Journal of Urban Economics 67 (2010) 61–77

Contents lists available at ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

Local labor force education, new business characteristics, and firm performance

Mark Doms [a,*], Ethan Lewis [b], Alicia Robb [c]

[a] Federal Reserve Bank of San Francisco, United States
[b] Dartmouth College, United States
[c] Kauffman Foundation and University of California, Santa Cruz, United States

ARTICLE INFO

Article history:
Received 22 June 2009
Revised 3 October 2009
Available online 24 October 2009

JEL classifications:
O12
R58

Keywords:
Startup firms
Entrepreneurship
Labor supply

ABSTRACT

It is often asserted that a highly educated workforce is vital to improving the competitive position of American businesses, especially by boosting entrepreneurship. To examine this contention, we use population Census data and a new panel data of startup firms, to examine how the education and skill level of the local labor force are related to the creation and success of new businesses. This paper studies relationship between education, entrepreneurship, and businesses outcomes, and considers simultaneously both the education of the entrepreneur and of the workforce where the entrepreneurs operate their businesses. Consistent with this simultaneous focus, our initial results indicate that more educated entrepreneurs tend to be located in metropolitan areas with more educated workforces. Moreover, highly educated areas have above average entrepreneurship rates. Finally, the level of education of entrepreneurs is strongly related to positive business outcomes, especially for college graduates compared to those with less than a four-year degree.

Published by Elsevier Inc.

1. Introduction

It has long been recognized that education may play an important role in economic growth (e.g., Barro, 1991), though identifying the exact channels has proved somewhat elusive. In the United States, for instance, highly educated metropolitan areas generally posted above average wage growth over the past several decades (see Beaudry et al., 2006; Glaeser and Saiz, 2003; and Fig. 1).[1] A firm-specific channel through which education may affect economic growth is through entrepreneurship, which itself has been repeatedly found to be associated with growth (e.g., Van Praag and Versloot, 2007). In this paper, we exploit the Kauffman Firm Survey (KFS), a new panel dataset with more than 4000 firms that began operations in 2004, to ask how education levels in local markets are related to entrepreneurship and business outcomes.

The relationship between education and entrepreneurship can be thought of in two, inter-related ways: the education of business owners and the average education in the local labor market. More educated markets have more educated entrepreneurs,[2] and previous research has found a strong association between the education of business owners and small business success. For instance, Fairlie and Robb (2008) document that businesses with more educated owners have higher sales and profits, are more likely to hire employees, and are more likely to survive.[3] Whether owner's education is related to business success because of self selection (higher-educated people may be more motivated or innately gifted in characteristics that would be beneficial to new businesses) and/or human capital (the education itself may be useful in starting and running a business) is not established by this literature.

Entrepreneurship may also benefit from a more educated local population. Educated workers appear to have better access to information (Wozniak, 2006) and are better at implementing new ideas (Bartel and Lichtenberg, 1987). Indeed, supplies of educated workers are associated with faster adoption of new technologies (Staiger and Skinner, 2005; Doms and Lewis, 2006) and production techniques (Lin, 2009).[4] In addition, various theories in urban economics describe so-called Marshallian externalities,

* Corresponding author.
E-mail addresses: mdoms@doc.gov (M. Doms), ethan.g.lewis@dartmouth.edu (E. Lewis), arobb@ucsc.edu (A. Robb).

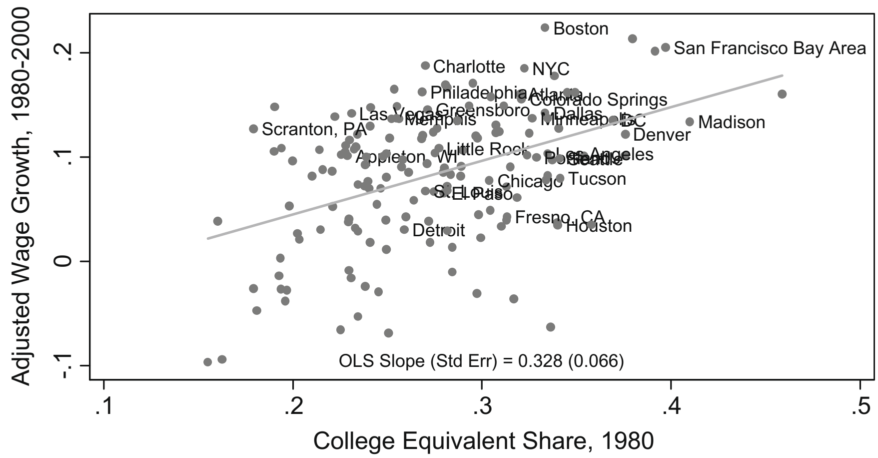
[1] There are many possible reasons why this result could emerge, not all of which involve a causal channel from education to growth (see Section 2.2 for more details). Among the possible causal channels are that highly educated areas were better able to take advantage of the information technology revolution (the Beaudry et al., 2006), while another is that highly educated areas are natural magnets for highly skilled industries (Beaudy et al., in preparation).

[2] As we will show below, the relationship between workforce and business owner education is close to one-for-one.

[3] See van der Sluis et al. (2008) for a recent review of the literature on the relationship between education and entrepreneurship, and Card (1999) for a review of the literature on the returns to education in the labor market.

[4] Schultz (1964, 1975) may have been the first to suggest that education improves one's ability to adapt to shocks, including the arrival of new technologies. Glaeser and Saiz (2003) provide empirical support for this view at a metropolitan level.

0094-1190/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.jue.2009.10.002



Data Source: 5 percent public-use Decennial Census (DC) of Population in 1980 and 2000. College equivalent share, the x-axis variable is the share with at least a four-year degree plus 1/2 those with 1-3 years college. Regression adjusted wage growth, the y-axis variable was computed by first regressing, separately in 1980 and 2000, the natural log of hourly wage on dummies fo five education levels, a quartic in potential work experience (age-yrs of ed-6), and dummies for female, foreign-born, and post 1950 birth, and dummies for each metropolitan area. The change in the estimated coefficient on each metropolitan area dummy between 1980 and 2000 is the dependent variable in the figure. Both college share and adjusted wages are computed using the sample age 16-65 with potential work experience>0 and survey 'person' weights (except the 1980 data have no survey weights).

Fig. 1. Adjusted wage growth, 1980–2000, and initial education.

how agglomerating skilled workers (among other things) lowers search costs for specialized skills, sometimes called "labor market pooling," and helps promote the spread of ideas, both of which could contribute to a more dynamic business environment. Labor market pooling has been found to be particularly important in high-tech industries like software (e.g., Freedman, 2008; Fallick et al., 2006).[5] The spread of ideas can be both important within industries (industry localization) as well as across industries. Jacobs (1969) describes how the exchange of existing ideas from disparate sources can lead to new ideas that help generate and sustain businesses.

In this paper, we explore the relationship between both area- and owner-level education and subsequent business performance. Previous research on entrepreneurship has not tended to distinguish between the effects of individual- and area-level education.[6] Regional analyses of business outcomes focuses tend to focus solely on regional education (for example, Acs et al., 2007), which, to be fair to the authors of many these studies, seems to be because owner characteristics are often absent from business data, while studies which use business data that do include owner characteristics have tended to only focus on owner education (for example, Bates, 1990; Astebro and Bernhardt, 2003; Fairlie and Robb, 2008; Van der Sluis et al., 2008).[7] One of the primary reasons for clearly making the distinction is that many government policies are directed towards the assistance of small businesses. In this paper we address to what extent the education of the local labor force is related to small business creation and performance. It could be that policies promoting and retaining a highly educated workforce could be at least, if not more, important than policies that attempt to more directly assist new businesses. In addition, it may help move towards determining why there is a relationship between owner education on business performance – is it causal? – something a recent review concluded is still not established (van de Sluis et al., 2008).

We use several datasets to examine these relationships. We observe outcomes of new businesses, revenues, profits, assets, and employment, in the Kauffman Firm Survey (KFS), a rich firm-level panel dataset of approximately 4000 businesses that began in 2004, which have been tracked through 2007. This survey includes owner characteristics as well as the location of the firm. Using the latter, we merge information on local labor market to the KFS, where the local market is defined as the consolidated metropolitan statistical area (CMSA).[8] The characteristics of the local labor market are constructed using the 2000 Decennial Census (DC). We also use information on the self-employed in the DC to create an alternative set of entrepreneurship outcomes. While self-employment and entrepreneurship are not one in the same, a comparative advantage of the DC data over the KFS data is a larger sample size; in 2000, our sample from the DC contains 357,000 full-time workers that are self-described as self-employed.

Lacking a convincing natural experiment, our approach to estimation is ordinary least squares (OLS) (and, as appropriate, methods for estimating non-linear associations).[9] As the decision to open, and the performance of, a business in a particular labor market may reflect unobserved factors correlated with owner or city education, these estimates cannot be taken as causal. A likely source of bias is what Combes et al. (2008) called "the endogeneous quality of labor bias," which in this case might be interpreted as that potential entrepreneurs in more educated areas are (unobservably) more talented. However, we take several steps to address the most obvious sources of potential endogeneity like this.

First, we control for detailed industry throughout the paper. Second, as the KFS data are a panel we include estimates with the dependent variable in first differences, which removes the

[5] More broadly, Rosenthal and Strange (2001)find that skilled industries tend to be more regionally concentrated than other industries, consistent with labor market pooling being most important in high-skill sectors.

[6] This has been much less of a problem in the related literature on urban agglomeration and human capital spillovers. For example, both Rauch (1993), Moretti (2004a) separately consider the effects of individual and aggregate education levels.

[7] For example, seemingly aware of this interpretation challenge, Glaeser and Kerr (2006) treat college share simply as a control variable in their regional analysis of manufacturing entry rates. In a related paper studying self-employment, Glaeser (2007) is able to and does make the distinction between individual-level and aggregate effects of education.

[8] The data section gives a richer description of CMSAs.

[9] We considered the strategy of historical metrics of local university density as instruments, the approach taken by, for example, Moretti (2004a). Local university density, even at long lags, is indeed strongly correlated with college share, and may really raise local labor supply of college graduates. However, in light of research suggesting a direct effect of university research on related outcomes like productivity and innovation (Andersson et al., 2009), these types of instruments are likely to be invalid. For what it is worth, however, the reduced form of the outcomes studied in this paper with a dummy for land-grant college are of the same sign as the OLS estimates.

influence of any fixed unobservable influences of business outcomes.[10] Third, we break the sample into high- and low-skill industries (based on their college share). To the extent that more skilled cities simply have higher quality entrepreneurs, this seems likely to improve business outcomes in both high- and low-skill industries; if so, the difference in effects between high- and low-skill industries may be closer to a causal effect. On the other hand, it might be the case that there are unobserved area attributes which raise the performance of high-skill industries in some locations more than low-skill industries (for example, university research); problematically, such forces would also tend to attract skilled labor, potentially generating a spurious relationship between college share and the success of high-skill industries. So our fourth strategy follows an approach similar to Rosenthal and Strange (2008) and controls for a full set of industry $\times$ metropolitan area effects, which we make possible by allowing local education mix to vary at a submetropolitan level (specifically, the workplace "PUMA," described later).

Even this last approach is unlikely to fully purge the data of the influence of third factors correlated with our regressors. The submetropolitan unit we examine, though smaller than a CMSA, tends to be roughly the size of a large city.[11] At this level of geography, it is plausible the same sources of endogeneity could still bias the results, even conditional on metropolitan area $\times$ industry effects. However, to the extent that these results are similar to the ones without metropolitan area $\times$ industry effects, it suggests that these third factors may not be a major source of bias. Finally, as most plausible endogeneity concerns (like the endogenous quality of labor bias) bias estimates upwards, finding no relationship would be an indication that there is unlikely to be a causal relationship; put another way, our estimates are probably an upper bound.[12]

Our analyses of these data produce four findings. First, unsurprisingly, more educated areas have more educated business owners, even within detailed industries. The educational attainment of primary business owners (KFS data) and self-employed workers (DC data) is strongly and positively related to the education of the local labor force both before and after controlling for industry detailed industry $\times$ occupation effects.[13] This reinforces our paper's motivation for considering jointly the influence of workforce and owner education.

Second, an area's average education level is positively associated with entrepreneurial activity, but this appears to be an individual-level, and not an aggregate phenomenon. The rate at which businesses turn over – both startups and deaths – increase with an area's college share. However, different data suggest that this is a compositional effect, not a "spillover" from education. More educated individuals are more likely to be self-employed, and conditional on that there is no additional positive association between self-employment and an area's college share.[14]

Third, in both the KFS data and using data on self-employment earnings in the DC, we find that the education of the business owner is associated with improved business outcomes, a result consistent with previous research. This association is not linear; performance is sharply higher among business owners with a four-year college degree.[15] We include some thoughts on why this might be in the discussion section.

