Who will have an entrepreneur as their neighbor?

A THESIS

Presented to

The Faculty of the Department of Economics and Business

The Colorado College

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Arts

By

Axel Bjerke

April 2015

Who will have an entrepreneur as their neighbor?

Axel Bjerke

April 2015

Economics

Abstract

Columbus on the boarder between Georgia and Alabama has gone from about 113 new firms in 2005 to 68 new firms in 2012, while Bend in Oregon had almost 830 new firms in 2007 or almost 0.006 new firms per capita. Which is the highest per capita for any city in the US between 2005-2012. These differences have been attributed to culture and other factors that are very hard to change and control. This thesis will investigate the effect of the presence of specific types of firms on entrepreneurship. Does a city want to attract big, small, tech or manufacturing firms to create more entrepreneurship?

The findings show that there are positive effects of attracting manufacturing, construction, retail and transportation firms. Big agricultural, health and accommodation firms all have a negative effect on the number of new firms started in the city.

KEYWORDS: Entrepreneurship, Firm births, Business creation, US cities

JEL CODE: O310

ON MY HONOR, I HAVE NEITHER GIVEN NOR RECEIVED
UNAUTHORIZED AID ON THIS THESIS

_____

Signature

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Innovation institutes, incubators and entrepreneurship alliances are popping up in cities across the United States. It's a hot topic in the current media and a lot of research have been written about the positive and negative effects of entrepreneurship; it's effect on job creation Decker, Haltiwanger, Jarmin and Miranda (2014), the commercial benefits Meewella and Sandhu (2012) and growth García (2014). The Global Entrepreneurship Monitor (GEM) has also worked for several years to create new international measures of entrepreneurship. Papers have been written on the determinants of entrepreneurship on a country level Freytag and Thurik (2007), on a city level Garcia (2013) and on the individual level Åstebro, Herz, Nanda and Weber (2014).

But the majority of the research on the determinants of entrepreneurship is using national level statistics or survey style data from individual entrepreneurs. Much fewer published papers have been written about the determinants across cities, especially concerning the United States, and only a handful about European cities or other parts of the world. Which is due to the problem of acquiring accurate data and due to the fact that Urban Economics and Entrepreneurship Economics are both smaller disciplines in economics. But cities might be the most interesting geographic scale to study when it comes to analyzing the determinants of start-up creation. This is due to the fact that cities are usually the level where innovative clustering happens and where polices are most effective.

There are many innovative cities in the United States, but the innovative mindset is seldom constant over a whole state or a country, but the effect of the region on the

individual is still very important. This makes the study of innovative countries or states too broad, and the study of individuals miss out the important effects of clustering.

Reynolds, Miller and Maki (1995) wrote a paper about entrepreneurship in the city, using data from 1976-88, but a lot has happened since 1988 and the paper also mentions how there was inconsistencies in the data and that some findings were contradictory to theoretical research. The European research by Tamásy (2006) and García (2012) show many interesting trends but a lot of the change was due to unobserved factors. There is also a lot of debate regarding how to measure entrepreneurship, should researchers count the number of new firms, the number of self-employed workers, people engaged in entrepreneurial activities or some completely different index. The problem shows up in the Glaeser's (2007) NBER working paper where self-employment rates where used, and many inconstancies with theory where found.

This thesis will try to add to current literature by investigating the effect of clustering of certain sizes and types of firms on entrepreneurship for all available Metropolitan Statistical Areas (MSA) in the United States. This is important both for policy makers and future researchers. Many cities are trying to attract big firms to benefit from knowledge spillovers, while others are trying to create start-up clusters using different types of incubators, but the actual effect of this haven't been studied in much detail. Future research will also benefit from the knowledge of what types of industries are important, what cities are outliers, biases of different measures and information about potential differences between Europe and the 1988 data compared to today.

Literature Review

The Measurement Problem

An entrepreneur is defined as one who organizes, manages, and assumes the risks of a business or enterprise, Merriam Webster (2015). But this is intrinsically hard to measure, since an investor might take the capital risk, the owner might be the one who organizes and one of the employees might be the one who actually manages the firm. Counting all entrepreneurs equally might then be misleading. The data used for measuring the positive effect of entrepreneurship, usually measure entrepreneurship using some proxy for this definition. Using measures that are similar to what research about the effects of entrepreneurship on society uses is important. Because the final goal is not just to find what promotes entrepreneurship, but rather what promotes entrepreneurship that leads to the previous studied positive effects.

Self-Employment. There are two main categories for measuring entrepreneurship, stock and flow, García (2014). Stock measures the number of entrepreneurs or the number of small firms as a proxy for entrepreneurship, and examples of flow measures would be the number of firm births or total venture capital spending in a year. Stock measures usually suffer from only showing net effects and flow measures are hard to come by. Data showing the number of self-employed people has been available in the US for a very long time and it fits well with the definition of an entrepreneur as someone who manages and assumes the risk of a business. The problem with this measure is that there is a great bias towards contractors, who might be working for someone, but are self-employed for administrative and taxation reasons, rather than the being "true" entrepreneurs. Half the variation in self-employment levels in US cities can be explained

by the fact that some industries and demographical compositions tend to have a structure promoting self-employment rather than being more entrepreneurial Glaeser (2007). This makes self-employment rates a dangerous measure of entrepreneurship since it's so readily available but very misleading.

Nascent Entrepreneurs. A very different approach is measuring nascent entrepreneurs, who are people in the process of staring up a firm or actively doing entrepreneurial activities. This is very different from the self-employment measure, in that it is much harder and ambiguous to measure, but it does not suffer from the same biases. The idea of measuring nascent entrepreneurs took off with start of the Global Entrepreneurship Monitor (GEM) project Wagner (2004), but data is so far only available for selected countries and only dating back to 1991. This is very different from the self-employment numbers who are available for all cities in the United States going very far back in time.

Nascent entrepreneur data usually focus on attitudes and experiences with entrepreneurship. The problem with this is the interviewees' biases and the ambiguity of questions concerning future expectations and opinions. Using nascent entrepreneurs as a measurement also suffers form a bias towards individuals that have ideas and are in the planning stage and might never actually start a business due to unforeseen barriers. Hence focusing on nascent entrepreneurs might miss the point of actually measuring successful entrepreneurs, and is hard to use in quantitative models.

Firm Births. The creation and destruction of firms has the benefit that it does not suffer from the bias of contractors as much as self-employment rates. It also measures the

result of entrepreneurship rather then the process of entrepreneurship. The problem with measuring firm birth is that all new firms are counted equally. Hence the successful firms will not stand out in a data set, and there is a loss of information about successful entrepreneurs. There is also no differentiation between what cities have more or less high growth firms. An unsuccessful firm with one employee is counted the same as a tech firm that doubles in size every year, which creates a bias towards cities with a volatile business climate, rather than a successful business climate Reynolds et al. (1995).

Indexes. There have been some entrepreneurship indexes created to try to work around the most common pitfalls, most notably the OECD-Eurostat's Entrepreneurship Indicators Program, the World Bank's Entrepreneurship Survey and the Global Entrepreneurship Index. These are the most comprehensive measures looking holistically at countries. They are unfortunately not available at a city level and are usually a composite measure. The problem is that the final score is dependent on what the weights attached to these different composite indicators are and it is not as clear-cut what one is actually measuring. Most of these measures are also created for ranking countries rather than using the data as a dependent variable. Since some of the composite indicators could be used as independent variables, instead of being part of the measure itself. But a lot of good research has been coming out from the use of these data sets, about the impacts of entrepreneurship on economic growth, the importance of agglomeration and individuals own perception of themselves as entrepreneurs Sternberg and Wennekers (2005).

