Online Appendix

Immigration and Entrepreneurship in the United States

Pierre Azoulay, Benjamin F. Jones, J. Daniel Kim, Javier Miranda

Appendix A: Data

U.S. Census Data

In this appendix, we describe the various data sets used in this study. Many of the data sets are Census-based products which are available to researchers through Census-approved projects and accessible through Federal Statistical Research Data Centers (FSRDC). Form W-2 data are currently accessible only by Census employees who have been granted access through approved internal projects.

The Longitudinal Business Database (LBD). The LBD is a panel dataset of all establishments in the U.S. with at least one paid employee (Jarmin and Miranda 2012). This dataset begins in 1976 and currently runs through 2015. The coverage includes all industries in the private non-farm sector and every state in the U.S. The LBD is sourced from administrative income and payroll filings and enhanced with other Census data sets, including the Economic Census and the Company Organization Survey. The LBD contains information on the firm size, firm age, location, payroll, legal form of entity, and other characteristics of the establishment. We define startups as *de novo* firms that have no prior activity at any of their establishments. The founding year is the year the firm first appears in the LBD. (Note that the Business Dynamics Statistics (BDS) of the Census includes young firms that are not de novo startups but rather can be the results of spin-outs and divestitures from existing firms, while our measure attempts to focus on true startups.)

Form W-2. Our annual individual earnings information are sourced from Form W-2, which is a tax form used to report income paid to employees for their services rendered. Employers are linked to the LBD based on their employer identification numbers (EIN). The W-2 database in the Census begins in 2005 and covers through 2016. Key variables in Form W-2 include income, social security taxes, and Medicare taxes.

The Survey of Business Owners (SBO). Information on the immigrant vs native-born nature of entrepreneurs is alternatively obtained from the 2012 Survey of Business Owners (U.S. Department of Commerce 2012). The SBO collects information about characteristics of the businesses and their owners from a representative sample of firms in the U.S. The random sample of businesses was selected from a list of all firms operating during 2012 with receipts of $1,000$ dollars or more. The SBO universe was stratified by state, industry, owner characteristics, and whether the company had paid employees in 2012. Large companies were selected with certainty. The remaining universe was subjected to stratified systematic random sampling. Each firm selected into the sample was asked the percentage of ownership, gender, ethnicity, race, and veteran status for up to four persons owning the largest percentages in the business. The final sample includes over 200,000 employer businesses in the SBO. Each firm in the SBO sample is assigned a weight equal to the reciprocal of the firm's probability of selection. Certainty cases are given a weight of one. Sample weights are used in the calculation of the results reported in the paper as frequency weights to return the population totals.

Census Numerical Identification System File (NUMIDENT). In order to define immigrant entrepreneurs, we use foreign-born status of individuals in the NUMIDENT. This Census database is originally sourced from the Social Security Administration (SSA) applications for Social Security Numbers (Form SS-5). Other person-level characteristics are contained in the NUMIDENT including gender, ethnicity, and date of birth.

The Patent Longitudinal Business Database Crosswalk (LPBD). The LPBD links patents data from the U.S. Patents and Trademark Office (USPTO) to firms in the LBD (Graham et al. 2018). This database begins in 2000 and extends to 2015. Though both application and grant years of the patent are observed, only granted patents are included in this sample. Other key variables include assignee location and type.

Fortune 500 Data

We collected founder and founding information for the firms listed in the 2017 edition of the Fortune 500 ranking. For each firm, we capture, whenever possible, the year of incorporation, the name of the founder, and his/her country of birth. This data collection builds on earlier efforts by the New American Economy Research Fund (2011, 2018) and the Center for American Entrepreneurship (2017).[i] We extend their analysis by including all founders for these firms, whether U.S.-born or immigrants.

This process is straightforward for many firms, particularly those that were founded in the recent past.[ii] For others, it is more challenging, since they are the offspring of many merged entities. Our approach is to walk back the genealogical tree of each firm to the earliest parent possible, and then to identify the founders of these parents.[iii] A firm will therefore have potentially many founders because it has multiple parents.

There are also particular cases where we do not include the firm. For some firms (particularly railroads and power utilities), there are very many mergers and it is not possible to trace the founders effectively. Further, web searches and biographical research occasionally do not enable us to ascertain the place of birth of any of the firm's original founders. If we cannot determine immigration status for any founder, the firm is dropped from the analysis. Separately, some firms listed in the Fortune 500 were not created through acts of entrepreneurship, but rather by government fiat (Fannie Mae is such an example; Delek U.S. holdings, the state-owned Israel oil company is another one). We exclude these firms from the analysis since they cannot be said to have founders in the traditional sense.[iv] Overall, the sample includes 449 firms and 730 founders for whom we can determine country of birth.

Post-1970 sample. As an additional check on the Fortune 500 analysis, we also consider firms founded since 1970. This includes 117 firms (and 223 founders with country of birth information) in the Fortune 500 ranking. We additionally focus on this time period for two reasons. First, the ability to identify founders —and to ascertain their country of birth —is greater when focusing on firms founded in the more recent past. Second, the recent subset may be most relevant to understanding links between entrepreneurship and immigration in a contemporary setting.

Population Data

The firm size distributions and rate of entrepreneurship measures are normalized by the population size of the relevant group (U.S.-born and immigrant individuals). To ascertain these population sizes we use two different methods, depending on the data source. We also consider robustness tests.

For the administrative data, we use the underlying, complete population of W-2 workers. All individuals with W-2's in the U.S. economy are matched to Census NUMIDENT to code U.S. born and foreign-born workers. This analysis covers these populations of workers from 2005-2010 to match with the founding years we study.