Fourth, conditional on owner's education, the average education level of local labor market has an ambiguous association with improved business outcomes. In the KFS data the association is usually positive, but it is never statistically distinguished from zero. As our estimation approach is likely biased towards finding effects, not finding an effect despite the bias may mean there is no effect to be found. On the other hand, the DC data on the self-employed suggest strong and separate roles for both entrepreneur and workforce education in business earnings. In both datasets there is at least some suggestion that this positive association is mainly driven by high-skill sectors.[16] This is unambiguously the case in the DC data, where college share has a significantly larger association with self-employment income in high- than in low-skill sectors. Furthermore, the estimated magnitudes of these effects are similar when industry $\times$ metro area controls are added (and college share varies at the lower level of geography). In the KFS data, four-year survival rates are higher in more educated areas for high-skill startups and not for low-skill startups, but neither relationship is statistically significant.

In addition to contributing to the literature on the determinants of entrepreneurship (including Acs et al., 2007; Glaeser and Kerr, 2006) we view this paper as potentially contributing to our understanding of the sources of human capital externalities in cities, the sort which may drive the relationship in Fig. 1. Our findings do not suggest a spillover from education directly to entrepreneurship, but leaves open the possibility that any wage externalities from education indirectly result from higher rates of entrepreneurship in more educated markets.[17] Perhaps more interestingly, the positive association between education and business success is concentrated in high-skill sectors suggests that a labor pooling channel – a thick market for skilled workers is valuable in fast-changing industries – is driving the association.

2. Data, motivation, and approach

2.1. Data

This paper investigates how both entrepreneur characteristics and local labor market characteristics affect business outcomes. There are many different ways to define a local labor market; for our main analysis we have chosen the largest definition, the so-called "Consolidated Metropolitan Statistical Area (CMSA)." We include 230 CMSAs. In most parts of the country, the "CMSA" is identical to the smaller definitions of metropolitan area, namely the Primary MSA (PMSA). This is true of 212 of our 230 metropolitan areas. In the other 18 cases, in densely populated parts of the country, the question of whether, say, Oakland and San Francisco (or, say, New York and northern New Jersey) should be treated as separate markets or the same market arises. We have chosen to use the largest definition in these cases. A key reason is to avoid some

[10] However, unobserved influences of business outcomes, especially for new businesses, may not in fact be fixed. For example, unobserved business talent may affect profit growth, not just levels. Extensive discussion of this issue appears in Section 2.3.

[11] The unit is the workplace public-use micro area (or "PUMA") of the DC. The geography of workplace is discussed more thoroughly below and in Rosenthal and Strange (2008).

[12] One exception is best understood in the Roback (1982) framework used in Rauch (1993) and Moretti (2004a): amenities which attract college-educated labor to an area would tend to reduce earnings in equilibrium.

[13] The result implies, say, that a self-employed taxi driver in San Francisco (a highly educated metropolitan area) is likely to have a higher educational attainment than a self-employed taxi driver in Hickory, North Carolina (an area on the lower end of the educational attainment spectrum).

[14] On the other hand, the relationship between self-employment and area college share is weaker than for business startup and death rates, even in the aggregate. So our results leave open some possibility that self-employment is an insufficiently strong proxy for entrepreneurship to detect the education spillovers.

[15] Though not emphasized, Van der Sluis's (2008) review of over 100 studies on entrepreneurship suggests that studies typically find this non-linearity in the association between entrepreneurial outcomes and education. See their Table 3.

[16] High-skill sectors are defined by the average college share in those industries nationally: see appendix table.

[17] Higher rates of entrepreneurship are certainly associated with faster productivity growth (Van Praag and Versloot, 2007), but there is no convincing evidence that shows a causal link.

arbitrariness in the division of dense areas into particular labor markets. It also serves as a compromise because of the limited sample size in one of our datasets. As a robustness check, however, we will also do some analysis at a lower level of geography, the workplace public-use micro area ("PUMA") level. This level of geography will be described in more detail in Section 4.

We use several datasets to examine issues surrounding education and entrepreneurship. The first dataset, the Kauffman Firm Survey (KFS), is a firm level survey, which consists of four years of longitudinal data. The second dataset, which comes from the five percent Public Use Microdata Sample (PUMS) from the Decennial Census (DC), contains demographic information on education levels and self-employment rates at the CMSA level. We also make use of tabulations by the Small Business Administration (SBA) giving the number of establishment "births" and "deaths" by metropolitan area.

The KFS is a survey of new businesses in the United States. This survey collected information on 4928 firms that started in 2004 and surveys them annually. This cohort is the first large national sample of firm startups that will be tracked over time. These data contain detailed information on both the firm and business owner(s). In addition to the 2004 baseline year data, there are three years of follow up data now available. Four additional years are planned. Detailed information on the firm includes three-digit NAICS industry, physical location, employment, sales, profits, intellectual property, and financial capital (equity and debt) used at start-up and over time. Information on up to ten business owners per firm includes age, education, work experience, previous startup experience, and gender, race, and ethnicity.[18]

We use the confidential dataset because the public use microdata contain geographical detail only at the Census region level, while the confidential version has detail to the zip code. This research uses a subset of the data, those firms having data for all four years and those verified as going out of business (as opposed to not responding to the survey) over the 2004–2007 period. This reduces the sample size to 3974 businesses. Also, businesses not in CMSAs are dropped, which lowers the sample size to 3213. The method for assigning owner demographics at the firm level was to first define a primary owner. For firms with multiple owners (35% of the sample), the primary owner was designated by the largest equity share. In cases where two or more owners owned equal shares, hours worked and a series of other variables were used to create a rank ordering of owners in order to define a primary owner.[19]

Table 1a presents means and standard deviations of several variables from the KFS sample used in this paper. Mean employment in 2007 for the KFS was just under four employees, and average revenue, profits, and assets were about \$150,000, \$20,000, and \$114,000, respectively. A little over 70% of firms in the KFS survived through 2007. The primary owners in the KFS were highly educated, with more than half having at least a college degree.

One of the other datasets used in this paper is the Decennial Census (DC) of population, which identifies self-employed workers and is used to construct information on each area's labor force. Glaeser (2007) noted the difficulty in measuring entrepreneurship with self-employment data, because of the heterogeneity in what self-employment consists of.[20] For example, contract labor work in construction accounts for a large portion of self-employment, and a good portion of this is probably closer to wage labor than entrepreneurship. As such, we examine include only full-time (works at least 1500 h per year) in our definition of self-employed. We also examine further refinements of this definition. First, we examine self-employed with no wage or salary income, which may get closer to capturing "entrepreneurs" by removing many casual workers who work as contract laborers. In addition, we further divide the self-employed into those working in high- and low-skill industries, defined by the national college share in the industry, and a list of high- and low-skill industries is shown in the Appendix.[21] The high-skill sectors exclude, for example, services that are often contracted, like housekeeping.

Table 1a
Means of Kauffman Firm Survey variables.

Variables	Mean (1)	Standard deviation (2)
Firm outcomes:		
2007 Employment	3.7	13.15
2007 Revenue (\$)	149,978	289,278
2007 Profits (includes those with losses) (\$)	19,678	135,565
2007 Assets (\$)	113,548	233,425
2004–2007 Survival	71.2%	44
Primary owner education:		
Less than high school degree	1.9%	
High school degree	10.4%	
Some college	36.2%	
College degree	25.4%	
Some graduate school or graduate degree	26.2%	

Data source: Confidential microdata from the Kauffman Firm Survey.

Table 1b shows descriptive statistics on self-employment from the 2000 DC. These data say 6.4% of the population in 2000 was self-employed on a full time basis. 3.5% of individuals were full-time self-employed with no wage or salary income. Finally, about 2.5% of the population was full-time self-employed in high-skill industries, and the remaining 3.9% were in low-skill sectors.[22]

Statistics on the income of these subgroups of self-employed are also shown in Table 1b. Overall, the average self-employment income from those working full time was just under \$30,000, while those that had only self-employment income and no wage and salary income averaged more than \$45,000. Finally, self-employment income in high-skill sectors averaged about \$40,000, while in low-skill sectors averaged less than \$22,000.

At the individual level, we categorize workers into one of five mutually exclusive categories based on their highest educational achievement: less than high school, high school graduates, some college, college graduates, and more than college. Table 1b shows that self-employed workers are highly educated, with some 16% having a graduate degree, though not as educated as the business owners in the KFS. At the market level, we use a measure often used in research on skill-biased technological change, the so-called "college-equivalent share" (Katz and Murphy, 1992; Autor et al., 1998; Card and DiNardo, 2002). It is defined as the share of the full-time work force with at least 16 years of education plus half of the share of those with some college but no four-year degree.

[18] For more information about the KFS survey design and methodology, please see Robb et al. (2009). A public use dataset is available for download from the Kauffman Foundation's website and a more detailed confidential dataset is available to researchers through a data enclave provided by the National Opinion Research Center (NORC). For more details about how to access these data, please see http://www.kauffman.org/kfs.

[19] For more information on this methodology, see Robb et al. (2009).

[20] Glaeser (2007) also shows the high correlation of self-employment rates across areas measured in different narrow sectors, suggesting that the broad self-employment rate does capture some common phenomenon about an area.

[21] We defined high-skill industries by the college-equivalent share in each three-digit NAICS sector nationally (in our sample of metropolitan areas). In particular, sectors above the employment-weighted median of college-equivalent share of 0.422 were considered high skill. The other half of employment was considered to be in low-skill sectors. A list of three-digit NAICS sectors, their classification into low- and high-skill sectors and their college share is appears in the appendix table.

[22] These statistics were computed with a 50% random subsample of workers who did not meet our definition of full-time self employment, whose person weight was doubled to generate population representative statistics. The smaller sample was useful in facilitating estimates of the regressions in Table 2.

Table 1b
Means of Census variables.

	Mean (1)	Standard deviation (2)	Sample size (3)
Self-employment dummy			
Full-time (FT) (>1500 h/yr)	0.064	0.246	2,854,403
FT, no wages[a]	0.035	0.184	2,854,403
FT, high-skill sector[b]	0.025	0.156	2,854,403
FT, low-skill sector[b]	0.039	0.195	2,854,403
Self-employment income			
Full-time (>1500 h/yr)	28,561	52,593	356,806
FT, no wages[a]	45,262	59,562	195,893
FT, high-skill sector[b]	40,227	68,726	134,282
FT, low-skill sector[b]	21,178	37,237	222,524
Education of FT self-employed			
High school dropout	0.117	0.322	356,806
High school graduate	0.239	0.427	356,806
Some college	0.290	0.454	356,806
College degree	0.196	0.397	356,806
Graduate degree	0.157	0.364	356,806
Aggregate variables[c]			
College-equivalent share			
Metropolitan area level	0.457	0.059	230
Workplace PUMA level	0.443	0.089	1241

Notes: 2000 Decennial Census of population, 5% Public-Use Data file. Full-time self-employed workers are defined as those who are (a) age 16–65 (b) with at least one year of potential work experience (age – years of education – 6), (c) report working at least 1500 h in the past year and (d) report being self-employed. A 50% random subsample of individuals who did not meet this definition were included, and weighted double in the calculations to make the statistics population representative. Outside of this modification, census-provided person weights were used. College-equivalent share is defined as the share of workers with at least a four-year college degree plus one-half of those with 1–3 years of college attendance.

[a] Rows labelled "no wages" include in the definition of full-time self-employed the above definition plus the requirement that they report wage and salary income in the past year equal to zero.

[b] High-skill sectors are defined to be three-digit NAICS sectors with a college-equivalent share above the employment-weighted median of 0.422. Low-skill sectors are below this median. See appendix table for complete industry classification.

[c] Figures show mean and standard deviation of college-equivalent share using estimates of the number of full-time self-employed in each metropolitan area or workplace PUMA as weights. Figures are similar to two decimal places when using other weights, including population or the.

We will sometimes refer to this as just the "college share."[23] Weighted by full-time self-employment, college share averages 0.45, with a standard deviation of 0.06 across metropolitan areas and 0.09 across workplace PUMAs. The larger variation across PUMAs is helpful for identification, but it is unfortunately not feasible to use it with the smaller KFS data.