Firm birth and survival rates will be used in this thesis since it is readily available and similar to the data use by other researchers, so a comparison can be carried out. Also most research done on the benefits of entrepreneurship is done on simpler data such as

firm births, rather than indexes. Hence the key point will be to control for these weaknesses of the measure, to try to investigate if it suffers from some of the drawbacks mentioned.

Individuals as Entrepreneurs

Opportunity. There are three basic factors that have to be analyzed when looking at the emergence of entrepreneurship; the personal characteristics of the individuals in the area, the economic environment, that provide the opportunity that make the entrepreneurial activity possible, and the law and the culture of the area have to allow for entrepreneurship Cuervo (2005). Hence an opportunity for entrepreneurship has to be present, and there have to be individuals that see this opportunity for creating new business and these individuals have the have education and experience to carry out the idea, in a society that allows for it both legally and socially.

Education. The entrepreneur need the knowledge to both see the opportunity and create the firm, this knowledge can come from two sources, formal education or life experience. The effect of education rates has been extensively researched but with inconclusive results, and the theoretical backing is not fully developed Tamasy (2006). Higher education rates are a reasonable indicator for knowledge, but this can have many different effects. It gives the knowledge necessary to become and entrepreneur, but it is highly dependent on the educational structure. Some institutions give graduates the means necessary to become entrepreneurs, while other educational systems create graduates that have the knowledge to become employed but not develop their own firms Cuervo (2005). There is also the negative effect of uncertainty avoidance Freytag and

Thurik (2006), some people with an education and good job have more to lose if the quit their job to become full time entrepreneurs. The results regarding the effect of high school education rates have been ambiguous. The self-employed are usually better educated and they can estimate the risks of entrepreneurship better Caliendo, Fossen and Kritikos (2007). But there is a problem of causality, do entrepreneurs educate themselves more or are more educated people more likely to become entrepreneurs?

Specific industry knowledge and university degrees have been shown to be more important than a general high school degree in the creation of an entrepreneur, especially looking at the effects on risk aversion and finding opportunities for entrepreneurship. The proportion of the population with a high school diploma and higher, and the proportion of the population with a bachelors degree and higher can then be used as a proxy for different types of formal knowledge. These measures are readily available and widely used and are good for looking at general trends keeping in mind their weaknesses.

Experience. This is the second type of the knowledge and its much more informal. Experience being much more informal makes it harder to measure, if interview data is not used. It includes prior experience with entrepreneurship and specific knowledge about a market. Being a multiple entrepreneur or being surrounded by entrepreneurs have a significant positive effect on both firm creation and the firms success rate, these benefits are significant even if the last venture was not successful Tamasy (2006) and García (2014). This is because the entrepreneurs have more knowledge about how to run a business, and most entrepreneurs have more knowledge about the risk. Also a person's inherent riskiness does not effect their decision to become

entrepreneurs nearly as much, if it's not the individual's first venture. Hence experience has the effect of breaking some of the social and mental barriers of creating a star-up. This is in contrast to people who always have been employed by a firm, where their attitudes towards risk play a significant role Caliendo et al. (2009). Hence experience has an effect on your knowledge and you attitude towards risk, giving it a very positive effect on the likelihood of one starting a firm. But there is still the problem of how to measure experience if survey data is not available, but one can measure the "cities'" experience by either measuring the number of small firms or the number of self-employed people. Another interesting measure of the cities' experience would be to measure the number of startups in the past 6-10 years, which is a readily available measure. The benefit is that this index shows how much experience the entrepreneur is surrounded by even if she doesn't have the experience herself. The problem with this measure is that it doesn't do so on an individual level, and hence the people might have startup experience from another city, but come to silicon valley to start their second firm. One also has to keep in mind that the prior papers have been focusing on the individuals experience rather then the cities experience, making this a different proxy for experience, which have not been heavily scrutinized.

Barriers to Entry. Making it difficult to enter the market is a classical microeconomic concept that hinders perfect competition and entrepreneurs to prevail. Entering the market for the manufacturing of DNA sequencing machinery is almost impossible, even if the entrepreneur has the education, experience and the idea to do so. It is too expensive and the administrational hoops are many to jump through. Two main factors that where found to especially impact the choice of the entrepreneur is the rule of

law and the size of the state sector along with taxation and welfare Aidis, Estrin and Mickiewicz (2009). These ideas make intuitive sense; having weak laws concerning intellectual property, unfair competition and complicated laws around creating firms leads to increasing barriers to entry due to the initial cost of the start-up and more required experience. The size of the state sector crowds out the entrepreneur or sometimes even makes it illegal for the entrepreneur to enter the market. Taxation can have two effects, the deterring effect of having to pay taxes for creating a firm, or the positive effect from contractors preferring to be self-employed due to tax incentives. But tax evasion as a form of entrepreneurship doesn't lead to as many positive effects. Excessive welfare benefit policies also increase the opportunity cost of starting up a new business, which leads to few entrepreneurs. Barriers to entry are in general very hard to measure on a city level, but are usually consistent across a state and time, and are therefore rather easy to controlled for if it's not the main aim of the research.

Unemployment. Theory and empirical models are not conclusive regarding the effect of unemployment on entrepreneurship. It can either lead to more self-employment due to fact that it is an option to normal employment. But it has also ben shown to have a negative effect since people are less willing to quit their job to start their own company, if they are unsure about the job market Golpe and van Stel (2007). Hence the question should be divided up in two, the effect of being unemployed and the effect of high unemployment rates. Being unemployed increases your likelihood of becoming an entrepreneur especially for people with more experience and education Golpe and Van Stel (2008). But this is not true on a city level, where higher rates of unemployment lead to fewer firm births or no effect, due to the increased risk in the job market and not being

able to control for unemployment being highly correlated with adverse economic situations Reynolds et al. (1995).

Gender and Age. There is a difference of risk averseness between male and females. This leads to women being less likely to become entrepreneurs in Germany Wagner (2006). It has also been shown in other studies that more men are likely to become entrepreneurs rather then women due to cultural, evolutionary and social factors García (2014) and Tamásy (2006). The effect of age on entrepreneurial activity has also been studied, but a lot of this can be controlled by accounting for experience and schooling. But in general the most entrepreneurial age group is males aged 25 to 44 according to Tamásy (2006). Hence age and gender are important factors when studying entrepreneurship and it has been widely used when studding the individual's choice, because the effect on the attitudes and knowledge. But it is not shown to have the same importance when looking at a macro scale of a city since the age and gender effect are due to other observed effect and not inherent to age or gender, but rather due to the social differences between genders and age groups.

Cities as Drivers of Innovation

Cities or rather Metropolitan Statistical Areas (MSAs) are very interesting to study for innovation. There are large differences in attitudes, culture and opportunity across a country, state or province, hence controlling for effects such as culture and other unobservable factors might be very difficult. Not knowing anything about the individuals' location might also create misleading conclusions, due to the fact that two people with the same attributes in different cities will act very differently towards their

environment. Limited research has been done about cities as innovative clusters especially in the USA, but many pro-entrepreneurship policies have been implemented, such as incubators and entrepreneurship competitions. Hence further research is needed about the city, as a driver of entrepreneurship.

Agglomeration. Firms try to spread out to create local monopolies, but the factors pulling firms together geographically usually overpowers their desire to spread out. Many firms share inputs like labor, raw materials and machinery, making them cluster where it is available. Take the film industry, it is is much easier to find and actor, studio and filmmaker in the labor market surrounding Los Angeles than in Wayne Nebraska. This makes it much cheaper to make a movie there than in any other place, because the firm can enjoy community-based economies of scale without having to be big, O'Sullivan (2012). But this reaches a limit when the price of land increases and makes firms that are big enough to have for example their own studio, actors and prop to locate outside city centers. The small, new and volatile firms on the other hand needs the input clusters and will locate in the center of the cluster. The effect is that we see more firm births in city centers than outside these clusters. Diseconomies of scale occurs if the demand for land goes up significantly, but also if there is two much competition leading to a limiting factor of the size of cities.

