BEYOND METROPOLITAN STARTUP RATES:
Regional Factors Associated with Startup Growth

Yasuyuki Motoyama, PhD and
Jordan Bell-Masterson
Ewing Marion Kauffman Foundation

January 2014

Ewing Marion
KAUFFMAN
Foundation

BEYOND METROPOLITAN STARTUP RATES:
Regional Factors Associated with Startup Growth

Yasuyuki Motoyama, PhD and
Jordan Bell-Masterson
Ewing Marion Kauffman Foundation

January 2014

The authors would like to thank Ian Hathaway, Dane Stangler, Arnobio Morelix, Michelle St. Clair, Lacey Graverson, and Barbara Pruitt.

BEYOND METROPOLITAN STARTUP RATES:
Regional Factors Associated with Startup Growth

EXECUTIVE SUMMARY

Understanding what fosters—and hinders—firm formation and growth at the metropolitan level across the United States is a challenge. Entrepreneurship can be measured by a variety of indicators, and they each can tell somewhat different stories. Furthermore, because entrepreneurship can refer to the growth of firms from a startup stage to mid- or large-scale, no one dataset covers the full range of companies that fall in this category.

This report contributes to the Kauffman Foundation's recent series of analyses on the rate of business creation in metropolitan areas. Going beyond identifying metropolitan areas with higher rates of entrepreneurship, we analyze what regional factors are associated, or unassociated, with entrepreneurial activity. Understanding what drives entrepreneurship at the regional level—especially high-growth business creation—will help policymakers and entrepreneurship supporters know where to invest their efforts.

We examine entrepreneurship activity at 356 metropolitan areas in the United States employing three sources: the Business Dynamics Statistics, the National Establishment Time-Series (NETS), and data on high-growth Inc. firms. This allows researchers to investigate the rates of entrepreneurship from multiple angles:

- the startup rate for all industries (BDS)
- the rate for high-tech sectors (NETS)
- the rate for high-growth firms (Inc.)

Key findings in this paper dispel some myths about what factors influence startup rates and growth in metro areas:

- Contrary to conventional understanding in literature, we find few significant factors that the public sector can affect. Despite billions of dollars in government research expenditures, the presence of research universities and patents are not associated with higher rates of entrepreneurship.
- The most significant factor by the public sector is related to education. High school and college completion is important when it comes to startup rates. However, while it is true that a high ratio of college graduates in a metropolitan area means more startups, a substantial high school completion rate will further increase the area's startup rate.

- The investment level of financial organizations, primarily by venture capitalists, in a metro area does not correlate to high startup activity. And VC-invested regions do not necessarily generate a higher ratio of startups. Policymakers should not rush to create public venture funds in the hope of creating more startups or a startup culture.
- High-tech sectors are not hotbeds for all kinds of startups – only for high-tech sectors. In other words, promoting high-tech entrepreneurship does not necessarily bring up the overall economy.
- Not surprising, but confirming, larger metropolitan areas tend to have higher entrepreneurial rates, possibly from the diversity and resilience of their economies.

We hope that this paper and the new compilation of metro-level data will serve as the first step in promoting more rigorous research about the dynamic relationships between startups and regional factors, and the relationships between different startup indicators, geographic factors, and others.

INTRODUCTION AND RESEARCH RATIONALE

This report will contribute to the Kauffman Foundation's recent series of analyses on the rate of business creation in metropolitan regions and in the nation as a whole. Recent research indicates that there is substantial variation in the business creation rates within states. Hathaway et al. (2013, 8) used the recently released series of Business Dynamics Statistics (BDS) and found substantial geographic variance in business creation rates within a state. For instance, the high level of business creation enjoyed by the state of North Dakota overall in 2010 was not reflected in the three metropolitan areas within the state. Similarly, California, Pennsylvania, and Kentucky had metropolitan areas with more than 20 percent increases in business creation rates, while other areas of the same states experienced substantial declines. This variation suggests the need for further analysis at a finer geographic scale. In this report, we delve into this metropolitan analysis, the smallest geographic scale in the BDS.

Going beyond efforts to identify the metropolitan areas with high or low ratios of business creation, this report also contributes additional analysis by identifying regional factors associated with those places with high levels of business creation. Previous work on this issue includes Konczal's (2013) analysis of levels of entrepreneurship by metropolitan areas over time using the BDS data. In addition, Stangler (2013), using National Establishment Time-Series (NETS) data, noted the high density of high-tech startup activities in small to medium metropolitan areas with first-class research universities, such as Boulder, Colo., Fort Collins, Colo., Ann Arbor, Mich., Lafayette, La., and Lexington, Ky. While Stangler cautioned that this pattern was uneven, it is important to determine if such a pattern persists in other small-to-medium metropolitan areas and if, indeed, regional indicators of research university activity correlate with high startup density.

METHODS: VARIABLES AND DESCRIPTIVE ANALYSIS

Measuring entrepreneurship is a challenge. While most people agree that entrepreneurship is related to starting a company, no single standard indicator for entrepreneurship exists. In the past, the Kauffman Foundation has supported an entrepreneurship index, the Kauffman Index of Entrepreneurial Activity, and has cooperated with the Census Bureau to create a geographic tabulation of the BDS at the metropolitan level. Motoyama and Konzcal (2013) demonstrated an only moderate correlation of these two indicators, 0.655, meaning that apparently more than a third of the variation cannot be explained by the other variable. It does not help to examine the self-employment rate, available from the American Community Survey, or the Decennial Census, because the majority of self-employment comes from micro-enterprises with no or few employees and with no prospect for growth. These assumptions can be validated by the low correlations with the Kauffman Index or the BDS, 0.34 or 0.04, respectively (Ibid.).[1] In short, the state of entrepreneurship can be measured by different indicators, and they indeed tell somewhat different stories. Furthermore, entrepreneurship also can refer to the growth of firms from a startup stage to mid- or large-scale. Starting a new business is only one of the steps in entrepreneurship, as entrepreneurs also have to scale the firm successfully.

In this paper, we start with an analysis of the startup rate based on the BDS data. Because of the aforementioned variations in the meaning and indicators of entrepreneurship, we also employ other indicators of entrepreneurship, NETS and Inc. firms, and analyze them comparatively. First, the BDS is created and maintained by the Census Bureau, which compiles data from the Business Register, payroll tax records from the Internal Revenue Service, the Company Organization Survey, Annual Survey of Manufacturers, and other Census-organized surveys (Census 2013). The BDS often is seen as the most comprehensive source of business creation data, although it does not necessarily offer 100 percent coverage, and its advantage is in tracking age and employment information at the establishment level, rather than the firm level. Information about firm age is critical because past academic studies have demonstrated that virtually all new jobs created in this country come from young firms (five years or younger) (Kane 2010; Haltiwanger 2012; Haltiwanger, Jarmin, and Miranda 2013). One disadvantage of BDS is that, in order to protect confidentiality, it does not allow detailed industry information when the geographic unit of analysis is smaller than a state. In short, the BDS is best for tracing business creation in the economy overall.