2.2. Motivation

A key motivation for examining the relationship between an area's education level and entrepreneurship is the above average wage growth in highly educated US metropolitan areas over the past several decades. As shown in Fig. 1, average wages, adjusted for individual level influences (education, experience, gender, nativity) grew systematically faster between 1980 and 2000 in areas with a higher initial college share, a fact that Glaeser and Saiz (2003) have previously pointed out.[24] In particular, the wages in the most highly educated areas, such as San Francisco, increased about 20% percentage points faster than areas not highly educated, such as Hickory, North Carolina. Notice this is not just the result of the fact that the returns to college rose between 1980 and 2000: wages in Fig. 1 are adjusted for the influence of individual-level education. Furthermore, one can look separately at wage growth among narrow education subgroups, and for each subgroup, wage growth is also strongly related to initial college share in their labor market.[25]

The relationship in Fig. 1 may be related to the urban wage premium (e.g., Glaeser and Mare, 2001). A rich set of empirical studies shows that in denser markets and, in particular, markets with greater concentrations of college-educated workers, observably similar workers are paid higher wages (e.g., Rosenthal and Strange, 2004, 2008; Fu and Ross, 2007; Rauch, 1993; Moretti, 2004a). As these studies are careful to point out, per se this does not imply college share (or density) has a causal impact on wages. Of particular concern for us is what Combes et al. (2008) call "the endogenous quality of labor bias," that higher quality workers differentially sort into more skilled areas.[26] Nevertheless, these studies tend to find that a (at least a small) wage premium associated with local workforce education is robust to a variety of instrument and control strategies, and therefore tend to come down on the side of there being a causal effect.[27,28] On the other hand, the mechanism which generates this premium is not firmly established by this research.

It is worth emphasizing that Fig. 1 examines wage growth, and so the endogeneity concerns are a bit different for it than for most of the urban literature, which tends to focus on wages levels. In particular, the fact that unobservably higher quality workers sort into more educated metro areas is not by itself a problem for the interpretation of Fig. 1. Instead, third factors which lead to unobserved worker quality to increase systematically faster in more educated areas are the problem for Fig. 1. This could occur, for example, if the amount of sorting is increasing over time.[29] On observables, at least, there is at least some suggestion that this may be occurring – college share increases slightly faster over this same period in areas with higher initial college share.[30] On the other hand, there is also reason to believe that at least some portion of this correlation is causal. For example, Glaeser and Saiz (2003) show a relationship like in Fig. 1 is

[23] Distilling the education distribution of an area into a single measure, such as the college-equivalent share, requires many assumptions. However, many of the results in this paper are robust to how education is measured; where the results do vary, it will be noted.

[24] A larger literature has documented the relationship between average education and population or employment growth, which is the main focus of Glaeser and Saiz (2003). Another example is Nardinelli and Simon (2002). A distinction between the two is that the relationship between average education and wage growth appears only to have emerged in recent decades (Beaudry et al., 2006).

[25] Beaudry et al. (2006) find that the relationship is steeper for college-educated workers, which they argue is because college share induces faster adoption of skill-complementary technology.

[26] How large this bias is not agreed upon in the literature. Research tends to find that control for worker fixed effects (Combes et al., 2008; Glaeser and Mare, 2001) substantially reduces, but does not eliminate, the urban wage premium. However, studies which have attempted to measure "unobserved" labor quality find that little evidence that labor quality differs systematically with density or college share. These studies include Wheeler (2006) and Moretti (2004a), who use the Armed Forces Qualifying Test in the NLSY, and Bacolod et al. (2009), who show that a long list of occupational skill metrics co-vary only weakly with an area's size.

[27] One exception is Acemoglu and Angrist (2000), who find that education spillovers are small when compulsory schooling laws are used as instruments for average education in the state. This study differs from others, however, in two important ways: first, compulsory schooling laws influence workforce education at levels below college. Second, their analysis is at the state level, which some research suggests may be too large an area to detect spillovers (e.g., Rosenthal and Strange, 2008).

[28] From a worker perspective, some evidence suggests systematically higher commuting times in more educated markets may allow wages to remain higher in equilibrium in these markets (Fu and Ross, 2007).

[29] It might also occur more educated markets attract workers who have steeper earnings profiles for other reasons (perhaps because they tend to invest more in upgrading their skills).

[30] A regression of the change in college share between 1980 and 2000 on initial college share produces an estimated coefficient (standard error) of 0.065 (0.036), which is positive and marginally significant. Moretti (2004a, 2004b) also finds a slightly upward sloping relationship between changes in college share and initial college share. On the other hand, this apparent divergence of college share across areas is not robust. In particular, it does not show up in logs, and is sensitive to the inclusion of "some college" workers.

also present in first differences.[31] They also provide mixed evidence that a concurrent rise in rents allows the increase in wages to be part of a stable equilibrium.

The relationship in Fig. 1 is also similar to the positive relationship between GDP growth and initial education that is present across countries (e.g., Barro, 1991). Bils and Klenow (2000) use a calibrated model to argue that this relationship is likely due to the effect of growth on education, rather than reverse, as growth provides an individual incentive to become more educated. However, this is likely to be less of a concern in the present context, inside a single country, where one's labor market is not as strongly tied to where one is educated.[32]

Furthermore, we do not take a stand on whether the relationship in Fig. 1 is causal. Instead, Fig. 1 provides motivation for trying to understand the mechanism which might lead more educated areas to have faster wage growth. In particular, instead of attempting to sort out the endogeneity issues, this paper looks for indirect evidence that greater entrepreneurship could be contributing to the relationship in Fig. 1. Glaeser (2007) shows that measures of entrepreneurship (firm size and self-employment rates) are correlated with faster population and employment growth. Entrepreneurship could contribute to wage growth through several channels. For instance, entrepreneurs may be more likely to adopt the most recent productivity-increasing technologies.[33] Also, entrepreneurs may be more innovative, developing new products and entering into new, higher productivity sectors.[34] Entrepreneurs may also be more successful at forming businesses, particularly in high-skill sectors, the thicker the local labor market for workers with specialized skills is, enjoying the benefits of so-called labor pooling. But the bottom line is if education contributes causally to wage growth through entrepreneurship, we ought to see greater rates of business formation and more productive new businesses forming in more educated markets. If it is all due to sorting on unobserved labor quality, or some other mechanism, we may see no such relationship.

To be sure, other factors might also bias OLS estimates of the relationship between college share and entrepreneurship outcomes. Reverse causality is one potential issue we will not be able to address – educated workers may seek out more "entrepreneurial" markets.[35] Endogenous quality of labor bias is a concern here, too – more educated cities may have more talented entrepreneurs. Industry mix is another potential confounder which has been found to be correlated with entrepreneurship rates (Glaeser and Kerr, 2006) and may have some correlation with education mix, so it will be important to control for industry mix.[36] Finally, as we show below it will be important not to confound the effects of business owners' skills with a "spillover" from the education of the local market.

2.3. Our approach

Our basic approach will consist of estimating relationships of the form:

$$Y_{ijc} = \alpha_j + \theta CS_c + \beta' X_{ijc} + u_{ijc} \tag{1}$$

where Y_{ijc} is an entrepreneurship outcome for person (or enterprise) i in three-digit NAICS industry j and CMSA c, α_j are industry effects, CS_c is college-equivalent share in CMSA c, and X_{ijc} is a vector of individual-varying covariates, and u_{ijc} are unobserved determinants of Y_{ijc}.

Previous research has found an association between owner education and the success of an individual business (e.g., Bates, 1990; Fairlie and Robb, 2008; Astebro and Bernhardt, 2003) so it will important for X_{ijc} to include owner education. Whether this is a causal relationship or not, however, has not been established (van der Sluis et al., 2008), and a variety of factors may account for this relationship: the knowledge and skills acquired through formal education may be useful for running a successful business; education may proxy for an owner's ability or send a positive signal to potential customers, lenders, and business suppliers; and education might simply be correlated with other traits that influence business success, such as access to social networks. Opportunity cost alone provides a powerful reason why education might be (non-causally) associated with business performance: in equilibrium, we should only expect those who expect to do better as entrepreneurs than in salaried jobs to remain as entrepreneurs.

Another (non-causal) reason why more educated business owners might be more successful, and one not usually considered in studies on the effect of owner education, is that highly educated business owners are more likely to have access to a highly educated local labor force. To oversimplify, if an area's entrepreneurs are drawn randomly from the local population, then areas with more educated populations are likely to have more educated business owners. Interestingly, this oversimplified description turns out to be not inconsistent with the empirical facts. Using the KFS data, a regression of owner's college completion on the college share in the surrounding area fails to reject a coefficient of one. Using the DC data, a regression of the college share among the self-employed on the college share in the labor market as whole also fails to reject a coefficient of 1. Fig. 2 plots this relationship for our sample of 230 metropolitan areas. It shows the two series are highly related (R^2 of 0.77).[37]

The strength of these empirical relationships does not imply that entrepreneurs are really just "random draws" from the local workforce. For example, perhaps industry mix drives the education of both workers and entrepreneurs – say, high tech areas attract both more educated entrepreneurs and more educated workers. In fact, however, this correlation is similar even within narrow industry and further within narrow industry $\times$ occupation cells. Fig. 3 is the same as Fig. 2 except it looks within narrow industry by occupation categories, which divide full-time self employed persons into 11,408 cells.[38] Even within these narrow cells, the college share among self-employed is highly correlated (R^2 = .75) with the workforce as a whole, and the slope is close to 1.

The point of Figs. 2 and 3 is that even in regressions with a large number of controls, owner education may pick up the influence of aggregate education and vice versa. In light of the fact that work-

[31] In particular, in their Table 6, they stack several censuses of population to estimate the relationship between adjusted wage growth and initial college share by decade, and condition on metro area and year effects, equivalent to differencing.

[32] In addition, their Table 3 makes it apparent that Bils and Klenow's (2000) reverse-causality finding derives from their assumption there are diminishing returns to education. This assumption is inconsistent with the fact that the wage-education relationship is highly convex in recent US data (e.g., Lemieux, 2006).

[33] Both across (Caselli and Coleman, 2001) and within (Doms and Lewis, 2006) countries, college share is associated with faster computer adoption, for example. This is another motivation for the present study.

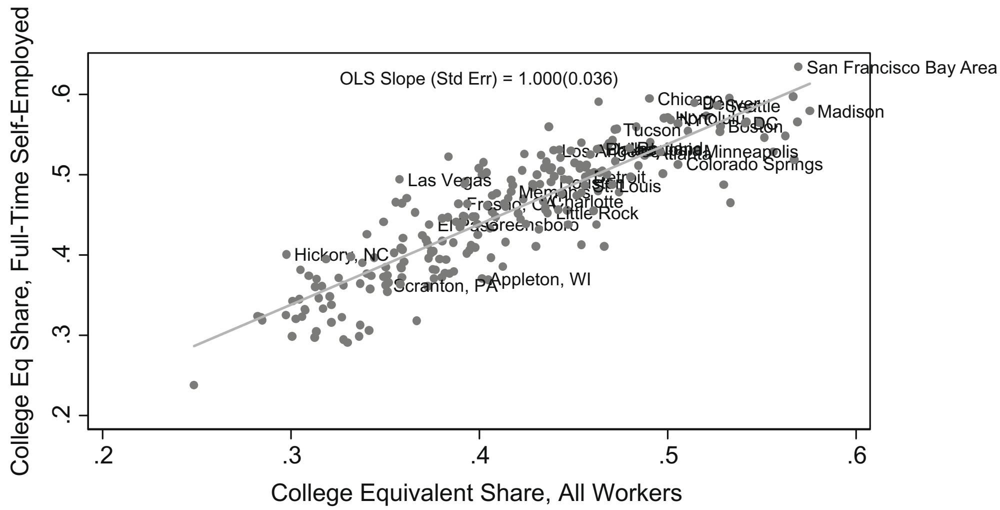
[34] This view is consistent with Glaeser and Saiz (2003) that describes how more educated cities are better at "reinventing" themselves in response to shocks (like the decline in manufacturing).

[35] Wozniak (2006) provides indirect evidence that educated workers are better informed about conditions in labor markets outside the one they currently are working in.

[36] It is worth noting that that the relationship between local education mix and industry mix is much weaker than many expect. Lewis (2004) finds education mix differences account for less than 10% of the differences in detailed industry mix across markets, or put another way, there are almost equally large differences in skill mix across markets within a given industry as in the markets a whole. Los Angeles, for example, has nearly double the proportion of high school dropouts in its overall workforce as the rest of the US Because LA's industry mix is similar to the rest of the US, even looking in a fairly high-skill industry, like retail banking, one finds Los Angeles has twice the proportion of high school dropouts as the rest of the country.

