Social Networks, Ethnicity, and Entrepreneurship

William R. Kerr and Martin Mandorff

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 nonwork relationships facilitate the acquisition of sector-specific skills and result in occupational stratification along ethnic lines via concentrated entrepreneurship.

Disclosure: 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.

Data Disclosure: The data used in this article are available online: Kerr, William, 2020, "Replication Data for: Social Networks, Ethnicity, and Entrepreneurship (Kerr and Mandorff, 2020)", https://doi.org/10.7910/DVN/ZUUHAM, Harvard Dataverse, V1, UNF:6:Kr7aMQGf04wkidnHWncA3A== [fileUNF].

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

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

Authors and Acknowledgements: William Kerr is a Professor at Harvard Business School and a Research Associate at the National Bureau of Economic Research. Martin Mandorff is an Economist at the Swedish Competition Authority. 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. Online Appendix can be found at http://jhr.uwpress.org/.

Downloaded from by guest on September 6, 2026. Copyright 2021

doi:10.3368/jhr.58.3.0719-10306R2

This open access article is distributed under the terms of the CC-BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0) and is freely available online at: http://jhr.uwpress.org

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

Downloaded from by guest on September 6, 2026. Copyright 2021

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 2018). 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 (for example, 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* (1997), 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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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 (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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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 (for example, 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 to stud-

[4] See Chung and Kalnins (2006); Fairlie (2008); Fairlie, Krashinsky, and Zissimopoulos (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, Kerr, and O'Connell (2013); Leyden and Link (2015); Kerr and Kerr (2020b); and Bennet and Chatterji (2019).

Downloaded from by guest on September 6, 2026. Copyright 2021

ies in sociology regarding entrepreneurial specialization and explanations like sojourner status, middleman minorities, discrimination in the labor market, social cohesion, social capital and networks, and 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 (for example, Patel and Vella 2013) and the importance of ethnic networks for immigrants (for example Munshi 2003; Beaman 2012). Social interactions are important in job referrals, searching, and hiring (for example, Granovetter 1973; Bayer, Ross, and Topa 2008; Neumark 2013), and the agglomeration literature describes how interactions can boost productivity (for example, Arzaghi and Henderson 2008; Glaeser and Gottlieb 2009). Whereas group-level differences tend to decay over time in a basic referral model, for example, 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.

[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, Savchenko, and Vella (2013) provide a review.

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

Downloaded from by guest on September 6, 2026. Copyright 2021

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 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. Because social interaction is ran-

Downloaded from by guest on September 6, 2026. Copyright 2021

dom 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):

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

Because 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 nonconvexities. 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. Because the outputs of both industries have unitary income elasticities, distributional aspects can be ignored when characterizing the

Downloaded from by guest on September 6, 2026. Copyright 2021

efficient outcome. The problem simplifies to choosing an industry distribution (X_A, X_B) that maximizes a representative utility function $U\left(Q_0\left(X_A, X_B\right), Q_1\left(X_A, X_B\right)\right)$. A marginal analysis is inappropriate because this is a nonconvex 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 $V\left(X_A, X_B\right) \equiv Q_1/Q_0$ as the ratio of industry outputs under the distribution (X_A, X_B). Along the curve with the kink $V\left(1,0\right)$ 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 $V\left(1,0\right)$ 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 $V\left(0,1\right)$, 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 $V\left(0,1\right)$, 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, because 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

Downloaded from by guest on September 6, 2026. Copyright 2021

efficient so long as industry 1 is small enough. Perfect complementarity simplifies the problem of solving for the optimal allocation, because 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 because $v \leq V(1,0)$. 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 because 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 special-

Downloaded from by guest on September 6, 2026. Copyright 2021

ization 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 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. Although our

[8] While our model does not depict competition or crowding-out among coethnic 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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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 nonwork 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 attract-

Downloaded from by guest on September 6, 2026. Copyright 2021

ing 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. Although 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

[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, Ross, and Topa (2008); Beaman (2012); and Cadena, Duncan, and Trejo (2015).

Downloaded from by guest on September 6, 2026. Copyright 2021

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.