Sharing of costumers is also a very important factor for some industries like restaurants. One usually finds all the restaurants in downtown and all the auto dealers along a specific road in every city. This is because the costumers want more variety, so hungry people go downtown to eat and then choose the restaurant when they get there. This gives unknown restaurants a chance to get more first time costumers, but paying

more for rent. Hence more new firms will locate in customer clusters, while bigger well known restaurants will locate in less expensive areas. This again leads to more volatile and new firms locating themselves in cities.

Knowledge spillovers are a key factor in the clustering of small competitive firms. This is usually due to labor pooling and matching. Imagine that there is only one firm that requires the skill you have, then you will work for them and get more paid if it goes well and less if the profits are smaller. But you can easily switch jobs if one firm goes out of business or does poorly, if there are hundreds of them. Because there might be another one who is more successful and can hire you. Going back to the film industry example, when one movie is finished then it is going to be much easier to find a job for another film in Hollywood, making actors (input) and film teams (firms) cluster. Then the knowledge one acquired making the last film will most defiantly be used making the next film, leading to clusters of knowledgeable workers O'Sullivan (2012). High-tech and volatile firms that need this extra expertise will then locate themselves in areas with similar firms, so that they can hire employees that have skills they got at another firm. Hence more firm births should occur in cluster areas where firms who depend on knowledge spillovers are more present.

Culture. It is rather hard to define culture in econometrics and even harder to measure it. But one definition is to say that culture is the collective mindset that differs one group from another Hofstede (2001). Defining culture in such a way makes it a bit easier to quantify, because then we just have to measure the differences across the cities instead of some sort of absolute value of knowledge, belief, ideas, norms etc.

There are many different theories concerning the effect culture have on entrepreneurship, some places and people have more entrepreneurship due to psychological traits making them create more entrepreneurs Fraytag and Thurik (2006). Progressive cultures are based on values that make people more entrepreneurial such as; autonomy, innovation, freedom of experimentation, acceptance of risks, support in taking initiatives and competitiveness Lumpkin and Dess (1996). All these factors in a progressive culture are lined up with the values needed to become an entrepreneur. This makes it more acceptable to become and entrepreneur, leading to more entrepreneurship where there would not have been any Etzioni (1987). Hence how society views entrepreneurship and it's status as something interesting and honorable instead of something for the ones that can't get a job, has an important impact the number of entrepreneurs.

An important note with this in mind is that culture is rather stable over time Fraytag and Thurik (2006). This makes it easy to control for, if the study isn't specifically focused on how to change the culture to become more accepting of entrepreneurs. But the general idea is to try to align the social norms with entrepreneurial ones, which is better studied in conjunction with sociology.

Past Results

The two papers on United States that most resemble what this thesis is trying to do is the 1995 paper Explaining Regional Variation in Business Births and Deaths: U.S. 1976-88 by Reynolds et al. (1995) and the NBER working paper Entrepreneurship in the City by Glaeser (2007). Most other econometric research focuses on country level data or the individual's decision to become and entrepreneur, the exception to this is Europe where

there has been some research done in the last ten years.

The paper by Reynolds et al. (1995) focuses on Metropolitan Labor Areas (MLA). The findings point towards the key factors being firm diversity, the presence of mid-career adults, volatile industries, employment flexibility and population growth. There is an absence of any positive effects arising from higher levels of education, availability of information services or technology clusters. But the authors also point to the problem of the hidden effects due to the large numbers of firm births and deaths that so regularly occurs in the retail industries. The theory and results may also differ from today due to a shift in technology availability between the 1980s and today.

The Glaeser (2007) paper uses self-employment rate and firm size. Which is less volatile and not as accurate as then the number of firm births, since firm closures cancel out firm births and it is also bias towards cities with self-employed contractors rather than entrepreneurs. The results are not as clear as in the Reynolds et al. (1995) paper, but points towards the importance of an appropriate labor force.

There are also two European papers that have carrying out similar research, the paper Determinants of Regional Entrepreneurship Dynamics in Contemporary Germany: A Conceptual and Empirical Analysis by Tamássy (2006) and the paper Analyzing the Determinants of Entrepreneurship in European Cities by García (2014). These papers recommend that policies should support city self-employment laws and tertiary education rates. But a lot is due to unobserved factors such as culture and attitudes, and can not be fully explained. These papers also point towards the dangers of using self-employment rates due to the bias towards contractors.

Theory

The Effect of Surrounding Firms

This thesis will focus how the size and the sector of the surrounding firms affect the choice of starting a new firm. The effect of the surrounding firms or the cluster effect, tie together many theories and is hence an interesting subject to study. Experience can't be directly measure but the number of successful start-ups in the city can. Making a scale for the barriers to market entry for every city and every sector would be a very time consuming project, but measuring the number of small firms to show how many other companies have crossed the barrier before might give an indication. Asking owners if they share the same inputs as their competitors will not be an easy task, but measuring the number of firms in the same city, sector and size shows if the owners are likely to benefit from being around companies with similar inputs and costumers.

Small firms. More small firms could have a positive effect on culture due to the fact that there will be learning opportunities from other entrepreneurs on how to start and run a firm. The laws and culture might also be more adapted to firm creation, this can include schools teaching entrepreneurship classes and the government's filing times and fees might be lower, leading to lower barriers to entry. It also means that entrepreneurship inputs are more readily available, such as venture capital and angel funding. The negative effect could be that the market is already saturated with small firms. The type of small firm might determine if the effect is positive or negative. It might be hard to start a new restaurant in New York for example, but Silicon Valley might be the place to be if one wants to be exposed to venture capital directed at the tech industry. Hence not all types of small firms have the same effect and indicate the same

thing.

There should then be a positive effect of more small firms in industries where knowledge spillovers and small firm input sharing is important. Many small firms might also lead to the negative effect of a crowded market and few opportunities to enter the market might be available. Hence a lot of small hairdressers or restaurants are unlikely to create any positive effect on business creation due to the crowding out. But many small construction companies might lead to more construction entreprnurs, due to the availability of renting expensive machinery and the knowledge of how to start a company.

Big firms. The effect of big firms has been widely debated. Having large monopolies crowed out small firms. The classic example is industrial farms crowding out smaller local farmers. Other large firms might create knowledge spillovers and create the job security needed for and entrepreneur to dear to start her on firm. The large firm can also give opportunities for a small firm to develop a patent and being able to sell it rather than having to invest in a lot of machinery. It could also have the opposite effect in that it buys out the small firms instead of letting it develop. Hence the type of large firm will be very important to control for, to be able to find out what large firms are good versus bad for the promotion of entrepreneurship. Hence there should be a negative effect on farming and other sectors that buy out the small family companies. But there should be positive effects of having more tech companies and similar since it creates knowledge spillovers and decrease the barriers to entry.

Firms in Different Sectors. Different sectors cluster for different reasons. Food processing plants need to be close to their inputs, but a retailers need to be closer to its

costumers. Hence different types of firms will cluster differently depending on what they do and what they are dependent on, the costs of transporting inputs or the competition for customers. This makes it important to look at what types of firms are dependent on clusters. One doesn't usually see a design or advertising studio in a small town because they need a cluster of firms wanting to advertise and costumers seeing the advertisement. But every city has multiple gas stations because they are much less dependent on large clusters and agglomeration.

Start-up history. Self-reinforcing effects are important in urban economics and they definitely apply to entrepreneurship. This is the idea of a growing cluster, one person starts a firm and gets rich, which makes everyone wanting to start a firm and get rich. So the cluster will grow and keep growing until it has reached diseconomies of scale. Hence there should be a point at which the area has so many start-ups that cost of locating your startup in a cluster offsets the benefits enjoyed by being there. The question is if USA has reached this point for any city and how this differs across different industries.