For the SBO data and the Fortune 500 data, the founding years of the firms extend back over many decades. For historical population estimates, we rely on numbers contained in U.S. censuses and collated by the Migration Policy Institute.[v] This data provides estimates of the immigrant population for each decade from 1850-2010 and annual estimates thereafter. These data explicitly include estimates of the unauthorized immigrant population.

Population weights. Since the immigrant population share changes over time, and the SBO and Fortune 500 data include a wide range of founding years, we calculate a weighted population over the relevant distribution of founding years. Specifically, for the firm size distributions, in each size bin × immigration status cell, we normalize the count

[i] See `https://startupsusa.org/fortune500/` and `https://research.newamericaneconomy.org/report/new-american-fortune-500-in-2018-the-entrepreneurial-legacy-of-immigrants-and-their-children/`.

[ii] Think for example of Hewlett-Packard: incorporated in 1939, with two founders, both native born. Or Google: incorporated in 1998, with two founders, one native-born, the other an immigrant.

[iii] For instance, American Airlines has two parents, Colonial Air Transport (incorporated in 1926, one native-born founder) and Robertson Aircraft Corporation (incorporated in 1921, two native-born founders).

[iv] A related example is that of Targa Resources. Warburg Pincus engineered a merger to create this firm in 2003, but it would be wrong to list as its founder Eric Warburg, who created the investment bank back in 1900.

[v] Available at `https://www.migrationpolicy.org/programs/data-hub/charts/immigrant-population-over-time`. Immigration figures for decades prior to 1970 stem from the work of Gibson and Campbell (1999).

of firms by the group's population. This population is the weighted averaged across the distribution of founding years of the firms in that size bin.

Entrepreneurship Rates. Comparing the administrative data and SBO approaches, note that rates of entrepreneurial entry have several differences in their calculation. First, there is a distinction between stocks and flows. The administrative data focuses on the flow of new business over the 2005-2010 interval, while the SBO looks at the stock of businesses at a point in time. Because the interval for the administrative data is short, the flow of new business formation will show a smaller rate of founders among the overall working population, whereas the SBO measure approximates the longer-run equilibrium rate. Second, the population normalizations are different. With the administrative data, we have the entrepreneurial rate among the workforce (W-2 workers). With the SBO data, we have the entrepreneurial rate among the entire population, which is substantially bigger than the total workforce and thus pushes down the SBO rate by comparison (and in contrast to the stock vs. flow issue which amplifies the SBO rate). Third, the measurement construct is different because the SBO allows us to see the top 4 owners of the firm (in 2012) and we look at those owners who were also reported as original founders in the survey. This means that we don't see founders for old firms or firms where the founders have exited from lead ownership roles. Despite these differences, we find that the relative entry rate of immigrants and native-born individuals (e_1^*/e_0^*) is very similar when using the administrative data or SBO data approach.

Unauthorized immigrant population. The census population data in each time period includes all individuals present in the U.S., regardless of citizenship or legal immigration status. In practice, demographers have long recognized that undocumented immigrants are less likely to participate in census surveys, a source of "coverage error" that is then corrected for in these population counts (Van Hook and Bachmeier 2013; Passel and Cohn 2018). Of note, disagreements regarding estimates of the undocumented immigrant population occur within a relatively narrow range.[vi]

While there is no obvious bias in these population estimates, one may nonetheless consider how sensitive the results in the paper could be to any under-count of the immigrant population. Specifically, how much would the underlying immigrant population need to be scaled up so that the aggregate employment in immigrant-founded firms, per immigrant in the population, declines to the equivalent measure for the native-born?

We find that the required immigrant population scaling is at least 40-60%, depending on the data set. To put this required under-count in context, estimates suggest approximately 45.6m foreign-born individuals in the U.S. in 2017, including approximately 10.5m unauthorized immigrants (Passel and Cohn 2018). Higher estimates suggest as many as 12 million unauthorized immigrants (Kamarck and Stenglein 2019). Under-counting the total immigrant population by 40-60% would mean that unauthorized immigrants total 30 million or more individuals, in comparison to standard estimates of 10.5-12 million. There is no evidence that immigration could be understated by anything close to this magnitude. Overall, immigration on net appears to be a net job creator in the U.S. economy when including unauthorized immigrants.

[vi]Despite using slightly different data and assumptions, estimates from the Pew Research Fund, the Department of Homeland Security, and the Center for Migration Studies have never differed by more than 1 million people, less than 10% of the total unauthorized population.

Appendix B: Additional Results

Fortune 500 firms, Post 1970

As an additional view of the Fortune 500 data, Appendix Figure B1 repeats Figure 3 but now focusing on the (2017) Fortune 500 firms that are founded from 1970 onward. This subset includes 123 firms. Studying firms founded since 1970 has two advantages: The founder information is more comprehensive for this subset and the recent subset is more relevant to contemporary immigration outcomes. As can be seen, the results are similar.

Slope Calculations

We calculate the firm size distribution slope parameters and their standard errors for all definitions in all figures. We follow Newman (2005) to use best practices for power law distributions, which is a maximum likelihood method.[vii] The slopes and their standard errors are reported in Table B2. We emphasize the SBO as this comes closest to the long-run distribution for the whole support of firm size. The slopes are extremely similar using Definitions 1, 2, or 3—varying slightly around 1.78. They are still statistically significantly different given the scale of the data. As another comparison, we also examined the power law slope in the whole LBD, and find a slope of 1.74.