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 2006-10 and 2014-18. 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 (for example, combining England, Scotland, Wales, and nonspecific 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]

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

Downloaded from by guest on September 6, 2026. Copyright 2021

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 {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 aged 22-70 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.

[11] In the IPUMS data, self-employment is assigned when it is the main activity of an individual (for example, not capturing academics who consulting part-time). The definition includes both owners of employer 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 (for example, 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 (for example, Rauch, 2001; Agrawal, Kapur, and McHale, 2008). An extension later in the paper considers variation over metropolitan areas.

Downloaded from by guest on September 6, 2026. Copyright 2021

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, although 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 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.

[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 (for example, Andersson et al. 2014; Andersson, Burgess, and Lane 2014; Åslund, Hensvik, and Skans 2014; Kerr, Kerr, and Lincoln 2015; Kerr and Kerr 2020c).

Downloaded from by guest on September 6, 2026. Copyright 2021

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

[14] Our NBER working paper (Kerr and Mandorff 2018) focuses on 77 groups that have a minimum of 10 observations in at least one industry.

Downloaded from by guest on September 6, 2026. Copyright 2021

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 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 (for example, 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 aged 22-70 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.

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

[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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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.

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 aged 36-55, share who are aged 55-70 (reference group is aged 22-35), share who

Downloaded from by guest on September 6, 2026. Copyright 2021

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]

[17] Considered as a distribution, 90.1% of ethnic groups have a realized overage that exceeds the

Downloaded from by guest on September 6, 2026. Copyright 2021

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 (that is, $OVER2_l$ 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]

Appendix Tables 3a and 3b further test the relationships of relative size and isolation on entrepreneurial clustering by using nonparametric 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 we 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.

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

Downloaded from by guest on September 6, 2026. Copyright 2021

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 nonwhite 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 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, although 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

Downloaded from by guest on September 6, 2026. Copyright 2021

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 (for example, 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 percentage 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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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. Although 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

Downloaded from by guest on September 6, 2026. Copyright 2021

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. Although 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, although 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.

Downloaded from by guest on September 6, 2026. Copyright 2021

3.7 IV Empirical Tests

We next consider IV specifications to test against reverse causality concerns (for example, 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 (for example, 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

Downloaded from by guest on September 6, 2026. Copyright 2021

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 (for example, 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

Downloaded from by guest on September 6, 2026. Copyright 2021

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

[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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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 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

[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 (for example, "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.

[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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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.

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 7% increase in annual earnings. For wage workers, the relationship is measured to be 4%.

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

[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 (Manso 2016; Dillon and Stanton 2017). 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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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

Downloaded from by guest on September 6, 2026. Copyright 2021

self-employment industries, and 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 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 (for example, 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

[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.

[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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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 coethnic 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. Construction has a higher reported reliance on colleagues, and 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. Although some industries of self-employment concentration

[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 (specifically, 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 significant 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.

Downloaded from by guest on September 6, 2026. Copyright 2021

are spatially distributed by nature (for example, 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, 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, and 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

[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.

[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 (for example, 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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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 (for example, 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, although 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-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, Krashinsky,

Downloaded from by guest on September 6, 2026. Copyright 2021

and Zissimopoulos 2010).

Downloaded from by guest on September 6, 2026. Copyright 2021

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(2): 258-69.

Amuedo-Dorantes, Catalina, and Cynthia Bansak. 2012. “The Labor Market Impacts of Mandated Employment Verification Systems.” *American Economic Review* 102(3): 543-48.

Andersson, Fredrik, Monica Garcia-Perez, John Haltiwanger, Kristin McCue, and Seth Sanders. 2014. “Workplace Concentration of Immigrants.” *Demography* 51(6): 2281-306.

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* 35(4), 373-92.

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.

Arzaghi, Mohammad, and J. Vernon Henderson. 2008. “Networking off Madison Avenue.” *Review of Economic Studies* 75(4): 1011-38.

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): 405-42.

Åstebro, Thomas, Holger Herz, Ramana Nanda, and Roberto Weber. 2014. “Seeking the Roots of Entrepreneurship: Insights from Behavioral Economics.” *Journal of Economic Perspectives* 28(3): 49-70.