While the new firm formation rate in the BDS has been steady historically, there was a substantial decline during the recent Great Recession; we are now at a historically low firm formation rate. Hathaway et al. (2013) demonstrated that 2006 was the most recent peak, and the rate has declined for the following four consecutive years. Only in 2011 did we start to see a modest recovery. For our analysis, we look at the new firm formation rate averaged between 2010 and 2011.[2]

[1] As a reference, the Kauffman Index is calculated from the Current Population Survey. See Fairlie (2013) for details.

[2] As a reference, we conducted the regression models with the average of 2006 and 2009 and obtained

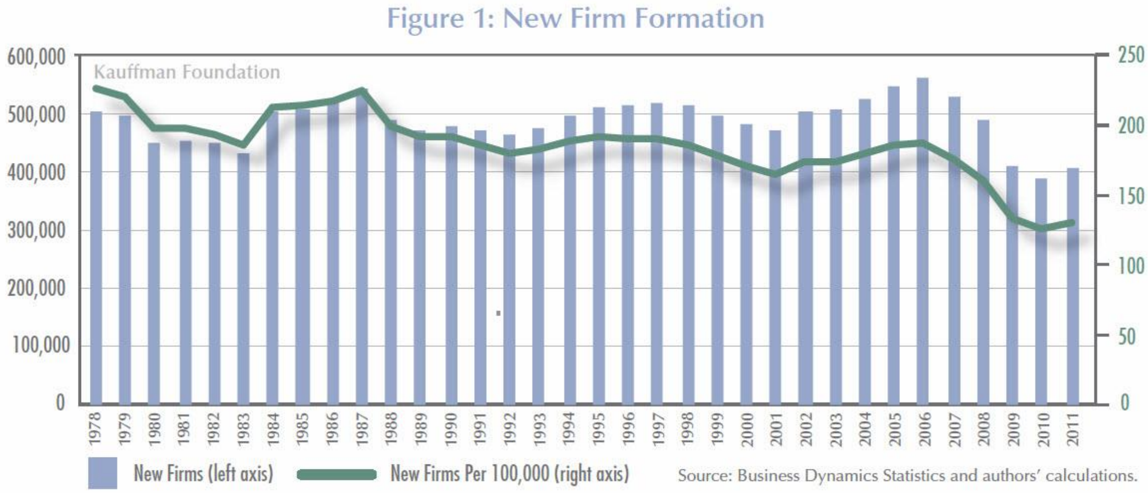

Figure 1: New firm formation by the BDS
Source: Hathaway et al. (2013, 3).

We used the NETS data as a second indicator of entrepreneurship. Dun and Bradstreet, a private data firm, collects firm and establishment data each year, and Walls and Associates converts it to time-series data. Its coverage includes private and public businesses (both employer firms and the self-employed), sole proprietors, nonprofit, and government establishments. Since this is a micro-level database, we can tabulate at any geographic level and specify any industrial groups. A downside of these data is that the coverage may not be as systematic as that of BDS. In addition, data collected by a private firm are less reliable for firm-specific information, such as employment and revenue.

As the NETS database allows us to disaggregate by industry, we use it to examine the startup creation rate for the high-tech sectors and the information and communication technology (ICT) sectors, a subset of the former. We use the definition of high-tech sectors provided by the Bureau of Labor Statistics (Hecker 2005), also used by Hathaway (2013). Broadly, they are the ICT, pharmaceutical, aerospace, engineering services, and scientific research and development sectors. Figure 2 demonstrates that this indicator fluctuates yearly, and we take the average of 2009 and 2010.

essentially the same result.

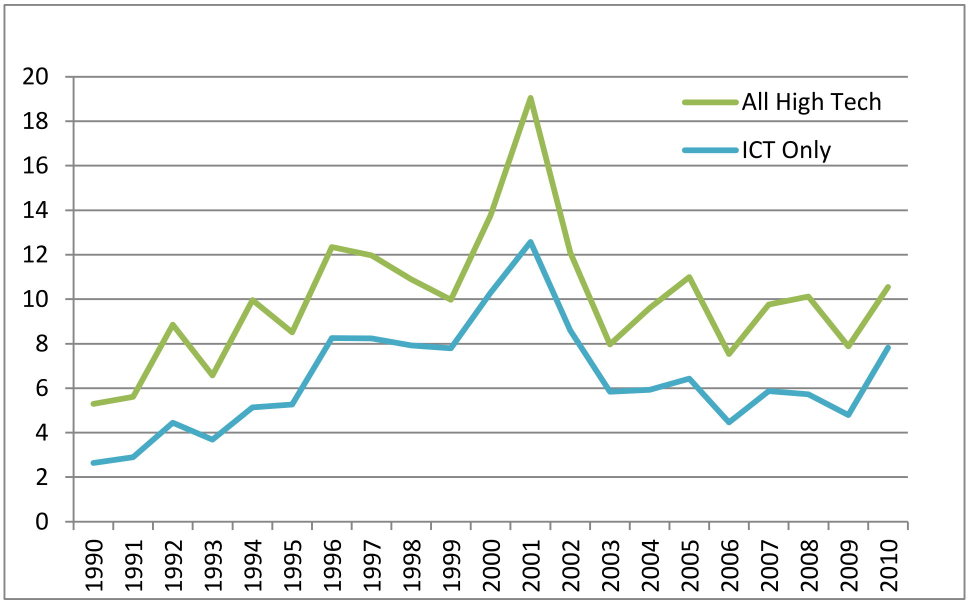

Figure 2: New firm formation by NETS, per 100,000 people
Source: Hathaway (2013).

Our third indicator for entrepreneurship is Inc. 500/5000 firms. Motoyama and Danley (2012, 7) provided more detail about these data. In essence, the Inc. list ranks fast-growing private firms with more than two million dollars in annual revenue, based on revenue growth in the previous three years. From the Inc. list, we select firms that achieved high growth based on the Organisation for Economic Cooperation and Development's definition: more than 20 percent revenue growth for three consecutive years (i.e., 72.8 percent growth over three years). Between 2,700 and 2,800 firms per year qualify in this category, and we call these firms the Inc. high-growth firms.