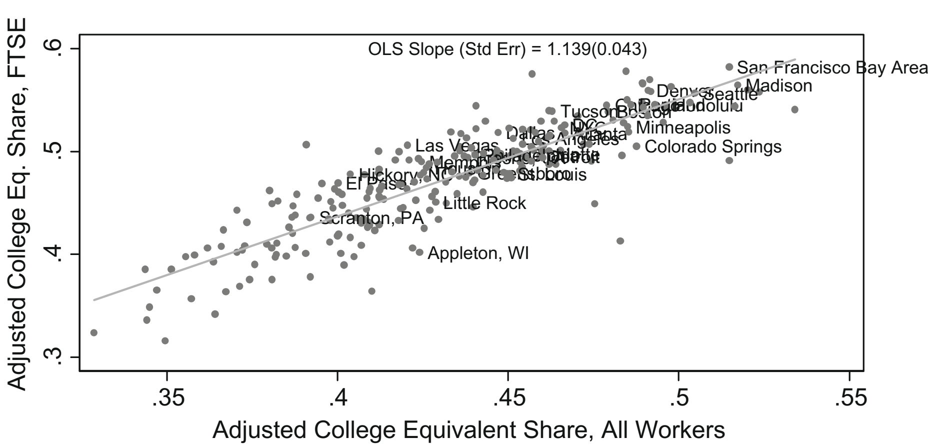
[37] This high correlation also implies that it will be difficult to separately identify the effects of owner and workforce education in these data.

[38] The industry and occupation categories are the most detailed 2000 DC categories. This divides the broader workforce into 51,489 cells.



Data Source: 5 percent public-use 2000 Decennial Census (DC) of Population. Each point represents a consolidated metropolitan statistical area (CMSA). College equivalent share is the share of workers in the CMSA with a four year college degree or great er plus one half of the share with 1-3 years of college education. Calculated in the sample of employed persons aged 16-65 with (age - years of education - 6) > 0. X-axis shows college equivalent share among all these workers. Y-axis shows college equivalent share among those who report being self employed and working at least 1,500 hours in the past year. Computed using DC 'person' weights.

Fig. 2. College share: full-time self-employed (FTSE) vs. all workers.



Data Source: 5 percent public-use 2000 Decennial Census (DC) of Population. Each point represents a consolidated metropolitan statistical area (CMSA). College equivalent share is the share of workers within each CMSA's average census industry x occupat ion cell with a four year college degree or greater plus one-half of the share with 1-3 years of college education. This was calcu lated as follows: in the sample of employed persons aged 16-65 with (age - years of education - 6) > 0, college equivalent share was com puted in each detailed census industry x occupation x cmsa cell, then weighted up to the CMSA level using by national employment in t he industry x occupation cells. (The employment weights were estimated using the DC for our sample of 230 CMSAs.) X-axis shows college equivalent share among all workers, which was weighted using total employment. Y-axis shows college equivalent share among those who report being self- employed and working at least 1,500 hours in the past year -- what we call full-time self-employed which was weighted using estimated full-time self employment. All figures estimated using DC 'person' weights.

Fig. 3. College share: full-time self-employed (FTSE) vs. all workers, within industry × Occupation.

force and entrepreneur education are so highly related, it seems appropriate to consider them jointly rather than separately whenever the data allow, so X_{ijc} in Eq. (1) must include individual education. Where the data do not allow their separate consideration, coefficients will be interpreted cautiously.

This does not put an end to concerns about unobserved heterogeneity driving our estimates of (1). A number influences in u_{ijc}, including unmeasured entrepreneurial talent and local business climate, could be correlated with X_{ijc} and CS_c. We cannot eliminate these influences, but we do what we can with these data to minimize them. In the KFS data, we take advantage of the fact that we have multiple observations on the same plant. In particular, we employ the beginning and end dates of the KFS panel. Now, as discussed at more length below, X_{ijc} and CS_c have limited time-series variation. So a simple first-difference estimate is unlikely to produce much. However, the panel structure can still help us get closer to unbiased estimates if we are interested in growth outcomes. To see this, reconsider (1) with a time dimension, i.e.

$$Y_{ijc1} = \alpha_{j1} + \theta_1 CS_c + \beta_1' X_{ijc} + u_{ijc1} \tag{1.1}$$

$$Y_{ijc2} = \alpha_{j2} + \theta_2 CS_c + \beta_2' X_{ijc} + u_{ijc2} \tag{1.2}$$

where $t = 2$ is the end and $t = 1$ is the beginning date of the panel. CS_c and X_{ijc} are modeled without time subscripts, in light of their limited time-series variation across the short period we observe them; individual characteristics like race and education, in particular, have little or no variation. The effect of these variables may, however, vary over time – β and θ have time subscripts. In particular we imagine in most cases β and θ will rise in magnitude over time, as productive attributes have a cumulative impact on new businesses. A major aim of a new business is usually to expand its market opportunities, and each business decision taken by an enterprise with a good manager may expand its market share relative to another enterprise with a bad manager. Learning may also play a role: more educated entrepreneurs may be better at gathering the information necessary to successfully run their business, which would result in a growing relationship between education and firm outcomes. These cumulative impacts can be recast as an impact of the right-hand side characteristics on the change in Y by differencing (1.2) and (1.1):

$$\Delta Y_{ijc} = \tilde{\alpha} + \tilde{\theta}CS_c + \tilde{\beta}'X_{ijc} + \Delta u_{ijc} \qquad (2)$$

where $\tilde{\alpha} \equiv \Delta\alpha = \alpha_2 - \alpha_1$, etc. The $\sim$ coefficients tells us the association between stock variables and change outcomes. These are of direct interest for the reasons described above, and because a key motivation for this study is the association between college share (in levels) and wage growth (Fig. 1).

Estimation of (2) is also motivated by the possibility that omitted variables bias is smaller than in (1) because the right-hand side variables' covariance with Δu is smaller than with u. To see the conditions under which this holds, we rewrite u_{ijct} as a factor model $u_{ijct} = \phi_t \varepsilon_{ijct}$, where ϕ_t is an unobserved time-varying factor loading on unobservables ϵ_{ijct}. In light of this error structure the error in (2) can be written as,

$$\Delta u_{ijc} = \tilde{\phi}\bar{\varepsilon}_{ijc} + \bar{\phi}\Delta\varepsilon_{ijc} \qquad (3)$$

$\bar{\varepsilon}_{ijc}$ and $\bar{\phi}$ are averages over the two periods, e.g., $\bar{\varepsilon}_{ijc} = 0.5*(\varepsilon_{ijc2} + \varepsilon_{ijc1})$. The potential advantage of (2) over (1) stems from the plausible possibility that variation in $\Delta\varepsilon_{ijc}$ is small – unobserved factors, like observed factors, evolve slowly – which may help reduce omitted variables bias. To be an unambiguous improvement over (1) requires also requires that the effect of the unobservables not vary much over time ($\phi_2 \approx \phi_1$ so $\tilde{\phi}$ is small). Working against this is our argument that observed factors have a time-varying impact on business outcomes, which suggests that unobserved factors might, too.

In some cases we also allow the effects to vary between high- and low-skill sectors. This is of direct interest, but again also provides an opportunity to potentially reduce bias in an analogous way. Think of t as indexing high-and low-skill sector rather than periods: we can compare plants in different sectors, but in the same CMSA. So long as, for example, unobserved business talent has roughly the same impact on high- and low-skill sectors, then the difference in impact of college share on high-and low-skill sectors will be better identified than each effect separately. To try to rule out the possibility that there are unobservable sector $\times$ CMSA effects biasing these estimates, in the DC data we perform estimates which include unrestricted sector $\times$ CMSA effects:

$$Y_{ijcp} = \alpha_{jc} + \theta CS_{cp} + \beta' X_{ijcp} + u_{ijcp} \qquad (4)$$

where α_{jc} is now a full set of industry $\times$ CMSA controls. This is identified when we allow college share to vary at the PUMA level (explained below), indexed by *p*.

3. Are more educated metropolitan areas more entrepreneurial?

We now turn to the question, are more educated markets more entrepreneurial? Several different measures of "entrepreneurship" more or less say "yes." Panel A of Fig. 4 shows that rates of business establishment formation (Panel A) is positively related to an area's college share in the latest available data, 2005.[39] Importantly this is not just a mechanical effect of higher rate of population growth in these markets (documented by Glaeser and Saiz, 2003), either: Panel B shows that the rate of business death is also higher in more educated markets. Instead, educated markets are more dynamic in the true sense of having greater change and turnover in businesses.

Panel A of Fig. 5 plots self-employment rates in the same year, calculated using the American Community Survey (ACS, via Ruggles et al., 2008), against college share in 2000 (calculated in the DC). The ACS is basically the same as the 2000 Census but is a smaller survey taken annually.[40] Fig. 5 shows that self-employment rates also weakly increase with college share, though an OLS estimate of the slope, shown at the top of the plot, is not statistically significant. Even this weak positive relationship is not present if one restricts the definition of "self-employed" to exclude those with wage income, shown in Panel B of Fig. 5. However, Panel A masks that college share is associated with significantly higher rates of self-employment in high-skill sectors (Panel C) and lower rates in low-skill sectors (Panel D).

The relationship between self-employment and education is not, as it turns out, a metropolitan area-level phenomenon. More educated people are more likely to be entrepreneurs, and once this is taken into account, the metropolitan area-level relationship disappears. This is shown in Table 2, which presents linear probability models using the DC where the dependent variable is a dummy for self-employment (1 = self employed, 0 = not) on four education categories of the individual (high school dropout is excluded) plus our metropolitan area-level measure of college share. Using the most basic definition of full-time self employed, in column (1), the coefficients on the individual-level education variables are significant, but there is no relationship with college share beyond this. Adding additional controls, in columns (2) and (3), and more restrictive definitions, in the remaining columns, do not revive the metropolitan area level positive relationship, and, if anything, a weak negative relationship surfaces.[41]

Table 2 is one last reminder that metropolitan area-level associations may reflect an aggregate or "spillover" effect (as they are sometimes interpreted in papers which use aggregate data) or merely reflect compositional differences, in this case, that more educated individuals are more likely to become entrepreneurs. In the regressions in the remainder of the paper we will examine associations with owner and area education simultaneously.

4. Education and business performance

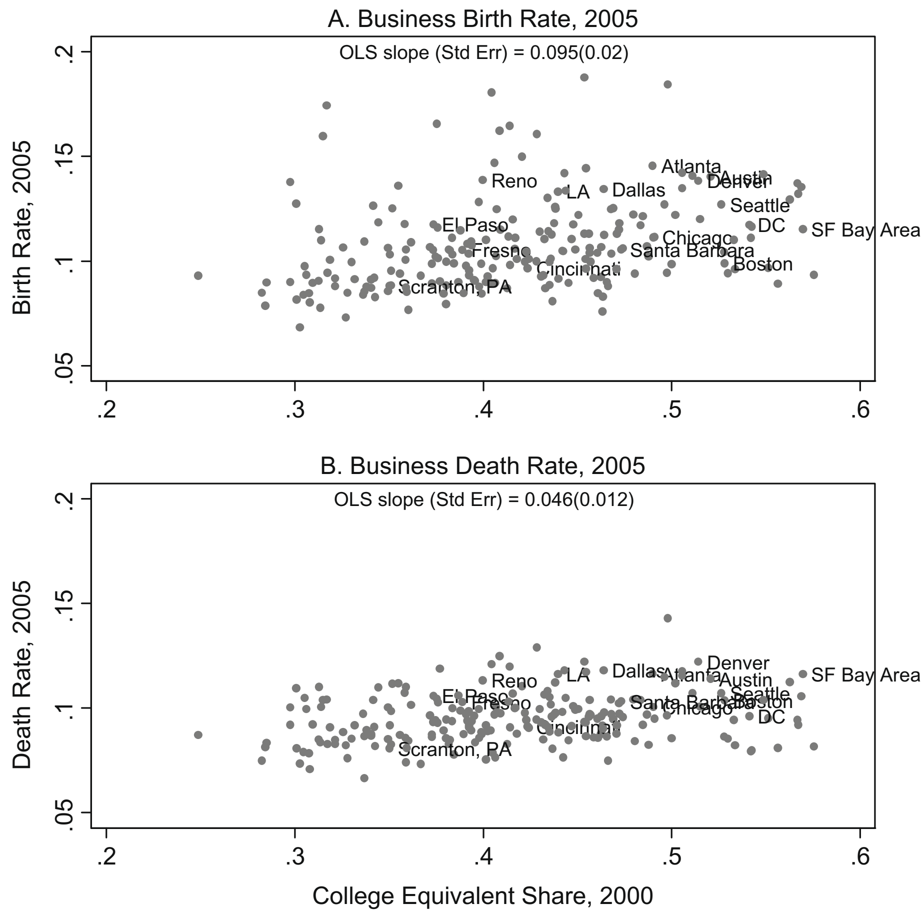
4.1. Findings

Using the KFS, we investigate whether owner education and the education level of metropolitan areas are positively correlated with a variety of outcomes we use to measure business performance. We will examine a variety of outcomes, including the logs of 2007 employment, revenue, profits, and assets, as well as the growth rates of these outcomes over the 2004–2007 period. We also investigate survival over the 2004–2007 period. In all cases, firm and owner characteristics are measured in the first wave of data (2004) and standard errors are calcu-

[39] The business formation rate is the count of new establishments divided by the number of existing ones, estimated from confidential Census establishment data by the Small Business Administration (SBA). The count of new establishments overstates the rate at which new businesses are formed, since some new establishments are created by existing businesses. The SBA data unfortunately do not distinguish between the two types of establishment openings. However, the rate of formation of very small establishments is also correlated with college share.