Data and Methodology

Firm Births

The American Census Bureau has created the Business Dynamics Statistics data set from the Longitudinal Business Database. This data set gives information about the number of firms and establishments and their entry and exit rates as well as there job creation rate, sorted by the firm's age group for every city across the US between 1977 and 2012. This makes it possible to see how many firms are born every year and how many of them survive. The data does unfortunately not divide the firms up into industries, but this could possibly be acquired through the Federal Statistical Research Data Centers. This was unfortunately not possible due to time constrains, but it could be a great opportunity for further research. This research focused on how many firms where created every year and how many survived.

Figure 4.1 Firm birth vs 6-10 year survival

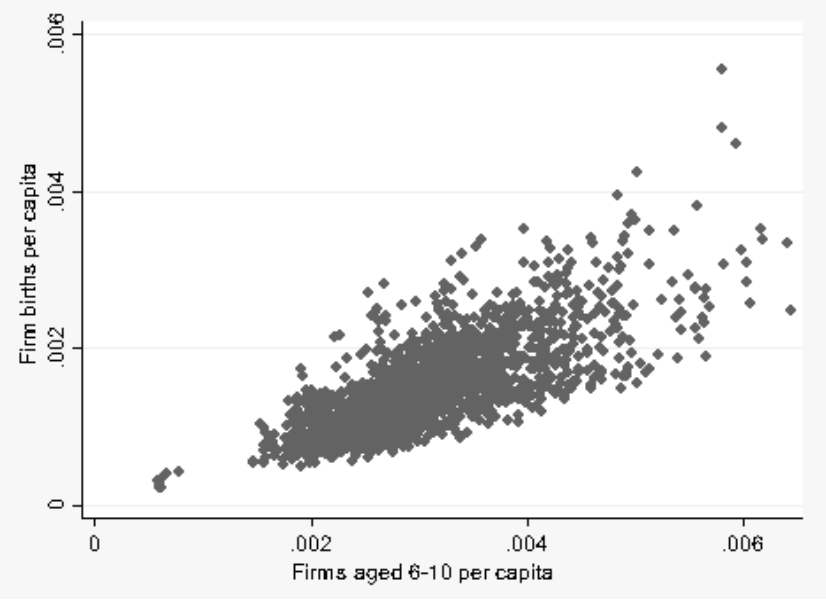


Figure 4.2 Small firms vs Firm births

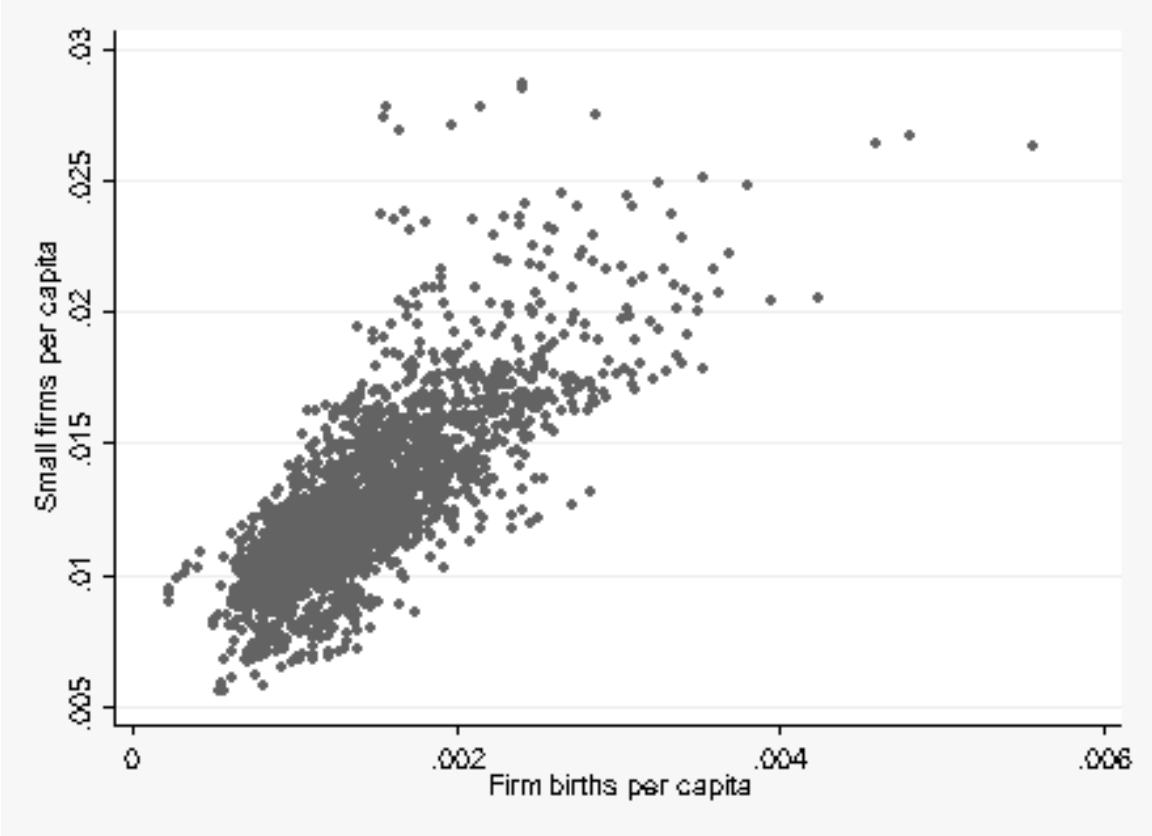


Figure 4.1 show how the 6 -10 year survival differ form the firm births, this is very interesting since there is an increasing variance showing that there are different survival rates across the cities. Figure 4.2 show how using small businesses as a measure of entrepreneurship is biased towards Miami and other areas where retail and small firms that does not intend to grow are very present.

Using firm births means that there will be a bias towards more volatile industries rather than successful entrepreneurs. This has to be taken into account for in the final analysis. It is still better and more interesting than using self-employment rates since there won't be a problem of as much information being hidden behind firms entering and exiting cancelling each other. But this data doesn't look at the nascent entrepreneur to see

why or why not she chose to carry through with the idea. The data is also not as holistic as an entrepreneurship index since it doesn't look at different types of firm births and their exact effect. But it's a straightforward measure that is simple and easy to interpret, which makes it a much clearer analysis. Firm births is the direct outcome of an entrepreneur taking a risk, organizing and managing a firm, may it be a successful firm or not.

The total number of firm births it's not a very graphically pleasing or sensible use of the data since the MSA New York-Northern New Jersey-Long Island has a population of about 19 million people and the mean population of the 308 US cities used is about 750'000 people.