The core analysis in this paper is identifying regional factors associated with the rate of firm creation or growth. We employ multivariate regression analysis. As noted, we have three separate dependent variables: the BDS, NETS, and Inc. For independent variables, we draw analytical frameworks from two broad streams of literature. The first comprises studies about regional knowledge spillover conducted by Jaffe, Trajtenberg, and Henderson (1993), Feldman and Florida (1994), Audretsch and Feldman (1996, 2004), and Feldman and Audretsch (1999). These studies determined that innovation is a function of university research, industrial research, networks of related businesses, and human capital (Cohen and Klepper 1991, 1992). The second stream of studies includes the innovative milieu theory (Castells 1989; Castells and Halls 1994) and the cluster theory by Porter (1994, 1998, 2000). This literature adds research and expenditures by government and its laboratories to our list of independent variables.

We recreate these variables from various sources at the metropolitan level. Table 1 summarizes the types of variables and their sources. For more detailed methodological notes about these variables, please see Appendix 1.

Table 1:

#	Name	Source	Year(s)
Dependent Variables			
A)	BDS business creation: firms with 0–1 age normalized by the total number of firms	BDS	2010–11
B)	NETS business creation: firms with high tech and ICT firms / 100,000 people	NETS	2009–10
C)	Inc. firms / million population / 1 million people	Inc. magazine	2010–12
Independent Variables			
1)	Location quotient (LQ) of high-tech sector	Milken Institute	2007
2)	Patents per capita	USPTO (2013)	2008–10
3)	Government R&D: NIH grants & SBIR grants per capita	NIH for NIH grants; SBA TECH-Net for SBIR grants	2008–10
4)	Investment by financial organizations, such as Venture Capital (VC)	CrunchBase	2009–11
5)	University research: count of Carnegie Research I universities and their research expenditures	Carnegie Classification of Institutions of Higher Education	2010
6)	Education: high school and college completion rates, college attendance rate	American Community Survey (Census 2013b)	2011
7)	Population: population in 2011 and growth between 2006 and 2010	Census	2006–11

The BDS follows the 2009 definition of metropolitan areas created by the Office of Management and Budget (OMB), resulting in 366 metropolitan areas. We omitted ten metropolitan areas that are not consistent with different sources of data, leaving 356 metropolitan areas for our analysis.

REGRESSION ANALYSIS

Figure 3 plots histograms of our dependent variables. The BDS data are slightly skewed to the right, but the standard Ordinary Least Square (OLS) can still robustly explain. The NETS data is highly skewed to the right, and we employ the log form, which sets the distribution closer to a bell shape, which is appropriate for OLS.[3] The Inc. data also are highly skewed to the right. Unlike the NETS data, the Inc. data have many metropolitan areas with a zero value,[4] so we cannot employ the log form. We construct zero-inflated negative binomial models (ZINB)[5] by following Zuur et al. (2009) and Statistical Consulting Group at UCLA (2013).

[3] There was only one case in the ICT sector in which the value was zero, which presents a problem for regression because log(0) is minus infinity. We omitted this case.

[4] There are 106 cases in which a metropolitan area has a zero value in this Inc. list.

[5] We also used a zero-inflated poisson model, tested by the log-likelihood test, and concluded that the zero-inflated negative binomial models fit better in all models.

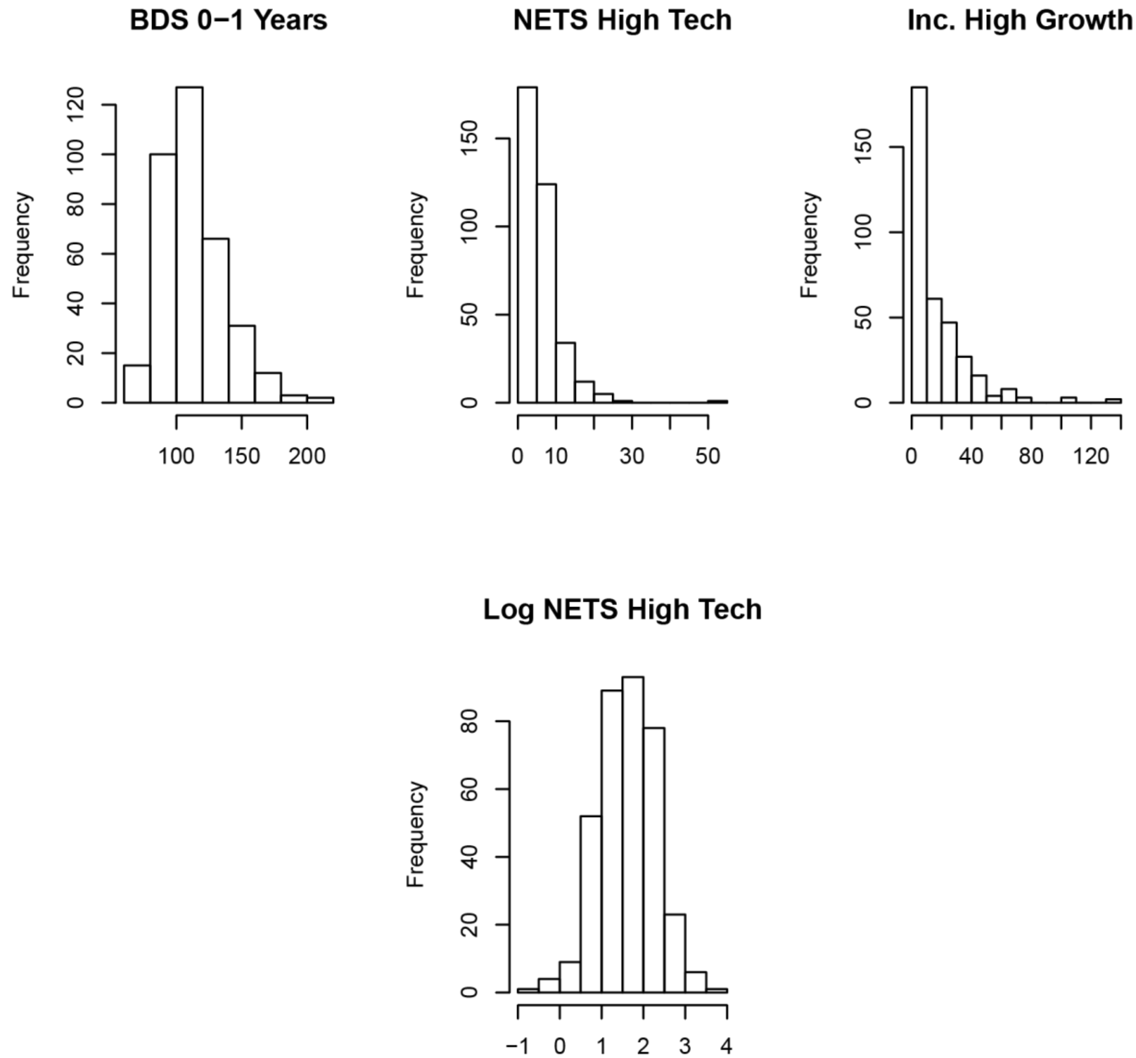

Figure 3: Three dependent variables and log form of NETS.

