their 1%/99% levels and transformed to have unit standard deviation for interpretation. Regressions include 131 observations, are weighted by log ethnic group counts, and report robust standard errors. Column 2 considers the metric that uses all employed workers for the ethnic group, Column 3 compares industry-level overages only to rates of other immigrant groups, Column 4 drops ethnic-industry settings where fewer than three observation counts exist, Column 5 excludes new arrivals to America during the prior five years, and Column 6 excludes the taxi cab industry. Panel B controls for the traits of the ethnic group in 2000: share who are 36-55 years old, share who are 55-70 years old (reference group is aged 22-35), share who are married, share who speak English well, share who have some college education, and share who have a college degree or higher (reference group is high school or less). +++, ++, and + indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 2: Robustness checks on OLS estimations

	Baseline estimation	Without sample weights	Without winsorization	Including fixed effects for origin continent	Using median regression format	Using bootstrapped standard errors	Including expected overage control
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	A. Baseline estimation without controls						
Inverse of log ethnic group size	0.582+++	0.612+++	0.571+++	0.481+++	0.322+++	0.612+++	0.384+++
(small groups have larger values)	(0.076)	(0.082)	(0.077)	(0.070)	(0.063)	(0.086)	(0.089)
Isolation of ethnic group	0.325+++	0.331+++	0.329+++	0.279+++	0.220+++	0.331+++	0.334+++
	(0.076)	(0.081)	(0.077)	(0.074)	(0.073)	(0.088)	(0.074)
Adjusted R-Squared value	0.373	0.368	0.364	0.507	0.198	0.368	0.435
	B. Including controls for ethnic group's traits						
Inverse of log ethnic group size	0.465+++	0.488+++	0.460+++	0.453+++	0.353+++	0.488+++	0.286+++
(small groups have larger values)	(0.079)	(0.085)	(0.079)	(0.079)	(0.080)	(0.092)	(0.091)
Isolation of ethnic group	0.447+++	0.441+++	0.450+++	0.367+++	0.390+++	0.441+++	0.486+++
	(0.094)	(0.100)	(0.095)	(0.107)	(0.088)	(0.107)	(0.084)
Adjusted R-Squared value	0.455	0.452	0.454	0.529	0.262	0.452	0.517

Notes: See Table 1. Columns 2-6 provide robustness checks on the baseline specification. Regressions in Columns 5 and 6 are unweighted and should be referenced against Column 2. Column 5 reports pseudo R-squared values. Column 7 adds a control for the expected overage level for an ethnicity based upon 100 Monte Carlo simulations with the number of observations in the sample.

Table 3: OLS estimations of overage metric designs

	Weighted average overage across all industries [OVER1]	Weighted average overage using three largest industries for ethnic group [OVER2]	Average of three largest overage ratios for ethnic group [OVER3]	Largest overage ratio for ethnic group [OVER4]
	(1)	(2)	(3)	(4)
		A. Baseline estimation without controls		
Inverse of log ethnic group size (small groups have larger values)	0.582+++	0.458+++	0.578+++	0.526+++
	(0.076)	(0.076)	(0.068)	(0.077)
Isolation of ethnic group	0.325+++	0.423+++	0.234+++	0.160++
	(0.076)	(0.086)	(0.079)	(0.068)
Adjusted R-Squared value	0.373	0.298	0.335	0.280
		B. Including controls for ethnic group's traits		
Inverse of log ethnic group size (small groups have larger values)	0.465+++	0.335+++	0.507+++	0.466+++
	(0.079)	(0.084)	(0.083)	(0.085)
Isolation of ethnic group	0.447+++	0.531+++	0.342+++	0.258+++
	(0.094)	(0.100)	(0.089)	(0.099)
Adjusted R-Squared value	0.455	0.380	0.364	0.291

Notes: See Table 1. Estimations consider variations in the overage metric design.

Table 4: OLS estimations of panel changes from 1980 - 2018

	Baseline panel estimation [OVER1]	Including control for expected overage from simulations	Including control for linear time trend in 1980 overage level
	(1)	(2)	(3)
	A. Baseline estimation without controls		
Inverse of log ethnic group size	0.970+++	0.568+++	0.207++
(small groups have larger values)	(0.264)	(0.186)	(0.084)
Isolation of ethnic group	0.342+	0.242++	0.197++
	(0.190)	(0.119)	(0.093)
Ethnic group and year FE	Yes	Yes	Yes
Adjusted R-Squared value	0.484	0.593	0.735
	B. Including controls for ethnic group's traits		
Inverse of log ethnic group size	0.948+++	0.501++	0.258+
(small groups have larger values)	(0.285)	(0.208)	(0.151)
Isolation of ethnic group	0.300+	0.228+	0.165+
	(0.179)	(0.119)	(0.092)
Ethnic group and year FE	Yes	Yes	Yes
Adjusted R-Squared value	0.500	0.602	0.742

Notes: See Table 1. Estimations describe the OLS panel relationship between industry concentration for ethnic entrepreneurship and ethnic group size and in-marriage isolation from 1980, 1990, 2000, 2010 and 2018 combining Censuses and the American Community Survey. The analysis considers 107 ethnic groups with full panel data, for 535 observations. Regressions are weighted by log ethnic group counts and report standard errors clustered by ethnic group. Column 2 adds a control for the expected overage level for an ethnicity and year based upon 100 Monte Carlo simulations with the number of observations in the sample. Column 3 adds a linear time trend in the 1980 overage level.

Table 5: OLS estimations with variations on ethnic group isolation metric

	Weighted average overage across all industries					
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Inverse of log ethnic group size (small groups have larger values)	0.582+++ (0.076)	0.718+++ (0.103)	0.466+++ (0.070)	0.507+++ (0.086)	0.488+++ (0.084)	0.624+++ (0.118)
Isolation of ethnic group using in-marriage rate in United States	0.325+++ (0.076)					0.331+++ (0.109)
Residential segregation in the United States		0.302+++ (0.066)				0.030 (0.112)
Genetic distance (country)			0.153++ (0.071)			0.189++ (0.081)
Linguistic distance (country)				0.054 (0.047)		-0.081 (0.075)
Religious distance (country)					0.151++ (0.076)	0.051 (0.086)
Adjusted R-Squared value	0.373	0.325	0.288	0.271	0.288	0.415
Observations	131	131	131	113	113	112
	B. Including controls for ethnic group's traits					
Inverse of log ethnic group size (small groups have larger values)	0.465+++ (0.079)	0.597+++ (0.103)	0.378+++ (0.077)	0.302+++ (0.084)	0.303+++ (0.086)	0.440+++ (0.104)
Isolation of ethnic group using in-marriage rate in United States	0.447+++ (0.094)					0.439+++ (0.125)
Residential segregation in the United States		0.335+++ (0.078)				0.063 (0.110)
Genetic distance (country)			0.103 (0.082)			0.115 (0.094)
Linguistic distance (country)				0.110+ (0.056)		-0.019 (0.067)
Religious distance (country)					0.149 (0.102)	0.060 (0.113)
Adjusted R-Squared value	0.455	0.394	0.349	0.392	0.391	0.509
Observations	131	131	131	113	113	112

Notes: See Table 1. Column 1 repeats the baseline estimation with social isolation measured through the 2000 in-marriage rate for the ethnic group. Column 2 uses average residential segregation of ethnic group across Public Use Micro Areas within Metropolitan Statistical Areas that have two or more PUMAs. Columns 3-6 use the average genetic, linguistic, and religious distances from home countries to the weighted ethnic composition of the United States as measured by Spolaore and Wacziarg. Columns 3-6 cluster standard errors by 120 groups from the Spolaore and Wacziarg data.

Table 6: IV estimations

	Instrumenting with predicted ethnic group size from gravity model and in-marriage rates in the United Kingdom in 1991			Instrumenting with predicted ethnic group size from gravity model and average of in-marriage rates in the United Kingdom in 1991 and Spain in 2011		
	First stage for size	First stage for isolation	Second stage	First stage for size	First stage for isolation	Second stage
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Baseline estimation without controls			
Instrument for size	0.648+++	-0.118		0.622+++	-0.023	
	(0.064)	(0.116)		(0.066)	(0.121)	
Instrument for isolation	-0.135+	0.540+++		-0.142+	0.490+++	
	(0.076)	(0.097)		(0.075)	(0.089)	
F-Statistic	52.2	17.4		48.2	18.9	
Inverse of log ethnic group size			0.459+++			0.469+++
			(0.130)			(0.140)
Isolation of ethnic group			0.316++			0.419+++
			(0.125)			(0.119)
Exogeneity test p-value			0.140			0.023
			B. Including controls for ethnic group's traits			
Instrument for size	0.503+++	-0.081		0.496+++	-0.054	
	(0.077)	(0.067)		(0.079)	(0.064)	
Instrument for isolation	-0.077	0.358+++		-0.078	0.313+++	
	(0.069)	(0.072)		(0.052)	(0.063)	
F-Statistic	21.7	12.7		23.2	12.8	
Inverse of log ethnic group size			0.294++			0.342++
			(0.143)			(0.174)
Isolation of ethnic group			0.484+++			0.728+++
			(0.153)			(0.209)
Exogeneity test p-value			0.237			0.006

Notes: See Table 1. Estimations describe the IV relationship between industry concentration for ethnic entrepreneurship and ethnic group size and in-marriage isolation. Instruments are the predicted ethnic group size from gravity model and in-marriage rates in UK 1991 or an average of the in-marriage rates in UK 1991 and Spain 2011. The null hypothesis in Wu-Hausman exogeneity tests is that the instrumented regressors are exogenous. Regressions have 130 and 129 observations, respectively, as UK and Spain are excluded when used in the instrument. Regressions cluster standard errors by UK 1991 dataset ethnic groups.

Table 7: OLS estimations of individual incomes and group concentration

	Log yearly income in 2000				
	(1)	(2)	(3)	(4)	(5)
	A. Self-employed individuals				
Percentage of individual's group who are self-employed in the industry	0.069+++ (0.009)	0.030+++ (0.009)	0.029+++ (0.010)	0.024+++ (0.008)	0.015+ (0.009)
Percentage of individual's group who are self-employed		0.100+++ (0.008)	0.101+++ (0.008)		
Percentage of individual's group who are working in the industry			0.002 (0.017)		0.035++ (0.014)
Person-level Traits FE	Yes	Yes	Yes	Yes	Yes
MSA-Industry FE	Yes	Yes	Yes	Yes	Yes
Ethnicity FE				Yes	Yes
Adjusted R-Squared value	0.244	0.250	0.250	0.260	0.260
Observations	49,026	49,026	49,026	49,026	49,026
	B. Wage workers				
Percentage of individual's group who are self-employed in the industry	0.045+++ (0.012)	0.018 (0.012)	0.012 (0.013)	0.002 (0.007)	-0.004 (0.007)
Percentage of individual's group who are self-employed		0.059+++ (0.007)	0.062+++ (0.007)		
Percentage of individual's group who are working in the industry			0.020 (0.013)		0.021++ (0.009)
Person-level Traits FE	Yes	Yes	Yes	Yes	Yes
MSA-Industry FE	Yes	Yes	Yes	Yes	Yes
Ethnicity FE				Yes	Yes
Adjusted R-Squared value	0.252	0.254	0.254	0.266	0.266
Observations	355,441	355,441	355,441	355,441	355,441

Notes: Estimations describe the OLS relationship between log yearly income of individuals and entrepreneurial activity of their ethnic group. Sample is taken from 2000 Census IPUMS. Sample includes immigrant males who arrived into the United States at age 16 or later and who are aged 22-70 in 2000. Estimations include fixed effects for the following person-level traits (category counts in parentheses): age (5), age at immigration for migrants (5), education (4), and English language fluency (2). Regressions use person weights and cluster standard errors by ethnic group.

Table 8: OLS estimations of industry size and group size

	Self-employed members of ethnic group				Wage workers of ethnic group			
	Log average industry size in terms of self-employed	Log average industry size in terms of total workers	Log size of largest industry measured by self-employed count	Log size of largest industry measured by total worker count	Log average industry size in terms of self-employed	Log average industry size in terms of total workers	Log size of largest industry measured by self-employed count	Log size of largest industry measured by total worker count
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Baseline estimation without controls							
Inverse of log ethnic group size	-0.172+++	-0.148+++	-0.427+++	-0.411+++	-0.062	-0.036	-0.311	-0.221++
(small groups have larger values)	(0.058)	(0.050)	(0.129)	(0.134)	(0.048)	(0.027)	(0.212)	(0.092)
Isolation of ethnic group	-0.189+++	-0.148+++	-0.700+++	-0.695+++	-0.055	-0.031	0.094	-0.150+
	(0.051)	(0.042)	(0.117)	(0.120)	(0.039)	(0.022)	(0.194)	(0.080)
Adjusted R-Squared value	0.115	0.101	0.208	0.199	0.007	0.007	0.011	0.046
	B. Including controls for ethnic group's traits							
Inverse of log ethnic group size	-0.107++	-0.083++	-0.189+	-0.125	0.026	0.022	-0.171	-0.119
(small groups have larger values)	(0.047)	(0.040)	(0.108)	(0.115)	(0.034)	(0.020)	(0.207)	(0.080)
Isolation of ethnic group	-0.347+++	-0.298+++	-0.886+++	-0.939+++	-0.179+++	-0.101+++	-0.461++	-0.437+++
	(0.055)	(0.048)	(0.144)	(0.142)	(0.036)	(0.021)	(0.201)	(0.097)
Adjusted R-Squared value	0.578	0.582	0.451	0.443	0.600	0.582	0.437	0.459

Notes: See Table 1.

Table 9: Sources of advice in 2016 Annual Survey of Entrepreneurs

Source of advice	Accommodation and food services (NAICS 72)	Construction (NAICS 23)	Average for other NAICS 2-digit sectors	Ratio of NAICS 72 to average in other sectors
Customers	12.5	8.3	6.5	1.92
Family	15.9	12.9	9.0	1.76
Friends	12.0	10.8	7.1	1.70
Other	8.4	5.3	5.2	1.61
Suppliers	12.4	12.5	10.8	1.15
Government	1.8	2.4	1.6	1.14
Colleagues	42.7	49.8	39.0	1.10
Employees	13.7	12.6	13.6	1.01
Advisors	75.9	79.6	85.2	0.89

Notes: Tabulation of employment-weighted share of businesses reporting source of advice.

Table 10: OLS estimations at metropolitan level for size

	Baseline MSA-level estimation with local size measure [OVER1]	Including control for expected overage from simulations	Including control for expected overage from simulations and ethnicity fixed effects
	(1)	(2)	(3)
		A. Baseline estimation without controls	
Inverse of log ethnic group size	0.420+++	0.209+++	0.115+++
(small groups have larger values)	(0.023)	(0.034)	(0.026)
Isolation of ethnic group	0.063++	0.066++	
(national measure)	(0.026)	(0.027)	
MSA FE	Yes	Yes	Yes
Ethnic group FE			Yes
Adjusted R-Squared value	0.175	0.209	0.247
Observations	6,649	6,649	6,649
		B. Including controls for ethnic group's traits	
Inverse of log ethnic group size	0.384+++	0.165+++	0.115+++
(small groups have larger values)	(0.020)	(0.026)	(0.026)
Isolation of ethnic group	0.087+++	0.095+++	
(national measure)	(0.023)	(0.023)	
MSA FE	Yes	Yes	Yes
Ethnic group FE			Yes
Adjusted R-Squared value	0.193	0.228	0.247
Observations	6,649	6,649	6,649

Notes: See Table 1. Estimations describe the OLS panel relationship between industry concentration for ethnic entrepreneurship and ethnic group size at a metropolitan level. The sample includes metropolitan areas for an ethnicity where self-employment activity is observed. Regressions are weighted by log ethnic group counts in the metropolitan area and report standard errors clustered by ethnic group. Variables are winsorized at their 10%/90% values to guard against outliers. Column 2 adds a control for the expected overage level for an ethnicity and metropolitan area based upon 100 Monte Carlo simulations with the number of observations in the sample. Column 3 adds a fixed effect for ethnicities.