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.

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(6): 1150-80.

Beaman, Lori. 2012. “Social Networks and the Dynamics of Labor Market Outcomes: Evidence from Refugees Resettled in the US.” *Review of Economic Studies* 79(1): 128-61.

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

Bennett, Victor, and Aaron Chatterji. 2019. “The Entrepreneurial Process: Evidence from a Nationally Representative Survey.” *Strategic Management Journal*: 1-31. https://doi.org/10.1002/smj.3077.

Bhattacharjee, Yudhijit. 2018. “How Indian Americans Came to Run Half of All U.S. Motels.” *National* Geographic, September 4, https://www.nationalgeographic.com/culture/2018/09/south-asia-america-motels-immigration/.

Birley, Sue. 1985. “The Role of Networks in the Entrepreneurial Process.” *Journal of Business Venturing* 1(1): 107-17.

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(3): 955-88.

Bisin, Alberto, Giorgio Topa, and Thierry Verdier. 2004. “Religious Intermarriage and Socialization in the United States.” *Journal of Political Economy* 112(3): 615-64.

Downloaded from by guest on September 6, 2026. Copyright 2021

Borjas, George. 1987. “Self-selection and the Earnings of Immigrants.” *American Economic Review* 77(4): 531-53.

_____. 1992. “Ethnic Capital and Intergenerational Mobility.” *Quarterly Journal of Economics* 107(1): 123-50.

_____. 1995. “Ethnicity, Neighborhoods and Human Capital Externalities.” *American Economic Review* 85(3): 365-90.

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*, ed. Barry Chiswick and Paul Miller, 1197-259. Amsterdam: North Holland.

Chiswick, Barry. 1978. “The Effect of Americanization on the Earnings of Foreign-born Men.” *Journal of Political Economy* 86(5): 897-921.

Chung, Wilbur, and Arturs Kalnins. 2006. “Social Capital, Geography, and Survival: Gujarati Immigrant Entrepreneurs in the U.S. Lodging Industry.” *Management Science* 52(2): 233-47.

Deming, David. 2017. “The Growing Importance of Social Skills in the Labor Market.” *Quarterly Journal of Economics* 132(4): 1593-640.

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

Dillon, Eleanor, and Christopher Stanton. 2017. “Self-employment Dynamics and the Returns to Entrepreneurship.” HBS Working Paper 17-022, https://www.nber.org/papers/w23168, accessed December 2020.

Durlauf, Steven, and Marcel Fafchamps. 2006. “Social Capital.” In *Handbook of Economic Growth*, ed. Philippe Aghion and Steven Durlauf, 1639-99. Amsterdam: North Holland.

Durlauf, Steven, and Yannis Ioannides, 2010. “Social Interactions.” *Annual Review of Economics* 2(1): 451-78.

Elfring, Tom, and Willem Hulsink. 2003. “Networks in Entrepreneurship: The Case of High-technology Firms.” *Small Business Economics* 21: 409-22.

Faggio, Giulia, and Olmo Silva. 2014. “Self-Employment and Entrepreneurship in Urban and Rural Labour Markets.” *Journal of Urban Economics* 83(1): 67-85.

Fairlie, Robert. 2008. “Estimating the Contribution of Immigrant Business Owners to the U.S. Economy.” Small Business Administration, Office of Advocacy Report.

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*, ed. Josh Lerner and Antoinette Schoar, 179-208. Chicago: University of Chicago Press.

Downloaded from by guest on September 6, 2026. Copyright 2021

Fairlie, Robert, and Magnus Lofstrom. 2013. “Immigration and Entrepreneurship.” In *The Handbook on the Economics of International Migration*, ed. Barry Chiswick and Paul Miller, 877-911. Amsterdam: North-Holland Publishing.

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(2): 225-45.

Furtado, Delia. 2010. “Why Does Intermarriage Increase Immigrant Employment? The Role of Networks.” *B.E. Journal of Economic Analysis & Policy* 10(1): Article 101.

Furtado, Delia, and Nikolaos Theodoropoulos. 2011. “Interethnic Marriage: A Choice between Ethnic and Educational Similarities.” *Journal of Population Economics* 24(4): 1257-79.