Overall, the models fit well. We use only nine independent variables, but achieve an adjusted R-square of approximately 0.5 or higher for all models. The first set of regressions with the BDS data provides somewhat surprising results. There are only two significant factors: population in log form and the population increase. The college completion rate is significant but only negatively, meaning that the higher the ratio of college graduates, the lower the ratio of startups. This result, however, is only at the 95 percent level, so we do not consider it definitive. All other variables are insignificant: the presence of high-tech sectors, government R&D in SBIR or NIH, investment by financial institutions, patents, and research universities.

Table 2: Regression Results with the Startup Rate by BDS

	Model 1		Model 2		Model 3		Model 4		Model 5	
(Intercept)	-29.931		-28.810		-29.094		-33.762		-22.096	
Log(Population)	10.582	**	10.434	**	10.484	**	9.671	**	9.734	**
Pop. Increase 2006–10	362.871	**	361.295	**	364.066	**	364.689	**	359.754	**
High-tech LQ	2.375		2.958		2.323		1.261		1.557	
SBIR	0.020				0.019		0.017		0.016	
NIH			0.001							
Investment	0.002		0.002		0.002		0.002		0.002	
Patents	0.000		0.000		0.001		0.000		0.000	
Research I	0.283		0.341		0.432		-1.831		0.313	
Research I expenditure										
College completion	-0.335	*	-0.314	*	-0.311	*				
HS completion							0.087			
College attendance									-0.130	
DF	344		344		344		344		344	
F-stat	56.75		56.51		56.65		55.29		55.29	
Adj. R-sq	0.5589		0.5578		0.5584		0.5524		0.5524	

Note: ** 99 percent significance level; * 95 percent significance level

The regression result from the high-tech sectors with NETS gives a different perspective. We find that the population size and increase are significant, consistent with the BDS regression. However, in these data, the presence of high-tech sectors (LQ) also is significant. Government expenditure on research via NIH, patents, and investment by financial institutions still are insignificant, and SBIR is sporadically significant only at the 95 percent level. The college completion rate and high school completion rates are significant, but college attendance is not. Essentially the same results were found when using only the ICT sector in the NETS.

Table 3: Regression Results for NETS All High-Tech Firms

	Model 1		Model 2		Model 3		Model 4		Model 5	
(Intercept)	-1.08071		-1.03767		-1.02631		-3.82058		-1.44819	
Log(Population)	0.11234	**	0.10140	**	0.10404	**	0.17791	**	0.18749	**
Pop. Increase 2006–10	5.32544	**	5.22734	**	5.46306	**	5.77356	**	5.58458	**
High-tech LQ	0.24995	**	0.28109	**	0.21878	**	0.33340	**	0.32646	**
SBIR	0.00109				0.00177	*	0.00212	*	0.00211	*
NIH			-0.00005							
Investment	-0.00003		-0.00002		-0.00004		-0.00004		-0.00005	
Patents	0.00002		0.00002		0.00002		0.00006		0.00007	
Research I	-0.00339		0.00217							
Research I expenditure					-0.11387		0.09889	*	0.07453	
College completion	0.03084	**	0.03447	**	0.03353	**				
HS completion							0.02780	**		
College attendance									0.00299	
DF	344		344		344		344		344	
F-stat	57.62		56.90		58.50		44.02		41.85	
Adj. R-sq	0.5627		0.5595		0.5665		0.4944		0.4815	

Last, we present the results of zero-inflated models[6] with Inc. data. Zero-inflated binomial models are two-stage models, and the first zero-inflation model tests which factors are associated to explain the probability of zero (note that all coefficients except the intercept are negative). For these data, the only significant variable at the 99 percent level is log of population, and its negative coefficient means that the smaller the population, the more likely there will be zero high-growth firms.

A second analysis uses the count model to explain which factors are associated when metropolitan areas have more than one high-growth firm. Again, population size and increase are significant, and so is the presence of high-tech sectors. SBIR, NIH, investment by financial institutions, patents, and the presence of research universities are insignificant. The college completion rate still is significant, and the college attendance rate is significant at the 95 percent level, but the high school completion rate is not significant.

[6] We use pscl and lmtest packages of R. Additionally, since the zeroinfl function of R only uses counts, we use integers of the dependent variables.

Table 4: Regression Results for Inc. High-Growth Firms

1) Zero-Inflation Model

	Model 1		Model 2		Model 3		Model 4		Model 5	
(Intercept)	24.9260		24.8676		25.0060		30.0678		24.7977	
Log(Population)	-1.9373	**	-1.9348	**	-1.9533	**	-1.9617	**	-1.9694	**
Pop. Increase 2006-10	-6.5962		-6.5208		-6.2896		-7.1926		-6.6924	
High-tech LQ	-0.8720		-0.9013		-0.9156		-1.0650		-1.0585	
SBIR	-0.0026				-0.0004		-0.0007		-0.0001	
NIH			-0.0002							
Investment	-0.0008		-0.0012		-0.0009		-0.0007		-0.0010	
Patents	-0.0006	*	-0.0006	*	-0.0006	*	-0.0007	*	-0.0007	**
Research I	-0.0084		-0.0086							
Research I expenditure					-0.2787		-0.4403		-0.4739	
College completion	-0.0326		-0.0304		-0.0270					
HS completion							-0.0584			
College attendance									-0.0022	

2) Count Model

	Model 1		Model 2		Model 3		Model 4		Model 5	
(Intercept)	0.0002		-0.0225		0.0255		-2.5816		-0.6586	
Log(Population)	0.1071	*	0.1027	*	0.1006	*	0.1970	**	0.2041	**
Pop. Increase 2006-10	3.4270	**	3.3122	**	3.5098	**	3.7478	**	3.7961	**
High-tech LQ	0.4198	**	0.4148	**	0.3738	**	0.4948	**	0.4943	**
SBIR	0.0000				0.0008		0.0013		0.0014	
NIH			-0.0001							
Investment	0.0001		0.0001		0.0001		0.0001		0.0001	
Patents	-0.0001		-0.0001		-0.0001		0.0000		0.0000	
Research I	-0.0263		-0.0227							
Research I expenditure					-0.1814		0.0846		-0.0482	
College completion	0.0389	**	0.0425	**	0.0418	**				
HS completion							0.0251			
College attendance									0.0095	*
Log(theta)	1.1310	**	1.1370	**	1.1420	**	0.9337	**	1.0040	**
Theta	3.099		3.118		3.133		2.701		2.729	
# of iterations	37		36		39		32		37	
Log-likelihood	-1074		-1074		-1073		-1088		-1088	

DISCUSSION

The three sets of regressions yield somewhat different results, but they present an overarching theme. In all cases, the two population variables are significant: the population size (in a log form) and the population increase. This supports the simple, inter-related relationship between population and startups: areas with larger populations and more population growth will have higher ratios of startups. This conclusion also suggests regional divergence, as smaller and non-growing regions tend to have the lower startup ratios, holding all other factors constant.