Online Appendices for Social Networks, Ethnicity, and Entrepreneurship

William Kerr and Martin Mandorff

Empirical Appendix Tables

Table A1a: Ethnic group tabulations

Ethnic group	Self-Empl. OVER1	Self-Empl. OVER2	Self-Empl. OVER3	Self-Empl. OVER4	All workers OVER1	In-marriage rate	Self-Empl. share	All worker count	Population count
Nepal	40.7	30.0	67.6	157.5	4.2	43%	5%	4,784	8,935
Senegal	37.0	67.6	81.4	123.4	5.6	62%	19%	4,689	7,351
Zimbabwe	36.5	39.5	124.2	228.8	2.6	43%	9%	3,482	8,955
Yemen	36.3	55.2	57.6	61.6	6.5	88%	20%	6,196	10,642
Ethiopia	34.1	56.8	63.1	65.7	8.9	53%	12%	23,962	52,577
Eritrea	31.8	47.6	52.0	61.7	5.6	67%	14%	6,167	12,843
Somalia	27.6	40.8	44.9	53.8	4.3	78%	6%	9,092	18,326
Oceania, ns/nec	25.5	14.9	47.8	113.3	3.1	39%	3%	5,216	16,053
India: Gujarati	25.3	48.8	69.5	83.7	3.1	91%	19%	54,867	117,871
Sudan	25.1	28.6	42.2	48.1	3.0	73%	6%	6,607	10,671
Cape Verde	24.8	38.2	68.0	105.4	2.6	68%	7%	6,154	18,176
Uganda	24.7	43.7	55.9	67.8	2.9	56%	14%	4,357	9,620
Bermuda	24.0	33.3	49.3	89.8	2.9	30%	9%	2,081	11,238
Bangladesh	21.6	32.0	46.6	49.8	5.4	80%	11%	36,267	58,529
Sierra Leone	20.3	25.0	39.9	44.9	4.4	54%	9%	6,537	15,101
Kuwait	18.2	29.7	33.5	42.4	3.4	57%	13%	6,555	12,490
St. Kitts-Nevis	16.9	21.2	23.7	34.0	2.1	62%	4%	2,942	8,735
Saudi Arabia	15.9	24.7	41.9	50.9	2.8	37%	10%	4,022	9,725
Liberia	15.7	14.7	41.9	119.6	2.6	45%	6%	10,617	28,936
American Samoa	15.6	19.8	24.0	31.7	2.5	44%	4%	3,565	21,894
Ghana	15.5	26.7	40.9	42.3	2.8	63%	8%	27,176	51,921
South America, ns	15.1	12.2	33.2	43.4	2.5	31%	10%	3,178	11,686
Cameroon	14.5	16.5	24.4	32.1	3.1	58%	9%	5,045	9,627
Nigeria	13.6	29.9	37.2	40.2	3.9	63%	14%	53,654	102,763
Chaldean	13.1	29.2	42.1	74.2	6.1	84%	24%	12,996	31,038
Dominica	12.8	10.5	30.6	40.4	1.9	58%	6%	4,488	12,282
Bahamas	12.2	12.1	29.4	45.9	2.0	41%	8%	4,373	19,404
Tanzania	11.9	15.2	31.9	42.7	3.4	49%	20%	3,785	8,908
Haiti	11.4	22.5	34.2	36.5	2.6	72%	8%	114,602	295,424
Americas, ns	11.2	9.2	24.0	27.8	2.3	40%	11%	2,488	7,029
Partial Nordic Region	11.1	1.4	58.8	71.0	2.2	16%	16%	4,571	14,527
Singapore	10.7	11.0	19.4	28.8	2.6	39%	8%	5,316	14,470
Belgium	10.6	2.3	66.8	95.7	2.1	26%	14%	6,636	23,591
Morocco	10.4	7.5	47.6	59.9	3.4	27%	14%	14,515	29,812
Pakistan	10.3	20.2	23.2	29.7	5.1	76%	17%	85,400	144,403
Afghanistan	10.2	15.6	26.4	27.1	4.7	78%	19%	12,573	28,075
Polynesia	9.1	11.4	15.9	18.0	2.3	68%	12%	8,727	22,766
India: Punjabi	9.0	14.7	24.4	25.6	6.0	94%	21%	35,325	68,412
Cyprus	8.5	4.7	26.0	33.5	2.8	43%	21%	3,301	7,638
Africa, ns/nec	8.4	20.2	25.1	27.7	2.4	42%	12%	49,046	104,639
Antigua-Barbuda	8.2	7.8	24.0	26.4	2.0	26%	10%	4,359	13,980
Baltic States	7.9	14.7	34.0	48.0	1.7	41%	17%	7,138	27,865
Dominican Republic	7.6	15.2	21.2	26.9	2.7	60%	11%	162,086	458,705
Indochina, ns	7.6	14.5	21.5	27.7	2.4	68%	23%	21,790	43,819
Iraq	7.4	6.6	24.2	46.1	2.5	58%	18%	18,494	32,852
Jordan	7.3	13.3	20.9	21.9	3.2	68%	27%	18,945	32,794
Korea	7.0	14.1	27.5	33.7	3.9	53%	34%	185,099	574,104
Sri Lanka (Ceylon)	6.8	10.7	14.4	14.5	2.4	48%	16%	9,546	19,756
Melanesia	6.7	7.9	22.9	26.3	1.9	69%	9%	8,338	22,856
Cambodia	6.7	8.8	13.6	30.5	2.1	74%	14%	29,578	89,044
Syria	6.6	6.1	28.1	37.4	4.6	53%	29%	16,533	27,237
Former USSR	6.6	13.4	22.4	35.6	2.9	50%	17%	10,763	24,733
Bulgaria	6.5	6.6	17.1	19.8	2.1	51%	16%	11,236	24,332
Norway	6.4	4.5	19.0	29.1	2.7	24%	16%	6,307	20,178
Grenada	6.3	3.1	27.2	38.7	2.4	54%	13%	6,568	19,247
Burma (Myanmar)	6.2	9.5	20.9	24.6	2.0	60%	11%	10,730	23,431
Kenya	6.0	11.1	23.2	25.8	2.5	52%	13%	13,896	30,690
Northern Ireland	6.0	2.4	30.1	35.9	2.3	23%	17%	2,900	8,258
Greece	5.7	8.2	13.1	13.4	2.7	61%	31%	52,382	125,152
Other Caribbean	5.4	3.3	17.7	32.0	1.7	40%	11%	6,803	23,826
Europe, ns.	5.3	3.2	22.4	26.5	2.0	32%	16%	6,471	23,750
St. Vincent	5.3	2.0	14.2	18.3	2.3	36%	8%	5,404	15,237
Panama	5.3	3.5	29.4	40.3	1.4	25%	8%	19,406	93,243
Armenian	5.1	7.8	25.0	29.8	3.5	81%	28%	38,206	93,455
Denmark	4.9	1.5	20.3	23.8	2.1	12%	21%	7,153	20,760
Thailand	4.9	7.8	10.1	10.6	2.1	31%	14%	25,131	111,254
Austria	4.9	2.7	23.2	26.5	2.1	26%	22%	8,528	37,079
India: Other	4.5	7.5	10.2	12.3	3.7	67%	10%	313,091	564,570
U.S. Virgin Islands	4.5	2.0	17.5	25.1	1.6	43%	9%	8,487	34,759
New Zealand	4.4	5.0	20.4	26.1	1.9	24%	15%	7,980	19,230

Table A1a: Ethnic group tabulations

Ethnic group	Self-Empl. OVER1	Self-Empl. OVER2	Self-Empl. OVER3	Self-Empl. OVER4	All workers OVER1	In-marriage rate	Self-Empl. share	All worker count	Population count
Albania	4.3	4.7	8.5	9.1	2.4	85%	10%	11,550	24,248
China	4.2	7.8	10.1	10.5	2.9	77%	12%	289,651	653,687
Turkey	4.1	5.2	17.4	22.9	2.7	46%	19%	26,617	57,076
Barbados	4.1	2.6	13.8	15.0	2.0	46%	6%	12,694	37,166
Paraguay	4.1	3.1	12.3	15.9	2.6	42%	26%	3,144	7,843
Vietnam	4.0	6.5	11.7	12.9	2.3	73%	12%	268,627	718,423
St. Lucia	4.0	2.3	11.3	18.0	2.1	51%	12%	3,802	9,885
Belize/British Honduras	3.9	2.6	15.3	30.6	1.8	44%	9%	9,194	29,934
Lebanon	3.9	7.4	18.2	23.4	2.3	49%	25%	33,995	64,639
Northern Africa	3.8	5.7	11.9	12.5	2.5	48%	17%	48,650	87,005
Laos	3.8	2.4	15.5	22.1	2.2	80%	7%	48,346	141,048
Romania	3.6	1.3	36.5	62.0	1.4	47%	18%	34,760	86,443
Sweden	3.6	1.6	15.4	19.6	2.0	17%	17%	11,498	34,566
Croatia	3.6	3.5	15.8	27.7	1.5	50%	18%	10,851	25,577
Malaysia	3.6	4.3	8.6	9.5	2.5	35%	11%	16,103	38,138
Indonesia	3.6	3.8	12.1	16.4	1.7	24%	11%	18,160	50,484
Switzerland	3.6	3.1	16.7	23.6	2.0	21%	18%	11,996	31,231
Ecuador	3.5	5.7	14.8	16.3	2.0	39%	9%	91,506	212,967
Guam	3.5	3.0	7.9	9.9	1.7	29%	5%	7,834	43,745
Bolivia	3.4	2.3	26.9	42.3	1.9	30%	10%	13,893	37,384
Ukraine	3.4	5.1	13.5	14.5	1.6	57%	14%	59,433	143,265
Italy	3.3	4.8	9.5	9.7	1.8	57%	24%	87,551	289,037
Uruguay	3.3	1.4	13.5	14.4	1.6	24%	22%	8,297	19,269
Taiwan	3.3	5.9	8.8	9.1	2.2	46%	18%	80,135	239,620
Hong Kong and Macau	3.2	4.6	8.4	11.0	2.2	66%	14%	47,605	154,484
Brazil	3.2	4.9	23.3	25.9	1.6	36%	15%	59,408	154,828
Guyana/British Guiana	3.2	4.3	12.9	16.1	2.1	59%	9%	55,565	151,927
USSR/Russia	3.0	5.3	12.7	13.8	1.9	61%	15%	97,769	249,585
France	2.9	2.0	20.8	24.1	1.8	29%	14%	36,805	130,567
Iran	2.8	2.3	11.1	16.1	1.9	55%	28%	85,202	178,670
Venezuela	2.8	2.4	10.7	13.8	1.4	33%	14%	27,603	76,541
Philippines	2.7	4.3	10.1	13.2	1.9	50%	7%	304,598	1,027,398
Japan	2.6	3.2	6.5	7.7	1.9	31%	14%	79,389	303,281
South Africa (Union of)	2.6	2.9	10.0	15.1	2.0	34%	20%	19,762	47,921
Israel/Palestine	2.5	2.2	12.9	19.6	1.9	52%	30%	35,990	82,664
Trinidad and Tobago	2.5	2.5	9.7	13.1	1.5	44%	10%	46,055	141,913
Netherlands	2.4	1.7	7.9	8.9	1.8	23%	19%	20,333	64,956
Portugal	2.4	2.9	11.1	13.3	1.8	62%	15%	47,004	149,179
Costa Rica	2.4	1.6	9.5	14.5	1.5	25%	9%	19,433	52,472
Australia	2.4	1.8	8.1	8.5	2.1	27%	17%	16,336	48,237
Hungary	2.4	2.0	16.5	22.6	1.4	30%	24%	18,848	51,325
Yugoslavia	2.2	1.8	7.5	8.5	1.5	52%	9%	57,896	131,241
Guatemala	2.1	2.1	8.7	10.6	1.8	43%	8%	162,886	358,480
Peru	2.1	1.1	11.7	15.5	1.4	23%	12%	83,560	204,158
Argentina	2.1	1.9	9.2	11.5	1.4	27%	21%	35,789	91,664
Chile	2.0	1.2	7.4	8.2	1.4	19%	17%	23,556	58,260
Czech	2.0	1.5	15.7	20.8	1.3	31%	19%	17,862	50,681
Colombia	2.0	2.4	7.8	8.2	1.5	38%	13%	131,514	365,985
Ireland	2.0	2.3	6.2	6.8	1.4	43%	20%	41,981	100,409
Nicaragua	1.9	1.7	10.9	17.3	1.4	36%	11%	54,305	162,528
Honduras	1.9	2.1	4.7	5.4	1.7	42%	7%	87,059	210,264
El Salvador	1.9	2.0	4.1	7.2	1.7	54%	8%	260,256	630,779
Mexico	1.8	2.4	4.7	4.9	1.9	83%	8%	2,764,037	6,335,953
United Kingdom	1.7	1.9	4.8	5.0	1.5	26%	16%	157,918	519,789
Jamaica	1.7	2.1	7.2	7.9	1.5	50%	11%	124,948	409,092
Puerto Rico	1.7	1.5	6.9	14.0	1.3	62%	7%	217,852	807,876
Spain	1.7	1.3	6.1	11.5	1.5	32%	17%	22,939	77,558
Poland	1.6	1.8	5.2	5.5	1.5	55%	15%	117,444	307,017
Germany	1.6	1.0	9.0	14.8	1.3	33%	17%	98,598	725,051
Canada	1.5	1.7	3.9	5.6	1.4	25%	17%	151,273	567,555
Cuba	1.4	1.3	4.2	4.4	1.3	56%	17%	196,375	566,413

Notes: See Table 1.