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): 207-22.

Furtado, Delia, and Stephen Trejo. 2013. “Interethnic Marriages and Their Economic Effects.” In *International Handbook on the Economics of Migration*, ed. Amelie Constant and Klaus Zimmermann, 276-92. Northampton, MA: Edward Elgar.

Ghani, Ejaz, William Kerr, and Stephen O'Connell. 2013. “Local Industrial Structures and Female Entrepreneurship in India.” *Journal of Economic Geography* 13(6): 929-64.

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.

Glaeser, Edward, Bruce Sacerdote, and José Scheinkman. 1996. “Crime and Social Interactions.” *Quarterly Journal of Economics* 111(2): 507-48.

Glaeser, Edward, and José Scheinkman. 2002. “Non-market Interaction.” In *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress,* ed. Mathias Dewatripont, Lars Peter Hansen, and Stephen Turnovsky, 339-70. Cambridge, UK: Cambridge University Press.

Granovetter, Mark. 1973. “The Strength of Weak Ties.” *American Journal of Sociology* 78(1): 1360-80.

Greve, Arent, and Janet Salaff. 2003. “Social Networks and Entrepreneurship.” *Entrepreneurship Theory and Practice* 28(1): 1-22.

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

Hoang, Cleste. 2015. “The Fascinating Story Behind Why So Many Nail Technicians Are Vietnamese.” *Takepart,* May 5, http://www.takepart.com/article/2015/05/05/tippi-hedren-vietnamese-refugees-nail-industry.

Hunt, Jennifer. (2011). “Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa.” *Journal of Labor Economics* 29(3): 417-57.

Kennedy, Ruby. 1944. “Single or Triple Melting-pot? Intermarriage Trends in New Haven, 1870-1940.” *The American Journal of Sociology* 49(4): 331-9.

Downloaded from by guest on September 6, 2026. Copyright 2021

Kerr, Sari Pekkala, and William Kerr. 2017. “Immigrant Entrepreneurship.” In *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, ed. John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar, 187-250. Chicago: University of Chicago Press.

_____. 2020a. “Immigrant Entrepreneurship in America: Evidence from the Survey of Business Owners 2007 & 2012.” *Research Policy* 49(3): 103918.

_____. 2020b. “Immigrant Networking and Collaboration: Survey Evidence from CIC.” In *The Roles of Immigrants and Foreign Students in U.S. Science, Innovation, and Entrepreneurship*, ed. Ina Ganguli, Shulamit Kahn, and Megan MacGarvie, 173-206. Chicago: University of Chicago Press.

_____. 2020c. “Whose Job Is It Anyway? Co-Ethnic Hiring in New U.S. Ventures.” *Journal of Human Capital*, forthcoming.

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): S147-86.

Kerr, William. 2013. “U.S. High-Skilled Immigration, Innovation, and Entrepreneurship: Empirical Approaches and Evidence.” NBER Working Paper 19377. https://www.nber.org/papers/w19377, accessed December 2020.

_____. 2020. “Replication Data for: Social Networks, Ethnicity, and Entrepreneurship (Kerr and Mandorff, 2020).” https://doi.org/10.7910/DVN/ZUUHAM, Harvard Dataverse, V1, UNF:6:Kr7aMQGf04wkidnHWncA3A== [fileUNF].

Kerr, William, and Martin Mandorff. 2018. “Social Networks, Ethnicity, and Entrepreneurship.” NBER Working Paper 21597. https://www.nber.org/papers/w21597, accessed December 2020.

Kihlstrom, R., and Jean-Jacques Laffont. 1979. “A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion.” *Journal of Political Economy* 87(4): 719-48.

Kim, Yoonha, and John Morgan. 2018. “Hidden Gems? Cultural Barriers and the Self-employment of High Skilled U.S. Immigrants.” Unpublished working paper.

Kuznets, Simon. 1960. “Economic Structure and Life of the Jews.” In *The Minority Members: History, Culture, and Religion*, ed. Louis Finkelstein, 1597-666. Philadelphia, PA: Jewish Publication Society of America.