Calibration

Corollary 3 provides the basis for a calibration, quantifying the implications of immigration for economy-wide wages and per-capita income. Specifically, consider equation (9), repeated here for the wage comparative statics:

$$\frac{d\log w^*}{dn_1} = \theta \frac{a_1^\gamma - a_0^\gamma}{a_0^\gamma n_0 + a_1^\gamma n_1} \tag{10}$$

Rewriting the right hand side in ratios of the form $(a_1/a_0)^\gamma$ and using (8), we can express the ratios in terms of relative entrepreneurial entry rates from the immigrant and native populations:

$$\frac{d\log w^*}{dn_1} = \theta \frac{e_1^*/e_0^* - 1}{[e_1^*/e_0^* - 1]n_1 + 1} \tag{11}$$

For a sufficiently fat tailed power law, where the slope is less than 2 (which we find in all data sets and founder definitions—see Table B2) the relevant case for calibration as a_{max} becomes large is $\theta = 1 - \beta$. Taking observed values of $e_1^*/e_0^* = 1.8$ and $n_1 = 0.13$ in the administrative data, and $\beta = 2/3$ to match the labor share, we find $\frac{d\log w^*}{dn_1} = 0.24$, as reported in the text. By Corollary 3, the comparative static for per-capita income is the same.

Wages in Immigrant vs. Native Founded Firms

The empirical results investigate employment and the firm size distribution for native-founded and immigrant-founded firms. The empirical analysis in turn follows our conceptual framework, where individuals have heterogeneous entrepreneurial acumen and, for simplicity, homogeneous labor.

Of additional interest may be the wages for the jobs that these founders create and how these wages compare between immigrant and native-founded firms. The administrative data, with which we have integrated the W-2 records for every individual working in these firms, provides an additional opportunity to examine wages in a systematic fashion. We run OLS regressions of the form

$$\log(w_i) = \alpha + \beta ImmigrantFounded_f + \gamma X_i + \theta Z_f + \varepsilon_i$$

where w_i is the individual worker i's annual W-2 earnings from employer f, X_i is a vector of the individual worker characteristics, and Z_f is a vector of the firm's characteristics. Individual worker characteristics include fixed effects for age and indicators for gender and for being foreign-born. Firm characteristics include fixed effects for founding year, fixed effects for county, and fixed effects for sector using NAICS 4-digit industry codes. Results are presented in Table B1 and are discussed in Section 5.2.

[vii] In particular, see equations [5] and [6], as well as Appendix B in Newman [2005].

Patenting

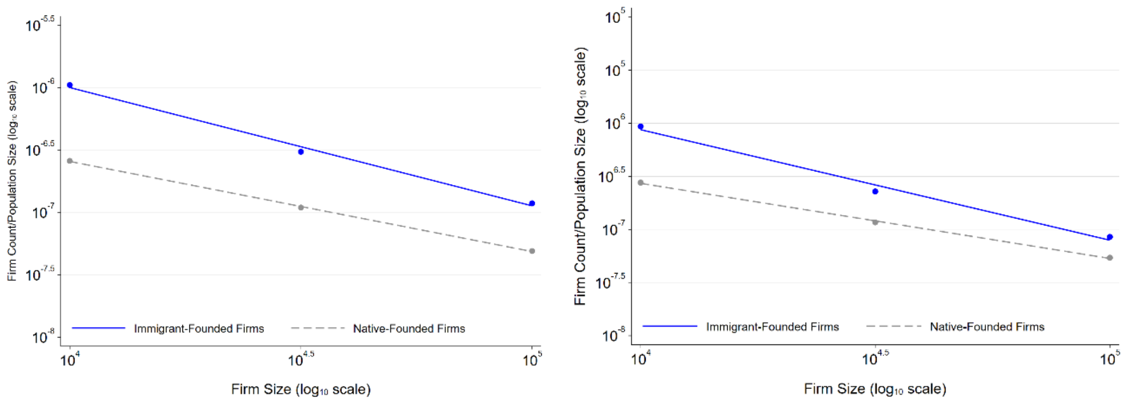
Figure 4 presents the patenting rate by firm size, comparing immigrant-founded and native-founded firms, in the administrative data. This analysis uses Definition 1 for defining immigrant firms. That is, we consider firms as immigrant-founded if at least 1 of the founders is an immigrant. Appendix Figure B2 repeats the analysis but uses Definition 2 instead. In this definition, the firm is an immigrant-founded firm only if the highest-wage individual in the founding team is an immigrant. (By construction, now the native-founded firms include some founding teams that include immigrants.) Under this definition, immigrant-founded teams still have a higher rate of patenting in each size bucket, although the difference is not as large.

Timing

Studies of mass migration events in U.S. history find large and persistent long-run gains in income per-capita in regions that experienced greater immigration (Tabellini 2020, Sequeira et al. 2020). Other studies investigate shorter-run outcomes from migration events, and often do not find negative employment or wage effects. For example, Card (1990) examines the five-year period after the Mariel Boatlift. While final output demand from immigrant workers is immediate, and pushes toward neutral employment effects in general equilibrium (as in the model of Section 3), an interesting question is whether the entrepreneurial orientation appears over a short enough time period to be of relevance to shorter-run outcomes.