Figure 4.3 Total new firms vs Total population

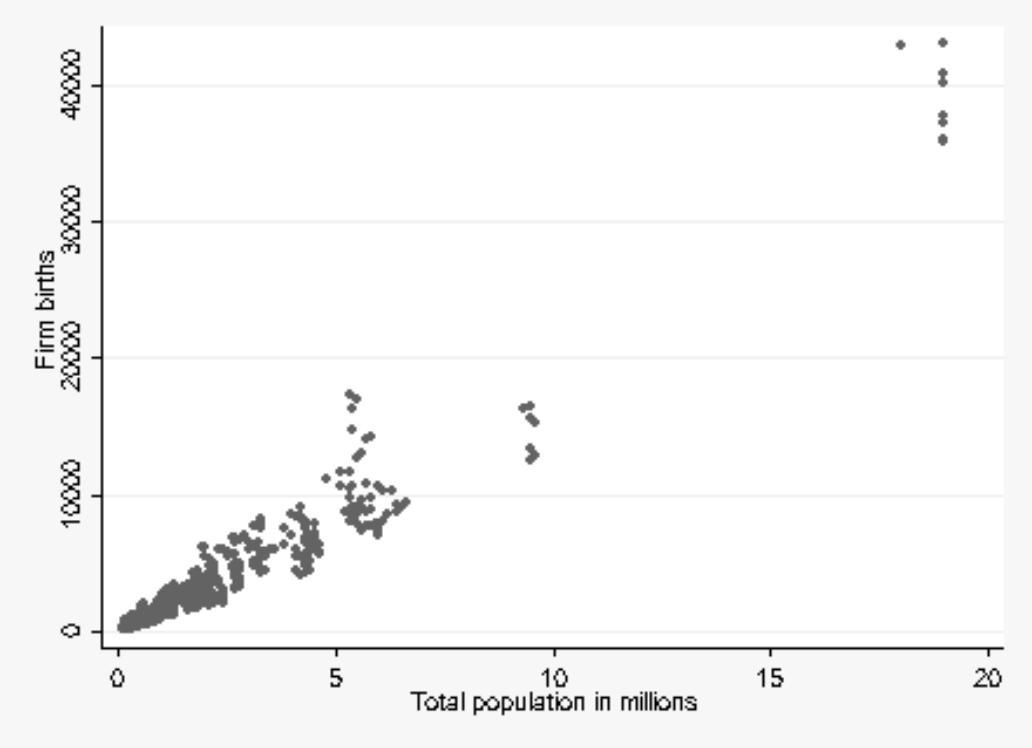
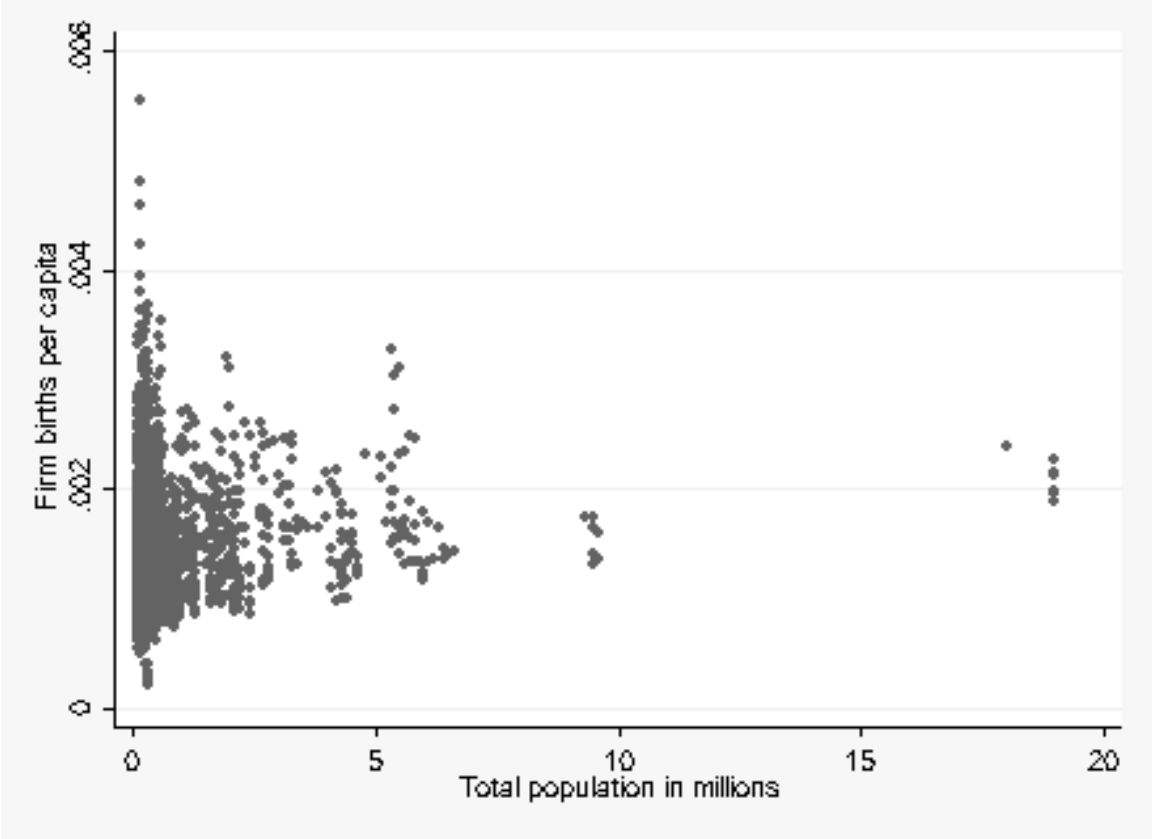


Figure 4.4 Firm births per capita vs Total population

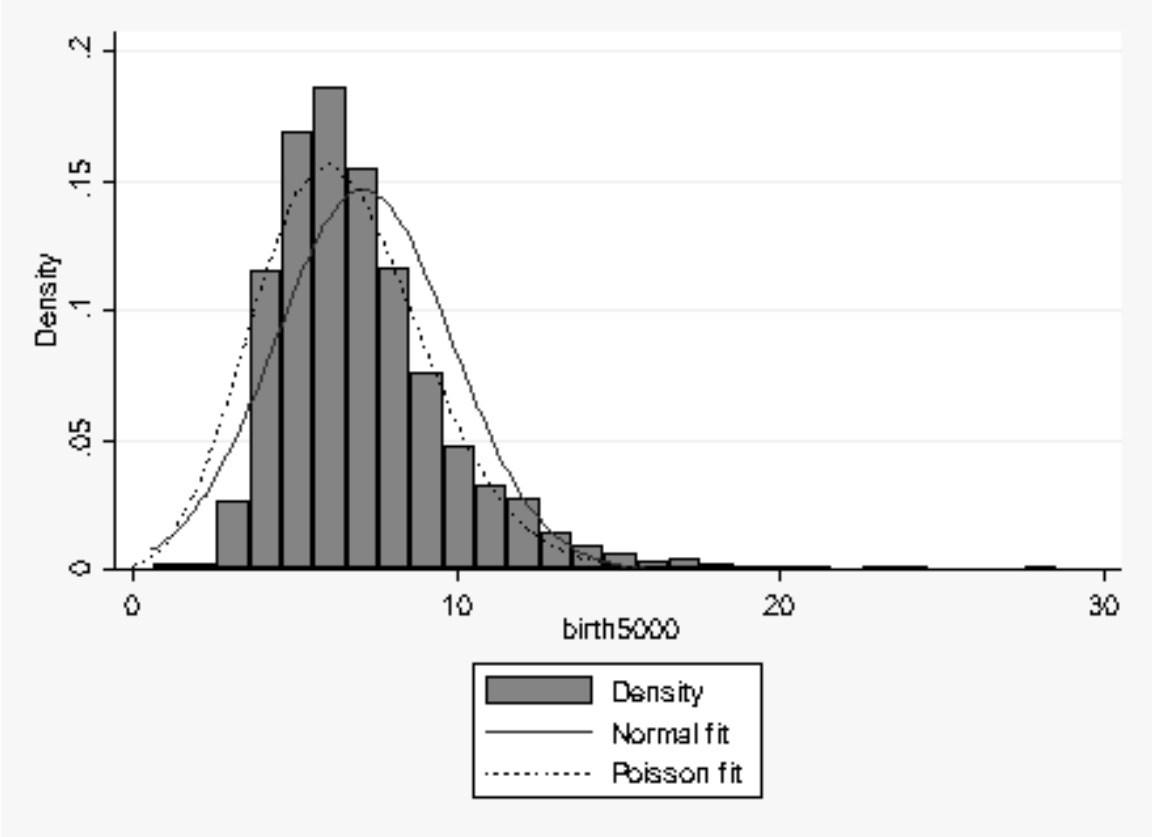


Using firm births per capita as in figure 4.4 is more sensible since we can see how much entrepreneurship the city has per citizen and it is not as skew towards the large cities.