[40] For comparability to the SBA data, full-time self-employment rates are calculated in Fig. 4 using the 2005 ACS, rather than the DC used in the rest of the paper. The ACS is, however, is designed to be comparable to the 2000 DC, and similar results are obtained with the 2000 DC. In addition, the means of the four full-time self-employed variables are similar in the 2005 ACS and in the DC. Means for the DC are in Table 1b.

[41] Glaeser (2007) and Glaeser and Kerr (2006) also find a weak negative relationship between entrepreneurship and local average education conditional on controls.



Data Source: For business birth and death rates - U.S. Small Business Administration, Office of Advocacy, based on data provided by the U.S. Census Bureau, Statistics of U.S. Businesses. Downloaded from http://www.sba.gov/advo/research/data.html. Birth rate and death rate are computed as the count of establishment births deaths, respectively, divided by beginning-of-year establishment counts. College equivalent share is computed as the share of population aged 16-65 with age-yrs of ed-6>0 with at least a four-year college degree plus one-half of those with 1-3 years college using the 2000 Census of Population 5% public-use microdata file. Share computed with census person weights.

Fig. 4. Business dynamics and college share.

lated to be robust to arbitrary error correlation within metropolitan area.[42]

Table 3 shows estimates of Tobit regressions on the logs of employment, revenue, profits, and assets in 2007 on individual and aggregate education variables and controls.[43] The omitted dummy for education level is less than high school graduation. These results show that owner education coefficients are all positive and, in most cases, monotonic and statistically significant. Despite the monotonicity, a linear "years of education" variable could not capture these relationships: measured by revenue, employment, and assets, it is owners with college and graduate degrees that have much larger businesses. Table 3 in Van der Sluis et al.'s (2008) review of more than 100 entrepreneurship studies shows that most also find this non-monotonicity at college.

The aggregate college share variable coefficient is also positive in all but the employment regression. Unfortunately the standard errors are large and confidence intervals include both zero and large effects. Thus, we are unable to say in these regressions whether the education of a metropolitan area's workforce matters for business performance as measured by the levels of these four outcomes.

The other independent variables in these regressions included primary owner's race, ethnicity, age (and age squared), years of work experience, and average hours worked in a week. Firm level controls include legal form and two-digit NAICS industry codes. The results for the various other owner characteristics that are controlled for in the models are consistent with previous research in this area.

Table 4 examines firm survival and the relationships of these same covariates. Results from a probit model, which have been transformed to marginal effects, are shown.[44] Column (1) shows estimates for all firms in the sample. Again, survival is monotonically increasing in the primary owner's education. The effects on all of the owner education dummies are statically significant and monotonic, but there is not the non-linear relationship present in Table 3. While the coefficient on the college share variable is also positive, it is not statistically significant. The other control variables were identical to the previous models (primary owner's race, ethnicity, age (and age squared), years of work experience, and average hours worked in a week, as well as the firm's legal form and two-digit NAICS industry code). Columns (2) and (3) split the sample into, respectively, high- and low-skill sectors. Interestingly, the owner education effects all load onto the low-skill industry variables; nothing is significant in high-skill industries and the point estimates are actually negative. College share is not significant in either regression, but its coefficient is positive in high-skill sectors and negative in low-skill sectors.[45]

---

42 To use STATA's terminology, errors are clustered at the metropolitan level.

43 Negatives and zeros were set to 1 before taking logs; the lower censoring point is zero.

44 The conversion to marginal effects was done automatically by the STATA command DPROBIT.

45 We find similar patterns college share effects for the outcomes in Table 3 (not shown). College share generally has a positive association in high-skill sectors and negative in low-skill sectors. In no case are the coefficients, or even the difference in coefficients, statistically distinguished from zero.

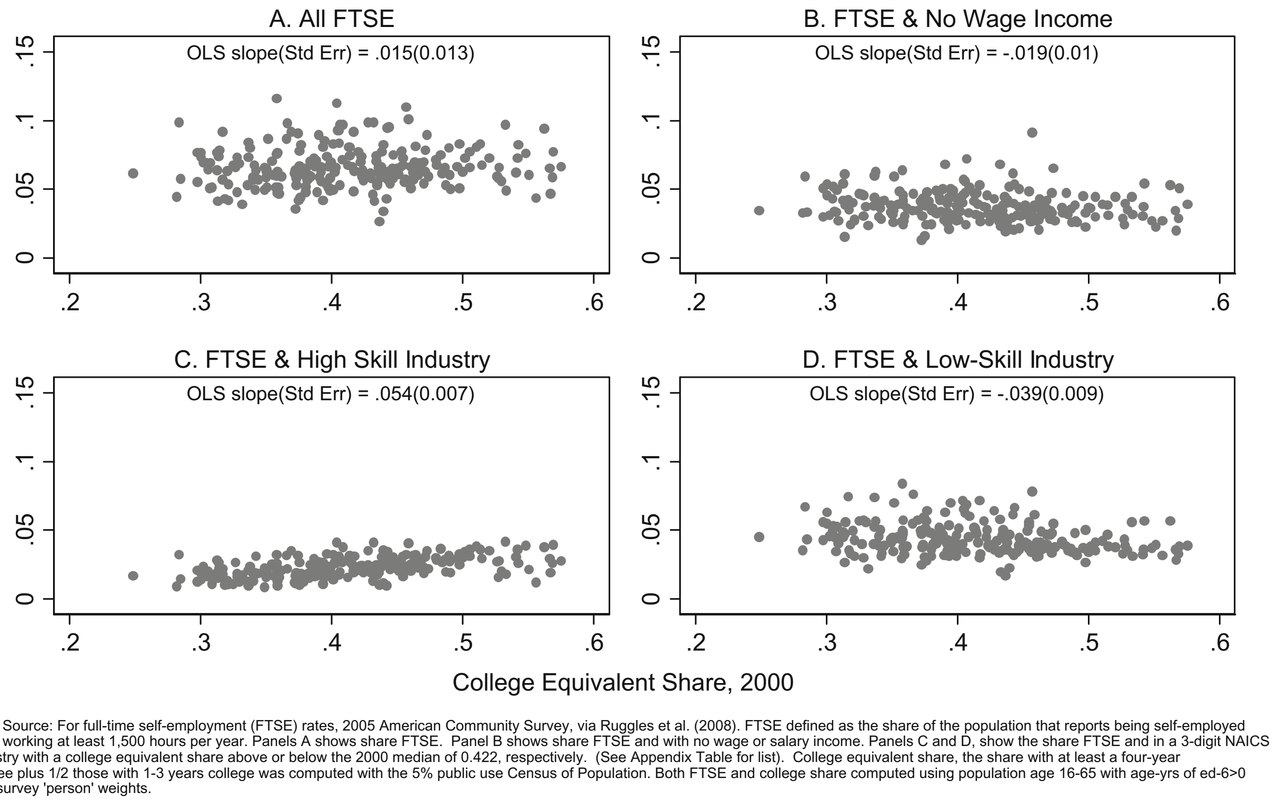


Fig. 5. Full-time self employment (FTSE) rate, 2005 vs. college-equivalent share, 2000.

Finally, we use the KFS to examine the growth rates of employment, revenue, profits, and assets over the 2004–2007 period. As described in Section 2, these outcomes are of direct interest, but also these regressions may suffer from less omitted variables bias than ones where outcomes are measured in levels. The results are shown in Table 5. The control variables used in Table 5 are identical to those used in previous models. The coefficients on primary owner are mainly positive in all of the models. However, only one of the coefficients is statistically significant (and that is at the 10% level) in any of the models.

The large standard errors on the KFS results, especially for college share, make the results uninformative. So another way we attempt to measure business performance is with data on the income from self-employment activities in the DC. While self-employment is not an ideal measure of entrepreneurial activity, the decision to be self-employed does involve risk taking. Self-employment data are widely available and are often used to proxy for entrepreneurial activity (see e.g., Blau 1987; Storey 1991; van Stel 2005). For our purposes, the main advantage with the self-employment measure is the much larger sample size. Table 6a studies the self-employment income for all individuals who report being self-employed on a full time basis. It also examines subgroups described earlier.

Column (1) of Table 6a shows the results for all full-time (>1500 h per year) self-employed workers. More educated individuals earn more as self-employed workers.[46] Demographically similar workers with exactly a four-year college degree earn $3500 more per year from self-employment than those with exactly a high school degree. Interestingly, high school dropouts seem to earn more from self-employment than those with slightly more education. However, this may reflect low-skill workers' mix of involvement in casual contract labor and other low skill jobs. When those with any labor income are dropped from the sample, in column (2), self-employment earnings become monotonic in education. In high-skill sectors, in column (3), the only significant education premium accrues to those with a four-year degree or more. Regardless of the sample used, there is a sharp upward break in earnings for those with at least a four-year degree, similar to the non-monotonicity found in the KFS outcomes in Table 3.

In addition to the showing that more educated individuals have higher self-employment earnings, Table 6a reveals otherwise similar workers in more educated CMSAs have higher self-employment earnings. This is shown in the fifth row of the table. This parallels a similar aggregate relationship for wages (Rauch, 1993; Moretti, 2004a). The association is substantially stronger in high-skill sectors, shown in column (3), than in low-skill sectors, shown in column (4). In fact, the point estimate for low-skill sectors, in addition to be less than one-third as large as for high-skill sectors, is not statistically distinguished from zero.[47] These data start to suggest that any "spillover" from education is mostly concentrated in high-skill sectors. However, before giving it that interpretation, we note that some previous research has found that agglomeration effects in general and human capital spillovers in particular are larger for college-educated workers.[48] So the greater response of high-skill sectors may not really be a sector effect, rather, it may simply reflect that that high-skill sectors have more college graduates,

[46] This result could arise for a variety of reasons. One could be that higher educated workers (and hence higher wage workers) are more likely to receive proportionately more non-wage benefits if they instead worked for employer firms. Therefore, in equilibrium, the self-employment wage premium would be larger for more educated workers.

[47] Interestingly, education is less important at an individual level in low-skill sectors as well.

[48] Examples include Wheeler (2001) and Fu and Ross (2007). In addition, Bacolod et al. (2009) find agglomeration effects are larger for more educated individuals, and people with higher "cognitive" and "people" skills. In contrast, Moretti (2004a) finds a smaller effect of college share on college graduates than other workers, which he interprets as evidence of a combination of a spillover and labor supply effects.

Table 2
Determinants of full-time self employment.