Lazear, Edward. 1999. “Culture and Language.” *Journal of Political Economy* 107(6): 95-126.

_____. 2005. “Entrepreneurship.” *Journal of Labor Economics* 23(4): 649-80.

Leyden, Dennis, and Albert Link. 2015. “Toward a Theory of the Entrepreneurial Process.” *Small Business Economics* 44(3): 475-84.

Lucas, Robert. 1978. “On the Size Distribution of Business Firms.” *Bell Journal of Economics* 9(2): 508-23.

Mandorff, Martin. 2007. “Social Networks, Ethnicity, and Occupation.” Ph.D. diss. University of Chicago.

Manso, Gustavo. 2016. “Experimentation and the Returns to Entrepreneurship.” *Review of Financial Studies* 29(9): 2319-40.

Downloaded from by guest on September 6, 2026. Copyright 2021

Millman, Joel. 1997. *The Other Americans: How Immigrants Renew Our Country, Our Economy, and Our Values.* London: Penguin Books.

Moris, Regan. 2015. “How Tippi Hedren Made Vietnamese Refugees into Nail Salon Magnates.” *BBC,* May 2, https://www.bbc.com/news/magazine-32544343.

Munshi, Kaivan. 2003. “Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market.” *Quarterly Journal of Economics* 118(2): 549-99.

Neumark, David. 2013. “Ethnic Hiring.” In *International Handbook on the Economics of Migration*, ed. Amelie Constant and Klaus F. Zimmerman, 193-213. Cheltenham: Edward Elgar Publishing.

Orrenius, Pia, and Madeline Zavodny. 2016. “Do State Work Eligibility Verification Laws Reduce Unauthorized Immigration?” *IZA Journal of Migration* 5(1): 1-17.

Patel, Krishna, Yevgeniya Savchenko, and Francis Vella. 2013. “Occupational Sorting of Ethnic Groups.” In *International Handbook on the Economics of Migration*, ed. Amelie Constant and Klaus F. Zimmerman, 227-41. Cheltenham: Edward Elgar Publishing.

Patel, Krishna, and Francis Vella. 2013. “Immigrant Networks and Their Implications for Occupational Choice and Wages.” *Review of Economics and Statistics* 95(4): 1249-77.

Rauch, James. 2001. “Business and Social Networks in International Trade.” *Journal of Economic Literature* 39(4): 1177-203.

Rosenthal, Stuart, and William Strange. 2012. “Female Entrepreneurship, Agglomeration, and a New Spatial Mismatch.” *Review of Economics and Statistics* 94(3): 764-88.

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): 306-30.

Spolaore, Enrico, and Romain Wacziarg. 2016. "Ancestry, Language and Culture." In *The Palgrave Handbook of Economics and Language*, ed. Victor Ginsburgh and Shlomo Weber, 174-211. London: Palgrave Macmillan UK.

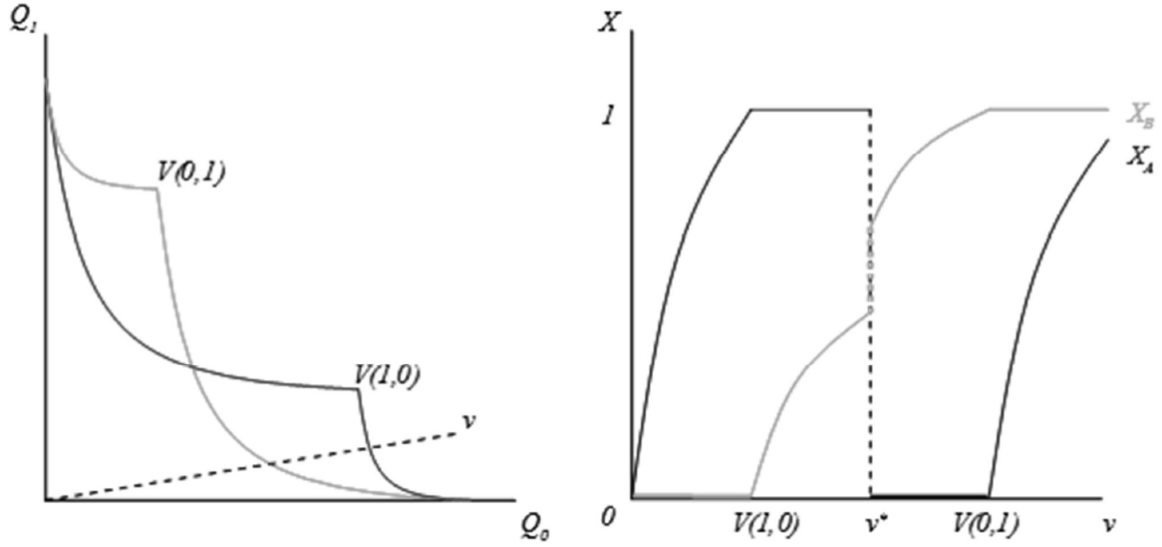
Virani, Aarti. 2012. “Why Indian Americans Dominate the US Motel Industry.” *Wall Street Journal,* June 11, https://blogs.wsj.com/indiarealtime/2012/06/11/why-indian-americans-dominate-the-u-s-motel-industry/?mg=prod/accounts-wsj.