A Poisson regression is also appropriate since the number of firm births can never be less then zero and it not a continuous variable, half a firm cannot be started. But the mean has to equal the variance for a Poisson regression to be correct. A negative binomial regression can also be used if the data is over dispersed, which it is, but it's a slower model to work with so it will only be used as a final confirmation of the results. The number of firm births, to make it more meaningful, will then be divided by the total population. The firm births per capita will also be multiplied by 5'000, so the mean is 7.1

and the variance is 7.4 and a Poisson can be used. The mean and variance can be made exactly equal, but the difference between having the mean and variance exact compared to 7.1 and 7.4 is not noticeable in the final regression and the measure of the number of firm births per capita for 5'000 people in the city makes more intuitive sense then for 4809.2 etc. Figure 4.5 shows the distribution of the number of firm births in a histogram.

Figure 4.5 Histogram of adjusted firm births



Independent Variables

The independent variables used are from different data sets available through the American Census Bureau. The number of firms by size and sector for every MSA is from the County Business Patterns data set. The American Community Survey (ACS) gives data on employment, population, education, incomes and more, but the ACS is only

available on a MSA level dating back to 2005, which is the limiting factor for extent of this regression. Earlier data is available from other sources using a different methodology and it is seldom in a readily available digital format for an MSA level. Many MSA's change their names and geographical configuration making it not as straight forward to determine what MSA number to use for what metropolitan area. Hence the years used will be 2005 to 2012 which is eight years of data, this makes it very reasonable to control for unobserved effects such as culture. The time span of eight years also makes it so the removal and redefinition of cities is to a minimum. Still 58 cities out of the 366 available through the Business Dynamics Statistics set had to be removed due to gaps in the data and the rearranging of MSA compositions. But no significant changes occurred when regressions where tested for the inclusion and exclusion of modified versions of some of the MSAs where used, hence they where excluded for simplicity.

High school and Bachelor graduation rates for the population above 25 years old will be used to control for the effect of formal education. Unemployment will also be an easy measure to obtain to check for its discussed effect. Total percentage of male and female will be used to see if the gender effects are present across cities. The access to high speed Internet is only available from 2013 but can be a very interesting measure for the future. Some state specific variables like the venture capital spending per state was also collected as well as a power diffusion index, none of the state measures where significant due to the small changes over time and the difference across cities in the same state. Hence these measures will not be used in the final regression.

The main variables used and collected are the number of firms by sector and size in every city. The firm sizes by sectors are divided up into four groups, more than 1000

employees (big), more than 4 employees (mid), 1 to 4 employees (small) and total number of firms (all). This is to be able to measure the cluster effects of different types of firms, these are all measured in per capita. The variable firms aged 6 to 10 is also used to control for the cities experience with start-ups, and the entrepreneurship community effect. Hence all factors mentioned in the theory is trying to be either be controlled for because it's not changing significantly or included in as a variable being a proxy for it, except the ambiguous problem of measuring entrepreneurship opportunity. This could be somewhat controlled for by total population to show how many available customers, but some competitiveness or inefficiency index would be more appropriate, this is unfortunately not available over this time scale and city level.

Table 4.1 shows some interesting outliers in the data. This shows that certain cities have clusters of different industries. This allows for a great future opportunity in investigating how certain exceptional clusters are created. But it can lead to certain biases that have to be investigated when using total firm births for all industries.

Table 4.1 Outliers in the data set

Miami	Exceptionally many small firms compared to firm births.
Boulder	Exceptionally many health, education and scientific firms.
Fayetteville-Springdale -Rogers	Exceptionally many management firms, compared to firm births.
Santa Fe	Exceptionally many small education firms.
Trenton-Ewing city & Pittsfield	Exceptionally few small education firms.
Ocean city	Exceptionally many accommodation, food and retail firms
Elkhart-Goshen	Exceptionally many larger manufacturing firms
Laredo	Exceptionally many transportation firms

Poisson and Negative Binomial Regression

The dependent variable is manipulated as explained earlier so it is distributed as a Poisson, figure 4.5. The data is over eight years and hence it is panel data. The basic Poisson regression works using a numerical method to find the maximum likelihood of function 4.1.

$$log\big(E(Y|x)\big) = a + b\,x \tag{4.1}$$

The data used in this regression allows for the use of a fixed effect model controlling for the culture and laws that have not changed by a significant amount during this period. Stata's xtpoisson data package will be used for this regression. The actual method for carrying out maximum likelihood on panel data is a bit more complicated but can be studied in Woolrdige (1999). The method is readily available with some limitations; "-xtpoisson, fe- the likelihood function conditions on the sum of the counts in each panel" (Stata Technical Support 2015). What this means is that the fixed effects and anything that doesn't change significantly over time is accounted for, but any information about the fixed effects are lost, hence there isn't an acceptable way to check for the autocorrelation of the residuals and to investigate the size of the fixed effects.

Another important note for the observant reader is that the data might now look like it's normally distributed, since the data is based on dividing two integer variables making it continuous. A model using a normal distribution might feel more appropriate. But the data set is still not a normally distributed since the variable is always above zero and still shaped like a Poisson due to its skewers as shown in figure 4.5. This makes the Poisson regression more appropriate than a normal distribution and Ordinary Least Squares (OLS) regression.

Using the raw number of firm births makes it over dispersed leading to artificially high significance levels. A negative binomial model could be used for dealing with over dispersion. The negative binomial regression is slower to calculate, and usually runs in to problems of over specification, not being concave and running iterations for hours. It also has less options available compared to the Poisson, and hence the Poisson will be primarily used and the negative binomial regression will be use to check that the results are similar and the assumptions around using the Poisson is correct. The two final regression models should be very similar. Since the negative binomial is in concept calculated in the same way as the Poisson, and suffers from the same problems.