While our regression models cannot explain why so, we explore several possibilities here. We find that this conclusion is similar to those of studies of regional resilience that investigate the types of metropolitan areas that recover more quickly after recession. These studies found that, holding other variables constant, larger metropolitan areas show greater resilience (Blumenthal et al. 2009; Chapple and Lester 2010; Hill et al. 2012). This may be because larger metros have more service sectors, such as retail stores, restaurants, and other enterprises, which are small and tend to be startups. While this explanation is consistent with the BDS model, which includes all kinds of startups, it does not explain the results from the NETS and Inc. models. A second possibility is that larger metros have more diverse sectors, and such diversity brings in more business and startup opportunities. This interpretation may apply to all three models. More research that can model factors suggested by these two possibilities is needed.

Contrary to the conclusions of most earlier studies, we find few significant factors that the public sector can affect. Government research expenditure through NIH and university research expenditure are uniformly insignificant. SBIR is significant only at the 95 percent level and only sporadically with limited models of NETS high-tech firms. A conservative approach would not include it as a significant factor. This conclusion would be intuitive if the NIH expenditure, which is exclusively health-related, was not associated with startups in all sectors or Inc. high-growth firms, but our NETS models focused specifically on high-tech sectors, including the pharmaceutical and medicine manufacturing sectors. Thus, research expenditures on health and biotech do not necessarily translate in startups even in the respective sector. In sum, science- and R&D-oriented government expenditure is not linked to startup activities in all sectors or even high-tech sectors.

Our models indicate that the most significant factor that the public sector may affect is related to education. Although the college completion rate is significant only at the 95 percent level with the BDS models, it is significant at the 99 percent level for the other two dependent variables. Thus, it is fair to claim that a population with more adults (between the ages of twenty-five and forty-four) who complete college will produce more startups. Note that the college attendance ratio is a flow variable, and the attendance ratio is not significant in most cases. In contrast, the ratio of adults who have completed college is a stock variable. Often, people suggest that startups are created by young college drop-outs or very recent college graduates in the tech sector. The results of our

NETS models suggest that this is not the case, but that the cumulative stock of people with college degrees matters for startups.

When academic studies discuss educational attainment or human capital, college completion often is taken as the minimum indicator of high skill. We extended this measurement to high school completion and found it significant for the ratio of startups in high-tech sectors in the NETS models. High school completion, therefore, is important, in addition to college completion. As a reference, the college and high school completion rates are not well correlated, at only 0.27 with Pearson correlation. (See Figure 4 for the distribution.) Thus, when policymakers discuss high school and college completion, they should not assume that one will automatically lead to the other, but think about how to connect these two factors effectively. Put another way, it is true that high numbers of college graduates in a metropolitan area will be good for startups there. These graduates are, however, only one piece of the puzzle. A substantial high school completion rate will further increase the area's startup rate.

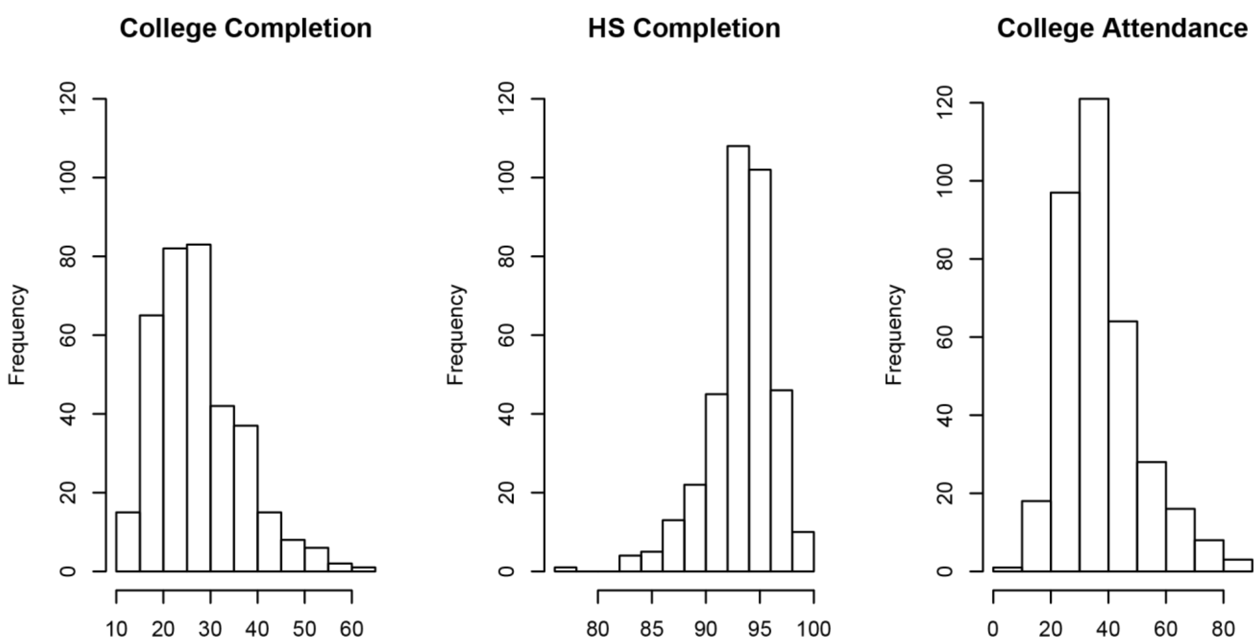

Figure 4: Histogram of college completion rates, high school completion rates, and college attendance rates.