Table A1b: Ethnic group tabulations

Ethnic group	Industry with most self-employed workers	Industry of max overage for self-employed	Industry with most workers
Nepal	Retail trade, n.s.	Textile mill products	Educational institutions
Senegal	Taxicab service	Apparel, fabrics, and notions	Eating and drinking places
Zimbabwe	Physicians & health practitioners	Leather and leather products	Educational institutions
Yemen	Grocery stores	Grocery stores	Grocery stores
Ethiopia	Taxicab service	Taxicab service	Taxicab service
Eritrea	Taxicab service	Taxicab service	Taxicab service
Somalia	Taxicab service	Taxicab service	Taxicab service
Oceania, ns/nec	Retail trade, n.s.	Petroleum products	Eating and drinking places
India: Gujarati	Hotels and motels	Hotels and motels	Hotels and motels
Sudan	Taxicab service	Misc. merchandise stores	Grocery stores
Cape Verde	Construction	Elementary and secondary schools	Construction
Uganda	Hotels and motels	Drug stores	Hospitals
Bermuda	Construction	Food stores, n.e.c.	Construction
Bangladesh	Taxicab service	Taxicab service	Eating and drinking places
Sierra Leone	Taxicab service	Taxicab service	Nursing and personal care facilities
Kuwait	Grocery stores	Gasoline service stations	Grocery stores
St. Kitts-Nevis	Taxicab service	Taxicab service	Construction
Saudi Arabia	Grocery stores	Jewelry stores	Educational institutions
Liberia	Taxicab service	Residential care facilities	Educational institutions
American Samoa	Services to dwellings	Taxicab service	Construction
Ghana	Taxicab service	Taxicab service	Hospitals
South America, ns	Construction	Professional/photographic equipment	Construction
Cameroon	Taxicab service	Drug stores	Hospitals
Nigeria	Taxicab service	Taxicab service	Hospitals
Chaldean	Grocery stores	Liquor stores	Grocery stores
Dominica	Construction	Direct selling establishments	Construction
Bahamas	Construction	Hospitals	Construction
Tanzania	Retail trade, n.s.	Liquor stores	Educational institutions
Haiti	Taxicab service	Taxicab service	Eating and drinking places
Americas, ns	Construction	Electrical repair shops	Construction
Partial Nordic Region	Construction	Paper and allied products	Construction
Singapore	Business services, n.e.c.	Food stores, n.e.c.	Educational institutions
Belgium	Legal services	Residential care facilities	Educational institutions
Morocco	Construction	Bus service and urban transit	Eating and drinking places
Pakistan	Taxicab service	Taxicab service	Taxicab service
Afghanistan	Taxicab service	Taxicab service	Eating and drinking places
Polynesia	Landscaping	Communications	Construction
India: Punjabi	Taxicab service	Taxicab service	Taxicab service
Cyprus	Eating and drinking places	Book and stationery stores	Eating and drinking places
Africa, ns/nec	Taxicab service	Taxicab service	Eating and drinking places
Antigua-Barbuda	Construction	Professional/photographic equipment	Construction
Baltic States	Construction	Personnel supply services	Construction
Dominican Republic	Taxicab service	Taxicab service	Construction
Indochina, ns	Grocery stores	Shoe stores	Grocery stores
Iraq	Grocery stores	Liquor stores	Construction
Jordan	Grocery stores	Grocery stores	Grocery stores
Korea	Laundry, cleaning, and garment services	Laundry, cleaning, and garment services	Eating and drinking places
Sri Lanka (Ceylon)	Physicians & health practitioners	Physicians & health practitioners	Educational institutions
Melanesia	Landscaping	Hotels and motels	Construction
Cambodia	Eating and drinking places	Retail bakeries	Eating and drinking places
Syria	Physicians & health practitioners	Liquor stores	Hospitals
Former USSR	Construction	Jewelry stores	Construction
Bulgaria	Construction	Elementary and secondary schools	Educational institutions
Norway	Construction	Child care services	Construction
Grenada	Construction	Misc. merchandise stores	Construction
Burma (Myanmar)	Eating and drinking places	Professional/photographic equipment	Eating and drinking places
Kenya	Physicians & health practitioners	Hotels and motels	Educational institutions
Northern Ireland	Construction	Elementary and secondary schools	Construction
Greece	Eating and drinking places	Eating and drinking places	Eating and drinking places
Other Caribbean	Construction	Educational institutions	Construction
Europe, ns.	Construction	Social services, n.e.c.	Construction
St. Vincent	Construction	Laundry, cleaning, and garment services	Construction
Panama	Construction	Shoe stores	Construction
Armenian	Construction	Leather and leather products	Construction
Denmark	Construction	Paper and allied products	Construction
Thailand	Eating and drinking places	Eating and drinking places	Eating and drinking places
Austria	Construction	Apparel and other finished textile products	Construction
India: Other	Physicians & health practitioners	Hotels and motels	Computer and data processing services
U.S. Virgin Islands	Construction	Apparel and accessory stores, except shoe	Construction

Table A1b: Ethnic group tabulations

Ethnic group	Industry with most self-employed workers	Industry of max overage for self-employed	Industry with most workers
New Zealand	Construction	Transportation equipment	Construction
Albania	Construction	Food stores, n.e.c.	Eating and drinking places
China	Eating and drinking places	Eating and drinking places	Eating and drinking places
Turkey	Construction	Leather and leather products	Educational institutions
Barbados	Construction	Educational services	Construction
Paraguay	Construction	Miscellaneous vehicle dealers	Construction
Vietnam	Misc. personal services	Misc. personal services	Electrical machinery and equipment
St. Lucia	Construction	Catalog and mail order houses	Construction
Belize/British Honduras	Construction	Banking	Construction
Lebanon	Construction	Gasoline service stations	Eating and drinking places
Northern Africa	Eating and drinking places	Taxicab service	Eating and drinking places
Laos	Agricultural production, crops	Textile mill products	Machinery and computing equipment
Romania	Construction	Residential care facilities	Construction
Sweden	Construction	Furniture and home furnishings	Construction
Croatia	Construction	Residential care facilities	Construction
Malaysia	Eating and drinking places	R&D and testing services	Eating and drinking places
Indonesia	Eating and drinking places	Furniture and home furnishings	Eating and drinking places
Switzerland	Misc. professional services	Farm supplies	Educational institutions
Ecuador	Construction	Taxicab service	Construction
Guam	Construction	Misc entertainment and recreation services	Construction
Bolivia	Construction	Residential care facilities	Construction
Ukraine	Construction	Taxicab service	Construction
Italy	Construction	Barber shops	Construction
Uruguay	Construction	Gasoline service stations	Construction
Taiwan	Eating and drinking places	Wholesale trade, n.s.	Electrical machinery and equipment
Hong Kong and Macau	Eating and drinking places	Textile mill products	Eating and drinking places
Brazil	Construction	Private households	Construction
Guyana/British Guiana	Construction	Metals and minerals, except petroleum	Construction
USSR/Russia	Construction	Taxicab service	Construction
France	Construction	Retail bakeries	Eating and drinking places
Iran	Construction	Apparel, fabrics, and notions	Eating and drinking places
Venezuela	Construction	Residential care facilities	Construction
Philippines	Physicians & health practitioners	Nursing and personal care facilities	Hospitals
Japan	Eating and drinking places	Museums, art galleries, and zoos	Eating and drinking places
South Africa (Union of)	Physicians & health practitioners	Metals and minerals, except petroleum	Computer and data processing services
Israel/Palestine	Construction	Sewing, needlework, and piece goods stores	Construction
Trinidad and Tobago	Construction	Child care services	Construction
Netherlands	Construction	Drugs, chemicals, and allied products	Educational institutions
Portugal	Construction	Retail bakeries	Construction
Costa Rica	Construction	Leather and leather products	Construction
Australia	Construction	Theaters and video rental	Educational institutions
Hungary	Construction	Drugs, chemicals, and allied products	Construction
Yugoslavia	Construction	Furniture and home furnishings	Construction
Guatemala	Construction	Private households	Construction
Peru	Construction	Nursing and personal care facilities	Construction
Argentina	Construction	Child care services	Construction
Chile	Construction	Chemicals and allied products	Construction
Czech	Construction	Shoe stores	Construction
Colombia	Construction	Taxicab service	Construction
Ireland	Construction	Farm-product raw materials	Construction
Nicaragua	Construction	Bus service and urban transit	Construction
Honduras	Construction	Private households	Construction
El Salvador	Construction	Private households	Construction
Mexico	Construction	Landscaping	Construction
United Kingdom	Construction	Educational institutions	Construction
Jamaica	Construction	Misc. merchandise stores	Construction
Puerto Rico	Construction	Paper and allied products	Construction
Spain	Construction	Alcoholic beverages	Construction
Poland	Construction	Museums, art galleries, and zoos	Construction
Germany	Construction	Museums, art galleries, and zoos	Construction
Canada	Construction	Furniture and home furnishings	Construction
Cuba	Construction	Shoe stores	Construction

Notes: See Table 1.

Table A2a: Pairwise correlations of various overage metrics

	Sample	Metric	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	Self-employed	Weighted average overage ratio across all industries [OVER1]	1							
(2)		Weighted average overage ratio in three largest industries [OVER2]	0.925	1						
(3)		Average of three largest overage ratios for ethnic group [OVER3]	0.887	0.802	1					
(4)		Largest overage ratio for ethnic group [OVER4]	0.826	0.668	0.920	1				
(5)	All workers	Weighted average overage ratio across all industries [OVER1]	0.693	0.749	0.600	0.467	1			
(6)		Weighted average overage ratio in three largest industries [OVER2]	0.547	0.656	0.456	0.305	0.889	1		
(7)		Average of three largest overage ratios for ethnic group [OVER3]	0.568	0.652	0.539	0.418	0.897	0.801	1	
(8)		Largest overage ratio for ethnic group [OVER4]	0.515	0.602	0.501	0.391	0.833	0.724	0.946	1

Notes: Table displays correlations between ethnic group overage measures calculated on both self-employment and industry total employment. All correlations are significant at a 5% level.

Table A2b: Pairwise rank correlations of various overage metrics

	Sample	Metric	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	Self-employed	Weighted average overage ratio across all industries [OVER1]	1							
(2)		Weighted average overage ratio in three largest industries [OVER2]	0.865	1						
(3)		Average of three largest overage ratios for ethnic group [OVER3]	0.894	0.707	1					
(4)		Largest overage ratio for ethnic group [OVER4]	0.858	0.647	0.969	1				
(5)	All workers	Weighted average overage ratio across all industries [OVER1]	0.817	0.809	0.667	0.601	1			
(6)		Weighted average overage ratio in three largest industries [OVER2]	0.509	0.583	0.375	0.282	0.762	1		
(7)		Average of three largest overage ratios for ethnic group [OVER3]	0.739	0.768	0.665	0.619	0.849	0.521	1	
(8)		Largest overage ratio for ethnic group [OVER4]	0.696	0.738	0.628	0.587	0.802	0.476	0.968	1

Notes: See Table A2a. Table displays rank correlations between ethnic group overage measures calculated on both self-employment and industry total employment. All correlations are significant at a 5% level.

Table A3a: OLS estimations of overage metric designs and non-parametric forms without controls

	Weighted average overage across all industries [OVER1]	Weighted average overage using three largest industries for ethnic group [OVER2]	Average of three largest overage ratios for ethnic group [OVER3]	Largest overage ratio for ethnic group [OVER4]
	(1)	(2)	(3)	(4)
(0,1) Indicator: ethnic size in smallest third x	1.758+++	1.458+++	1.377+++	1.225+++
(0,1) Indicator: ethnic isolation in highest third	(0.451)	(0.447)	(0.404)	(0.443)
(0,1) Indicator: ethnic size in smallest third x	1.178+++	0.693+++	0.943+++	0.899+++
(0,1) Indicator: ethnic isolation in middle third	(0.220)	(0.252)	(0.209)	(0.249)
(0,1) Indicator: ethnic size in smallest third x	1.132+++	0.728++	1.509+++	1.378+++
(0,1) Indicator: ethnic isolation in lowest third	(0.296)	(0.348)	(0.336)	(0.326)
(0,1) Indicator: ethnic size in middle third x	0.844+++	0.946+++	1.011+++	0.829+++
(0,1) Indicator: ethnic isolation in highest third	(0.211)	(0.281)	(0.249)	(0.195)
(0,1) Indicator: ethnic size in middle third x	0.429+	0.345	0.735+++	0.787+++
(0,1) Indicator: ethnic isolation in middle third	(0.244)	(0.307)	(0.250)	(0.245)
(0,1) Indicator: ethnic size in middle third x	0.039	-0.170	0.208	0.205
(0,1) Indicator: ethnic isolation in lowest third	(0.111)	(0.173)	(0.211)	(0.151)
(0,1) Indicator: ethnic size in largest third x	0.289+	0.396	0.321	0.228
(0,1) Indicator: ethnic isolation in highest third	(0.167)	(0.258)	(0.241)	(0.165)
(0,1) Indicator: ethnic size in largest third x	-0.073	-0.107	-0.094	-0.077
(0,1) Indicator: ethnic isolation in middle third	(0.096)	(0.174)	(0.166)	(0.112)
(0,1) Indicator: ethnic size in largest third x (0,1) Indicator: ethnic isolation in lowest third	Excluded group			
Adjusted R-Squared value	0.302	0.182	0.265	0.224

Notes: See Tables 1 and 3. Effects are measured relative to largest and least isolated ethnic groups.

Table A3b: OLS estimations of overage metric designs and non-parametric forms with controls

	Weighted average overage across all industries [OVER1]	Weighted average overage using three largest industries for ethnic group [OVER2]	Average of three largest overage ratios for ethnic group [OVER3]	Largest overage ratio for ethnic group [OVER4]
	(1)	(2)	(3)	(4)
(0,1) Indicator: ethnic size in smallest third x (0,1) Indicator: ethnic isolation in highest third	1.638+++ (0.399)	1.299+++ (0.407)	1.343+++ (0.401)	1.194++ (0.463)
(0,1) Indicator: ethnic size in smallest third x (0,1) Indicator: ethnic isolation in middle third	0.943+++ (0.184)	0.424+ (0.226)	0.795+++ (0.212)	0.773+++ (0.226)
(0,1) Indicator: ethnic size in smallest third x (0,1) Indicator: ethnic isolation in lowest third	1.079+++ (0.259)	0.682++ (0.296)	1.488+++ (0.317)	1.350+++ (0.316)
(0,1) Indicator: ethnic size in middle third x (0,1) Indicator: ethnic isolation in highest third	0.859+++ (0.241)	0.847+++ (0.309)	1.002+++ (0.282)	0.867+++ (0.272)
(0,1) Indicator: ethnic size in middle third x (0,1) Indicator: ethnic isolation in middle third	0.414+ (0.211)	0.292 (0.269)	0.711+++ (0.231)	0.777+++ (0.233)
(0,1) Indicator: ethnic size in middle third x (0,1) Indicator: ethnic isolation in lowest third	-0.049 (0.096)	-0.236 (0.145)	0.154 (0.191)	0.134 (0.140)
(0,1) Indicator: ethnic size in largest third x (0,1) Indicator: ethnic isolation in highest third	0.683+++ (0.212)	0.726+++ (0.271)	0.670++ (0.259)	0.561++ (0.231)
(0,1) Indicator: ethnic size in largest third x (0,1) Indicator: ethnic isolation in middle third	0.163 (0.160)	0.108 (0.206)	0.153 (0.197)	0.128 (0.164)
(0,1) Indicator: ethnic size in largest third x (0,1) Indicator: ethnic isolation in lowest third	Excluded group			
Adjusted R-Squared value	0.419	0.303	0.335	0.269

Notes: See Tables 1 and 3. Effects are measured relative to largest and least isolated ethnic groups. Estimations include controls for ethnic group traits.