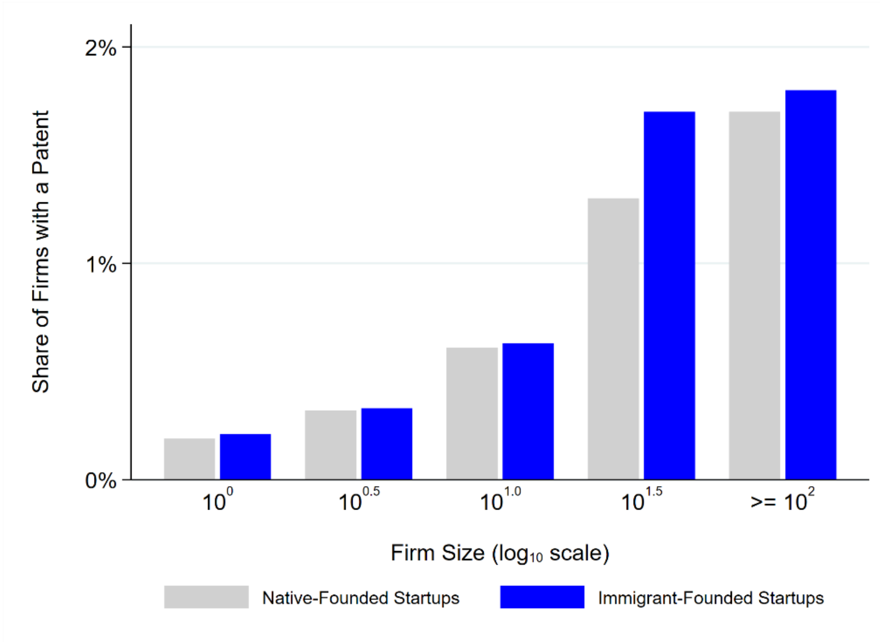
To study this question, we integrated the American Community Survey, which provides year of entry for U.S. immigrants. For this Census survey, we then focused on individuals who entered the United States in 2005, providing a ten-year window through 2015 where we can comprehensively ascertain labor force participation, wage employment, and new business creation. We find that, among the immigrants in the sample who founded a firm within ten years, roughly half did so within a window of five years, with the highest rate appearing in year 2. This indicates that new venture dynamics appear to be relevant even in the short run. We treat these findings as suggestive and leave systematic study of dynamics of labor market entry and new venture creation to future work.

Figure B1
Immigrant and Native-Born Entrepreneurship:
Firm Size Distributions using the Fortune 500 firms founded post-1970



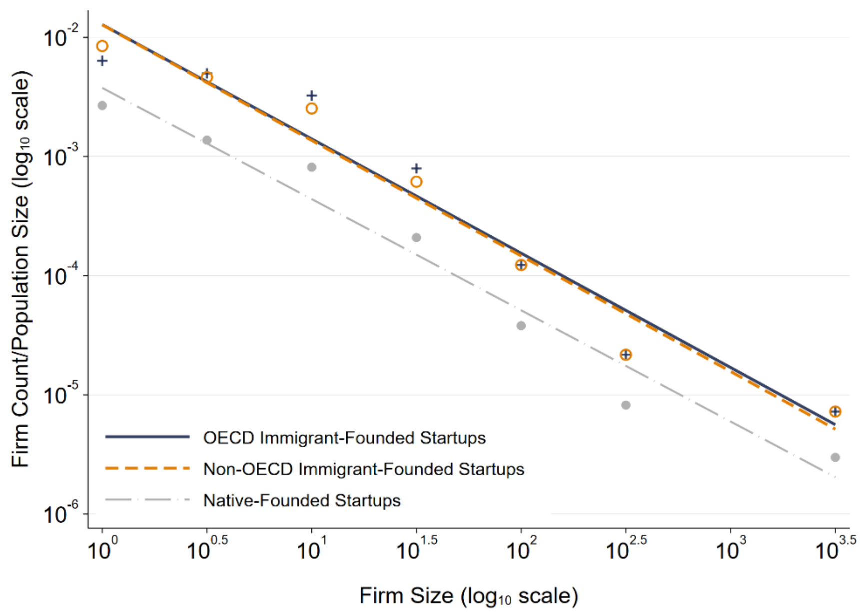
Notes: Each panel consider the firm size distribution, distinguishing between immigrant-founded and native-founded firms. The x-axis is the log of firm size measured as current total employment in the firm, using the 2017 Fortune 500. The y-axis is the log count of firms of a given size, with the count normalized by the population size for the relevant group (immigrant or native born). The population measure is an average of the immigrant or native-born population in the decade of founding, weighted by the number of firms founded in that decade. The plotted measures correspond to Corollary 2. Panel A counts a firm as immigrant-founded if any of the founders are immigrants (Definition 1 in the text). Panel B assigns firms to immigrant and non-immigrant proportionally based on the mix of immigrant and native-born founders of the initial business (Definition 3). Definition 2 is not available for the Fortune 500, as discussed in text.

Figure B2
Technology-Based Immigrant and Native-Born Entrepreneurship using Definition 2



Notes: Using W-2 and LBD data combined with patenting records from the USPTO, this figure shows the share of firms in each firm size bin that own at least one patent, distinguishing between native-founded versus immigrant-founded startups, for all firms in the US between 2005 and 2010. Immigrant-founded startups are identified using Definition 2, which equals 1 if the highest paid founder is foreign-born. Firms are grouped into five bins along the *x*-axis based on the number of employees five years after founding. The difference in firm size binning relative to Figure 4 is due to Census disclosure rules requiring the minimum number of firms represented in each cell.