Downloaded from by guest on September 6, 2026. Copyright 2021

Downloaded from by guest on September 6, 2026. Copyright 2021

Figure 1: Model depiction of entrepreneurial specialization

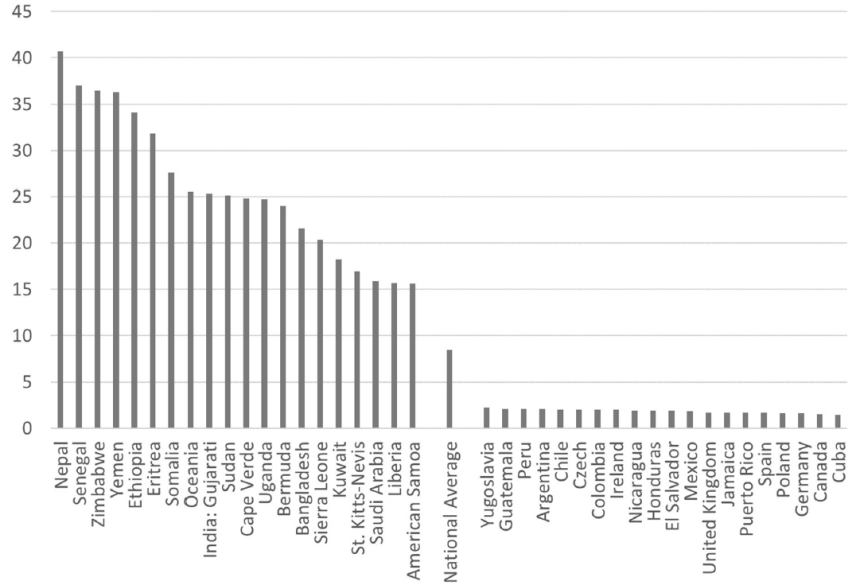
Notes: Left panel shows production possibilities with specialized occupational distributions. The ray *v* is the preference parameter over goods in a Leontief utility function. Along the curve with kink *V(1,0)*, all entrepreneurs belong to group A (below kink) or all members of group A are entrepreneurs (above). Along the curve with 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 convex productivity in interactions. The minority group A specializes as entrepreneurs if that sector is small enough.