Dependent Variable is a dummy for...	Full-time self-employed (FTSE)			FTSE and no wage income	FTSE in high-skill sector	FTSE in low-skill sector
	(1)	(2)	(3)	(4)	(5)	(6)
High school graduate	0.00884***	−0.00103	0.00988***	0.00326***	−0.00240***	0.0123***
	(0.00100)	(0.000864)	(0.000777)	(0.000525)	(0.000256)	(0.000666)
Some College	0.00756***	0.000279	0.0187***	0.00388***	−0.000336	0.0190***
	(0.00103)	(0.00126)	(0.000846)	(0.000818)	(0.000309)	(0.000748)
College degree	0.0175***	0.00212	0.0300***	0.00461***	0.00806***	0.0219***
	(0.00136)	(0.00163)	(0.000902)	(0.000851)	(0.000509)	(0.000661)
Graduate degree	0.0501***	0.0225***	0.0588***	0.0192***	0.0479***	0.0109***
	(0.00238)	(0.00249)	(0.00170)	(0.00113)	(0.00158)	(0.000705)
College-equivalent share	−0.0181	−0.00673	−0.0353***	−0.0248*	−0.0247***	−0.0106
	(0.0144)	(0.0139)	(0.0116)	(0.0143)	(0.00715)	(0.00846)
Black		−0.0370***	−0.0238***	−0.0124***	−0.00663***	−0.0172***
		(0.00142)	(0.00137)	(0.000685)	(0.000691)	(0.000785)
Asian		−0.0116***	−0.000987	0.00145	−0.00756***	0.00658***
		(0.00330)	(0.00272)	(0.00161)	(0.000910)	(0.00233)
Hispanic		−0.0258***	−0.0233***	−0.0116***	−0.00242***	−0.0209***
		(0.00190)	(0.00124)	(0.000668)	(0.000531)	(0.00120)
Female		−0.0483***	−0.0411***	−0.0185***	−0.0231***	−0.0179***
		(0.00124)	(0.000862)	(0.000646)	(0.000561)	(0.000433)
Foreign-born		0.0152***	0.00492**	0.00125	−0.00181*	0.00673***
		(0.00209)	(0.00202)	(0.00164)	(0.000999)	(0.00148)
Age		0.00653***	0.00717***	0.00393***	0.00170***	0.00547***
		(0.000169)	(0.000145)	(0.000137)	(7.01e−05)	(0.000107)
Age squared		−5.53e−05***	−6.04e−05***	−3.40e−05***	−1.07e−05***	−4.97e−05***
		(2.17e−06)	(1.83e−06)	(1.22e−06)	(8.02e−07)	(1.39e−06)
R-squared	0.003	0.028	0.090	0.055	0.111	0.106
Observations	2,854,403	2,854,403	2,854,403	2,854,403	2,854,403	2,854,403
Three-digit NAICS?	No	No	Yes	Yes	Yes	Yes

Data source: 2000 Decennial Census of Population, 5% Public-Use Data files. Dependent variable in columns (1)–(3) is a dummy variable equal to 1 if the person reports: (a) being age 16–65 with at least one year of potential work experience (age – years of education – 6) (b) being self-employed, and (c) working 1500 h or more per year. This is what we call "full-time self-employed." In column (4) the dependent variable is equal to if the person is full-time self-employed and reports no wage or salary income. The dependent variable in column (5) is a dummy variable equal to 1 if the person is full-time self employed and works in a three-digit NAICS sector which has a college-equivalent share above the (employment weighted) median of 0.422. The dependent variable in column (6) is a dummy variable equal to 1 if the person is full-time self employed and works in a three-digit NAICS sector which has a college-equivalent share below the (employment weighted) median of 0.422. See appendix table for classification of industries. Estimation sample includes all full-time self-employed workers plus a 50% random subsample of those who do not meet the definition, whose weight is doubled to make estimates population representative. Estimation method is by OLS using census sampling weights (doubled when necessary). College-equivalent share, which is the sum of the share of workers with at least a four-year college degree plus 1/2 of the share with 1–3 years of college education, is computed at the CMSA level using the sample of employed persons age 16–65 with at least one year of potential work experience, again using modified census sampling weights. Standard errors in parentheses robust to arbitrary error correlation within CMSA and heteroskedasticity.
* $p < 0.1$ significance level.
** $p < 0.05$ significance level.
*** $p < 0.01$ significance level.

and college graduates' self-employment earnings (possibly) respond more to local college share than non-college graduates. However, this alternative story does not appear to be what drives the sector differences. Column (5) and (6) adds an interaction between college share and whether or not the individual is a college graduate (four-year degree or more). It shows that the association between self-employment earnings and college share are indeed larger for four-year college graduates (last row of table), but the association for col-

Table 3
Kauffman Firm Survey: Tobits on firm outcomes.

Coefficient	Log of 2007 revenue	Log of 2007 employment	Log of 2007 profits	Log of 2007 assets
High school graduate	2.681	1.296**	4.254**	2.999*
	(1.877)	(0.632)	(1.961)	(1.691)
Some college	1.912	1.207**	2.313	3.274**
	(1.610)	(0.613)	(1.993)	(1.478)
College degree	4.375***	1.527**	4.287**	4.229***
	(1.620)	(0.597)	(1.946)	(1.461)
Graduate degree	4.818***	1.546***	5.128**	4.537***
	(1.635)	(0.591)	(2.137)	(1.420)
College educated share	1.230	−1.079	0.0277	0.105
	(4.307)	(0.940)	(4.259)	(2.881)
Observations	3121	3121	3121	3121

Standard errors in parentheses robust to clustering on CMSA.
Confidential Kauffman Firm Survey Microdata.
Other controls include: three-digit NAICS, legal form, and primary owner race, ethnicity, gender, age, age squared, years of previous industry experience, and hours worked in the business. Negatives and zeros were set to 1 before taking logs; the lower censoring point is zero.
College-equivalent share, which is the sum of the share of workers with at least a four-year college degree plus 1/2 of the share with 1–3 years of college education, is computed at the CMSA level using the sample of employed persons age 16–65 with at least one year of potential work experience, again using census sampling weights. Standard errors in parentheses robust to aribitrary error correlation within CMSA and heteroskedasticity.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Table 4
Kauffman Firm Survey: Dprobit on firm survival (2004–2007).

Coefficient	All industries survival (2004–2007)	High-skill industries survival (2004–2007)	Low-skill industries survival (2004–2007)
High school graduate	0.134**	−0.0493	0.151*
	(0.0552)	(0.200)	(0.0777)
Some college	0.163***	−0.108	0.187**
	(0.0571)	(0.184)	(0.0817)
College degree	0.194***	−0.0713	0.187**
	(0.0545)	(0.169)	(0.0751)
Graduate degree	0.203***	−0.0467	0.224***
	(0.0425)	(0.165)	(0.0621)
College educated share	0.0807	0.355	−0.268
	(0.169)	(0.259)	(0.297)
Observations	3066	1518	1203

Standard errors in parentheses robust to clustering on CMSA.
Confidential Kauffman Firm Survey Microdata.
Other controls include: three-digit NAICS, legal form, and owner race, ethnicity, gender, age, age squared, years of previous industry experience, and hours worked in the business. See appendix table for classification of industry.
College-equivalent share, which is the sum of the share of workers with at least a four-year college degree plus 1/2 of the share with 1–3 years of college education, is computed at the using the sample of employed persons age 16–65 with at least one year of potential work experience, again using census sampling weights. Standard errors in parentheses robust to aribitrary error correlation within CMSA and heteroskedasticity.
Dprobit fits maximum-likelihood probit models and is an alternative to probit. Rather than reporting the coefficients, dprobit reports the marginal effect, that is the change in the probability for an infinitesimal change in each independent, continuous variable and, by default, reports the discrete change in the probability for dummy variables.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Table 5
Kauffman Firm Survey: Tobits on firm outcomes for survivors.

Coefficient	2004–2007 Log employment growth	2004–2007 Log revenue growth	2004–2007 Log profit growth	2004–2007 Log assets growth
High school graduate	0.0360	−0.107	0.151	0.181
	(0.0946)	(0.227)	(0.152)	(0.219)
Some college	−0.00106	−0.128	0.173	0.0762
	(0.0896)	(0.217)	(0.153)	(0.150)
College degree	0.0608	0.0282	0.215	0.287*
	(0.0942)	(0.229)	(0.147)	(0.161)
Graduate degree	0.101	0.00697	0.236	0.292
	(0.0858)	(0.221)	(0.148)	(0.182)
College educated share	0.0189	−0.805*	−0.348	−0.632
	(0.212)	(0.486)	(0.353)	(0.629)
Observations	2166	2166	2166	2166

Standard errors in parentheses robust to clustering on CMSA.
Confidential Kauffman Firm Survey Microdata.
Other controls include: two-digit NAICS, legal form, and owner race, ethnicity, gender, age, age, squared, years of previous industry experience, and hours worked in the business.
College-equivalent share, which is the sum of the share of workers with at least a four-year college degree plus 1/2 of the share with 1–3 years of college education, is computed at the CMSA level using the sample of employed persons age 16–65 with at least one year of potential work experience, again using census sampling weights. Standard errors in parentheses robust to aribitrary error correlation within CMSA and heteroskedasticity.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

lege- and non-college graduates are both larger in the high-skill sector (column 5) than the low-skill sector (column 6).[49]

These findings are consistent with the association of self-employment earnings with college share being driven by labor pooling.[50] It is intuitive (and commonly asserted in urban research) that a more educated workforce is likely to be more specialized.[51] If this is the case, then firms in high-skill industries, which will require more specialized workers, will find it easier to sustain themselves and grow in thicker markets for those specialized skills.[52] To this point, more skill-intensive industries have previously been found to be more geographically concentrated than other industries (Rosenthal and Strange, 2001). Also, from both a worker and firm perspective, a bigger market for specialized skills would allow more rapid reallocation of workers to firms with more productive innovations, to paraphrase Fallick et al. (2006).[53] In addition to higher wage and productivity levels, then, this could lead to faster wage growth like was found in Fig. 1.

[49] To drive the point home, notice that the implied point estimate for college graduates in low-skill sectors, 22,764 (=10,631 + 12,133), is smaller than for non-college graduates in high-skill sectors, 29,532 (though the difference may not be statistically significant).

[50] Rosenthal and Strange (2004) identifies two particular benefits of labor pooling – matching and risk sharing – which are formalized in Duranton and Puga (2004). See also notes below.

[51] Rosenthal and Strange (2004) say "it is likely that skilled labor is more specialized than is unskilled labor" (p. 2158). Bacolod, Blum, and Strange (2009) say "cognitive workers may be more specialized" (p. 137). Although very intuitive, as far as we know, there is no research directly confirming this. Developing metrics of specialization would be a requirement for this. For example, Rosenthal and Strange cite a study by Baumgardner (1988) showing doctors perform a narrower range of tasks in larger markets.

[52] This is where the benefits of risk sharing come in Duranton and Puga (2004) formalize the conditions under which if firms face uncertainty about their output, then increasing the number of firms raises average profits and wages.

[53] In their conclusion, Fallick et al. say "frequent job-hopping facilitates the rapid reallocation of resources towards firms with the best innovations." Rosenthal and Strange (2004) cite this as an example of better matching of workers to firms.

Table 6a
Income of full-time self-employed, college share at the CMSA level.

Sample	All (1)	No wage income (2)	High-skill sectors (3)	Low-skill sectors (4)	High-skill sectors (5)	Low-skill sectors (6)
High school graduate	−549.4	748.2*	−1028	260.6	−1017	267.1
	(395.4)	(431.9)	(718.1)	(470.4)	(721.7)	(469.8)
Some college	−1154*	1942***	−444.7	265.4	−297.2	284.9
	(611.7)	(527.7)	(783.4)	(692.2)	(786.0)	(690.0)
College degree	2944***	13507***	7335***	1908***	−8324	−2970
	(598.0)	(644.0)	(904.3)	(681.5)	(6243)	(4083)
Graduate degree	17788***	38724***	24044***	4730***	8336	−182.3
	(1038)	(1390)	(1335)	(1050)	(5885)	(4346)
College equivalent Share	26998**	43930***	49729***	13898	29532***	12133
	(11905)	(8280)	(15833)	(10422)	(10531)	(9212)
College Eq. share *College graduate					34128**	10631
					(13540)	(8619)
R-squared	0.082	0.232	0.076	0.017	0.077	0.017
Observations	356,806	195,893	134,282	222,524	134,282	222,524
Three-digit NAICS?	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls?[a]	Yes	Yes	Yes	Yes	Yes	Yes

Data source: 2000 Decennial Census of population, 5% Public-Use Data files. Dependent variable is self-employment earnings for various subsamples of self employed. In column (1) the sample consists of persons who report (a) being age 16–65 with at least one year of potential work experience (age – years of education – 6) (b) being self-employed, and (b) working 1500 h or more per year. This is what we call "full-time self-employed. In column (2) the sample is full-time self-employed persons who report no wage or salary income. In columns (3) and (5) the sample is full-time self employed persons works in a three-digit NAICS sector which have a college-equivalent share above the (employment weighted) median of 0.422. In columns (4) and (6) the sample is full-time self employmed persons works in a three-digit NAICS sector which have a college-equivalent share below 0.422. See appendix table for classification of industries. Estimation method is by OLS using census sampling weights. College-equivalent share, which is the sum of the share of workers with at least a four-year college degree plus 1/2 of the share with 1–3 years of college education, is computed at the CMSA level using the sample of employed persons age 16–65 with at least one year of potential work experience, again using census sampling weights. Standard errors in parentheses robust to arbitrary error correlation within CMSA and heteroskedasticity.