Figure B3
OECD-Immigrant and Non-OECD Immigrant Entrepreneurship:
Firm Size Distributions using Administrative Data



<u>Notes</u>: Each panel consider the firm size distribution, distinguishing between OECD immigrant-founded, non-OECD immigrant-founded, and native-founded firms, for all U.S. firms in the Longitudinal Business Database founded in the 2005-2010 period. The x-axis is the log of firm size measured as total employment in the firm five years after founding. The y-axis is the log count of firms of a given size, with the count normalized by the number of workers from the relevant population (immigrant or native born). The plotted measures correspond to Corollary 2. A firm as immigrant-founded if any of the founding team members are foreign born (Definition 1 in the text). Country of birth is used to identify OECD versus non-OECD immigrant founders.

Table B1: Wages at Immigrant versus Native-Founded Firms

	DV = Ln(Annual Wages)					
	(1)	(2)	(3)	(4)	(5)	(6)
Immigrant-founded firm	0.041***	0.045***	-0.000	-0.040***	0.008***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(Firm size)						-0.029***
						(0.000)
Male				0.337***	0.245***	0.246***
				(0.001)	(0.001)	(0.001)
Foreign born				0.082***	0.124***	0.121***
				(0.001)	(0.001)	(0.001)
Constant	8.761***	–	–	–	–	–
	(0.001)					
Observations (Individuals)	14,640,000	14,640,000	14,640,000	14,640,000	14,640,000	14,640,000
R^2	0.000	0.001	0.010	0.131	0.230	0.231
Individual Age Fixed Effects				YES	YES	YES
Founding Year Fixed Effects		YES	YES	YES	YES	YES
County Fixed Effects			YES	YES	YES	YES
NAICS-4 Fixed Effects					YES	YES

Notes: This table shows a series of OLS regressions using log annual wages as the dependent variable. Sample consists of individuals employed by startups at five years after founding, distinguishing immigrant versus native-founded firms based on Definition 1. Standard errors in parentheses are clustered at the firm level. Constants are not reported for fixed effects regressions. Note that the constant in the first specification represents the mean of log wages, rather than the log of mean wage. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table B2: Slope of Power Laws for Firm-size Distributions

	Definition 1	Definition 2
Administrative Data (5-Year Old Firms)		
Immigrant Founded	1.70 (.001)	1.78 (.002)
Native Founded	1.76 (.0009)	1.72 (.008)
SBO Data		
Immigrant Founded	1.78 (.002)	1.78 (.005)
Native Founded	1.78 (.002)	1.78 (.005)
Fortune 500		
Immigrant Founded	1.35 (.036)	–
Native Founded	1.17 (.007)	–
Administrative Data (All firms in the LBD)	1.74 (.003)	

Notes: The power law slope parameter and its standard error are calculated by maximum likelihood, following Appendix B in Newman (2005). The SBO data is the most appropriate sample for comparing the long-run firm size distribution between immigrant and non-immigrant founded firms. As a reference, the slope in the overall administrative data (the LBD) in the year 1.74 is provided in the last line of the table.

Appendix C: Proofs

Proof of Proposition 1

This proof proceeds in two steps. First, we consider how a shift in $f(a)$ influences the equilibrium a^*. Second, we consider how other equilibrium outcomes shift. To clarify the comparative statics, we will write $f(a|\theta)$, where θ is a parameter that affects the distribution of entrepreneurial talent. In particular, θ can be the fraction of immigrants in the economy. Equilibrium outcomes will in general be functions of θ.

Comparative Statics on a^*

To begin, we look at a^*. There are two key equations to develop the relevant comparative statics. The first equation comes from the free entry condition to entrepreneurship, defining a threshold value of entrepreneurial talent at which people start firms as opposed to being workers. This relationship is (4)

$$w = a^* (1-\beta)^{1-\beta} \beta^\beta$$

indicating that there is a monotonically increasing relationship $w(a^*)$.

The second equation comes from the resource constraint, which tells us that the number of entrepreneurs (E) and the number of workers (L) must add up to the total number of people, N. Given the distribution $f(a|\theta)$, we can then write the mapping between the threshold value for founding a firm, a^*, and the number of entrepreneurs as in (5)

$$\frac{E^*}{N} = \int_{a^*}^{\infty} f(a_i|\theta)\, da_i$$

where we are using the fact that anyone with $a_i \geq a^*$ will start a firm (because income as a founder then exceeds income as a worker).

Similarly, the number of workers at a given firm is $l_i^* = \left(\frac{\beta a_i}{w}\right)^{\frac{1}{1-\beta}}$,where a_i is the acumen of the founder. Integrating across all firms we have the total labor force L^*. We can then write that the resource constraint, $\frac{E^*}{N} + \frac{L^*}{N} = 1$, as

$$\int_{a^*}^{\infty} \left[1 + \left(\frac{\beta}{w}\right)^{\frac{1}{1-\beta}} a_i^{\frac{1}{1-\beta}}\right] f(a_i|\theta)\, da_i = 1 \tag{12}$$

This gives us our second function for $w(a^*)$. Using the entrepreneurial entry condition, (4), we can then rewrite this resource constraint to eliminate the wage and put everything in terms of a^*. Namely,

$$\int_{a^*}^{\infty} \left[1 + \frac{\beta}{1-\beta}\left(\frac{a_i}{a^*}\right)^{\frac{1}{1-\beta}}\right] f(a_i|\theta)\, da_i = 1 \tag{13}$$

To interpret this expression, note that we are counting up the number of people at each firm, which must sum to all the people in the economy. We have divided by N so we are counting people in terms of fractions of the population. The term in square brackets is the number of people associated with a given firm. The 1 in square brackets is the entrepreneur—every firm has 1 entrepreneur. The second term in square brackets, $\frac{\beta}{1-\beta}\left(\frac{a_i}{a^*}\right)^{\frac{1}{1-\beta}}$, is the number of workers at that firm, which is increasing in the acumen of the entrepreneur. The $f(a_i)$ then gives the mass of the founder population associated with that firm.