Table A4: IV estimations with in-marriage rates in Spain 2011

	Instrumenting with predicted ethnic group size from gravity model and in-marriage rates in Spain 2011			Instrumenting with predicted ethnic group size from gravity model and in-marriage rates in UK 1991 and Spain 2011		
	First stage for size	First stage for isolation	Second stage	First stage for size	First stage for isolation	Second stage
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Instrument for size	0.624+++	0.008		0.633+++	-0.087	
	(0.067)	(0.119)		(0.075)	(0.125)	
Instrument for isolation Spain	-0.092	0.312+++		-0.059	0.135	
	(0.067)	(0.103)		(0.068)	(0.083)	
Instrument for isolation UK				-0.114+	0.491+++	
				(0.064)	(0.088)	
F-Statistic	50.6	5.0		27.1	26.4	
Inverse of log ethnic group size			0.478+++			0.456+++
			(0.153)			(0.132)
Isolation of ethnic group			0.583+++			0.350+++
			(0.219)			(0.082)
Exogeneity test p-value			0.044			0.058
Overid test p-value						0.199
	B. Including controls for ethnic group's traits					
Instrument for size	0.519+++	-0.060		0.500+++	-0.073	
	(0.094)	(0.100)		(0.098)	(0.094)	
Instrument for isolation Spain	-0.035	0.164+		-0.023	0.051	
	(0.065)	(0.096)		(0.062)	(0.083)	
Instrument for isolation UK				-0.071	0.344+++	
				(0.073)	(0.063)	
F-Statistic	15.4	1.7		9.0	11.7	
Inverse of log ethnic group size			0.407+			0.296++
			(0.238)			(0.138)
Isolation of ethnic group			1.212+			0.525+++
			(0.638)			(0.145)
Exogeneity test p-value			0.014			0.074
Overid test p-value						0.110

Notes: See Table 6. Regressions with the Spain 2011 and dual instruments have 130 and 129 observations, respectively. Regressions cluster standard errors by the ethnic groups in the Spain 2011 dataset.

Table A5a: Robustness checks on IV estimations, UK IV only

	Baseline estimation	Without sample weights	Without winsorization	Using bootstrapped standard errors	Isolation IV only with expected overage control	Dual IV with expected overage control
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Inverse of log ethnic group size	0.459+++	0.480+++	0.443+++	0.480+++	0.384+++	0.082
(small groups have larger values)	(0.130)	(0.138)	(0.116)	(0.142)	(0.123)	(0.220)
Isolation of ethnic group	0.316++	0.308++	0.311++	0.308	0.298+	0.252
	(0.125)	(0.135)	(0.121)	(0.201)	(0.153)	(0.187)
	B. Including controls for ethnic group's traits					
Inverse of log ethnic group size	0.294++	0.295+	0.298++	0.295	0.297+++	-0.115
(small groups have larger values)	(0.143)	(0.157)	(0.124)	(0.279)	(0.106)	(0.224)
Isolation of ethnic group	0.484+++	0.456++	0.495+++	0.456	0.469+++	0.348++
	(0.153)	(0.180)	(0.159)	(0.354)	(0.130)	(0.170)

Notes: See Table 6.

Table A5b: Robustness checks on IV estimations, combined UK and Spain IV

	Baseline estimation	Without sample weights	Without winsorization	Using bootstrapped standard errors	Isolation IV only with expected overage control	Dual IV with expected overage control
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Inverse of log ethnic group size	0.469+++	0.487+++	0.454+++	0.487+++	0.449+++	0.062
(small groups have larger values)	(0.140)	(0.144)	(0.125)	(0.163)	(0.143)	(0.229)
Isolation of ethnic group	0.419+++	0.421+++	0.411+++	0.421+++	0.543+++	0.476+++
	(0.119)	(0.126)	(0.104)	(0.133)	(0.127)	(0.163)
	B. Including controls for ethnic group's traits					
Inverse of log ethnic group size	0.342++	0.344+	0.347++	0.344	0.343++	-0.082
(small groups have larger values)	(0.174)	(0.186)	(0.156)	(0.525)	(0.134)	(0.195)
Isolation of ethnic group	0.728+++	0.714+++	0.717+++	0.714	0.813+++	0.668+++
	(0.209)	(0.214)	(0.182)	(1.315)	(0.237)	(0.248)

Notes: See Table 6.

Table A6a: Robustness checks on IV estimations, UK IV only

	Baseline estimation with OVER1	Using total worker sample	Excluding natives from denominator shares	Imposing min counts on ethnic industry presence	Excluding new arrivals over the prior five years	Excluding the taxi industry
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Inverse of log ethnic group size	0.459+++	0.276+	0.606+++	0.582+++	0.323++	0.404+++
(small groups have larger values)	(0.130)	(0.164)	(0.154)	(0.156)	(0.152)	(0.137)
Isolation of ethnic group	0.316++	0.612+++	0.386+	0.347	0.535+++	0.497++
	(0.125)	(0.213)	(0.207)	(0.218)	(0.204)	(0.202)
	B. Including controls for ethnic group's traits					
Inverse of log ethnic group size	0.294++	0.043	0.381+++	0.373++	0.162	0.382+++
(small groups have larger values)	(0.143)	(0.153)	(0.144)	(0.145)	(0.139)	(0.138)
Isolation of ethnic group	0.484+++	0.700+++	0.373++	0.355++	0.712+++	0.613+++
	(0.153)	(0.146)	(0.152)	(0.176)	(0.158)	(0.189)

Notes: See Table 6.

Table A6b: Robustness checks on IV estimations, combined UK and Spain IV

	Baseline estimation with OVER1	Using total worker sample	Excluding natives from denominator shares	Imposing min counts on ethnic industry presence	Excluding new arrivals over the prior five years	Excluding the taxi industry
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Inverse of log ethnic group size	0.469+++	0.297+	0.613+++	0.593+++	0.344++	0.417+++
(small groups have larger values)	(0.140)	(0.178)	(0.160)	(0.164)	(0.165)	(0.141)
Isolation of ethnic group	0.419+++	0.798+++	0.462++	0.464++	0.718+++	0.618+++
	(0.119)	(0.198)	(0.210)	(0.217)	(0.181)	(0.179)
	B. Including controls for ethnic group's traits					
Inverse of log ethnic group size	0.342++	0.086	0.386++	0.390++	0.209	0.380++
(small groups have larger values)	(0.174)	(0.157)	(0.159)	(0.156)	(0.156)	(0.151)
Isolation of ethnic group	0.728+++	0.920+++	0.401++	0.446++	0.952+++	0.602+++
	(0.209)	(0.212)	(0.203)	(0.219)	(0.283)	(0.220)

Notes: See Table 6.

Table A7a: IV estimations of overage metric designs, UK IV only

	Weighted average overage across all industries [OVER1]	Weighted average overage using three largest industries for ethnic group [OVER2]	Average of three largest overage ratios for ethnic group [OVER3]	Largest overage ratio for ethnic group [OVER4]
	(1)	(2)	(3)	(4)
	A. Baseline estimation without controls			
Inverse of log ethnic group size	0.459+++	0.331++	0.392+++	0.394+++
(small groups have larger values)	(0.130)	(0.149)	(0.145)	(0.143)
Isolation of ethnic group	0.316++	0.483+++	0.179	0.093
	(0.125)	(0.145)	(0.163)	(0.127)
Exogeneity test p-value	0.140	0.145	0.122	0.276
	B. Including controls for ethnic group's traits			
Inverse of log ethnic group size	0.294++	0.125	0.160	0.242
(small groups have larger values)	(0.143)	(0.141)	(0.175)	(0.175)
Isolation of ethnic group	0.484+++	0.645+++	0.215	0.174
	(0.153)	(0.134)	(0.224)	(0.206)
Exogeneity test p-value	0.237	0.141	0.082	0.270

Notes: See Table 6.

Table A7b: IV estimations of overage metric designs, combined UK and Spain IV

	Weighted average overage across all industries [OVER1]	Weighted average overage using three largest industries for ethnic group [OVER2]	Average of three largest overage ratios for ethnic group [OVER3]	Largest overage ratio for ethnic group [OVER4]
	(1)	(2)	(3)	(4)
	A. Baseline estimation without controls			
Inverse of log ethnic group size	0.469+++	0.349++	0.411+++	0.409+++
(small groups have larger values)	(0.140)	(0.161)	(0.151)	(0.147)
Isolation of ethnic group	0.419+++	0.639+++	0.379++	0.256+
	(0.119)	(0.140)	(0.176)	(0.137)
Exogeneity test p-value	0.023	0.000	0.007	0.023
	B. Including controls for ethnic group's traits			
Inverse of log ethnic group size	0.342++	0.182	0.232	0.306+
(small groups have larger values)	(0.174)	(0.172)	(0.174)	(0.182)
Isolation of ethnic group	0.728+++	0.933+++	0.588++	0.501+++
	(0.209)	(0.231)	(0.271)	(0.193)
Exogeneity test p-value	0.006	0.002	0.025	0.055

Notes: See Table 6.

Table A8a: IV results with alternative gravity model designs for predicted size, UK IV only

	Baseline estimation	Including border in the gravity model	Including distance squared in the gravity model	Using distance and population as instruments	Using distance, population, and border as instruments	Using distance, population, and distance squared as instruments
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Inverse of log ethnic group size (small groups have larger values)	0.459+++ (0.130)	0.451+++ (0.125)	0.465+++ (0.132)	0.480+++ (0.135)	0.481+++ (0.136)	0.508+++ (0.139)
Isolation of ethnic group	0.316++ (0.125)	0.314++ (0.126)	0.316++ (0.125)	0.379+++ (0.113)	0.390+++ (0.101)	0.420+++ (0.100)
	B. Including controls for ethnic group's traits					
Inverse of log ethnic group size (small groups have larger values)	0.294++ (0.143)	0.297++ (0.136)	0.298++ (0.143)	0.343++ (0.159)	0.344++ (0.158)	0.350++ (0.158)
Isolation of ethnic group	0.484+++ (0.153)	0.484+++ (0.156)	0.485+++ (0.152)	0.644+++ (0.145)	0.638+++ (0.143)	0.665+++ (0.135)

Notes: See Table 6.

Table A8b: IV results with alternative gravity model designs for predicted size, combined UK and Spain IV

	Baseline estimation	Including border in the gravity model	Including distance squared in the gravity model	Using distance and population as instruments	Using distance, population, and border as instruments	Using distance, population, and distance squared as instruments
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Baseline estimation without controls					
Inverse of log ethnic group size (small groups have larger values)	0.469+++	0.463+++	0.475+++	0.488+++	0.489+++	0.517+++
	(0.140)	(0.138)	(0.140)	(0.143)	(0.144)	(0.145)
Isolation of ethnic group	0.419+++	0.416+++	0.422+++	0.428+++	0.439+++	0.501+++
	(0.119)	(0.120)	(0.118)	(0.120)	(0.117)	(0.118)
	B. Including controls for ethnic group's traits					
Inverse of log ethnic group size (small groups have larger values)	0.342++	0.344++	0.343++	0.384++	0.385++	0.389++
	(0.174)	(0.171)	(0.172)	(0.186)	(0.186)	(0.180)
Isolation of ethnic group	0.728+++	0.729+++	0.728+++	0.829+++	0.824+++	0.847+++
	(0.209)	(0.208)	(0.208)	(0.252)	(0.255)	(0.229)

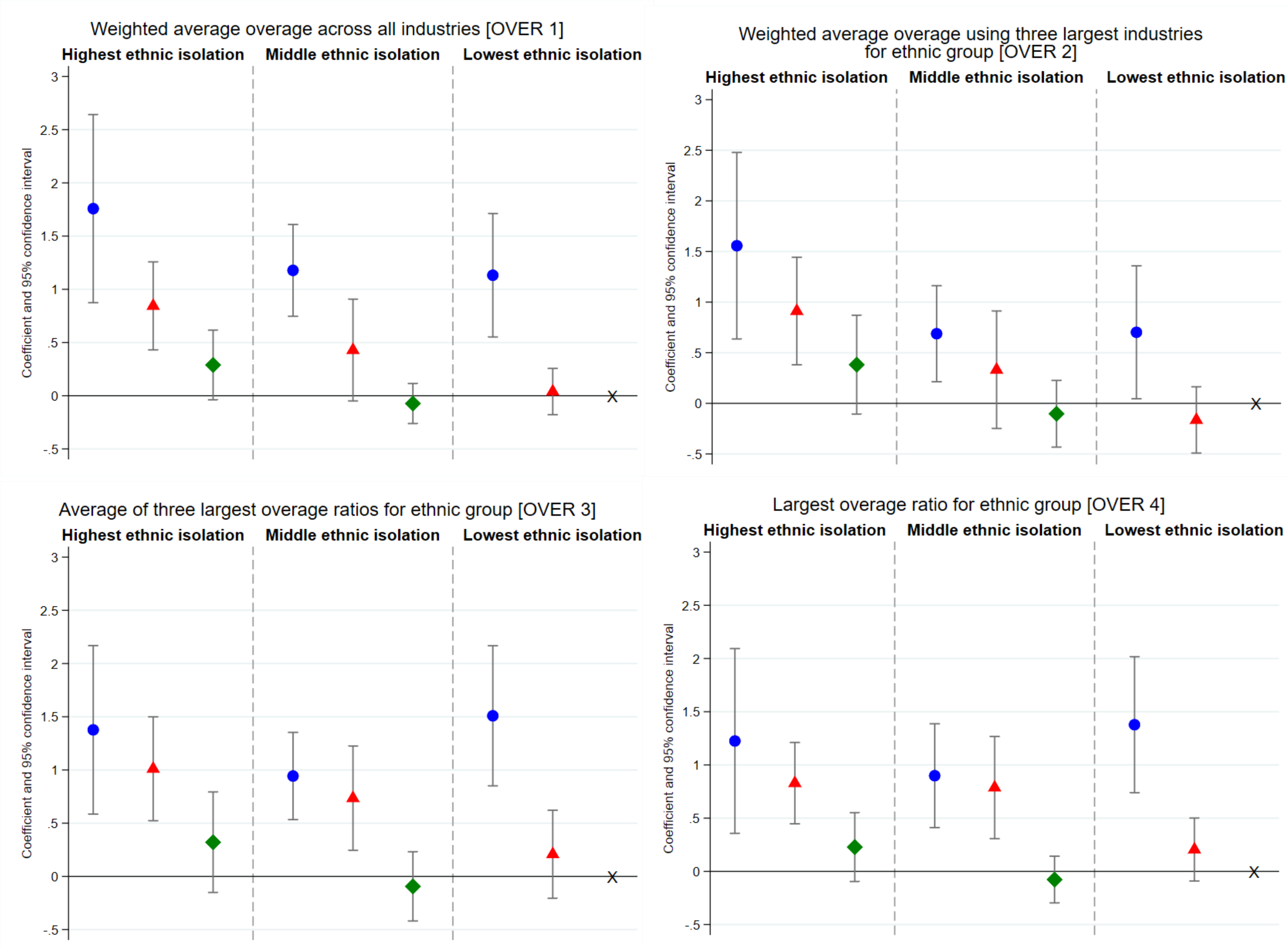
Notes: See Table 6.