[a] Demographic controls include dummies for race, gender, immigration status, age and age squared (see Table 2, which uses these same controls) as well as annual hours worked.

* $p < 0.1$ significance level.

** $p < 0.05$ significance level.

*** $p < 0.01$ significance level.

Table 6b
Income of full-time self-employed, college share at the PUMA level.

Sample	All (1)	All (2)	All (3)	High-skill sectors (4)	Low-skill sectors (5)
High school graduate	−670.7*	−231.1	−194.8	−619.7	607.3
	(395.4)	(430.0)	(433.9)	(714.7)	(499.8)
Some college	−1433**	−1313***	−1389***	−291.8	5.169
	(613.1)	(445.3)	(458.8)	(778.8)	(513.7)
College degree	2433***	2476***	2484***	−5924	−2678*
	(616.1)	(504.9)	(510.3)	(4868)	(1483)
Graduate degree	17180***	17295***	17105***	10396**	60.05
	(922.9)	(881.0)	(881.4)	(4154)	(1849)
College equivalent Share	25500***	22231***	21444***	23411***	6671***
	(5396)	(3495)	(3204)	(2816)	(1570)
College Eq. share *College graduate				27835***	9748***
				(10487)	(2879)
R-squared	0.083	0.090	0.125	0.113	0.077
Observations	356,806	356,806	356,806	134,282	222,524
Demographic controls?[a]	Yes	Yes	Yes	Yes	Yes
Three-digit NAICS?	Yes	Yes	Yes	Yes	Yes
CMSA effects?	No	Yes	Yes	Yes	Yes
CMSAx3-digit NAICS?	No	No	Yes	Yes	Yes

Data source: 2000 Decennial Census of Population, 5% Public-Use Data files. Dependent variable is self-employment earnings for various subsamples of self employed. In columns (1)–(3) the sample consists of persons who report (a) being age 16–65 with at least one year of potential work experience (age – years of education – 6) (b) being self-employed, and (b) working 1500 h or more per year. This is what we call "full-time self-employed. In column (4) the sample is full-time self-employed persons who work in a three-digit NAICS sector which has a college-equivalent share above the (employment weighted) median of 0.422. In column (5) the sample is full-time self employmed persons works in a three-digit NAICS sector which have a college-equivalent share below 0.422. See appendix table for classification of industries. Estimation method is by OLS using census sampling weights. College-equivalent share, which is the sum of the share of workers with at least a four-year college degree plus 1/2 of the share with 1–3 years of college education, is computed at the place-of-work PUMA level using the sample of employed persons age 16–65 with at least one year of potential work experience, again using census sampling weights. Standard errors in parentheses robust to arbitrary error correlation within cmsa and heteroskedasticity.

[a] Demographic controls include dummies for race, gender, immigration status, age and age squared (see Table 2, which uses these same controls) as well as annual hours worked.

* $p < 0.1$ significance level.

** $p < 0.05$ significance level.

*** $p < 0.01$ significance level.

Notice that this finding also narrows the type of unobserved heterogeneity which could drive the results if it is not a causal relationship. In particular, if what accounts for higher self-employment earnings is the higher unobserved ability of self-employed individuals in more educated markets, it must be that this higher unobserved ability is (mainly) limited to those in high-skill industries, and not just the most educated individuals in those industries. A simpler selection story in which entrepreneurs in more educated areas are more able is not adequate to account for these findings.

One factor which could account for these findings is any set of third factors or natural advantage which raised the productivity of high-skill industries in particular metropolitan areas. For example, perhaps successful high-skill sectors are often spinning off of local university research in areas with top universities. Notice that this would potentially raise earnings in high-skill sectors while simultaneously raising demand for college-educated labor (which could attract college-educated labor to the area) generating a non-causal correlation between college share and pay in high-skill sectors. To address this, we follow an approach similar to Rosenthal and Strange (2008) and examine the influence of college share at the submetro area, specifically the workplace "PUMA," which allows the addition of controls for unobserved metro $\times$ industry effects. A "public-use micro area" or "PUMA" is the smallest geographic unit available in the (public-use version of the) Census, containing 100,000 residents.[54] It is considerably smaller than a metropolitan area; on average, in our sample, there are over five workplace PUMAS per metropolitan area. To give some indication of its size, within the metro area of Philadelphia, the city of Philadelphia is one workplace PUMA.[55]

Results using workplace PUMA college share are shown in Table 6b. Column (1) is the same as column (1) of Table 6a, except that college share is now at the PUMA level. The point estimate is nearly the same as before. Column (2) adds metropolitan area effects, which lowers the point estimate on college share by 13%. Interacting metro area effects with three-digit industry effects lowers the point estimate another 4%, to 21,444. So there is some evidence here of the type of unobserved heterogeneity described, but it appears that it may not be the whole story. Columns (5) and (6) break the sample into high- and low-skill industries as before, and also allow for a different effect for college-educated self-employed persons. The pattern is exactly the same as in Table 6a: for both more- and less-educated persons, self-employment pay is higher rises more in high- than in low-skill sectors with an increase in local college share. Interestingly, the point estimates in the low-skill sector are now significantly distinguished from zero, so the additional variation within metropolitan areas has allowed us to pin down the estimates more precisely.

Again, finding this relationship does not mean that there is a causal effect of college share on the productivity of self-employed workers. It is not implausible that there is some type of selection or unobserved third factor operating at the PUMA $\times$ industry level which is correlated with the PUMA being on average more educated. But two features of the findings are consistent with a role for labor pooling at least partly driving the relationship: (1) the effect is larger for high-skill sectors than for low-skill sectors, consistent with intuition that high-skill sectors are more specialized; and (2) the effect is nearly the same at the metropolitan level as at lower levels of geography where a much richer set of controls are included and consistent with Rosenthal and Strange's (2001) finding that labor pooling is powerful at both the metropolitan-level and lower levels of geography.

We further discuss the interpretation and implications of our findings in the next section.

4.2. Discussion

To summarize our findings, we consistently find strong associations with self-employment and business outcomes at an individual level, especially for college graduates relative to less educated business owners, but the additional aggregate effect of CMSA average education produces mixed evidence. CMSA college share is not associated with higher self-employment rates, has an imprecise relationship with business outcomes in the KFS data, and is strongly associated with self-employment earnings, especially in high-skill sectors. What accounts for these findings?

Starting with the individual-level association, there are a variety of causal (human capital) and non-causal (selection) reasons why education would be associated with business outcomes and entrepreneurship. One reason why education might improve business performance is if education increases a person's breadth of skills. Lazear's (2004, 2005) provides empirical support for a model in which successful entrepreneurs are "generalists." Potentially consistent with this description, we find the benefits of education appear to be concentrated among those who complete college, where a lot of broad learning tends to go on in the US Lazear's original test of this proposition was, in fact, to compare those who took broad to those who took narrow curricula in Stanford's MBA program. Another set of facts that support this are that, in Europe, where curricula tend to be narrower (Cascio et al., 2008) the premium to college completion is smaller, even relative to the wage lower returns to schooling in European countries (Van der Sluis et al., 2008).

Results on the aggregate education variable – college share – are sometimes significant and sometimes not. One explanation for this inconsistency is that it could be the CMSA is too large an area to sharply detect the effect of college share for some outcomes. Recent research on human capital spillovers suggest they largely operate below the level of the metropolitan area (Rosenthal and Strange, 2008), in some cases in an area perhaps even as small as a city block (Arzaghi and Henderson, 2008). If so, our coefficients on our CMSA-level metrics could be severely attenuated. An interesting exception that some research suggests is that the economies of scale in pooling together workers with specialized skills, operates effectively at the metropolitan area level (Rosenthal and Strange, 2001).[56] This may be one reason why we do detect an especially strong effect of college share on self-employment earnings in high-skill sectors in the DC, and that a similar magnitude effect is found at the PUMA level. Another possibility is that the inconsistencies in the results merely stem from the fact that the smaller KFS data do not provide us enough power to detect the effects of aggregate variables separately from individual ones.[57] Consistent with this, point estimates in KFS regressions are often the same sign as one obtained for self-employment earnings. For example, when survival was the outcome, the point estimate on college share was positive in high-skill sectors and negative for low-skill sectors, but neither was significant. Finally, it may be that there really is no causal effect of college share on the firm outcomes studied in this paper.

[54] PUMAs are defined in terms of residential population counts. Workplace PUMAs are either geographically identical to these residential PUMAs or combinations of more than one residential PUMA.

[55] Rosenthal and Strange (2008) report that a metropolitan areas typically fit with a circle with a radius of 25 miles, while a typical metropolitan PUMA fits within a circle with a radius of less than 5 miles.

[56] In contrast, the information spillover channels emphasized by Rauch (1993) and Moretti (2004b) appear in this research to operate primarily at a lower level of geography.

[57] On top of this, there is an identification problem: Section 2 showed area-level education and the education of self-employed individuals or business owners are highly collinear.

As we argued in the introduction, our estimation approach is likely biased towards finding effects, so not finding an effect despite the bias may mean there is no effect to be found.

These findings also circle back to one of the motivations for this paper: trying to explain why wage growth among observably similar workers seems to have been higher in more educated labor markets since 1980 (Fig. 1). Though not necessarily a "spillover" from education, since it may operate at the individual level (Table 2), it does seem that educated labor markets are more dynamic in the true sense of having greater rates of change: the rates of both plant births and deaths are higher (Fig. 4) in more educated areas. And this could enhance wage growth because it "facilitates the rapid reallocation of resources toward the firms with the best innovations" to quote Fallick et al.'s (2006) description of the benefits of frequent "job-hopping." A related paper to Fallick et al.'s, in fact, finds job hopping is associated with faster wage growth at the individual level (Freedman, 2008). So if there is a spillover from education to wage growth, it may partly operate indirectly, by creating a more dynamic business environment.

5. Conclusion

This paper studies relationship between education, entrepreneurship, and businesses outcomes, and unlike most of the previous entrepreneurship research in this area, considers simultaneously both the education of the entrepreneur and of the workforce where the entrepreneurs operate their businesses, as Rauch (1993) and Moretti (2004a) previously did for wage workers. Consistent with this simultaneous focus, our initial results indicate that more educated entrepreneurs tend to be located in metropolitan areas with more educated workforces. Moreover, highly educated areas have above average entrepreneurship rates. Finally, the level of education of entrepreneurs is strongly related to positive business outcomes, especially for college graduates compared to those with less than a four-year degree.

This paper also presents some indirect evidence that more educated markets grow faster potentially partly as a result of having a more dynamic business environment.[58] Workers, who are more frequently reallocated to new businesses in more educated markets, appear to share in the gains of these more entrepreneurial environments.

Acknowledgments

We received many helpful comments from Stuart Rosenthal, Harvey Goldstein, Kevin Lang, Enrico Moretti, Jed Kolko, an anonymous referee, and many participants at the Western Regional Science Association, the NBER Entrepreneurship Working Group, and the NBER Cities and Entrepreneurship Conference. We thank the Ewing Marion Kauffman Foundation for financial support.

Appendix A

Industry Classification.

NAICS code	NAICS name	College equiv. share	NAICS code	NAICS name	College equiv. share
High-skill sectors			*Low-skill sectors (continued)*		
611	Educational services	0.744	422	Wholesale, nondurables	0.386
541	Professional/Scientific services	0.738	453	Miscellaneous retail stores	0.385
514	Information Svcs/Data processing	0.695	532	Rental/Leasing	0.384
923	Public administration	0.681	713	Amusement/Gambling	0.382
55	Management of companies	0.672	442	Furniture stores	0.378
712	Museums/Historical sites	0.651	333	Machinery manufacturing	0.377
813	Religious/Grantmaking/Civic Svcs	0.634	335	Electrical Equip/Appliance manufacturing	0.368
N/A[a]	Banking/Investment	0.628	488	Transportation support	0.366
711	Performing arts	0.617	623	Nursing/Residential care	0.358
511	Publishing	0.615	323	Printing	0.348
928	National security	0.605	561	Admin support Svcs	0.339
622	Hospitals	0.599	115	Agriculture support	0.337
512	Motion pictures/Recording	0.588	482	Rail transportation	0.328
621	Ambulatory health care Svcs	0.587	444	Building equipment stores	0.318
211	Oil and gas extraction	0.586	721	Accommodation	0.309
524	Insurance	0.581	322	Paper manufacturing	0.304
334	Computer/Electronic prdct manufacturing	0.577	441	Motor vehicle parts/Dealers	0.301
443	Electronics/Appliance stores	0.571	N/A[a]	Not specified manufacturing	0.298
921	Executive, legislative govt support	0.567	485	Ground passenger transport	0.297

(continued on next page)

[58] The channels through which education contributes to a more dynamic business environment include a compositional channel – more educated individuals are more likely to start businesses – a "labor pooling" channel – in thicker markets, its easier to find workers with the appropriate skills – and a information sharing channel – productive ideas get transmitted more quickly in an educated population. We provide direct evidence for the first channel in this paper, and other research suggests the labor pooling channel is especially important at the metropolitan area level studied in this paper.