While this kind of individual information is not available in the BDS or NETS data, the Kauffman Firm Survey data indicated that many entrepreneurs do not have college degrees. In fact, the percentage of entrepreneurs with high school degrees or less education is 15.3 percent. If we include those with technical degrees (6.7 percent), some college with no degree (22.5 percent), and associate degrees (8.4 percent), the total is 52.6 percent (Kauffman Foundation et al. 2008).[7] This group represents more

[7] The Kauffman Firm Survey captures all kinds of startup firms. As a reference, our internal research suggests that entrepreneurs who have high school degrees or less education comprise about 4.5 percent of Inc. firms.

than half of the entrepreneur population, and an exclusive focus on college education obscures the significant role played by entrepreneurs without college degrees. The positively significant results for both the high school and college completion rates indicate that there are many factors involved in educational attainment, and a simplistic, linear assumption that more education (i.e., college completion) leads to more entrepreneurship is probably not true. Unfortunately, the education variables in our models do not capture these segments between high school completion and college completion, and further research is needed to assess the more mechanical ties between educational attainment and entrepreneurship.

While further investigation of this education variable is necessary, note that we measured all college graduates in all areas of study. The recent policy debate has focused primarily on STEM (Science, Technology, Education, and Math) education, but our models indicate that college graduates in other disciplines also are likely to contribute to the creation of startups, given that 88 percent of college graduates are from non-STEM fields (Ryan 2012).

The LQ of high-tech sectors also is significant, meaning that the presence of other high-tech firms in the area is associated with higher startup rates. This conclusion makes intuitive sense, as the NETS models examined the startup ratio of high-tech firms. At the same time, we should keep in mind that a higher LQ does not yield a higher startup ratio for all firms in the BDS models. High-tech sectors, then, are not hotbeds for all kinds of startups, but only for high-tech sectors.

In each of our models, investment by financial institutions does not correlate with startup ratios. This is consistent with our earlier analysis at the state level (Motoyama and Danley 2012). Furthermore, it complements the micro-level assessment of venture capital's low return by Bradley et al. (2012), which found that approximately 80 percent of venture capitals (VCs) do not produce returns even at the level of general stock investments, 3 percent per annum. In other words, VCs do not yield above-market returns, and VC-invested regions do not necessarily generate higher ratios of startups. Policymakers should not rush to create public venture funds in the hope of creating more startups or a startup culture.[8]

It is worth noting the insignificance of various science and R&D-related factors: SBIR, NIH, the presence of research universities, and patents. These factors often are modeled as a result of a series of assumptions based on the linear model of development. These begin with the idea that the government allocates research expenditures for technology firms, health-related research, and university research. In the process, it follows that patents are filed to protect intellectual property rights of inventors, and then such stock of knowledge and intellectual property rights will eventually trickle down to the private sector. Yet, the insignificance of all these factors in virtually all of our models suggests that these assumptions do not hold true for the creation of startups and firm growth.

[8] For further discussion about the failure of public venture funds, see Lerner (2009).

Again, while our regression models do not indicate why the linear model is not working, we explore possible interpretations. Since this linear model of development assumes that technology, especially new technology driven from scientific research, creates a *push* in the economy, it does not incorporate other dimensions of innovations, business creation, and opportunity identification. Entrepreneurs may identify business opportunities and create new businesses possibly because of the market *pull* (Schmookler 1966). In fact, several emerging qualitative studies have found that users have become entrepreneurs (Shah, Smith, and Reedy 2012) and that successful entrepreneurs are skilled at finding market niches and keen on maintaining close relationships with customers and suppliers (Motoyama et al. 2013). These patterns of business creation are not and should not be mutually exclusive, but the market pull or other patterns may be responsible for the overwhelming majority of business creation, instead of the technology push. If this is true, factors closely reflecting the linear model would not be effective in influencing the overall economic activity at the metropolitan level.

If this linear model does not function, we must reconsider the conventional policy practice of regional development, which follows the linear model by establishing science parks, university technology transfer offices, research expenditures through universities, and publicly funded venture capital funds (Plosila 2004; Mayer 2007; Sa, Greiger, and Hallacher 2008; Lerner 2009). This review would take a broader view of the interaction between government, universities, and firms in their efficiency and effectiveness in creating new businesses, innovations (which by definition includes commercialization), and firm growth.

CONCLUDING REMARKS

Our models analyzed the patterns of entrepreneurship with different measures of business creation and growth in almost all metropolitan areas, using the major variables discussed in previous studies. These models are statistically significant and robust, despite their simplicity. However, we have not incorporated the spatial autocorrelation, i.e., how proximate areas may affect each other. We have observed some clustering of high startup rates in nearby metros, such as Provo-Ogden-Salt Lake City and Boulder-Denver. Our next step, therefore, is to conduct a geographically weighted regression.

We plan to open-source these startup data at the metropolitan level and make all the data publicly available for academic and non-academic researchers. We hope that this paper and the available data will serve as the next step in promoting more rigorous research about the dynamic relationships between startups and regional factors, and relationships between different startup indicators, geographic factors, and others.

REFERENCES

Audretsch, David B., and Maryann P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* no. 86 (3):630–640.

Audretsch, David B., and Maryann P. Feldman. 2004. "Knowledge Spillovers and the Geography of Innovation." In *Handbook of Regional and Urban Economics*, edited by Vernon Henderson, Peter Nijkamp, and Jacques-Francois Thisse, 2713–2739. Amsterdarm: Elsevier.

Blumenthal, Pamela, Harold L. Wolman, and Edward Hill. 2009. "Understanding the economic performance of metropolitan areas in the United States." *Urban Studies* no. 46 (3):605–627.

Bradley, Harold S., Diane Mulcahy, and Bill Weeks. 2012. We have met the enemy ... and he is us: Lessons from twenty years of the Kauffman Foundation's investments in venture capital funds and the triumph of hope over experience. In *Kauffman Foundation Research Series*. Kansas City, Mo: Kauffman Foundation.

Castells, Manuel. 1989. *Informational City: Information Technology, Economic Restructuring, and the Urban-Regional Process*. Oxford: Blackwell.

Castells, Manuel, and Peter Geoffrey Hall. 1994. *Technopoles of the world: the making of twenty-first-century industrial complexes*. London: Routledge.

Census Bureau. 2013a. *Overview - Business Dynamics Statistics, footnote 1*. Census Bureau 2013 [cited Nov. 12, 2013]. Available from http://www.census.gov/ces/dataproducts/bds/overview.html#footnote1.

Bureau, Census. 2013b. *American Community Survey data on educational attainment*. Census 2013 [cited Nov. 1, 2013]. Available from http://www.census.gov/hhes/socdemo/education/data/acs/index.html.