The core result is then seen directly. By inspection, the term in square brackets is strictly positive. Therefore, if you increase the mass of $f(a_i)$ for all points $a_i > a^*$, then the value of the integral would rise. The only way for the integral value to remain constant is therefore for a^* to rise. And if a^* rises, then the wage has to rise, per (4).

More formally, one can takes the comparative statics for $a^*(\theta)$ using Leibniz's Rule. Differentiating (13) with respect to θ, we find that

$$a^{*\prime}(\theta) = \frac{1-\beta}{f(a^*(\theta)) + \frac{1}{a^*(\theta)}\frac{L^*}{N}} \int_{a^*(\theta)}^{\infty} \left[1 + \frac{\beta}{1-\beta}\left(\frac{a_i}{a^*(\theta)}\right)^{\frac{1}{1-\beta}}\right] \frac{df(a_i|\theta)}{d\theta} da_i$$

By inspection, the sign of $a^{*\prime}(\theta)$ depends on the sign of the integral. One can then generate necessary and sufficient conditions for the comparative statics by evaluating the integral for known probability distributions and shifts in these distribution. However, since the term in square brackets is strictly positive, we can also develop simple sufficient conditions that generalize across $f(a)$. In particular, consider the comparative static on the share of immigrants in the economy, defined as $\theta = n_1 = N_1/N$. The population distribution of entrepreneurial acumen is $f(a_i) = (1-n_1) f_0(a_i) + n_1 f_1(a_i)$ and thus

$$\frac{df(a_i|n_1)}{dn_1} = f_1(a_i) - f_0(a_i)$$

It then follows that

$$\begin{aligned} a^{*\prime}(n_1) &> 0 \text{ if } f_1(a_i) > f_0(a_i) \text{ for all } a_i \geq a^* \\ a^{*\prime}(n_1) &= 0 \text{ if } f_1(a_i) = f_0(a_i) \text{ for all } a_i \geq a^* \\ a^{*\prime}(n_1) &< 0 \text{ if } f_1(a_i) < f_0(a_i) \text{ for all } a_i \geq a^* \end{aligned}$$

which correspond to the three cases in the text and the first part of Proposition 1, as was to be shown.

The comparative statics on other equilibrium quantities are then as follows.

Comparative Statics on w^*

From (4), the equilibrium wage w^* is monotonically increasing in a^*. Hence, the effect of increased immigration on equilibrium wages has the same sign as the comparative statics for a^*, as was to be shown.

Comparative Statics on Y^*/N

From the income side, we can write GDP per capita, $y = Y/N$, as

$$y = \int_{a_m}^{a^*} wf(a_i)da_i + \int_{a^*}^{\infty} \pi_i f(a_i)da_i$$

Using Leibniz's rule, we have

$$\begin{aligned} \frac{\partial y}{\partial \theta} &= \frac{\partial w}{\partial \theta}\int_{a_m}^{a^*} f(a_i)da_i + w\int_{a_m}^{a^*} \frac{\partial f(a_i)}{\partial \theta}da_i + wf(a^*)\frac{\partial a^*}{\partial \theta} \\ &+ \int_{a^*}^{\infty} \frac{\partial \pi_i}{\partial \theta} f(a_i)da_i + \int_{a^*}^{\infty} \pi_i \frac{\partial f(a_i)}{\partial \theta}da_i - \pi(a^*) f(a^*)\frac{\partial a^*}{\partial \theta} \end{aligned}$$

Noting that $w = \pi(a^*)$, the third and sixth terms cancel. Further, the first and the fourth terms will also cancel. In particular, the first term solves as

$$\frac{\partial w}{\partial \theta}\int_{a_m}^{a^*} f(a_i)da_i = \frac{\partial w}{\partial \theta}\frac{L^*}{N}$$

For the fourth term, from the envelope theorem we have $\frac{\partial \pi_i^*}{\partial \theta} = -l_i^* \frac{\partial w}{\partial \theta}$. This integral thus solves as

$$\int_{a^*}^{\infty} \frac{\partial \pi_i}{\partial \theta} f(a_i)da_i = -\frac{\partial w}{\partial \theta}\frac{L^*}{N}$$

which cancels with the first integral.

The comparative statics on income per capita thus simplify to

$$\frac{\partial y}{\partial \theta} = w\int_{a_m}^{a^*} \frac{\partial f(a_i)}{\partial \theta}da_i + \int_{a^*}^{\infty} \pi_i \frac{\partial f(a_i)}{\partial \theta}da_i$$

Now, consider the case of a right shift in the distribution $f(a_i)$, where $\frac{\partial f(a_i)}{\partial \theta} > 0$ for all $a_i \geq a^*$. Noting that $\pi_i(a^*) = w$ and $\pi_i > w$ for all $a_i > a^*$, it follows that,

$$\int_{a^*}^{\infty} \pi_i \frac{\partial f(a_i)}{\partial \theta}da_i > \int_{a^*}^{\infty} w\frac{\partial f(a_i)}{\partial \theta}da_i$$

and therefore

$$\frac{\partial y}{\partial \theta} > w\int_{a_m}^{a^*}\frac{\partial f(a_i)}{\partial \theta}da_i + w\int_{a^*}^{\infty}\frac{\partial f(a_i)}{\partial \theta}da_i = w\int_{a_m}^{\infty}\frac{\partial f(a_i)}{\partial \theta}da_i = 0$$

Thus income per-capita is increasing with a right shift in the distribution of entrepreneurial acumen, as was to be shown. Similar reasoning gives the other two cases.