Table A9: OLS estimations of individual incomes and group concentration

	Log yearly income in 2000					
	(1)	(2)	(3)	(4)	(5)	(6)
(0,1) Self employed worker	0.049+++	0.014	0.015	0.027	0.024	0.020
	(0.013)	(0.013)	(0.013)	(0.020)	(0.020)	(0.019)
Percentage of individual's group who are self-employed in the industry	0.057+++ (0.013)	0.027++ (0.012)	0.020 (0.013)			
x (0,1) Self employed worker	0.001	-0.003	0.000		0.004	-0.027
	(0.012)	(0.010)	(0.013)		(0.014)	(0.017)
Percentage of individual's group who are self-employed		0.059+++ (0.007)	0.062+++ (0.007)			
x (0,1) Self employed worker		0.042+++	0.042+++			0.080+++
		(0.008)	(0.010)			(0.015)
Percentage of individual's group who are working in the industry			0.023+ (0.012)			
x (0,1) Self employed worker			-0.000			0.029
			(0.015)			(0.021)
Person-level Traits FE	Yes	Yes	Yes	Yes	Yes	Yes
MSA-Industry FE	Yes	Yes	Yes			
MSA-Industry-Ethnicity FE				Yes	Yes	Yes
Adjusted R-Squared Value	0.246	0.248	0.248	0.308	0.308	0.308
Observations	404,467	404,467	404,467	404,467	404,467	404,467

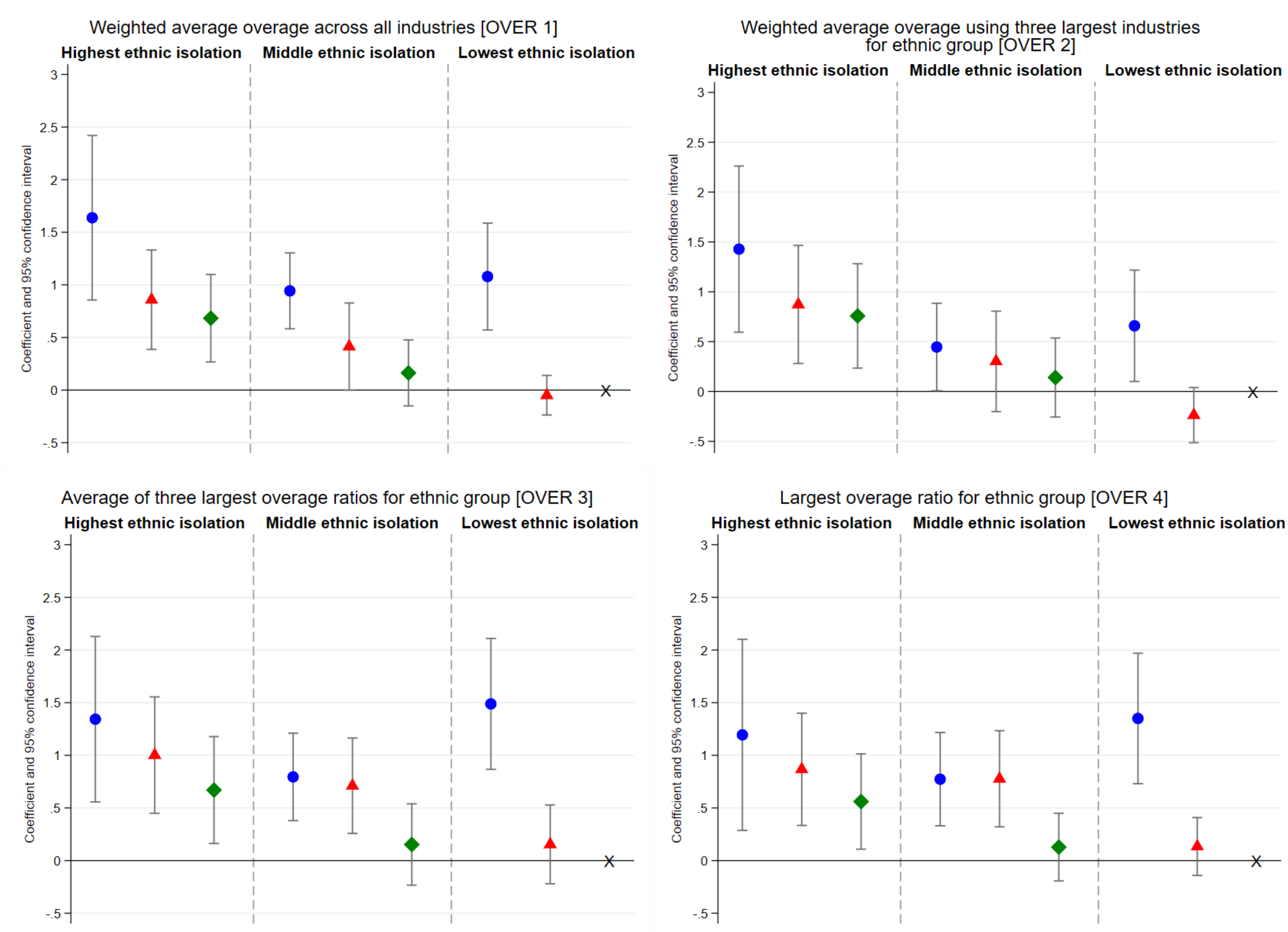
Notes: See Table 7.

App. Figure 1a: Non-parametric estimations without controls



Notes: See App. Table 3a. Ethnic size in smallest third; Ethnic size in middle third; Ethnic size in largest third

App. Figure 1b: Non-parametric estimations with controls



Notes: See App. Table 3b. ● Ethnic size in smallest third ▲ Ethnic size in middle third ◆ Ethnic size in largest third

Online Appendix: Theory

The theory in this paper consists of two fundamental building blocks. First, social interactions and production are complementary. Second, different social relationships are not close substitutes for one another. The former is analyzed in the main text, and this appendix begins with additional discussion. We then consider pricing equilibrium and social networks with endogenous matching. The numbering of assumptions and propositions continues from the main text.

1 Discussion of Baseline Model

1.1 Quality and Convex Productivity

In addition to the quantity of social interactions with other self-employed entrepreneurs, the quality of these interactions could also matter for productivity. Let individual productivity for self-employed entrepreneurs in industry 1 increase both in the quantity and average productivity of other entrepreneurs in the sector of the same group. Write this as

$$\theta = \phi + \delta X_l \overline{\theta}, \tag{1}$$

where $\phi > 0$ is a productivity term, $0 < \delta < 1$ is a social multiplier, X_l is the fraction of entrepreneurs in group l, and $\overline{\theta}$ is the average productivity of these entrepreneurs. Solving for equilibrium productivity by setting θ equal to $\overline{\theta}$, individual productivity in group l is a function:

$$\theta\left(X_l\right) = \frac{\phi}{1 - \delta X_l}. \tag{2}$$

Under these conditions, productivity is convex in the degree of specialization when taking both the quantity and the quality of interaction into account.[1] With this result in mind, we make the following assumption:

Assumption 1B *Productivity of self-employed entrepreneurs in industry* 1 *is convex in specialization:* $\theta'' > 0$.

Assumption 1B allows a full characterization of the efficient solution without having to resort to explicit functional form. We discuss further below. Convex productivity gives the following result:

[1]This specification highlights the differences from a standard interaction model. The standard model is generally specified so that individual productivity is a function of a group-specific term ϕ and the discounted mean of the group, $\delta\overline{\theta}$. Solving $\theta = \phi + \delta\overline{\theta}$, interaction exacerbates the difference in ϕ across groups, $\theta = \frac{\phi}{1-\delta} > \phi$, but the degree of specialization X_l has no effect on productivity.

Lemma *If productivity is convex, both groups never work in both industries.*

Proof: Assume by contradiction that an efficient distribution (X_A, X_B) exists where $0 < X_l < 1$ for $l = \{A, B\}$. Consider a marginal change ϵ in the ethnic composition of self-employed entrepreneurs in industry 1 while holding fixed the overall number of said entrepreneurs M (and therefore also the outputs of both industries). Taking the derivative of Q_1 with respect to ϵ, and evaluating it at $\epsilon = 0$:

$$\frac{\partial Q_1}{\partial \epsilon}\left(X_A + \frac{\epsilon}{N_A}, X_B - \frac{\epsilon}{N_B}\right) = \theta(X_A) + X_A\theta'(X_A) - \theta(X_B) - X_B\theta'(X_B) \tag{3}$$

Since (X_A, X_B) is efficient, and since X_l is interior, this derivative has to be zero.[2] But with convex productivity the derivative is zero only at $X_A = X_B$, which is the global minimum. This contradicts efficiency. ■

The efficient economy aims for maximum ethnic homogeneity in self-employed entrepreneurship in industry 1. Ruling out that both groups work in both sectors implies that only the specialized distributions along the two curves depicted in Figure 1 of the main text could possibly coincide with the transformation frontier. The shape of the entire transformation frontier can therefore be deduced by tracing out the maximum of the two curves in that figure.

Proposition 2 *If productivity is convex, there is a cutoff value v^* such that for $v < v^*$, the minority group specializes as self-employed entrepreneurs in industry 1, whereas for $v > v^*$, the majority specializes.*

Proof: Direct from Proposition 1 and Lemma proofs with convexity. ■

The right panel of Figure 1 of the main text also shows how the degree of specialization varies with the size of industry 1, as governed by v, and the cutoff value v^* for majority group specialization. The greater the value of v, the greater is the demand for industry 1 and the more people work in it. As industry 1 increases in size, the interaction externality generates a characteristic discrete jump from one type of equilibrium to another. At the point v^*, where many from group B have also joined self-employed entrepreneurship in industry 1, the economy abruptly moves from minority specialization to majority specialization.

[2] If the derivative is nonzero, then the output of industry 1 could increase while keeping the output of industry 0 constant. By subsequently increasing the number of workers in industry 0 marginally, a Pareto improvement is feasible, thus contradicting efficiency.

1.2 The Case of Non-Convex Productivity

To see that convexity is needed for the Lemma on ethnic homogeneity to hold, consider a non-convex production function where a threshold fraction must work as self-employed entrepreneurs in industry 1 for interaction to have value: $\theta > 0$ if $X_l \geq b$ and zero otherwise. This specification violates the assumption that productivity is strictly increasing in the degree of specialization. Then, if the demand for industry 1 output is so great that a single group cannot satisfy it entirely, $v > V(0,1)$, and if in addition $V(b,b) < v < V(b,1)$, efficiency requires that both ethnic groups work in both industries, contradicting the Lemma.

To see why, consider what would happen if one of the groups specialized completely. In this case the non-specialized group's degree of specialization would be positive but below b, causing the self-employed industry 1 entrepreneurs in that group to have zero productivity. If, however, the industrial distribution was unspecialized instead, with $X_A = X_B$, then self-employed industry 1 entrepreneurs in both groups would be as productive as those in the most productive group were under the alternative. Clearly this would be Pareto superior, contradicting the Lemma. This special case shows how the Lemma fails for non-convex productivity, and how in this case the qualitative features of specialization will depend on specific functional form assumptions. Recall however that the results for both $v \leq V(1,0)$ and $v = V(0,1)$ are more general and apply both for convex and non-convex productivity. This condition is less important for the remaining model discussion.

2 The Price Equilibrium

The model in the main text characterizes the efficient outcome. The focus now turns to the competitive outcome. An equilibrium analysis will yield two insights into how social interaction affects distribution over industries. First, it shows how stratifying forces act to make groups more and more different, and second, how group earnings are positively related to the degree of specialization.

To see how social interaction works as a stratifying force, begin by introducing time into the analysis, with $t = 0, 1, ..., \infty$. Dynamics are built into the model by making the interaction effect work with a lag. Denote by X_l^t the degree of specialization in period t for group l, and let self-employed individual entrepreneurial productivity in industry 1 in period t be a function $\theta\left(X_l^{t-1}\right)$. This one-period lag specification for the interaction effect could easily be generalized to a distributed lag. Interaction now effectively works as a form of social capital, with the group's self-employment activities in the previous period benefiting individual productivity today. Let p_1^t and p_0^t be the prices of industry 1 output and industry 0 output respectively. Entrepreneurial earnings in industry 1 are $y_{1,l}^t = p_1^t \theta\left(X_l^{t-1}\right)$ and worker earnings in industry 0 are $y_{0,l}^t = p_0^t$. Competitive industrial choice is straightforward to derive in this setting; defining the relative price of industry 0 output to industry 1 output as $p^t = \frac{p_0^t}{p_1^t}$, an individual in group l joins industry 1 as a self-employed entrepreneur if

$$\theta\left(X_l^{t-1}\right) \geq p^t \tag{4}$$

and favors being a worker in industry 0 if $\theta\left(X_l^{t-1}\right) \leq p^t$. Since individuals have identical skills, aggregate labor supply for group l is discontinuous, with:

$$X_l^t = \begin{cases} 1 & \text{if } \theta\left(X_l^{t-1}\right) > p^t \\ [0,1] & \text{if } \theta\left(X_l^{t-1}\right) = p^t \\ 0 & \text{if } \theta\left(X_l^{t-1}\right) < p^t. \end{cases} \tag{5}$$

Avoid for now the knife-edge *unspecialized* case where $X_A^{t-1} = X_B^{t-1}$. Since there is a single price of labor, p^t, at least one of the two groups A and B must then be in a corner:

$$\left(X_A^t, X_B^t\right) = \begin{cases} \left(X_A^t = 1, 0 < X_B^t\right) \text{ or } \left(X_A^t \leq 1, X_B^t = 0\right) & \text{if } X_A^{t-1} > X_B^{t-1} \\ \left(0 < X_A^t, X_B^t = 1\right) \text{ or } \left(X_A^t = 0, X_B^t \leq 1\right) & \text{if } X_A^{t-1} < X_B^{t-1} \end{cases} \tag{6}$$

In equilibrium, supply must satisfy (6) and production must meet demand so that markets clear. Because of perfect complementarity, meeting demand reduces to satisfying $v = V\left(X_A^t, X_B^t\right)$. The resulting equilibrium distribution is unique. To see why, take the case when group l is more specialized than group l' in the previous period, with $X_l^{t-1} > X_{l'}^{t-1}$. Given that at least one of the two groups must be in a corner according to (6), the equilibrium distribution must either be of the type $\left(X_l^t, 0\right)$ or of the type $\left(1, X_{l'}^t\right)$. Since the function V is strictly increasing in both arguments, it follows that $V\left(1, X_{l'}^t\right) > V\left(X_l^t, 0\right)$. Only one distribution can consequently make V equal to v.

The equilibrium distribution is therefore uniquely determined by the distribution in the previous period. Continuing to avoid the knife-edge unspecialized case, define a function ϕ that maps every previous distribution into a new distribution:

$$\left(X_A^t, X_B^t\right) = \phi\left(X_A^{t-1}, X_B^{t-1}\right) \tag{7}$$

Next, proceed to characterize stationary equilibrium distributions. Like other equilibrium distributions, stationary distributions must satisfy (6) and must meet demand. Following the same argument as above, based on V being strictly increasing in both arguments, it follows that there is a stationary equilibrium where each of the two groups specializes. Denote the stationary distribution as $\left(X_A^A, X_B^A\right)$ when the minority specializes, and the stationary distribution as $\left(X_A^B, X_B^B\right)$ when the majority specializes.

Finally, returning for a moment to the unspecialized knife-edge case where $X_A^{t-1} = X_B^{t-1}$, this type of initial condition is of measure zero and therefore not elaborated on. Note only that since V is strictly increasing in both arguments, there can only be one such stationary unspecialized equilibrium distribution. Denote that equilibrium distribution as $\left(X_A^U, X_B^U\right)$. In the unspecialized case, although there is only one stationary equilibrium, the uniqueness of equilibria no longer applies. To summarize, there are consequently three stationary equilibrium distributions: two specialized, $\left(X_A^A, X_B^A\right)$ and $\left(X_A^B, X_B^B\right)$, and one unspecialized, $\left(X_A^U, X_B^U\right)$. Figure A1 shows the two specialized equilibria, as well as the knife-edge equilibrium, when v is less than $V(1,0)$.