Chapple, Karen, and T. William Lester. 2010. "The resilient regional labour market? The US case." *Cambridge Journal of Regions, Economy and Society* no. 3 (1):85–104.

Cohen, Wesley M., and Steven Klepper. 1991. "Firm size versus diversity in the achievement of technological advance." In *Innovation and technological change: An international comparison*, edited by Zoltan Acs and David B. Audretsch, 183–203. Ann Arbor: University of Michigan Press.

Cohen, Wesley M., and Steven Klepper. 1992. "The anatomy of industry R&D intensity distributions." *American Economic Review* no. 82 (4):773–799.

Fairlie, Robert W. 2013. Kauffman Index on Entrepreneurial Activity. In *Kauffman Foundation Research Paper Series*. Kansas City, Mo.

Feldman, Maryann P., and Richard Florida. 1994. "Geographic Sources of Innovation: Technological Infrastructure and Product Innovation in the United States." *Annals of Association of American Geographers* no. 84 (2):210–229.

Feldman, Maryann P., and David B. Audretsch. 1999. "Innovation in cities: Science-based diversity, specialization and localized competition." *European Economic Review* no. 43:409–429.

Haltiwanger, John. 2012. "Job creation and firm dynamics in the U.S." *Innovation Policy and the Economy* no. 12:17–38.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. 2013. "Who creates jobs? Small vs. large vs. young." *Review of Economics and Statistics* no. 95 (2):347–361.

Hathaway, Ian, Jordan Bell-Masterson, and Dane Stangler. 2013. The return of business creation. In *Kauffman Foundation Research Series*. Kansas City, Mo: Kauffman Foundation.

Hecker, Daniel, E. 2005. "High-technology employment: A NAICS-based update." *Monthly Labor Review* (July):57–72.

Hill, Edward, Travis St. Claire, Howard Wial, Harold L. Wolman, Patricia Atkins, Pamela Blumenthal, Sarah Ficenec, and Alec Friedhoff. 2012. "Economic shocks and regional economic resilience." In *Urban and regional policy and its effects*, edited by Margaret Weir, Nancy Pindus, Howard Wial and Harold L. Wolman. Washington, D.C.: Brookings Institution Press.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Location of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* no. 63 (3):577–598.

Kauffman Foundation, Janice Ballou, Tom Barton, David DesRoches, Frank Potter, E.J. Reedy, Alicia Robb, Scott Shane, and Zhanyun Zhao. 2008. Kauffman Firm Survey: Results from the baseline and first follow-up surveys. Kansas City, Mo: Ewing Marion Kauffman Foundation.

Kane, Tim. 2010. The importance of startups in job creation and job destruction. In *Kauffman Foundation Research Series*. Kansas City, Mo: Kauffman Foundation.

Konczal, Jared. 2013. The most entrepreneurial metropolitan area? In *Kauffman Foundation Research Paper Series*. Kansas City, Mo: Kauffman Foundation.

Lerner, Joshua. 2009. *Boulevard of broken dreams: Why public efforts to boost entrepreneurship and venture capital have failed and what to do about it*. Princeton: Princeton University Press.

Mayer, Heike. 2007. "What is the role of the university in creating a high-technology region?" *Journal of Urban Technology* no. 14 (3):33–58.

Motoyama, Yasuyuki, and Brian Danley. 2012. The ascent of America's high-growth companies: An analysis of the geography of entrepreneurship. In *Kauffman Research Paper Series*. Kansas City, Mo: Kauffman Foundation.

Motoyama, Yasuyuki, and Jared Konczal. 2013. "How can I create my favorite state ranking?" *Journal of Applied Research in Economic Development* no. 10 (Sept):1–17.

Motoyama, Yasuyuki, Brian Danley, Jordan Bell-Masterson, Kate Maxwell, and Arnobio Morelix. 2013. Leveraging regional assets: Insights from high-growth companies in Kansas City. In *Kauffman Foundation Research Paper*. Kansas City, Mo: Kauffman Foundation.

Plosila, W.H. 2004. "State science- and technology-based economic development policy: History, trends and developments, and future directions." *Economic Development Quarterly* no. 18 (2):113–126.

Porter, Michael E. 1994. "The Role of Location in Competition." *Journal of the Economics of Business* no. 1 (1):35–39.

Porter, Michael E.. 1998. "Clusters and the New Economics of Competition." *Harvard Business Review* no. 76 (6):77–90.

Porter, Michael E. 2000. “Location, Competition, and Economic Development: Local Clusters in a Global Economy.” *Economic Development Quarterly* no. 14 (1):15–34.
Ryan, Camille. 2012. Field of degree and earnings by selected employment characteristics: 2011. In *American Community Survey Briefs*. Washington, D.C.: Census Bureau.
Sa, Creso M., Roger L. Geiger, and Paul M. Hallacher. 2008. “Universities and state policy formation: Rationalizing a nanotechnology strategy in Pennsylvania.” *Review of Policy Research* no. 25 (1):3–19.
Schmookler, Jacob. 1966. *Invention and Economic Growth*. Cambridge, Mass.: Harvard University Press.
Shah, Sonali K., Sheryl Winston Smith, and E.J. Reedy. 2012. Who are user entrepreneurs? Findings on innovation, founder characteristics, and firm characteristics. In *Kauffman Foundation Research Paper Series*. Kansas City, Mo: Ewing Marion Kauffman Foundation.
Stangler, Dane. 2013. Path-dependent startup hubs: Comparing metropolitan performance. In *Kauffman Foundation Research Series*. Kansas City, Mo: Kauffman Foundation.
UCLA. 2013. *Zero-inflated negative binomial regression*. Statistical Consulting Group 2013 [cited Nov. 18, 2013]. Available from http://www.ats.ucla.edu/stat/r/dae/zinbreg.htm.
USPTO. 2013. *Patenting in technology classes, breakout by origin, U.S. metropolitan and micropolitan areas*. U.S. Patent and Trademark Office 2013 [cited Nov. 19, 2013]. Available from http://www.uspto.gov/web/offices/ac/ido/oeip/taf/cls_cbsa/explan_cls_cbsa.htm.
Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly Saveliev, A., and Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer.

APPENDIX 1: EXPLANATION OF DATA COLLECTION

INVESTMENT DATA FROM CRUNCHBASE

We have cautiously used data from CrunchBase.com. This dataset is based on grassroots, voluntary input. We observe that calculating startups or the startup ratio may be less reliable using these data because, for instance, any software consultant could claim to be a non-employee firm startup with no profits. However, the investment information seems to be highly reliable because people do not typically claim false investment information for their own company or others, particularly when it is necessary to reveal the sources of investments, such as the exact name of the VCs or other financial institutions. Furthermore, investment information generally has a citation of some news story.