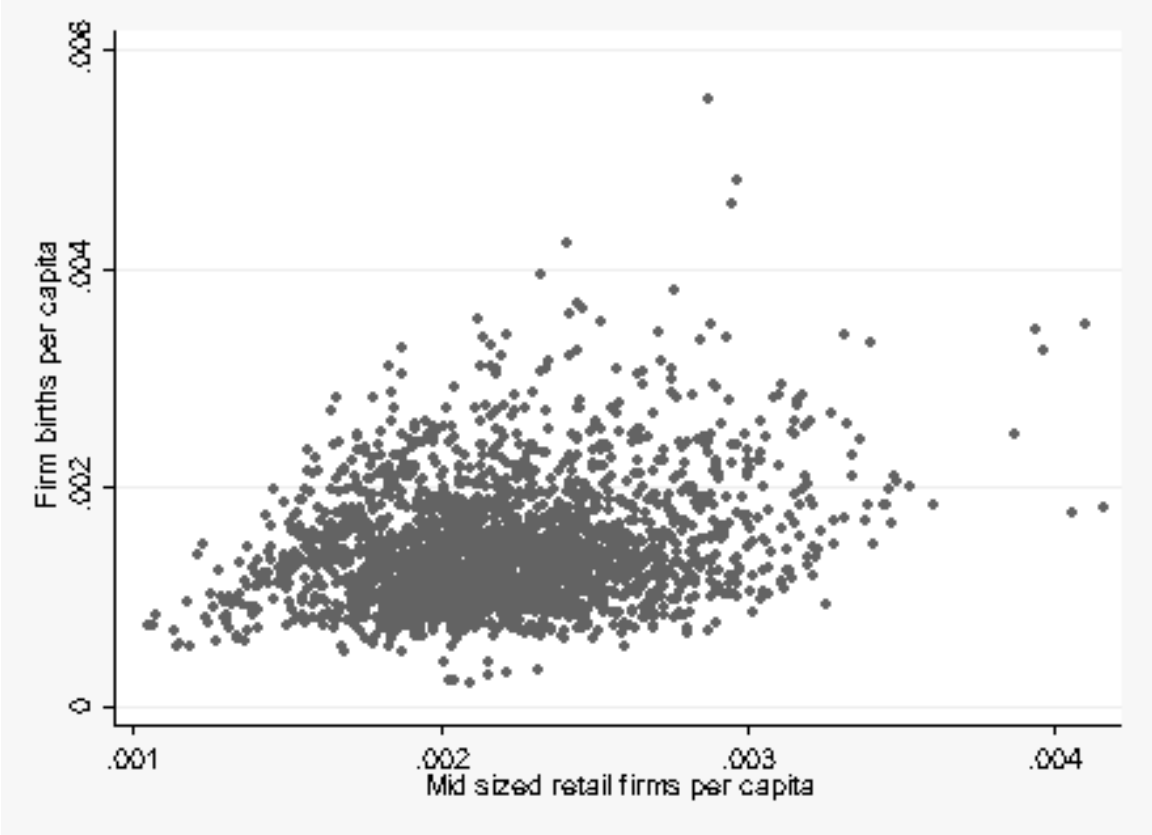
Using per capita data makes it under dispersed, which would show artificially low significance levels for the independent variables. Under dispersed models are very rare and would require completely different modeling capabilities and methods. Hence using a Poisson with a mean and variance that is manipulated is the most reasonable approach, together with the negative binomial regression for the final model.

The variance of the data has to be accounted for before any meaningful regression is carried out. The clearest way to show the variance is through a graph, figure 4.6, this graph shows the relation between small tech firms and firm births, but most scatter plots relating the independent variables and firm birth have a similar shape. This makes intuitive sense, since more firms per capita allows for more variance, having more firms naturally allows for more dispersion, due to the fact that a 10 percentage change for a city with many firms is much bigger than the percentage change for a city with few firms, counting in absolute terms.

The observed information estimator matrix is the usual option to control for

variance for a maximum likelihood regression. But the observations are independent so the more commonly used Huber-White-sandwich estimator can also be used for the Poisson. The Negative binomial can only use observed information estimator matrix to adjust the standard errors with respect to increasing variance. The results differ a bit between the two estimators since the observed information estimator matrix method seems to be more sensitive towards correlated variables used in this regression. This leads to a double underestimation since correlated variables have usually a worse than expected significance result.

Figure 4.6 Graph showing variance



Results and Analysis

The total number of small firms and the number of firms that are 6 to 10 years are highly correlated with many variables and the number of firm birth. This would be is interesting to use to create other models to compare with, but it has no significance in this model. Some firms have exceptionally many or few small firms in certain sectors compared to the rest of the cities, table 4.1 shows a list of these interesting outliers to keep track. No significant change occurred when these cities where included or excluded, so they where included in the model

The first model result was as expected; the more volatile sectors like construction and retail a very significant and dominates the regression. The next step is than to remove all insignificant variables that have very little theoretical backing and test for the more interesting ones. The result using robust standard errors are shown in Table 5.1.

Table 5.1 Poisson Results, robust standard errors

Variables	Coefficients	z-value	
Unemp	-1.769	-14.38	***
Inc	-0.00591	-6.66	***
HS	-0.0430	-8.64	***
BS	0.00179	2.91	***
Big Agg	-29909	-7.03	***
Mid Health	-380.3	-7.12	***
Mid Acc & Food	-202.6	-3.41	***
Mid Tech	-1271	-4.67	***
HS * Tech	15.47	5.13	***
Mid Manuf	332.2	4.16	***
Mid Const	439.5	10.23	***
Small Transp	293.7	3.50	***
Mid Retail	272.8	6.85	***
Wald chi^2	5185		

Note: *** = 1%; ** = 5% : * = 10% significance

All the independent variables are in per capita, HS is the percentage high school graduates and above, BS is the percentage bachelor graduates and above, HS*Tech is a the number of technical and professions firms times the percentage high school graduates and above. Inc is the income per capita in $1000. The results are a bit more sensitive to variables that are correlated if one uses the observed information estimator matrix (oim) to control for changing variance, shown in Table 5.2.

Table 5.2 Poisson Results, oim error calculation

Variables	Coefficient	z-value	
Unemp	-1.769	-4.11	***
Inc	-0.00591	-1.98	**
HS	-0.0430	-2.77	***
BS	0.00179	0.68	
Big Agg	-29909	-0.61	
Mid Health	-380.3	-2.4	**
Mid Acc & Food	-202.6	-1.29	
Mid Tech	-1271	-1.59	
HS * Tech	15.47	1.74	*
Mid Manuf	332.2	1.64	*
Mid Const	439.5	6.27	***
Small Transp	293.7	1.39	
Mid Retail	272.8	2.38	**
Wald chi^2	664		

Note: *** = 1%; ** = 5% : * = 10% significance

These tables do not make intuitive sense as the normal OLS regressions do. Since the model is;

$$log\big(E(Y|x)\big) = x_1(Umenp) + x_2(Inc) + x_3(HS) + \cdots + x_{13}(Mid\ Retail) \quad (5.1)$$

This means that what is calculated in the tables is the difference between the logs of the expected value. A more intuitive way to interpret the results is using incidence rate ratios (IRR). The incidence rate ratio means that the coefficient in table 5.3 can be interpreted as the change in the in dependent variable by a factor of the coefficient if the dependent variable changes by 1. This is why the coefficient of big agricultural firms is zero in table 5.3. If you get one more agricultural firm per capita (which is absurdly many), then you wouldn't have any more new firms starting. Hence table 5.4 is more meaningful when all the dependent variables are multiplied by 5000.

Table 5.3 Poisson IRR, robust standard errors

Variables	IRR	z-value	
Unemp	0.17	-14.38	***
Inc	0.99	-6.66	***
HS	0.96	-8.64	***
BS	1.00	2.91	***
Big Agg	0	-7.03	***
Mid Health	7.1 * 10^-166	-7.12	***
Mid Acc & Food	1.04 * 10^-88	-3.41	***
Mid Tech	0	-4.67	***
HS * Tech	5.23 * 10^6	5.13	***
Mid Manuf	1.9 * 10^144	4.16	***
Mid Const	7.2 * 10^190	10.23	***
Small Transp	3.5 * 10^127	3.5	***
Mid Retail	3.1 * 10^118	6.85	***
Wald chi^2	5519		

Note: *** = 1%; ** = 5% : * = 10% significance

Table 5.4 Adjusted Poisson IRR, robust standard errors

Variables	IRR	z-value	
Unemp	0.1704514	-14.38	***
Inc	0.9941071	-6.66	***
HS	0.9579274	-8.64	***
BS	1.001794	2.91	***
Big Agg	0.0025242	-7.03	***
Mid Health	0.9267651	-7.12	***
Mid Acc & Food	0.9602915	-3.41	***
Mid Tech	0.7755084	-4.67	***
HS * Tech	5238032	5.13	***
Mid Manuf	1.068704	4.16	***
Mid Con	1.091872	10.23	***
Small Trans	1.060498	3.5	***
Mid Retail	1.056085	6.85	***
Wald chi^2	5185		

Note: *** = 1%; ** = 5% : * = 10% significance

Let's take a look at Colorado Springs in 2012 to understand the results better, it had a around 668'000 people in the MSA and 430 construction firms with more than 4 employees (3.2 per capita times 5000) and about 1000 new firms. Having one more construction firm per capita times 5000 (or 134 more construction firms in total) would increase the number of firm births to 1092 firms. Trying to calculate the effect like this is obviously silly, but it's a good way to show what the numbers mean. The main idea is to compare the different effects of the different sectors and investigate if the sectors have a positive or negative effect.