2.1 Industrial Stratification

Our next analysis shows that the dynamic system in (7) converges to a stationary specialized equilibrium, so long as the interaction externality is not too strong. This analysis only examines unspecialized initial conditions, which establishes convergence on measure one. Consider what happens to the aggregate production of industry 1 when one (infinitesimal) person in group l becomes a self-employed entrepreneur in that industry. First, aggregate production increases by an amount equal to the individual productivity of that person, $\theta(X_l)$. In addition, all other self-employed entrepreneurs in industry 1 from group l benefit from the interaction externality when socializing with this new entrepreneur. Individual productivity therefore increases by $\frac{1}{N_l}\theta'(X_l)$ for all $X_l N_l$ self-employed industry 1 entrepreneurs in group l. Consequently, the internalized effect on aggregate production of one person joining the self-employed entrepreneurial sector of industry 1 is $\theta(X_l)$, and the external effect is $X_l\theta'(X_l)$. Assume that the external effect is smaller than the internal effect.[3]

Assumption 2 *The internal effect dominates:* $\theta'(X_l)\,X_l < \theta(X_l)$.

This condition is satisfied if productivity is concave in X_l, but it also holds for some convexity as long as $\theta(0) > 0$. To see why the assumption is needed for the system to be stable, consider the extreme case when group A has no mass at all, with $N_A = 0$. Since the derivative of V with respect to X_A^t is zero in this case, group A can be ignored altogether in the general equilibrium analysis. There is then a single stationary level of specialization for group B; denote this value as X_B^*.

Consider a perturbation in period t so that the majority starts out with too many entrepreneurs in industry 1, $X_B^t > X_B^*$, shown in Figure A2. Such a deviation boosts the interaction effect in period $t+1$ relative to the stationary equilibrium, $\theta(X_B^t) > \theta(X_B^*)$. With perfect complementarity, the outputs of both industry 0 and industry 1 must therefore increase relative to their stationary equivalents. Increasing the output of industry 0 requires an increase in the number of workers in that industry, and consequently, a decrease in the number of self-employed entrepreneurs in industry 1 to

[3]We thank Rachel Soloveichik for this interpretation of Assumption 2.

below the stationary value X_B^*. With fewer of these entrepreneurs in period $t+1$ than the stationary number, the tables turn in period $t+2$, so that the interaction effect now is reduced to below that in the stationary equilibrium. Reducing the production of industry 0 and industry 1 in period $t+2$ in response, the number of industry 0 workers in period $t+2$ has to decrease and the number of self-employed industry 1 entrepreneurs has to increase relative to the stationary equilibrium. These reversals repeat every period in cobweb-style dynamics.[4]

The question of whether the system is stable reduces to whether the number of self-employed entrepreneurs in industry 1 in period $t+2$ is less than the number of such entrepreneurs in period t, so that the degree of specialization in group B gets closer and closer to the stationary value X_B^* over time. Using the derived direction of the change in industry 1 production, $Q_1^{t+1} > Q_1^{t+2}$, this latter inequality can be equivalently expressed, after multiplying and dividing the left-hand side by X_B^t and dividing both sides by $X_B^{t+1} N_B$, as:

$$X_B^t \frac{\theta\left(X_B^t\right)}{X_B^t} > X_B^{t+2} \frac{\theta\left(X_B^{t+1}\right)}{X_B^{t+1}} \tag{8}$$

Given that productivity is not too convex, as stipulated by Assumption 2, it follows that $\frac{\theta(X_l)}{X_l}$ is strictly decreasing in X_l. Since $X_B^t > X_B^{t+1}$, equation (8) then establishes that $X_B^t > X_B^{t+2}$. This proves convergence and the stability of group B's degree of specialization around X_B^*.

Having established stability in the case of $N_A = 0$, the same example also serves to show how the stratifying force comes into play. Let group B be in its stable state, with $X_B^t = X_B^*$, and perturb the minority's industry distribution so that $X_A^t > X_B^*$. Since group B is so much greater in size than group A, the former is unaffected by the perturbation and the price continues to be locked in at $p^{t+1} = \theta\left(X_B^*\right)$. The interaction effect in period $t+1$, generated by the perturbation in period t, then results in everyone in group A becoming more productive as self-employed entrepreneurs in industry 1 than as workers in industry 0, with $\theta\left(X_A^t\right) > p^{t+1}$. Group A's degree of specialization consequently jumps from X_A^t to $X_A^{t+1} = 1$, and the distribution stays in this stratified state forever. This stratification result is extended later for the general case of any population size of the two groups, and it follows that for $l \in \{A, B\}$ and $l' \in \{A, B\}$:

Proposition 3 *Initial differences result in long-run specialization: If group l is more specialized than group l' initially, $X_l^0 > X_{l'}^0$, then group l specializes in the long run and the limiting distribution is $\left(X_A^l, X_B^l\right)$.*

[4]The flip-flopping character of the equilibrium distribution is a result of the one-period lag specification for the interaction effect. The distribution would change more gradually with a more general specification allowing for distributed lags.

Proof: Consider the equilibrium sequence of industry distributions:

$$\left(\left(X_A^1, X_B^1\right), \left(X_A^2, X_B^2\right), ...\right) \tag{9}$$

If one group l is more specialized than the other group l' initially, $X_l^0 > X_{l'}^0$, supply in (5) requires that the equilibrium sequence begins in one of the following three ways:

$$\left(\left(X_l^1, X_{l'}^1\right), \left(X_l^2, X_{l'}^2\right), ...\right) = \begin{cases} \left((<1, 0), ...\right) \\ \left((1, \geq 0), (1, \geq 0), ...\right) \\ \left((1, \geq 0), (<1, 0), ...\right). \end{cases} \tag{10}$$

The proof proceeds by establishing that the sequence converges to $\left(X_A^l, X_B^l\right)$ in each of these three cases. Define the variable $\lambda\left(X_l\right) \equiv \frac{\theta(X_l)}{X_l}$ for $X_l > 0$. From Assumption 2 it follows that $\lambda'\left(X_l\right) < 0$. Proceed to establish convergence:

Case 1 $X_l^1 < 1$ *and* $X_{l'}^1 = 0$.

Show first that group l' stays out of entrepreneurship in industry 1 for good. By contradiction: if not, then there exists a time t where $X_{l'}^{t+1} = 0$ and $X_{l'}^{t+2} > 0$. Since supply must satisfy (6) it then follows that $X_l^{t+1} > 0$ and $X_l^{t+2} = 1$. The change in the output of industry 1 can then be written as:

$$Q_1^{t+2} - Q_1^{t+1} = N_l\left(\theta\left(X_l^{t+1}\right) - X_l^{t+1}\theta\left(X_l^t\right)\right) + X_{l'}^{t+2}N_{l'}\theta\left(X_{l'}^{t+1}\right). \tag{11}$$

This difference is strictly positive if the first term is positive. Clearly this is the case if $X_l^{t+1} \geq X_l^t$. If, instead, $X_l^{t+1} < X_l^t$, then again focusing on the first term:

$$\begin{aligned} \theta\left(X_l^{t+1}\right) - X_l^{t+1}\theta\left(X_l^t\right) &= \lambda\left(X_l^{t+1}\right)X_l^{t+1} - X_l^{t+1}\lambda\left(X_l^t\right)X_l^t \\ &= X_l^{t+1}\left(\lambda\left(X_l^{t+1}\right) - \lambda\left(X_l^t\right)X_l^t\right) > 0. \end{aligned} \tag{12}$$

This establishes that $Q_1^{t+2} > Q_1^{t+1}$. Since the output production of both industries must move in the same direction to clear the market, because of perfect complementarity, it follows that the output of industry 0 also increases from $t+1$ to $t+2$. This in turn requires that the number of workers in industry 0 increases, or equivalently, that the number of self-employed entrepreneurs in industry 1 decreases:

$$X_l^{t+2}N_l + X_{l'}^{t+2}N_{l'} < X_l^{t+1}N_l + X_{l'}^{t+1}N_{l'}. \tag{13}$$

Since $X_l^{t+2} = 1$ and $X_{l'}^{t+1} = 0$, this inequality can be simplified as $N_l + X_{l'}^{t+2}N_{l'} < X_l^{t+1}N_l$. This inequality is a contradiction and establishes that group l' stays out of self-employed entrepreneurship in industry 1 for good. The stationary equilibrium must consequently be of the form $\left(X_l^l, 0\right)$.

Assume first that $X_l^t > X^*$, in which case it is easy to show that $Q_1^{t+1} > Q_1^l > Q_1^{t+2}$ as well as $X_l^{t+1} < X_l^l < X_l^{t+2}$. Since $Q_1^{t+1} > Q_1^{t+2}$ it follows that:

$$\begin{aligned} X_l^{t+1} N_A \theta\left(X_l^t\right) &> X_l^{t+2} N_A \theta\left(X_l^{t+1}\right) \\ X_l^{t+1} \lambda\left(X_l^t\right) X_l^t &> X_l^{t+2} \lambda\left(X_l^{t+1}\right) X_l^{t+1} \\ X_l^t \lambda\left(X_l^t\right) &> X_l^{t+2} \lambda\left(X_l^{t+1}\right). \end{aligned} \tag{14}$$

The last line implies that $X_l^t > X_l^{t+2}$. The exact same argument, but with reverse inequalities, can be made for $X_l^t < X_l^l$. Therefore, having established that $X_l^t > X_l^{t+2} > X_l^l$ when $X_l^t > X_l^l$, and vice versa when $X_l^t < X_l^l$, it has been shown that X_l^t approaches the stationary equilibrium value X_l^l over time. This establishes convergence in Case 1.

Case 2 *$X_l^1 = 1$, $X_{l'}^1 \geq 0$, $X_l^2 = 1$ and $X_{l'}^2 \geq 0$.*

Show first that in this case, group l stays specialized for good. By contradiction: if not, then there exists a time t when $X_l^t = 1$, $X_l^{t+1} = 1$ and $X_l^{t+2} < 1$. Since supply must satisfy (6), it follows that $X_{l'}^{t+2} = 0$. The change in the output of industry 1 can be written as

$$Q_1^{t+2} - Q_1^{t+1} = N_l \left(X_l^{t+2} \theta(1) - \theta(1)\right) - X_{l'}^{t+1} N_{l'} \theta\left(X_l^t\right) < 0. \tag{15}$$

Since the supply of output of both industries must move in the same direction to clear the market, it follows that the output of industry 0 also decreases, which requires that the number of self-employed entrepreneurs in industry 1 increases:

$$X_l^{t+2} N_l + X_{l'}^{t+2} N_{l'} > X_l^{t+1} N_l + X_{l'}^{t+1} N_{l'}. \tag{16}$$

Since $X_{l'}^{t+2} = 0$ and $X_l^{t+1} = 1$, this inequality can be rewritten as $X_l^{t+2} N_l > N_l + X_{l'}^{t+1} N_{l'}$, which is a contradiction. This establishes that group l stays specialized in industry 1 for good. The stationary equilibrium must consequently be of the form $\left(1, X_{l'}^l\right)$. By the same argument as in Case 1, the sequence can be shown to approach the stationary equilibrium value $X_{l'}^l$ over time, both if $X_{l'}^t > X_{l'}^l$ and if $X_{l'}^t < X_{l'}^l$. This establishes convergence in Case 2.

Case 3 *$X_l^1 = 1$ and $X_{l'}^1 \geq 0$ and $X_l^2 < 1$ and $X_{l'}^2 = 0$.*

By the same argument in Case 1, it follows that group l' stays out of entrepreneurship in industry 1 permanently. Repeating the arguments in Case 1, convergence can then be established also in Case 3.

Consequently, in all three cases there is convergence. ■

This also implies that the stationary unspecialized equilibrium $\left(X_A^U, X_B^U\right)$ is unstable. If the minority group is slightly more specialized initially, then the economy converges to minority specialization $\left(X_A^A, X_B^A\right)$, and if the opposite is true, then the economy converges to majority specialization $\left(X_A^B, X_B^B\right)$. Over time, social segregation amplifies initial group differences.

2.2 Initial Conditions and Multiple Groups

Depending on the initial conditions, as is clear from Proposition 3, either of the two groups A and B can specialize as self-employed entrepreneurs in industry 1. Social interaction amplifies initial differences, but it does not explain why they are there to begin with. The difference in group size has some implications for what initial conditions to expect, however.

Consider an economy with more than two groups. As before, the group with more self-employed entrepreneurs in industry 1 initially will specialize in the long run. If the initial industrial distribution is subject to randomness, one of the smaller groups is likely to be the most specialized initially. To see why, let the initial distribution be generated by random draws, where each person becomes a self-employed entrepreneur in industry 1 with probability ρ.[5] This probability structure results in the same expected initial degree of specialization for all groups, but since the population size varies across groups, the variance in the degree of specialization also varies. The smallest groups have the largest variance, and therefore, the smallest groups are most likely to exhibit the lowest and also the greatest initial degrees of specialization. Consequently, with the smallest groups the most likely to specialize initially, as interaction amplifies initial differences over time, the smallest groups are also the most likely to specialize in the long run.

2.3 Assimilation

Our model does not feature assimilation of immigrants and their offspring and thus yields permanent social and industrial segregation. In our framework, assimilation would reduce the social isolation of an ethnic group (or some members of it) to the majority group. Our framework then predicts the industry choices of the assimilated individuals to look like those of the majority, especially if another ethnic group shows strong social isolation.

[5] These draws can be partially correlated within groups with the assumption that the correlation is the same for every group.

2.4 Heterogeneity and Earnings

Social complementarities also have implications for earnings. To examine how interaction effects would show up in earnings data, it is necessary to move away from the framework of identical skills. Returning to a static environment, endow each person i with entrepreneurial skills relevant to self-employment in industry 1, $s_1(i)$, and with another set of skills necessary for industry 0, $s_0(i)$. Self-employed entrepreneurial earnings in industry 1 are now a function of both interactions and skills. Denote the earnings of individual i in group l when she is a self- employed entrepreneur in industry 1 as $y_1(X_l, i) = p_1\theta(X_l) s_1(i)$, and when she is a member of industry 0 as $y_0(i) = p_0 s_0(i)$. Defining the ratios $s \equiv \frac{s_1}{s_0}$, $p \equiv \frac{p_0}{p_1}$, and $q \equiv p\frac{y_1}{y_0}$, the earnings-maximizing industry choice of individual i is to consider becoming a self-employed entrepreneur in industry 1 if:

$$q(X_l, i) \geq p \tag{17}$$

and to consider working in industry 0 if $q(X_l, i) \leq p$. Here the term $q(X_l, i) = \theta(X_l) s(i)$ summarizes the individual's comparative advantage in self-employed entrepreneurship in industry 1, at parity prices, as a function of social interaction and skills.

When individuals have different skills, the character of the price equilibrium depends crucially on the marginal self-employed entrepreneur and how her comparative advantage changes as more and more untalented people also become entrepreneurs in industry 1. If the benefits of interaction are weak and the marginal entrepreneur "deteriorates" as more intrinsically untalented people enter the industry, then the economy reduces to a standard Roy model, or sorting model, with a unique unspecialized equilibrium. Only if the interaction effect is strong enough to overcome skill heterogeneity can interaction change the character of the equilibrium.

Without loss of generality, order individuals from the greatest to the smallest comparative advantage in industry 1-style entrepreneurship, so that the skill ratio is decreasing in i, $s'(i) \leq 0$. The marginal entrepreneur is then the individual indexed by $i = X_l$, and her comparative advantage is $q(X_l, X_l)$. To prevent the economy from reducing to a sorting model, assume that the interaction effect trumps heterogeneity:

Assumption 3 *Interaction dominates at the margin:* $\frac{d}{dX_l} q(X_l, X_l) > 0$.