To confirm the validity of these data, we first aggregated the investment information only from financial institutions and excluded non-financial institutions and non-attributed entities, which aggregated to $24.86 billion in 2012. We should note that “financial institutions” is as close as we can get to disaggregating VC from the rest of the investment information, but some of the financial institutions still include other entities, such as angel groups. The aggregated “fin-org” investment amount is 7.8 percent lower than the amount found in the MoneyTree Report by PricewaterhouseCoopers and the National Venture Capital Association. The underreporting ratio has been fairly consistent, with 4.1 percent in 2009 and again 7.8 percent in 2010. Thus, the fluctuation of reporting under CrunchBase is small and consistent.

More importantly, we aggregated the investments in CrunchBase at the state level, the smallest geographic unit that the MoneyTree Report provides, and analyzed the correlation. The correlation is 0.99 for all years since 2008. We further omitted California, the largest recipient, and the correlation still holds from 0.93–0.98 in all years. Thus, we concluded that this information is reliable, particularly for regression analysis, which calculates how a difference from one unit to another unit leads to the difference in the dependent variable.

Available at: http://developer.crunchbase.com/.

LOCATION QUOTIENT FROM THE MILKEN INSTITUTE

The Milken Institute produced “Tech Pole” reports to describe where North America’s high-tech economy was located and concentrated, and we source the 2007 report. Tables made available online report wages, share of North American wages, employment, share of North American employment, tech pole scores, and Location Quotient (LQ). We are concerned only with LQ for this paper.

In Milken’s words, “Location quotient (LQ) measures the concentration of high-tech employment or wages as a percentage of a metro’s total employment or wages and then compares it to the average for all of North America. An LQ of 1.0, for example, matches the North American average, while an LQ of 2.0 states that the metro's concentration in high-tech industries is twice as large." Milken splits up their data by industry, including such varied categories as "Aerospace Product and Parts Manufacturing" and "Internet Search Providers and Web Search Portals." We use their aggregated table, however: "Total High Tech—TOTHT."

Available at:
http://www.milkeninstitute.org/nahightech/nahightech.taf?rankyear=2007&type=naics&naics=TOTHT.

PATENTS FROM THE USPTO

The United States Patent and Trademark Office (USPTO) keeps official government statistics on U.S. patents. Through 2011, the USPTO keeps data broken out by MSA, assigning location based on the residence of the first-named inventor.

In general, patents can be one of six types, but the MSA dataset kept by the USPTO only shows data on the most common: utility patents. These often are referred to as "patents for invention," the most closely associated with the common conceptualization of a patent, and account for "approximately 90 percent of the patent documents issued by the USPTO in recent years."

Available at:
http://www.uspto.gov/web/offices/ac/ido/oeip/taf/cls_cbsa/allcbsa_gd.htm and
http://www.uspto.gov/web/offices/ac/ido/oeip/taf/patdesc.htm.

GRANTS DATA FROM THE NIH

The National Institutes of Health (NIH) is the second-biggest U.S. public spender on scientific R&D, next to the Department of Defense (DoD). In recent years, the NIH has spent about $29– $30 billion a year on R&D expenditures.

We obtained complete data on these expenditures from the NIH's ExPORTER data exporting tool. The NIH give these data by city and state, so we developed a Python script to geocode and aggregate up to the MSA level.

There were discrepancies between the data we assembled from ExPORTER and some of the NIH's aggregate reports, so we inquired with an NIH analyst, who was very helpful. He listed a number of reasons he thought could be responsible, including bad contracts data, timing of the reports, and the fact that the tables frequently are updated as grants are changed, completed, etc. His conclusion was that the NIH's data ultimately are somewhat soft, but that we should trust what we had through ExPORTER as long as we acknowledged that it would not be perfect.

Available at:
http://exporter.nih.gov/ExPORTER_Catalog.aspx and
http://grants.nih.gov/grants/funding/ac_search_results.htm.

SBIR/STTR DATA FROM THE SBA

Each year, federal agencies that spend more than $100 million in external R&D are required to allocate 2.5 percent of their budgets to Small Business Innovation Research (SBIR) grants. Eleven agencies meet these criteria and participate in the program (e.g., the Department of Agriculture, the Department of Homeland Security, the EPA, etc.). The Small Business Administration runs a free and publicly available search engine called TECH-*Net*, which holds data on SBIR/STTR grants across all eleven participating agencies. These data, like those obtained from the NIH, are provided only at the city and state level, so we again used a Python script to geocode and aggregate up to the MSA level.

Except for a .005 percent discrepancy in 2010, checks against annualized, national aggregate data at sbir.gov showed identical numbers from 2000–2010 (sbir.gov tools were not ideal for mass exporting of the granular data we needed, however).

Available at: http://tech-net.sba.gov/tech-net/public/dsp_search.cfm.
The above and more information: http://www.sbir.gov/about/about-sbir and
http://www.sbir.gov/about/about-sttr.

EDUCATION DATA FROM CARNEGIE

The Carnegie Foundation produces the Carnegie Classification, which has long been the gold-standard framework for describing U.S. higher education. The Carnegie Foundation made updates to the Classification system in 2010, and we draw on the associated data from this latest framework for our paper. Our version was downloaded in September 2013, which coincided with the month of Carnegie's most recent update at the time: noted as "Updated on September 13, 2013" in the file itself.

The data contain "All accredited, degree-granting colleges and universities in the United States represented in the National Center for Education Statistics IPEDS system." Finally, and importantly, "All-inclusive classifications are time-specific snapshots of institutional attributes and behavior based on data from 2008 and 2010. Institutions might be classified differently using a different timeframe. Individual classifications are not updated with more recent data."

Available at: http://classifications.carnegiefoundation.org/resources/.
The above and more information:
http://classifications.carnegiefoundation.org/resources/faqs.php.

POPULATION

We source population data from the Bureau of Economic Analysis (BEA), which has useful statistics at the MSA level because they adjust to the fluid OMB definitions on the fly. In their own words, “When OMB adds a new statistical area, BEA creates a time series for it starting in the earliest year even though it may not have had any urban area at the time. Similarly, when OMB changes the definition of a statistical area, BEA recreates the time series for that area, using the new definition for every year in the time series, published at the scheduled release date of the dataset.” This allows us to make smooth comparisons across years despite shifting OMB definitions, which is why we prefer this population data set.

Available at:
http://www.bea.gov/regional/downloadzip.cfm (under CA1-3)
With additional notes on construction:
http://www.bea.gov/regional/docs/msalist.cfm