Appendix A: Factors associated with Inc. 5000 high-growth firm regional density

This Appendix to "High-growth firms and cities in the US: An analysis of the Inc. 5000" contains details on the regression analysis referenced in the main body text that assesses the regional factors associated with Inc. 5000 high-growth firm (I5HGC) density (company entries in the Inc. 5000 lists between 2011 and 2017 per one million residents by metropolitan area).

To assess which factors are associated with regional I5HGC density, a linear regression analysis (OLS) was conducted. The explained or dependent variable is the number of I5HGC entries per one million residents (I5HGC density) during the period, among the 303 US metropolitan areas for which a full set of control variables could be collected.

The impact of nearly two dozen explanatory or independent variables on I5HGC density were tested, and each were considered based on a vast research literature (see this footnote for a complete list).[1] Most of these factors (20 of 23) held statistically significant bivariate correlations with the density measure, but failed to hold statistically meaningful relationships when considered jointly with other factors. The explanatory variables in the table below represent the most robust of these—they maintain impact and statistical significance under different permutations and variable choices.

Because the explanatory variables are measured in different units and because we are most concerned with *relative* drivers of our explained variable ("what matters and what matters most?"), the regression model presented here contains standardized variables. In a standardized regression, the coefficient associated with each explanatory variable indicates the standard deviation change in the explained variable (in this case, I5HGC density for each metro area) occurring from a one standard deviation change in that explanatory variable, holding other factors constant. For example, a coefficient of 0.5 indicates a one standard deviation increase (decrease) in that explanatory variable is associated with a 0.5 standard deviation increase (decrease) in the explained variable.

This simple model does a pretty good job—collectively, explaining almost two-thirds of the variation in I5HGC density across the 306 metros for which all explanatory variables are available. In the first iteration of the model, each of the five explanatory variables are statistically significant at one percent (very high significance). In the second model, those robust relationships hold with the introduction of three additional variables—those three are also statistically significant but are less robust (lower levels of significance).

Standardized OLS Regression of Inc. 5000 High-Growth Company Density on Explanatory Variables for 306 Metropolitan Areas (2011-2017)

Explanatory Variables	(1)	(2)
Bachelor's degree % of emp.	0.425***	0.386***
High-tech ind. % of emp.	0.417***	0.341***
Prime-age e-ship (35-44yrs) % of pop.	0.259***	0.192***
Firm entry rate (base-year)	0.208***	0.268***
Patents per 1m residents	-0.219***	-0.247***
Creative class occ. % of emp.	--	0.162*
Manufacturing ind. % of 1980 emp.	--	0.136**
% of adult population married	--	0.0746*
Constant	0.0121	0.0124
Observations	306	306
R-squared	0.631	0.652

**** p<0.01, ** p<0.05, * p<0.1*

Source: Brookings analysis of Inc. Magazine and Center for American Entrepreneurship data

Model #1

Five key variables stand out as being especially important. First is the share of workers with college degrees, a factor that has consistently been found to be important for driving regional entrepreneurship more broadly.[2] A deep pool of well-educated workers is particularly important here given the concentration of knowledge-intensive industries among the I5HGCs.

Second is the share of workers employed in high-tech industries. To begin, many of the I5HGCs are in high-tech industries themselves—research establishes the link between the overall presence of an industry in a region and local business formation rates in that same industry (a relationship that holds more so for high-tech and knowledge-intensive activities).[3] Also, the most represented non-high-tech industries among the I5HGCs are in areas that are important inputs to high-tech production, and have themselves become relatively high-tech (e.g. advertising and marketing). Research shows that high-tech supply chains exhibit a high degree of geographic proximity.[4]

Third is the share of the population of prime entrepreneurship age (35 to 44 years). Research shows that the relationship between starting a business and founder age exhibits an inverted-U shape—increasing as one ages up to a point (as one gains more experience and wealth) but then declines thereafter (as one becomes more risk-averse and moves closer to retirement). Mid-career professionals are in that sweet spot, and empirical studies consistently establish that relationship (though founders of high-growth firms tend to be slightly younger).[5]

Fourth is the overall rate of business formation in the region. Research has shown wide and persistent variation of business formation rates across different regions.[6] In other words, entrepreneurial regions tend to stay that way and the evolution of a region being more or less entrepreneurial occurs slowly. This is partly to do with culture (some areas value entrepreneurship more) and partly to do with experience (learning from one's own

experiences as a founder or at a startup company or from the experiences of other entrepreneurs is important to venture formation and growth).