This assumption implies that the solid line in Figure A3 is upward sloping. The equilibrium distribution (X_A, X_B) must be competitively supplied and enough output must be produced by both industries to meet demand. Using a similar line of reasoning as in the previous section, based on V being strictly increasing in both arguments, it follows from Assumption 3 that there are three equilibria: one unstratified, denoted $\left(X_A^U, X_B^U\right)$; one where the minority group A specializes, denoted $\left(X_A^A, X_B^A\right)$; and one where the majority group B specializes, denoted $\left(X_A^B, X_B^B\right)$.[6]

In the equilibrium where minority A specializes as self-employed entrepreneurs in industry 1, the mean earnings of members of group A are higher than the mean earnings of members of group B, and vice versa in the equilibrium where group B specializes. To see why, let $y = \max\left(y_0, y_1\right)$ be actual individual earnings, and denote mean group earnings as $\mu = \int_0^1 y di$.

Proposition 4 *Earnings covary with self-employed entrepreneurship in industry 1:* $\mu\left(X_l\right) > \mu\left(X_{l'}\right)$ *if* $X_l > X_{l'}$.

Proof: Since people sort into industries, mean earnings can be rewritten as

$$\mu\left(X_l\right) = \int_0^1 y_0\left(i\right) di + \int_0^{X_l} \left(y_1\left(X_l, i\right) - y_0\left(i\right)\right) di \tag{18}$$

Rearranging, the difference in mean earnings between the two groups is:

$$\mu\left(X_l\right) - \mu\left(X_{l'}\right) = \int_0^{X_{l'}} \left(y_1\left(X_l, i\right) - y_1\left(X_{l'}, i\right)\right) di + \int_{X_{l'}}^{X_l} \left(y_1\left(X_l, i\right) - y_0\left(i\right)\right) di \tag{19}$$

where both parts of the expression are positive. The first part is strictly positive due to the interaction effect, $\frac{\partial y_1(X_l,i)}{\partial X_l} > 0$, and the second part is positive because of sorting, $y_1\left(X_l, i\right) \geq y_0\left(i\right)$ for all $i \leq X_l$. ■

This unequivocal effect on mean earnings at the group level does not carry through to the industry level. Depending on the joint distribution of skills, mean earnings in either industry can increase or decrease as interaction increases self-employed entrepreneurial productivity in industry 1 and shifts people of different ability between industries. The effect of interaction on industry earnings is similar to the effect of changing skill prices, which cannot be signed for a general skill distribution (Heckman and Honore, 1990).

The difference in mean earnings, normalized in units of industry 0 output, is shown in Figure A4 for the equilibrium with minority specialization. The exact derivation is included below. The relative price of industry 0 to industry 1 outputs is always such that the marginal entrepreneur is indifferent between industries. Keeping track of whether the marginal entrepreneur is in group A or in group B depending on the industrial distribution, the equilibrium price can be expressed as:

$$p = \begin{cases} q\left(X_l, X_l\right) & \text{if } X_l > X_{l'} \text{ and } X_{l'} = 0 \text{, or } X_l < X_{l'} \text{ and } X_l > 0 \\ q\left(X_{l'}, X_{l'}\right) & \text{if } X_l > X_{l'} \text{ and } X_{l'} > 0 \text{, or } X_l < X_{l'} \text{ and } X_l = 0 \end{cases} \tag{20}$$

[6] Note that Assumptions 2 and 3, when combined, put both an upper and a lower bound on the interaction effect: $-\frac{d \ln s}{dX_l} < \frac{d \ln \theta}{dX_l} < \frac{1}{X_l}$.

When increasing the number of self-employed entrepreneurs in industry 1 in equilibrium with minority specialization, the relative price of industry 0 output to industry 1 output increases continuously as the marginal entrepreneur in group A becomes more and more productive. This increase in price continues until all As are self-employed entrepreneurs in industry 1. To expand industry 1's self-employed entrepreneurial sector further from the point where everyone in group A are entrepreneurs, the price has to drop discretely from $p = q(1,1)$ to $q(0,0)$, to lure the unproductive Bs into the sector as well. The earnings differential between groups A and B moves accordingly, as shown in Figure A4, increasing continuously until all As are self-employed entrepreneurs in industry 1, at which point earnings jump in response to the discontinuous drop in the relative price.

Derivation of Earnings Differential in Figure A4: Mean earnings denominated in terms of industry 0 outputs are:

$$\frac{\mu(X_l)}{p_0} = \int_0^{X_l} p^{-1}\theta(X_l)\, s_1(i)\, di + \int_{X_l}^{1} s_0(i)\, di. \tag{21}$$

Replace the relative price of industry 0 output to industry 1 output, $p = \frac{p_0}{p_1}$, with the comparative advantage of the marginal entrepreneur, q, since these two are equal in equilibrium. Denote the earnings differential as $\Delta(X_l, X_{l'}) \equiv \frac{\mu(X_l)-\mu(X_{l'})}{p_0}$. It can be expressed as:

$$\Delta(X_l, X_{l'}) = \int_0^{X_{l'}} q^{-1}\left(\theta(X_l) - \theta(X_{l'})\right) s_1(i)\, di + \int_{X_{l'}}^{X_l} \left[q^{-1}\theta(X_l)\, s_1(i) - s_0(i)\right] di. \tag{22}$$

For $X_l < 1$ and $X_{l'} = 0$, where $q = q(X_l, X_l)$, and $q(X_l, X_l) = \theta(X_l)\, s(X_l)$, differentiating with respect to X_l gives

$$\frac{\partial \Delta(X_l, 0)}{\partial X_l} = -s'(X_l)\, s(X_l)^{-2} \int_0^{X_l} s_1(i)\, di > 0. \tag{23}$$

For $X_l = 1$ and $X_{l'} = 0$, the drop in price from $q(1,1)$ to $q(0,0)$ results in a jump in the mean earnings differential equal to

$$\Delta(1,0)\big|_{p=q(0,0)} - \Delta(1,0)\big|_{p=q(1,1)} = \left(q(0,0)^{-1} - q(1,1)^{-1}\right)\theta(1) \int_0^1 s_1(i)\, di > 0. \tag{24}$$

For $x = 1$ and $X_{l'} > 0$, where $q = q(X_{l'}, X_{l'})$, differentiating with respect to $X_{l'}$ gives

$$\frac{\partial \Delta(1, X_{l'})}{\partial X_{l'}} = -\frac{dq}{dX_{l'}} q^{-2}\theta(1) \int_0^1 s_1(i)\, di + s'(X_{l'})\, s(X_{l'})^{-2} \int_0^{X_{l'}} s_1(i)\, di - 2 s_0(X_{l'}) < 0. \tag{25}$$

■

3 Relationships in a Social Network

Since interactions have been restricted to be random, the analysis has so far abstracted from changes in the social structure that could arise in response to the productive value of interaction. The most interesting question is whether the majority will split up into smaller social groups, formed around choice of industry, to capitalize on interaction. If such splinter groups could form *costlessly*, then social interaction would no longer be able to generate industrial stratification along ethnic lines.

By developing a utility-based theory of interaction, explicitly stating social preferences and characterizing the optimal social structure, this section shows that splinter groups will not arise so long as preferences are sufficiently diverse, and so long as different social relationships are not close substitutes for one another. Under these two premises it is costly to confine social interactions to within a small group since the quality of social matches deteriorates with decreasing group size.

The theory developed in this section is constructed around a standard marriage market as in Becker (1973). In addition to spousal matching, people are also related by birth, which yields a larger social structure where individuals are interrelated not just pairwise but in a social network. Since the social network is derived as the outcome of matching, the problem analyzed here is different in nature from the problems most commonly analyzed in the social network literature, for example in Jackson and Wolinsky (1996), which focuses on strategic interaction between identical agents.

3.1 The Marriage Market

Take a very large finite population $i = 1, ..., N$, which is divided into mutually exclusive and exhaustive *families* by birth, with each family consisting of $d > 3$ individuals. Every person i independently draws a trait t_i, which could be for example beauty or intelligence, uniformly distributed between zero and one:

Assumption 4 *Individual traits t_i are independent draws.*

The independence of the draw signifies what can be thought of as maximal diversity: even within families people have different traits.

Based on realized traits, each person is assigned a spouse. To simplify, there are no gender restrictions and spouses can belong to the same family.[7] Traits are assumed to be complementary inputs in marriage. A marriage between i and j yields utility $u(t_i, t_j)$, where the function u is symmetric and strictly increasing with a positive cross-derivative:

[7]Removing gender restrictions maps this problem into a one-sided assortative matching problem. One-sided assortative matching is used in a different context in Kremer (1993).

Assumption 5 *Inputs are complementary:* $u(t_i, t_j) = u(t_j, t_i)$, $u_1 > 0$, $u_2 > 0$ *and* $u_{1,2} > 0$.

Since different relationships produce different utility, social relationships are not perfect substitutes and there is an optimal matching of spouses. Assume that utility is transferable, in which case the efficient spousal matching has to maximize aggregate utility. Labelling individuals according to rank, so that $t_1 < t_2 < ...$,[8] it follows that the efficient matching is positively assortative: person one marries person two, person three marries person four, ..., and person $N-1$ marries person N. To see this, let the matching function v be symmetric and the cross-derivative positive. For traits $t_1 < t_2 < t_3 < t_4$, we show that the only efficient matching is (t_1, t_2) and (t_3, t_4). As in Becker (1973), we use a property of v when the cross-derivative is positive,

$$v(a, d) + v(c, b) < v(a, b) + v(c, d) \tag{26}$$

for $a < c$ and $b < d$. Take an arbitrary efficient matching (x_1, x_2) and (x_3, x_4), which is a permutation of the traits t_1, t_2, t_3 and t_4. Without loss of generality, relabel these traits pairwise so that $x_1 < x_2$ and $x_3 < x_4$. Also without loss of generality, relabel the pairs so that $x_1 < x_3$. This implies that $x_1 < x_3 < x_4$. Using the symmetry of v, the aggregate utility from the arbitrary efficient matching can be written as $v(x_1, x_2) + v(x_4, x_3)$. Since $x_1 < x_4$ it follows from (26) that $x_2 < x_3$, otherwise aggregate utility could be increased by interchanging x_2 and x_3, just as b and d were interchanged in (26). Consequently, with $x_1 < x_2 < x_3 < x_4$, the arbitrarily chosen efficient matching (x_1, x_2) and (x_3, x_4) is identical to the efficient matching (t_1, t_2) and (t_3, t_4).

3.2 Splinter Groups

Say that two people i and j are *related* if they are married and/or belong to the same family. Define a *splinter group* as a proper subset of the population where no one in the subset is related to anyone outside of that subset. Given an efficient assignment of spouses in a very large population where traits are independently distributed, it follows that:

Proposition 5 *The probability that splinter groups exist is zero.*

Proof: Define a d-regular multigraph with loops, where every vertex corresponds to a family, and every edge corresponds to a marriage. A splinter group is equivalent to an unconnected component of this graph. Assortative marriages on independent traits generate a random configuration of vertices. A random configuration is equivalent to

[8] Since having equal-valued traits, $t_i = t_j$, is of measure zero, this possibility is ignored.

a regular random multigraph, as defined in Janson et al. (2000). A regular random multigraph is asymptotically almost surely Hamilitonian for $d > 3$ (Janson et al. 2000). Connectivity follows from Hamiltonicity, which rules out the existence of unconnected components, and consequently, the existence of splinter groups. ■

A partial explanation for this result is that if person i marries person j, then because of the independence of traits, it is unlikely that anyone else in i's family marries into j's family as well. As the population grows larger, it becomes less and less likely that there is more than one marriage between the families of i and j. This "mismatch" prevents i and j, and their families, from socially isolating themselves from the larger population. The problem is more interesting than what this partial intuition conveys, however. The likelihood of more than one marriage between two particular families decreases as the population grows larger, but on the other hand, the number of families for whom this event could occur increases. If, for example, d had been equal to two, then these two effects would have balanced, so that small splinter groups would have formed even as the population approached infinity. This proof most likely also goes through for $d \geq 3$, since it really only needs connectivity and since connectivity is closely related to cubic graphs. The fourth edge is necessary in the case of multigraphs to ensure Hamiltonicity, but Hamiltonicity is stronger than connectivity.

In addition to the above proof, we can provide a more structured intuition for no splinter groups by using a branching tree to trace out relationships in the population. Let Σ be the set of all families. Define an arbitrary family in Σ as the singleton set $\sigma(0)$. Let $\sigma(1)$ be the set of families in $\Sigma/\sigma(0)$ with at least one family member married to someone in the original family $\sigma(0)$. Define $\sigma(2)$ as the set of families in $\Sigma/(\sigma(0) \cup \sigma(1))$ with at least one family member married to someone in $\sigma(1)$. Continuing by iteration to more and more distant relations, let $\sigma(r)$ be the set of families in $\Sigma/(\sigma(r-2) \cup \sigma(r-1))$ married to someone in $\sigma(r-1)$. The variable r denotes what is sometimes called the degree of separation between the initial family $\sigma(0)$ and the families in $\sigma(r)$. The degree of separation is a measure of the social distance between individuals; compare Milgram (1967). The collection of these sets, $\cup_{q=0}^{r} \sigma(q)$, constitutes a branching tree. The sets in this collection are mutually exclusive, but if there are splinter groups, the sets are not exhaustive even as $r \to \infty$. Denote by $s(r)$ the cardinality of the set $\sigma(r)$. Since each family in $\sigma(r)$ is composed of d family members, where at least one member in each family by definition is married into $\sigma(r-1)$, the expansion of the tree $\cup_{q=0}^{r} \sigma(q)$ is bounded by

$$s(r+1) \leq s(r)(d-1). \tag{27}$$

If equation (27) holds with equality, then as r increases $s(r)$ very soon encompasses the entire population. It turns out that the equation generally holds as an inequality, however. The reason for this slowdown is threefold. First, a person in $\sigma(r)$ could marry another person in $\sigma(r)$. Second, a family in $\sigma(r)$ could have more than one family member married to someone in $\sigma(r-1)$. Thirdly, several people in $\sigma(r)$ could marry into the same family. These three types of events combine to prevent each family in $\sigma(r)$ from contributing a full $d-1$ new families to $\sigma(r+1)$, and consequently cause (27) to hold as an inequality.

Applying the branching tree $\cup_{q=0}^{r} \sigma(q)$ to the efficient assortative matching, the branching tree is overwhelmingly likely to grow to encompass the entire population in the limit. Since the branching tree only expands to include people who are directly or indirectly related, this limit result is equivalent to Proposition 5 that there are no splinter groups. To see why the entire population is included in the limit, consider what would happen if it were not true, if the branching tree died out without having reached a positive fraction of the population. If this were the case, then $\sigma(r)$ would eventually have to grow arbitrarily small relative to the remainder set $\Sigma/(\sigma(r-2) \cup \sigma(r-1))$, and therefore the likelihood that someone in $\sigma(r)$ married someone else in $\sigma(r)$ rather than in the remainder set, or that several people in $\sigma(r-1)$ married into the same family in $\sigma(r)$ rather than in the remainder set, or that several people in $\sigma(r)$ married into the same family in the remainder set, must also grow arbitrarily small. But then equation (27) should hold as an equality, implying that $s(r+1) > s(r)$, which contradicts the premise that the branching tree died out without having reached the entire population. Consequently, everyone in the population is either directly or indirectly related, and there are no splinter groups.