Comparative Statics on Π^*/N

The equilibrium profit rate is such that $\Pi/Y = 1-\beta$. Thus comparative statics for profits per capita follow the direction as the comparative statics for income per capita, which are shown above.

Proof of Corollary 1

Define the total number of jobs created by a given group j as M_j. This count is the total number of founders from that group, E_j^*, plus the total number of wage workers in the firms these founders create, which we define as L^{j*}. We are interested in whether M_j exceeds the population size of the group, N_j. We have

$$\begin{aligned} M_j &= E_j^* + L^{j*} = N_j\int_{a^*}^{\infty} f_j(a_i)\,da_i + N_j\int_{a^*}^{\infty} l_i^*(a_i)\,f_j(a_i)\,da_i. \\ &= N_j\int_{a^*}^{\infty}\left[1+l_i^*(a_i)\right]f_j(a_i)\,da_i \end{aligned}$$

Comparing immigrants and the native born, we have

$$\frac{M_1}{N_1}-\frac{M_0}{N_0} = \int_{a^*}^{\infty}\left[1+l_i^*(a_i)\right](f_1(a_i)-f_0(a_i))da_i$$

By inspection, the integral is zero in case 1, less than zero in case 2, and greater than zero in case 3. Taking case 3, we have $M_1/N_1 > M_0/N_0$. Combining this with the population constraint $M_0+M_1 = N_0+N_1$ (total employment equals total population), it follows that $M_0/N_0 < 1$ and $M_1/N_1 > 1$ in case 3. Following this logic similarly for each case gives

$$\begin{aligned} &M_1 = N_1 \text{ in case 1.} \\ &M_1 < N_1 \text{ in case 2.} \\ &M_1 > N_1 \text{ in case 3.} \end{aligned}$$

as was to be shown.

Proof of Corollary 2

The firm size distribution within a given group follows from the founder acumen distribution, $f_j(a)$, and the relationship between founder acumen and firm size. From profit maximization, firm size is monotonically increasing in founder acumen. Specifically, a firm's employment is

$$l_i^* = \left(\frac{\beta a_i}{w}\right)^{\frac{1}{1-\beta}} = \frac{\beta}{1-\beta}\left(\frac{a_i}{a^*}\right)^{\frac{1}{1-\beta}} \tag{14}$$

where $a_i \geq a^*$.

Let the firm size distribution for group j be $g_j(l_i^*)$. Consider the case where $a < a_{max}$. Using the change-in-variables rule, the firm size distribution relates to the acumen distribution as

$$g_j(l_i^*) = \left|\frac{da_i(l_i^*)}{dl_i^*}\right| f_j(a_i(l_i^*)\,|a_i \geq a^*) \tag{15}$$

The first term in (15) is as follows. Inverting the monotonic relationship (14), we have

$$a_i\left(l_i^*\right) = a^*\left(\frac{1-\beta}{\beta}\right)^{1-\beta}\left(l_i^*\right)^{1-\beta}. \tag{16}$$

and the slope of acumen with firm size is then

$$\frac{da_i\left(l_i^*\right)}{dl_i^*} = (1-\beta)\,\frac{a_i\left(l_i^*\right)}{l_i^*}. \tag{17}$$

The second term in (15) is, from (3),

$$f_j\left(a_i\left(l_i^*\right)|a_i \geq a^*\right) = \frac{1}{E_j^*/N_j}\begin{cases}\gamma_j a_j^{\gamma_j}/a_i\left(l_i^*\right)^{\gamma_j+1} & \text{if } a_j \leq a_i\left(l_i^*\right) < a_{max}\\ \left(a_j/a_{max}\right)^{\gamma_j} & \text{if } a_i\left(l_i^*\right) = a_{max}\end{cases} \tag{18}$$

where E_j^*/N_j is the population share of founders – the total mass of $f_j(a_i)$ above a^*, which is

$$\frac{E_j^*}{N_j} = \left(\frac{a_j}{a^*}\right)^{\gamma_j} \tag{19}$$

Using (16), (17), (18), and (19) in (15) produces

$$g_j\left(l_i^*\right) = \gamma_j\left(1-\beta\right)\left(\frac{1-\beta}{\beta}\right)^{1-\beta}\left(l_i^*\right)^{-\gamma_j(1-\beta)-1}$$

Taking logs and differentiating produces the slope

$$s_j = \frac{d\log g_j\left(l_i^*\right)}{d\log l_i^*} = -\gamma_j\left(1-\beta\right)-1 \tag{20}$$

Scaling the firm size distribution by any constant produces the same power law slope in logs. Specifically, noting that the total number of entrepreneurs from this group is E_j^*, it follows that the frequency count of firms ($c_j(l_i^*) = E_j^* g_j\left(l_i^*\right)$) and the frequency count normalized by the group population size ($c_j(l_i^*)/N_j$) will also have this same power law slope in firm size, as was to be shown.