Appendix A (*continued*)

NAICS code	NAICS name	College equiv. share	NAICS code	NAICS name	College equiv. share
N/A[a]	Misc public sector	0.567	452	General merchandise stores	0.297
513	Broadcasting/Telecom	0.560	812	Personal/Laundary care services	0.283
325	Chemical manufacturing	0.558	327	Nonmetallic mineral manufacturing	0.276
522	Credit intermediation	0.548	331	Primary metal manufacturing	0.273
481	Air transportation	0.534	326	Plastics and rubber manufacturing	0.273
486	Pipeline transportation	0.524	311	Food manufacturing	0.268
N/A[a]	Unspecified utilities	0.506	332	Fabricated metal products manufacturing	0.266
531	Real estate	0.499	212	Mining	0.258
N/A[a]	Wholesale, unspecified	0.487	316	Leather products manufacturing	0.256
324	Petroleum products manufacturing	0.485	114	Fishing, hunting, trapping	0.252
451	Sporting Goods/Hobby stores	0.464	23	**Construction**	0.249
446	Health and personal care stores	0.463	113	Forestry and logging	0.248
624	Social assistance	0.462	112	Animal production	0.245
221	Utilities	0.456	445	Food and beverage stores	0.243
213	Mining support	0.445	722	Food service and drinking	0.240
454	Nonstore retailers	0.429	562	Waste management	0.226
312	Beverage/Tobacco manufacturing	0.428	313	Textile mills	0.226
336	Transportation equip manufacturing	0.422	321	Wood product manufacturing	0.226
			493	Wharehousing and Storage	0.226
Low-skill sectors			337	Furniture manufacturing	0.225
N/A[a]	Commercial leasing	0.421	N/A[a]	Unspecified metal manufacturing	0.224
487	Scenic transportation	0.420	447	Gasoline stations	0.224
483	Water transportation	0.419	484	Truck transportation	0.224
421	Wholesale, durables	0.418	811	Repair and maintenance	0.220
N/A[a]	Camera, sports, toy stores	0.411	315	Apparel manufacturing	0.213
339	Misc manufacturing	0.400	314	Textile mill products	0.207
491	Postal services	0.395	814	**Household services**	0.198
N/A[a]	Not specified retail	0.394	111	Crop production	0.197
492	Couriers and messengers	0.393	N/A[a]	Knitting mills	0.171
448	Clothing stores	0.386			

Data source: Public-use 2000 Census of Population. The subsample of workers employed in one of the 230 metropolitan areas used in our analysis is used to construct the table. Table ranks three-digit NAICS sectors on college equivalent share, defined as the sum of share of workers employed in each industry who report having completed at least four years of college plus one half of the share who report completing 1–3 years of college. This was computed using Census sampling weights. Industries were separated on this measure into "low-skill" "high skill" at the employment-weighted median. Put another way, roughly half of all workers are employed in the industries listed as "high-skill."

[a] *N/A:* These industry codes in the Census could not strictly be categorized into a three-digit NAICS industry, and were left as separate categories.

References

Acemoglu, Daron, Angrist, Joshua, 2000. How Large Are Human Capital Externalities? Evidence from Compulsory Schooling Laws. NBER Macroeconomics Annual, 15.

Acs, Zoltan, Catherine, Armington, Ting, Zhang, 2007. The Determinants of New-Firm Survival across Regional Economies. Discussion Papers on Entrepreneurship, Growth and Public Policy, #407. Max Planck Institute of Economics, Group for Entrepreneurship, Growth and Public Policy.

Andersson, Roland, Quigley, John M., Wilhelmsson, Mats, 2009. Urbanization, productivity and innovation: evidence from investment in higher education. Journal of Urban Economics 66, 2–15.

Arzaghi, Muhammad, Henderson, J.V., 2008. Networking off madison avenue. Review of Economic Studies 75, 1011–1038.

Astebro, Thomas, Bernhardt, Irwin, 2003. Start-up financing, owner characteristics and survival. Journal of Economics and Business 55 (4), 303–320.

Autor, David H., Katz, Lawrence F., Krueger, Alan B., 1998. Computing inequality: have computers changed the labor market? Quarterly Journal of Economics 113 (4), 1169–1213.

Bacolod, Marigee, Blum, Bernardo S., Strange, William C., 2009. Skills in the city. Journal of Urban Economics 65 (2), 136–153.

Barro, Robert J., 1991. Economic growth in a cross section of countries. Quarterly Journal of Economics 106 (2), 407–443.

Bartel, Ann, Lichtenberg, Frank, 1987. The comparative advantage of educated workers in implementing new technology: some empirical evidence. The Review of Economics and Statistics 69 (1), 1–11.

Bates, Timothy, 1990. Entrepreneur human capital inputs and small business longevity. The Review of Economics and Statistics 72 (4), 551–559.

Baumgardner, J.R., 1988. Physicians services and the division of labor across local markets. Journal of Political Economy 9 (5), 948–982.

Beaudry, Paul, Mark Doms, Ethan Lewis, 2006. The IT Revolution at the City Level: Testing a Model of Endogenous Biased Technology Adoption, mimeo.

Beaudy, Paul, Mark Doms, Ethan Lewis (in preparation). Endogenous industry location.

Bils, Mark, Klenow, Peter J., 2000. Does schooling cause growth? The American Economic Review 90 (5), 1160–1183.

Blau, David, 1987. A time series analysis of self-employment. Journal of Political Economy 95, 445–467.

Card, David, 1999. The causal effect of education on earnings. In: Ashenfelter, Orley, Card, David (Eds.), Handbook of Labor Economics, vol. 3A. Elsevier, Amsterdam, pp. 1801–1863.

Card, David, DiNardo, John E., 2002. Skill-biased technological change and rising wage inequality: some problems and puzzles. Journal of Labor Economics 20 (4), 733–783.

Cascio, Elizabeth, Clark, Damon, Godon, Nora, 2008. Education and the age profile of literacy into adulthood. Journal of Economic Perspectives 22 (3), 47–70.

Caselli, Francesco, Coleman, John, 2001. Cross-country technology diffusion: the case of computers. American Economic Review: Papers & Proceedings May.

Combes, Pierre-Philippe, Laurent Gobillon, Sébastien Roux, 2008. Estimating agglomeration economies with history, geology, and worker effects. CEPR Discussion paper no. 6728, February 2008.

Doms, Mark E., Lewis, Ethan G. 2006. Labor supply and personal computer adoption. FRB of Philadelphia working paper no. 06–10.

Duranton, Giles, Puga, Diego, 2004. Micro-foundations of urban agglomeration economies. In: Vernon Henderson, J., Thisse, Jacques-Francois (Eds.), Handbook of Urban and Regional and Urban Economics, vol. 4. Elsevier, North Holland, pp. 2063–2117.

Fairlie, Robert, Robb, Alicia, 2008. Race, families and business success: African-American, Asian-, and white-owned businesses. MIT Press, Cambridge.

Fallick, Bruce, Fleischmann, Charles, Rebitzer, James, 2006. Job hopping in silicon valley: some evidence concerning the micro-foundations of a high technology cluster. Review of Economics and Statistics 88 (3), 472–482.

Freedman, Matthew L., 2008. Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry. Journal of Urban Economics 64 (3), 590–600.
Fu, Shihe and Stephen Ross, 2007. Wage premia in employment clusters: agglomeration economies or worker heterogeneity. University of Connecticut working paper 2007-26R, June 2007.
Glaeser, Edward L. 2007. Entrepreneurship and the city. NBER working paper no. 13551.
Glaeser, Edward L., Kerr, William R., 2006. Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?, vol 15. Harvard University, Mimeo (November).
Glaeser, E.L., Mare, D.C., 2001. Cities and skills. Journal of Labor Economics 19 (2), 316–342.
Glaeser, Edward L., Albert Saiz, 2003. The Rise of the Skilled City. NBER working paper 10191.
Jacobs, Jane, 1969. The economy of cities. Vintage Books, New York.
Katz, Lawrence F., Murphy, Kevin M., 1992. Changes in relative wages, 1963–1987: supply and demand factors. Quarterly Journal of Economics 112, 291–303.
Lazear, Edward P., 2004. Balanced skills and entrepreneurship. American Economic Review Papers and Proceedings 94 (2), 208–211.
Lazear, Edward P., 2005. Entrepreneurship. Journal of Labor Economics 23 (4), 649–680.
Lemieux, Thomas, 2006. The mincer equation thirty years after schooling, experience, and earnings. In: Grossbard-Shechtman, S. (Ed.), Jacob Mincer, A Pioneer of Modern Labor Economics. Springer-Verlag.
Lewis, Ethan, 2004. How do local labor markets in the US adjust to immigration? Federal Reserve Bank of Philadelphia working paper, #04-01.
Lin, Jeffrey, 2009. Technological adaptation, Cities, and New Work. Mimeo, Federal Reserve Bank of Philadelphia working paper, July 2009.
Moretti, Enrico, 2004a. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. Journal of Econometrics 121 (1–2), 175–212.
Moretti, Enrico, 2004b. Human capital externalities in cities. In: Vernon Henderson, J., Thisse, Jacques-Francois (Eds.), Handbook of Urban and Regional and Urban Economics. Elsevier, North Holland.
Nardinelli, Clark, Simon, Cuirtis J., 2002. Human capital and the rise of American cities. Regional Science and Urban Economics 32, 59–96.
Rauch, James, 1993. Productivity gains from geographic concentration of human capital: evidence from the cities. Journal of Urban Economics 34 (3), 380–400.
Roback, Jennifer, 1982. Wages, rents and the quality of life. Journal of Political Economy 90 (6), 1257–1278.
Robb, A., Ballou, J., Barton, T., DesRoches, D., Potter, F., Reedy, E.J., Zhao, Z., 2009. An overview of the Kauffman Firm Survey: results from the 2004–2007 data. Kauffman Foundation.
Rosenthal, Stuart S., Strange, William C., 2001. The determinants of agglomeration. Journal of Urban Economics 50 (2), 191–229.
Rosenthal, Stuart S., Strange, William C., 2004. Evidence on the nature and sources of agglomeration economies. In: Vernon Henderson, J., Thisse, Jacques-Francois (Eds.), Handbook of Urban and Regional and Urban Economics. Elsevier, North Holland, pp. 2119–2171.
Rosenthal, Stuart S., Strange, William C., 2008. The attenuation of human capital spillovers. Journal of Urban Economics 64 (2), 373–389.
Steven Ruggles, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, Chad Ronnander, 2008. Integrated Public Use Microdata Series: Version 4.0 [machine-readable database]. Minnesota Population Center [producer and distributor], Minneapolis, MN.
Schultz, T.W., 1964. Transforming Traditional Agriculture. Yale University Press, New Haven.
Schultz, T.W., 1975. The value of the ability to deal with disequilibria. Journal of Economic Literature 13 (3), 827–846.
Staiger, Douglass, Jonathan Skinner, 2005. Technology Adoption From Hybrid Corn to Beta Blockers. NBER working paper #11251.
Storey, D.J., 1991. The birth of new firms – does unemployment matter? A review of the evidence. Small Business Economics 3, 167–178.
Van der Sluis, J., Van Praag, M., Vijverberg, W., 2008. Education and entrepreneurship selection and performance: a review of the empirical literature. Journal of Economic Surveys 22 (5), 795–841.
Van Praag, C.M., Versloot, Peter H., 2007. What is the value of entrepreneurship? A review of recent research. Small Business Economics 29 (4), 351–382.
Van Stel, A.J., 2005. COMPENDIA: harmonizing business ownership data across countries and over time. International Entrepreneurship and Management Journal 11, 105–123.
Wheeler, Christopher, 2001. Search, sorting, and urban agglomeration. Journal of Labor Economics 19, 879–899.
Wheeler, Christopher, 2006. Cities and the growth of wages among young workers: evidence from the NLSY. Journal of Urban Economics 60 (2), 162–184.
Wozniak, Abigail., 2006. Why Are College Graduates More Responsive to Distant Labor Market Opportunities? University of Notre Dame, Mimeo.