3.3 Implications for Productivity

The social network developed here allows more individual choice than the random interaction model analyzed earlier, since here industry choice can be made contingent on every aspect of the social structure. The main results from the random interaction model continue to hold nevertheless. A large group cannot align social relationships so as to maximize productivity in a small industry where social interaction and productivity are complementary, without incurring the cost of deteriorating social matches that comes from breaking up into smaller groups. This follows from the result that no splinter groups arise under first-best matching on social traits. Since the social choice set of ethnic minority groups is restricted anyway, these groups can limit their social interactions to a single industry at no alternative cost. Ethnic minorities are therefore well suited for social interaction-intensive industries.

A social network with the same properties could also be derived from a meeting technology where spouses meet and marry at random. The social structure derived here can therefore equally well be thought of as arising in a rigid environment where people meet randomly, as arising from efficient matching. Since randomness is likely to also have a role in who marries whom, this adds additional strength to the result.

Breaking up into smaller groups not only carries a social utility cost but also carries the cost of having to bypass naturally occurring random matching.

3.4 Future Model Extensions

An interesting extension for future work is to include both general and specific skills in the same framework. In such a model of spillovers between sectors, it should be possible to derive stratification in overall entrepreneurial activity as well as industry stratification between different forms of self-employed entrepreneurship at the same time. This would correspond to the current situation in the United States, where groups like the Koreans are strongly clustered in a few business sectors, while at the same time being overrepresented as self-employed owners in almost all other business activities as well.

4 Additional Appendix References:

Becker, Gary. 1973. A Theory of Marriage: Part I. *Journal of Political Economy* 81: 813-846.

Heckman, James and Bo Honore. 1990. The empirical content of the Roy model. *Econometrica* 58: 1121-1149.

Jackson, Matthew, and Asher Wolinsky. 1996. A strategic model of social and economic networks. *Journal of Economic Theory* 71: 44-74.

Janson, Svante, Tomasz Luczak, and Andrzej Rucinski. 2000. *Random Graphs*. New York: John Wiley.

Kremer, Michael. 1993. The O-ring theory of economic development. *Quarterly Journal of Economics* 108: 551-575.

Milgram, Stanley. 1967. The small world problem. *Psychology Today* 22: 61-67.

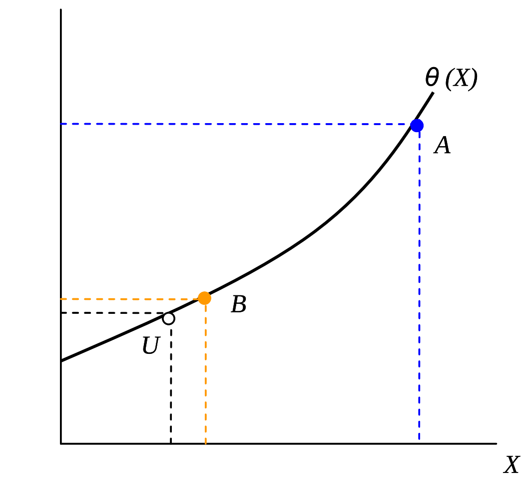


Figure A1. Individual productivity and the three stationary equilibria: one specialized equilibrium with minority specialization (A), one specialized equilibrium with majority specialization (B), and one unstratified equilibrium (U).

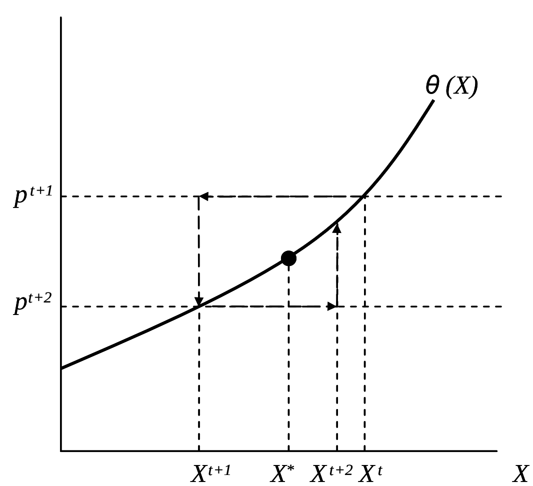


Figure A2. Stable dynamics when the internal effect dominates.

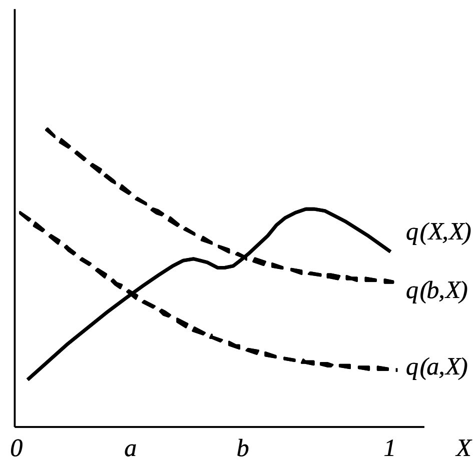


Figure A3. Sorting versus interaction effects in individual productivity. The dotted lines illustrate how the interaction effect raises productivity at all ability levels when specialization increases from *a* to *b*. The solid line shows the productivity of the marginal entrepreneur, for whom *i*=*X* at every level of *X*.

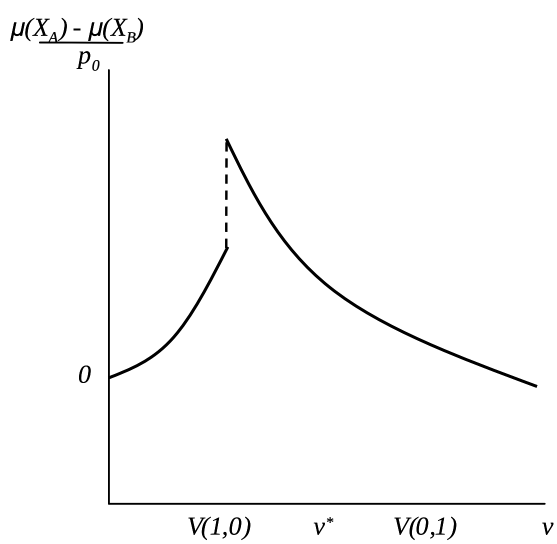


Figure A4. The difference in mean earnings between group *A* and group *B*, for different values of *v*, when minority group *A* specializes.

Online Appendix: Sociology Literature

Studies in sociology offer several theories and explanations for ethnic entrepreneurial specialization, including individuals' tie to their home country's culture, discrimination in the labor market, social cohesion theory, and cultural and/or religious traits and preferences. While it is beyond the paper's scope to offer a comprehensive review, this appendix introduces several connected theories as background for our study.

1 Connected Theories

A first set of work depicts the "sojourner status" of some migrant groups. These individuals are not planning to settle permanently in the host country to which they have temporarily moved and accordingly focus on the cultural heritage of their own ethnic group rather than assimilating into their host society (Siu, 1952). Sojourners may be inclined to seek portable occupations due to their expectation that they will return to their home country one day, and they have less incentive to invest in the business community outside of their ethnicity. Sojourners may be more reliant on their own group for business partnerships, hiring needs, and resources like capital investment and knowledge.

A connected "middleman minority" theory of ethnic employment is also based on individuals' tie with their home country. Middleman minorities are minority ethnic groups that tend to concentrate in intermediate occupations in their economies, facilitating transactions between one party and another. Examples of industries involving middleman minorities include trade and money lending (Blalock, 1967). Sojourner groups may end up in self-employment in middleman industries, which typically provide a portable livelihood requiring less fixed investment (Bonacich, 1973). Trust within the ethnic group may allow that group to compete successfully with native middleman businesses in the host country and encourage further specialization. Middleman minorities exhibit a preference for marriage within the ethnic group and established community institutions that reinforce the segregation of the ethnic group from those native to the host country (Bonacich, 1973). While important background, these theories have less direct application to our work that focuses on permanent migration and the industry specialization of ethnic entrepreneurs.

Whereas the initial theories depict a choice to become self-employed, others argue migrants can be forced into self-employment due to factors like discrimination that prevent them from accessing other labor opportunities (e.g., Wong, 1985) and the "blocked mobility hypothesis" (e.g., Light, 1972; Min, 1988a). Light and Gold (2000) write about the ethnic economy, the part of the labor market consisting of coethnic entrepreneurs and their coethnic employees. They summarize many advantages and disadvantages of ethnic economies for ethnic minorities, acknowledging the negative aspects of exploitation, discrimination, and inequalities while highlighting the positive benefits of increased opportunities for jobs, goods, and services that ethnic economies bring to an ethnic group in an economy.

The theory of "social cohesion" focuses on the direct and indirect social forces that act on members of a group to remain in a group. These forces include both individual group members' attitudes and behaviors and the interactions between group members, which influence members' attitudes and behaviors. Socially cohesive groups are often self-reinforcing; the greater the group cohesion, the more stable the group's membership and the more group members are motivated to reinforce attitudes and behaviors that maintain group membership. Friedkin (2004) and Fonseca et al. (2019) provide reviews. Forces aiding social cohesion like within-group trust, information sharing, and cooperation could also facilitate the self-employment specialization of the ethnic group.

The related concept of "social capital" also provides potential explanations for immigrant self-employment. Social capital refers to resources gained through group connections that can be used for economic, social, or cultural purposes. Individuals' can acquire social capital through participating in a group; social networks help generate reciprocity and trust. See Bourdieu (1986), Coleman (1988), and Putnam (1993) for varying definitions of social capital. As a member of an ethnic group develops more social capital, they may find it easier to start and maintain a business if it is in an industry others in their network are self-employed in, as they would have more access to information about business opportunities and more trust from potential business acquaintances and resources suppliers.

Concentrations of ethnic firms that occupy a particular urban space are known as ethnic enclaves. Studies of New York's Chinatown, Miami's Little Havana, and Los Angeles' Koreatown emphasize the importance of social networks for obtaining start-up capital, business information, and access to the labor force. Similarly, an ethnic niche emerges when a group takes prominence in a sector of employment, where members find jobs for each other through network chains, and when entry-level openings are filled by kin and friends. Portes (1998) provides a review of both concepts.

Weak ties can be important for job referrals (Granovetter, 1973). Immigrant entrepreneurs, on the other hand, have particular use of strong ties with kin and co-ethnics. When ties are deeply embedded within their networks, they are more likely to engage in the receipt and transmission of business support and information (Waldinger et al., 1990). Light (1972), Min (1988a), and Bates (1997) consider access to credit within ethnic networks. Bonacich and Light (1988) study how Koreans in Los Angeles are brought together for ritualized occasions (e.g., church) and how they afterwards exchange information about business conditions and techniques; Kim (1987) considers similar patterns in New York. A group's ability to exploit opportunities is linked to their internal organizational capacity. Ethnic groups with densely connected networks can support aspiring business owners through friends and family and through ethnic institutions like religious organizations. Aldrich and Zimmer (1986) provides an overview.

Several researchers have also emphasized the importance of cultural and religious distinctions within an ethnic group to how individuals organized themselves to specialize in entrepreneurship. See, for example, the work of Morris (1956) on Indians in East Africa, Winder (1963) on Lebanese in West Africa, and Bonacich and Light (1991) and Min (1988a-c, 1990) on Koreans in the United States. Botticini and Eckstein (2005) delve into Jewish economic history to explain the occupational selection of Jewish people, theorizing that the occupational selection of Jews was due to the religious and educational reforms, which brought about widespread literacy and thereby a comparative advantage in starting businesses in skilled occupations as new urban centers developed. Sharma (2019) provides a recent UK depiction of cultural and social reasons for Asian immigrant specialization as shopkeeper entrepreneurs in the United Kingdom.

2 Additional Appendix References:

Aldrich, Howard, and Catherine Zimmer. 1986. Entrepreneurship through social networks. In D.L. Sexton and R.W. Smiler, eds. *The Art and Science of Entrepreneurship.* Cambridge, MA: Ballinger, 3-23.

Bates, Timothy. 1997. Financing small business creation: The case of Chinese and Korean immigrant entrepreneurs. *Journal of Business Venturing* 12(2): 109-124.

Blalock, Hubert. 1967. *Toward a Theory of Minority Group Relations.* New York: John Wiley.

Bonacich, Edna. 1973. A theory of middleman minorities. *American Sociological Review* 38: 583-594.

Botticini, Marestella, and Zvi Eckstein. 2005. Jewish occupational selection: Education, restrictions, or minorities? *Journal of Economic History* 65(4): 922-948.

Bourdieu, Pierre. 1986. The forms of capital. In J.G. Richardson, eds. *Handbook of Theory and Research for the Sociology of Education.* New York: Greenwood Press.

Coleman, James. 1988. Social capital in the creation of human capital. *American Journal of Sociology* 94: S95-S120.

Fonseca, Xavier, Stephan Lukosch, and Frances Brazier. 2019. Social cohesion revisited: A new definition and how to characterize it. *Innovation: The European Journal of Social Science Research* 32(2): 231-253.

Friedkin, Noah E. 2004. Social cohesion. *Annual Review of Sociology* 30: 409-425.

Kim, Illsoo. 1987. The Koreans: small business in an urban frontier. *New Immigrants in New York*. New York: Columbia University Press.

Light, Ivan. 1972. *Ethnic Enterprise in America*. Berkeley, CA: University of California Press.

Light, Ivan, and Edna Bonacich. 1991. *Immigrant Entrepreneurs: Koreans in Los Angeles, 1965-1982*. Berkeley, CA: University of California Press.

Light, Ivan, and Steven Gold. 2000. *Ethnic Economies*. Bingley, UK: Emerald Group.

Min, Pyong Gap. 1988a. *Minority Business Enterprise: Korean Small Business in Atlanta*. Staten Island, NY: The Center for Migration Studies.

Min, Pyong Gap. 1988b. Korean immigrant entrepreneurship: A multivariate analysis. *Journal of Urban Affairs* 10: 197-212.

Min, Pyong Gap. 1988c. *Ethnic Business and Economic Mobility: Korean Immigrants in Los Angeles*. Prepared for Presentation at the Annual Meeting of the American Sociological Association, Atlanta.

Min, Pyong Gap. 1990. Problems of Korean immigrant entrepreneurs. *The International Migration Review* 24(3): 436-455.

Morris, Stephen. 1956. Indians in East Africa: A study in a plural society. *The British Journal of Sociology* 7: 194-211.

Portes, Alejandro. 1998. Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology* 24(1): 1-24.

Putnam, Robert D. 1993. The prosperous community. *The American Prospect* 4(13): 35-42.

Sharma, Babita. 2019. *The Corner Shop: Shopkeepers, the Sharmas and the Making of Modern Britain.* Hodder & Stoughton.

Siu, Paul. 1952. The sojourner. *American Journal of Sociology* 58(1): 34-44.

Waldinger, Roger, Howard Aldrich, and Robin Ward. 1990. *Ethnic Entrepreneurs: Immigrant Business in Industrial Societies*. London: Sage Publications, Inc.

Winder, R. Bayly. 1962. The Lebanese in West Africa. *Comparative Studies in Society and History* 4: 296-333.

Wong, Eugene. 1985. Asian American middleman minority theory: The framework of an American myth. *Journal of Ethnic Studies* 13(1): 51-88.