The final variable is the number of patents per capita. Readers should note that the sign on this is *negative*—meaning that, when controlling for the other factors discussed above, I5HGC density *decreases* as patent density *increases*. When correlating these two factors in isolation this is not the case—the sign is positive and significant. A few tests of the data indicate that it is the controlling for human capital and high-tech activity that changes this relationship—in other words, once we put each region on equal footing in terms of brains and techies, those that produce more patents are less likely to produce I5HGCs. There is no concrete explanation for this, but one might be that these regions are more dominated by large corporations or universities that are not entrepreneurial or where patents become barriers to entry for would-be entrepreneurs in those fields.

Model #2

The three additional variables added into the second iteration of the model opens up some interesting discussions. The first of these—the share of creative class workers (high-skilled professionals and creative types of many trades)—is intriguing because, even though it meets only the lowest of accepted statistical significance thresholds (10 percent), a relationship at all after controlling for workers with college degrees and those in high-tech (where there is a sizable overlap with creative class workers) is notable. So, again, after putting regions on even footing for brains and techies, those with a higher proportion of creative activity—beyond that already exhibited by college degree holders and high-tech workers—have a higher density of I5HGCs than those that don't.

Second is the share of workers in manufacturing in 1980. This one is a bit of a mystery given that regions concentrated in natural resources and heavy industry in the past see less entrepreneurial activity (as measured by business formation rates) today. However, after controlling for business formation rates and a host of other factors detailed above, those with a higher manufacturing base in 1980 (the peak of manufacturing employment nationally) produce a higher rate of I5HGCs today—whatever the cause. At the outset of this work, a relationship was expected, but in the opposite direction that we see here.

Third is the share of the adult population that is married. This is consistent with the broader research literature, whereby marriage provides a number of advantages to entrepreneurs (income safety net, unpaid labor, startup capital, etc.) and marriage rates at a community level are seen to be indicative of greater social capital (an important factor for entrepreneurial success is a local social capital).

Again, the relationship between these three factors and I5HGC density across metros is less robust statistically than the previous five explanatory variables when considered jointly—so they should be considered with some degree of caution. However, their existence is worth discussing and consideration for future work.

Endnotes

[1] For a summary of this literature, see Parker (2009), *The Economics of Entrepreneurship*, Cambridge University Press. Also see Acs and Armington, (2006), *Entrepreneurship, Geography, and American Economic Growth*, Cambridge University Press. These excluded regional variables include: the foreign-born percent of the population; median household income; unemployment rate; Gini coefficient (a measure of income inequality); voter participation rate; venture capital per capita; natural amenities (weather, topographic variation, proximity to water); net migration rate; university share of employment; industry diversity of employment (HHI); non-employer firms per capita; personal income per capita growth (5-year); population change (5-year); business formation rate thirty-years prior; and log of population.

[2] For a summary of this literature, see Chatterji, Glaeser, and Kerr (2013), "Clusters of Entrepreneurship and Innovation", *National Bureau of Economic Research*.

[3] See Porter (1998), *On Competition*, Harvard Business School Publishing; Gilbert, McDougall, and Audretsch (2008), "Clusters, Knowledge Spillovers and New Venture Performance: An Empirical Examination", *Journal of Business Venturing*; Gilbert, McDougall, and Audretsch (2006), "New Venture Growth: A Review and Extension", *Journal of Management*; Chatterji, Glaeser, and Kerr (2013), "Clusters of Entrepreneurship and Innovation", *National Bureau of Economic Research*.

[4] For more on this, see Chatterji, Glaeser, and Kerr (2013), "Clusters of Entrepreneurship and Innovation", *National Bureau of Economic Research*.

[5] For a summary of this literature, see Parker (2009), *The Economics of Entrepreneurship*, Cambridge University Press. See also Wadhwa, Aggarwal, Holly, and Salkever (2009), "The Anatomy of an Entrepreneur", *Kauffman Foundation*.

[6] For more on this, see Acs and Armington, (2006), *Entrepreneurship, Geography, and American Economic Growth*, Cambridge University Press; and Chatterji, Glaeser, and Kerr (2013), "Clusters of Entrepreneurship and Innovation", *National Bureau of Economic Research.*