Proof of Corollary 3

Consider the equilibrium rate of entrepreneurship within a given group. We have

$$e_j^* = \frac{E_j^*}{N_j} = \int_{a^*}^{\infty} f_j\left(a\right)da.$$

For the Pareto distributions, (3), with $\gamma_j = \gamma$ this integrates as

$$e_j^* = \left(\frac{a_j}{a^*}\right)^{\gamma}.$$

It follows directly that the ratio e_1^*/e_0^* is given by (8), as was to be shown.

Now consider the equilibrium wage. From the entrepreneurial entry condition (4) the equilibrium wage is linearly related to a^*, so we will first consider comparative statics in terms of a^*.

Take the general result in (13), which is the population resource constraint expressed by adding up the workers and founders over all firms. To perform this integral, note that the joint distribution of acumen in the relevant region for the whole population is

$$f\left(a\right) = \begin{cases}\gamma a_m^{\gamma}/a^{\gamma+1} & \text{if } a^* \leq a < a_{max}\\ \left(a_m/a_{max}\right)^{\gamma} & \text{if } a = a_{max}\end{cases} \tag{21}$$

where the Pareto scale parameter is $a_m = \left(a_0^{\gamma}n_0 + a_1^{\gamma}n_1\right)^{1/\gamma}$.[viii] Performing the integral (13) using this population-wide acumen distribution provides the following implicit expression for $a^\star$,

$$1 = \left(\frac{a_m}{a^*}\right)^{\gamma} + \frac{\frac{\beta}{1-\beta}}{1-\gamma(1-\beta)}\left(\frac{a_m}{a^*}\right)^{\gamma}\left(\left(\frac{a_{max}}{a^*}\right)^{\frac{1}{1-\beta}-\gamma} - \gamma(1-\beta)\right) \tag{22}$$

[viii] Note that this population-wide acumen distribution makes the empirically relevant and natural assumption that there is entrepreneurial entry from both immigrants and the native-born; i.e., $a^* \geq a_j$ for both sub-populations.

We can then perform comparative statics that depend on how this expression behaves as a_{max} becomes large.

Case 1. If $\gamma > \frac{1}{1-\beta}$, then the a_{max} term disappears as a_{max} becomes large. This leads to the explicit solution for a^* where

$$a^* = a_m \left(\frac{1-\gamma}{1-\gamma(1-\beta)} \right)^{1/\gamma} \tag{23}$$

Case 2. If $\gamma = \frac{1}{1-\beta}$, then for any a_{max} the expression (22) becomes

$$a^* = a_m \left(\frac{1}{1-\beta} \right)^{1/\gamma} \tag{24}$$

Case 3. If $\gamma < \frac{1}{1-\beta}$, then the a_{max} term grows large as a_{max} becomes large. It follows in (22) that the term $\left(\frac{a_m}{a^*}\right)^\gamma$ must go to zero. Thus for large a_{max} we can rearrange (22) as

$$a^* \approx \frac{\frac{\beta}{1-\beta}}{1-\gamma(1-\beta)} a_m^{\gamma(1-\beta)} a_{max}^{\gamma(1-\beta)-1} \tag{25}$$

Using the entrepreneurial entry condition (4) and the chain rule, we have

$$\frac{d \log w^*}{dn_1} = \frac{d \log a^*}{da_m} \frac{d \log a_m}{dn_1} \tag{26}$$

which is straightforward to compute for the three cases and the definition of a_m above, producing (9) in Corollary 3.

Finally, noting that the labor share of income is $\beta = w^* L/Y^*$, it follows that equilibrium GDP per capita follows the same comparative statics in n_1 as the wage.

Appendix References

Fortune 500 List of Companies. 2017. *Fortune Magazine.* Accessed at https://fortune.com/fortune500/2017/search/.

Gibson, Campbell, and Emily Lennon. 1999. "Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 1990." Working Paper No. 29*,* U.S. Bureau of the Census, Population Division.

Graham, Stuart J.H., Cheryl Grim, Tariqul Islam, Alan C. Marco and Javier Miranda. 2018. "Business Dynamics of Innovating Firms: Linking U.S. Patents with Administrative Data on Workers and Firms." *Journal of Economics & Management Strategy* **27**(3): 372-402.

Jarmin, Ron and Javier Miranda. 2012. "The Longitudinal Business Database". Center for Economic Studies Working Paper 02-17.

Kamarck, Elaine, and Christine Stenglein. 2019. "How many undocumented immigrants are in the United States and who are they?" *Brookings Policy 2020*. Available at https://www.brookings.edu/policy2020/votervital/how-many-undocumented-immigrants-are-in-the-united-states-and-who-are-they/.

Migration Policy Institute, "U.S. Immigrant Population and Share over Time, 1850-Present." Available at https://www.migrationpolicy.org/programs/data-hub/charts/immigrant-population-over-time (accessed October 1, 2019).

Newman, M.E.J. 2005. "Power Laws, Pareto Distributions, and Zipf's Law." *Contemporary Physics* **46**(5): 323-351.

Passel, Jeffrey S., and D'Vera Cohn. 2018. "U.S. Unauthorized Immigrant Total Dips to Lowest Level in a Decade." Pew Research Center. Available at https://www.pewresearch.org/ hispanic/2018/11/27/u-s-unauthorized-immigrant-total-dips-to-lowest-level-in-a-decade/.

Van Hook, Jennifer, and James D. Bachmeier. 2013. "How Well Does the American Community Survey Count Naturalized Citizens?" *Demographic Research* **29**(1): 1-32.

U.S. Department of Commerce, Bureau of the Census. 2012. "Survey of Business Owners (SBO)—Survey Results 2012." Available at https://www.census.gov/library/publications/ 2012/econ/2012-sbo.html.