Downloaded from by guest on September 6, 2026. Copyright 2021

Figure 2: Entrepreneurial specialization by ethnic group

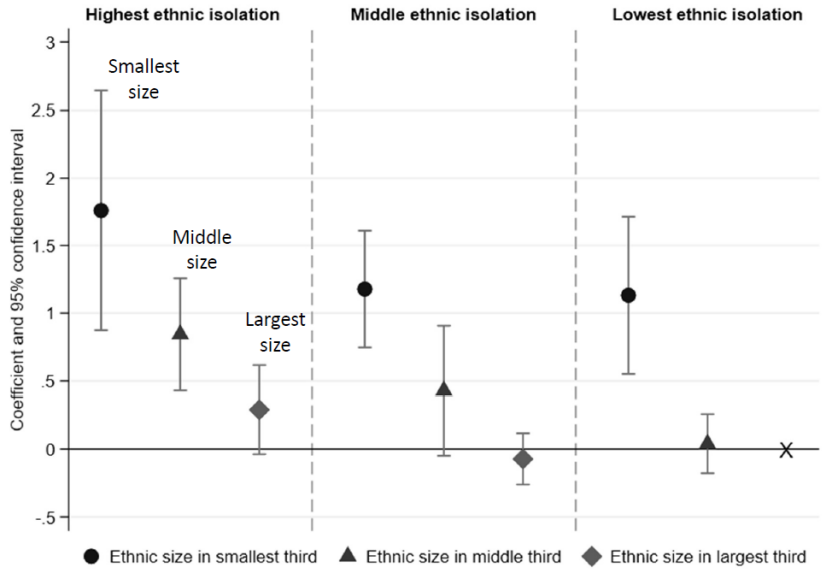
Notes: Figure shows the weighted average overage for the entrepreneurial concentration by ethnic immigrant group. The top 20 and bottom 20 values are shown, along with the national average.



Downloaded from by guest on September 6, 2026. Copyright 2021

Figure 3: Non-parametric estimations of specialization

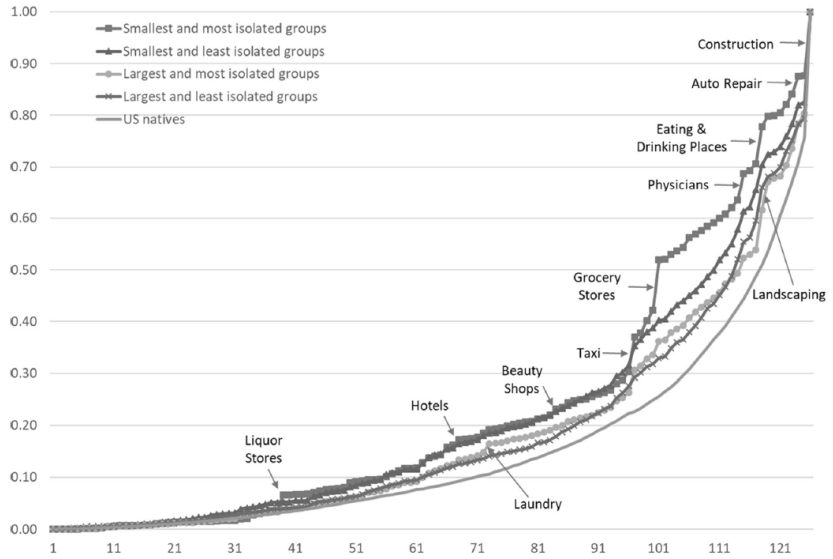
Notes: See Table 1. Ethnic groups are divided into equal-sized bins based upon group size and social isolation using terciles. Within each isolation triplet, groups are ordered smallest to largest as shown for the most isolated groups. Effects are measured relative to largest and least isolated ethnic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



Downloaded from by guest on September 6, 2026. Copyright 2021

Figure 4: Industry distributions of self-employment

Notes: Figure 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.



Downloaded from by guest on September 6, 2026. Copyright 2021

Table 1
OLS estimations of weighted average overage across all industries

	Baseline estimation with OVER1 (1)	Using total worker sample (2)	Excluding natives from denominator shares (3)	Imposing min counts on ethnic industry presence (4)	Excluding new arrivals over the prior five years (5)	Excluding the taxi industry (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.

Downloaded from by guest on September 6, 2026. Copyright 2021

Table 2
Robustness checks on OLS estimations

	Baseline estimation (1)	Without sample weights (2)	Without winsorization (3)	Including fixed effects for origin continent (4)	Using median regression format (5)	Using boot-strapped standard errors (6)	Including expected overage control (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.

Downloaded from by guest on September 6, 2026. Copyright 2021

Table 3
OLS estimations of overage metric designs

	Weighted average overage across all industries [OVER1] (1)	Weighted average overage using three largest industries for ethnic group [OVER2] (2)	Average of three largest overage ratios for ethnic group [OVER3] (3)	Largest overage ratio for ethnic group [OVER4] (4)
A. Baseline estimation without controls				
Inverse of log ethnic group size	0.582+++	0.458+++	0.578+++	0.526+++
(small groups have larger values)	(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	0.465+++	0.335+++	0.507+++	0.466+++
(small groups have larger values)	(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.