All the variables included are significant within a 99 percent confidence interval, using robust standard errors, but all variables loose some significance when using OIM especially the effect of big agricultural firms, percentage with bachelor's degrees, small

transportation firms, accommodation and food services and professional and technical services. The chance that it will fail the chi-squared test is still miniscule.

The negative binomial model using firm births (not per capita) produces similar results as the Poisson regression shown in table 5.4. Except that the regression does not converge when the effect of small professional and technical firms are included, due to over specification in the model. So the result for HS*Tech is very different, but it still has a positive effect. Robust standard error can not be used, so the method of observed information estimator matrix is used, which is the standard. Another interesting result is that the effect of an increased population is 1, hence no effect, also shown in figure 4.4.

Table 5.5 Adjusted Negative Binomial, oim standard errors

Variables	IRR	z-value	
Tot pop	1	8.86	***
Unemp	0.1693629	-19.42	***
Inc	0.9970768	-4.4	***
HS	0.9813324	-10.68	***
BS	1.004482	7.54	***
Big Agg	0.0050475	-2.32	***
Mid Health	0.9189268	-11.92	***
Mid Acc & Food	0.9561077	-6.32	***
HS * Tech	4.516585	6.2	***
Mid Manuf	1.06951	8.31	***
Mid Con	1.081572	24.96	***
Small Trans	1.052073	5.32	***
Mid Retail	1.038868	7.2	***
cons	1003.114	42.72	***
Wald chi2	5185		

Note: *** = 1%; ** = 5% : * = 10% significance

Conclusion

Insights

The IRR results produced by the negative binomial regression are the results that should be used for interpretation, but we don't know the size of the fixed effects. The results are that income has a negative effect, showing that there is a smaller chance that people with high incomes will start new firms. This can be due to the high opportunity cost of leaving a job, or because there might not be a reason for leaving ones job when the income is already high. Unemployment seems to fallow a similar trend, showing that people are more risk averse when the job market is bad. High school education rates have a negative effect, if the percentage of bachelor's degrees is controlled for. This might be due to the US educational environment, where a high school degree is focused on getting a job but a bachelor's degree might be focused on creating jobs. Construction, retail transportation and manufacturing firms are all volatile industries and that is most likely the main reason for the positive effects, but also due to agglomeration effects. Big agricultural, fishing and forestry firms have a strong negative effect, which can be explained by the type of economy such a firm creates. This can also be because it crowds out other firms or is located in an area specific for these resources. Firms having to do with accommodation and food services have a negative effect, this might be a sign of the effects of crowding out firm births, since there might already be so many hotels and restaurants in tourist areas like Myrtle Beach. The same crowding effect might be present for small professional and scientific firms, which has a negative effect. But it has a positive effect if it's combined with the total percent of high school graduates, so innovative tech clusters are only positive if there is a well educated population. The negative effect of

health care and assistance is most likely due to that these firms cluster around retirement areas. What was also interesting is that the number of firms aged 6 to 10 was not significant, showing that the type of firm that is present might be more important than just having a bunch of start-ups. It might also be due to a lot of correlation between this factor and others. Gender and age did not have a significant effect, which is a bit surprising, but this might be because it did not change much over time in the city. The population size had no effect, which might be due an increased crowding out effect. Finding the curve to fit that effect would be a very interesting future area of study.

The take away from this thesis is that the effects of knowledge, input and client sharing is very important, and it differs across sectors. Having a service oriented such as accommodation, health and small professional services has a negative effect, due to the crowding out effects. But clustering of construction, manufacturing, transportation, retail and large educated professional services have a positive effect, due to the community it creates.

Further research

There is a lot to be done in this area of economics, this analysis show that there are important effects on entrepreneurship originating from the type of surrounding firms. Further research should focus of the determinants of survival rates, and try to use the Longitudinal Business Database to differentiate what type of firm births are occurring. This data used a total number of firm births, but there are definitely important effects that need to by studied considering what type of start up benefit from what type of firms. This would be the next natural step in this research process since these trends where found. A

better regression analysis should also be used to be able to measure the fixed effect. This is to make sure that that the effects are not unimportant compared to what a change in educational structure or a change in culture could do. Finding out how much bang for the buck a city gets from increasing the number of construction firms compared to changing employment laws is a very important consideration that this regression does not show. Research on the general determinants of entrepreneurship should also differentiate between the effect of of service sectors and production, transportation and retail sectors.

References

Aidis, Estrin and Mickiewicz (2009) Entrepreneurial Entry: Which Institutions Matter? Institute for the Study of Labor, Discussion Paper 4123

Caliendo, Fossen and Kritikos (2009). Risk Attitudes of Nascent Entrepreneurs: New Evidence from an Experimentally-Validated Survey. Small Business Economics, 32 (2)

Cuervo (2005) Individual and Environmental Determinants of Entrepreneurship. The International Entrepreneurship and Management Journal, Volume 1, Issue 3311

Decker, Haltiwanger, Jarmin and Miranda (2014). The Role of Entrepreneurship in US Job Creation and Economic Dynamism. Journal Of Economic Perspectives, 28(3)

Etzioni (1987) Entrepreneurship, Adaptation and Legitimation. Journal of Economic Behavior and Organization Vol 8

Entrepreneur. (n.d.). In Merriam Webster Online, Retrieved March 5, 2015, from http://www.merriam-webster.com/dictionary/entrepreneur

Freytag and Thurik (2007). Entrepreneurship and its determinants in a cross-country setting, Journal of Evolutionary Economics Volume 17

García (2014). Analyzing the determinants of entrepreneurship in European cities. Small Business Economics, 42(1)

Glaeser (2007). Entrepreneurship and the City, NBER Working Paper, No. 13551. http://www.nber.org/papers/w13551

Golpe and van Stel (2007) "Self-employment and Unemployment in Spanish Regions. Frontiers of Entrepreneurship Research: Vol. 27: Issue. 20

Hofstede (2001). Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. 2nd Edition, Thousand Oaks CA: Sage Publications

Lumpkin and Dess1996. Clarifying the entrepreneurial orientation construct and linking it to performance. Academy of Management Review, 21: 135–172

Meewella and Sandhu (2012). Commercial Benefits of Social Entrepreneurship. World Review Of Entrepreneurship, Management And Sustainable Development, 8(3), 340357

O'Sullivan, (2012). Urban Economics 8th. New York: McGrawHill Irwin

Reynolds, Miller and Maki (1995). Explaining Regional Variation in Business Births and Deaths: U.S. 1976-88. Small Business Economics, 7(5), 389-407

Sternberg and Wennekers (2005). Determinants and Effects of New Business Creation Using Global Entrepreneurship Monitor Data (Special Issue), Small Business Economics, Vol. 24, Issue 3

Tamásy (2006). Determinants of Regional Entrepreneurship Dynamics in Contemporary Germany: A Conceptual and Empirical Analysis. Regional Studies Vol. 40.4

United States Census Bureau. (2015). doi: https://www.census.gov/ces/dataproducts /bds/data.html http://factfinder2.census.gov/

Wagner (2004). Nascent Entrepreneurs. Institute of Economics, University of Lueneburg, Discussion Paper 1293

Wooldridge (1999). Distribution-free estimation of some nonlinear panel data models. Journal of Econometrics 90: 77–97

Åstebro, Herz, Nanda and Weber (2014). Seeking the Roots of Entrepreneurship: Insights from Behavioral Economics. Journal Of Economic Perspectives, 28(3)