Downloaded from by guest on September 6, 2026. Copyright 2021

Table 4
OLS estimations of panel changes from 1980 - 2018

	Baseline panel estimation [OVER1] (1)	Including control for expected overage from simulations (2)	Including control for linear time trend in 1980 overage level (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.

Downloaded from by guest on September 6, 2026. Copyright 2021

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	0.582+++	0.718+++	0.466+++	0.507+++	0.488+++	0.624+++
(small groups have larger values)	(0.076)	(0.103)	(0.070)	(0.086)	(0.084)	(0.118)
Isolation of ethnic group using	0.325+++					0.331+++
in-marriage rate in United States	(0.076)					(0.109)
Residential segregation in the		0.302+++				0.030
United States		(0.066)				(0.112)
Genetic distance (country)			0.153++			0.189++
			(0.071)			(0.081)
Linguistic distance (country)				0.054		-0.081
				(0.047)		(0.075)
Religious distance (country)					0.151++	0.051
					(0.076)	(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	0.465+++	0.597+++	0.378+++	0.302+++	0.303+++	0.440+++
(small groups have larger values)	(0.079)	(0.103)	(0.077)	(0.084)	(0.086)	(0.104)
Isolation of ethnic group using	0.447+++					0.439+++
in-marriage rate in United States	(0.094)					(0.125)
Residential segregation in the		0.335+++				0.063
United States		(0.078)				(0.110)
Genetic distance (country)			0.103			0.115
			(0.082)			(0.094)
Linguistic distance (country)				0.110+		-0.019
				(0.056)		(0.067)
Religious distance (country)					0.149	0.060

					(0.102)	(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.

Downloaded from by guest on September 6, 2026. Copyright 2021

Downloaded from by guest on September 6, 2026. Copyright 2021

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 (1)	First stage for isolation (2)	Second stage (3)	First stage for size (4)	First stage for isolation (5)	Second stage (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.

Downloaded from by guest on September 6, 2026. Copyright 2021

Downloaded from by guest on September 6, 2026. Copyright 2021

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	0.069+++	0.030+++	0.029+++	0.024+++	0.015+
self-employed in the industry	(0.009)	(0.009)	(0.010)	(0.008)	(0.009)
Percentage of individual's group who are		0.100+++	0.101+++		
self-employed		(0.008)	(0.008)		
Percentage of individual's group who are			0.002		0.035++
working in the industry			(0.017)		(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	0.045+++	0.018	0.012	0.002	-0.004
self-employed in the industry	(0.012)	(0.012)	(0.013)	(0.007)	(0.007)
Percentage of individual's group who are		0.059+++	0.062+++		
self-employed		(0.007)	(0.007)		
Percentage of individual's group who are			0.020		0.021++
working in the industry			(0.013)		(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.

Downloaded from by guest on September 6, 2026. Copyright 2021

Downloaded from by guest on September 6, 2026. Copyright 2021

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 (1)	Log average industry size in terms of total workers (2)	Log size of largest industry measured by self-employed count (3)	Log size of largest industry measured by total worker count (4)	Log average industry size in terms of self-employed (5)	Log average industry size in terms of total workers (6)	Log size of largest industry measured by self-employed count (7)	Log size of largest industry measured by total worker count (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

Downloaded from by guest on September 6, 2026. Copyright 2021

Table 9
Sources of advice in 2016 Annual Survey of Entrepreneurs

Source of advice	Accommodation and food services (NAICS 72) (1)	Construction (NAICS 23) (2)	Average for other NAICS 2-digit sectors (3)	Ratio of NAICS 72 to average in other sectors (4)
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.

Downloaded from by guest on September 6, 2026. Copyright 2021

Table 10
OLS estimations at metropolitan level for size

	Baseline MSA-level estimation with local size measure [OVER1] (1)	Including control for expected overage from simulations (2)	Including control for expected overage from simulations and ethnicity fixed effects (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.