Entrepreneurial Reluctance: Academic Talent and Firm Creation in China

Chong-en Bai (Tsinghua)
Ruixue Jia (UCSD)
Hongbin Li (Stanford)
Xin Wang (CUHK)

Date: February 18, 2020

Abstract

Are academic talents are more or less likely to establish firms and why? Linking administrative data on college admission with information on the universe of firms in China, we investigate this question by examining who had established firms by their mid-30s as well as the success of the firms. We document that within colleges, individuals with higher college entrance exam scores are *less likely* to create firms. We allow for different interpretations of this fact in a conceptual framework and test these interpretations by exploiting data on firm success, personal background and business cycle. We find that (i) firms created by higher-score students are larger and more successful, (ii) the negative score-firm creation link is stronger for male (vs. female), for urban (vs. rural), and for those from better high schools, which is mirrored by a stronger positive score-wage link for these more advantaged groups in survey data, and (iii) the negative score-firm creation link is also stronger when one's home province experienced better economic growth in one's graduation year. These empirical patterns are most consistent with the hypothesis that scores are positively correlated with entrepreneurial ability but higher-score students are attracted away by non-entrepreneur sectors.

1 Introduction

Entrepreneurship has long been regarded as one of the most important input in economic development. The talent of entrepreneurs thus has important consequences on firm productivity and economic growth of a society (e.g., Bertrand and Schoar 2003; Bloom and Van Reenen 2007; Syverson 2011; Gennaioli et al. 2013). Are talented individuals in a society more likely to become entrepreneurs? There seem to exist two opposite answers to this question. On the one hand, a literature on the determinants of entrepreneurship argues that talent facilitates entrepreneurship especially in terms of incorporated business (as opposed to self-employment) while other traits like risk preferences also play a role (e.g., Levine and Rubinstein 2017, 2019); on the other hand, a different literature on talent allocation emphasizes that talents are often attracted by non-entrepreneurial sectors in many societies which can lead to misallocation of human capital (e.g., Murphy, Shleifer and Vishny 1991; Baumol, Litan and Schramm 2007).

In this paper, we investigate the link between talent and firm creation using administrative data in China. We study a specific and arguably the most important measure of academic talent in the Chinese society: one's score in the National College Entrance Exam (known as *"Gaokao"* or the highest exam) – a human capital measure tens of millions of Chinese families have invested massive resources in and is believed to determine the course of one's life. Specifically, we link college admission records in 1999-2003 with the universe of Chinese firms and their owners to examine who had created firms by 2015 and how their firms perform. Such administration data allows us to paint a comprehensive picture for both small and very large firms. Since one's exam score shapes one's college outcome, we focus on comparing individuals graduating from the same college, and present additional results unconditional on colleges or further considering majors.

Using a 20% random sample of the linked administrative data, we show that within colleges, higher-score students are less likely to create firms: a one standard deviation higher scores is associated with a 10% lower probability of creating a firm. This finding holds regardless of firm size, including restricting the definition of firms to the very large ones. We further investigate the within-college patterns after dealing with missing information in linking administrative datasets, excluding repeated exam takers, re-examining it by college ranks, and making comparisons within a college-major, all of which reveal a strongly negative link between scores and firm creation. In addition, we find that women are half as likely as men to create firms; rural individuals and those from lower-quality high schools are also less likely to create firms than their counterparts.

What explains the negative link between academic ability and firm creation? We formalize

three important explanations in a simple conceptual framework built on earlier work (e.g., Evans and Jovanovic 1989; Levine and Rubinstein 2019). Specific to our context, we allow for the mapping from exam scores to entrepreneurial ability to differ from the mapping from exam scores to wage ability. We also allow personal preferences like risk attitudes to play a role in occupation choice. The first explanation is the *"lower entrepreneurial ability"* hypothesis, which asserts a negative correlation between exam scores and entrepreneurial ability. For instance, one can assume that those with higher scores are the opposite type of Jack-of-all-trades (e.g., Lazear 2004) or that the exam-focused Chinese education system deters entrepreneurship. The second is the *"ability informativeness"* hypothesis, which assumes that scores are positively correlated with entrepreneurial ability and wage ability but is more informative for the latter. This hypothesis also includes the case of no significant correlation between score entrepreneurial ability. The third is the *"relative returns"* hypothesis, which also assumes that scores are positively correlated with entrepreneurial and wage abilities, but emphasizes that the relative returns to scores are higher for waged jobs. Based on this simple framework, we test the implications of different hypotheses. Four sets of analysis help us disentangle these interpretations.

First, the lower entrepreneurial ability hypothesis implies that firms created by higher-score individuals are smaller and less successful, a prediction we can test by examining how scores matter firm success. However, we face a challenge that estimating the score-firm success link conditional on entry of firms can be biased due to endogenous firm entry. To address this challenge, we employ the Heckman two-step correction (Heckman 1976), where we argue that conditional on colleges, the propensity of firm creation by peers within the same college-cohort provides a reasonable instrument to predict one's own firm creation. Conceptually, higher-score students may select into colleges with higher or lower propensity to create firms. But conditional on colleges, it is difficult for students to predict the firm creation propensity for a specific cohort. This is supported by data: within a college, firms per capita among peers in the same cohort is uncorrelated with one's own score but is a strong predictor for one's own firm creation. Employing the Heckman two-step method and estimating firm outcomes conditional on entry deliver similar patterns: firms created by higher-score individuals (within a college) have larger registered capital in 2015, are more likely to break provincial protection barrier, more likely to expand, and more likely to become publicly listed (a rare event indicating extreme success). We further employ firms per capita created by college-cohort peers from the same home province and that by college-cohort peers from other provinces as two instruments to illustrate that exclusion restriction is not likely to be a critical concern for the Heckman method. Taken together, these patterns all suggest a positive correlation between exam scores and entrepreneurial

ability.

While the “ability informativeness" hypothesis and the “relative returns” hypothesis can work together, they have different implications regarding personal background. In particular, we find strong heterogeneity in the negative link between academic ability and firm creation: the negative link is significantly stronger for the male (vs. the female), the urban (vs. the rural), and those from better high schools. Unless one assumes that higher scores are more informative for wage ability and even more so for the male, the urban, and those from better high schools, one cannot explain the patterns across personal background.

Moreover, we provide two additional pieces of evidence for the relative returns interpretation. Using College Graduate Survey data on the first job outcomes of college graduates we conducted during 2010-2015, we document a positive link between score (within colleges) and first-job wages. Moreover, this positive score-wage link is also stronger for the male, the urban, and those from better high schools, which mirrors the pattern in firm creation. In addition, we exploit business cycle variation in the year one graduated and find that the negative score-firm creation link is also stronger when one’s home province experienced better economic growth in one’s graduation year, which is likely due to the fact that better economic growth makes it even easier for better students to land a waged job. These findings are related to a literature documenting the importance of macroeconomic shocks at the time of college graduation for the outcomes in the long term (e.g., Oreopoulos, Von Wachter, and Heisz 2012; Shu 2016). Our focus is not on the level effect of macroeconomic shocks on entrepreneurship that has been discussed by a literature on the cyclicality of entrepreneurship (e.g., Shleifer 1986; Caballero and Hammour 1994), but how such shocks alter the negative link between academic talent and firm creation. We find that better economic growth actually amplifies the negative talent-firm creation link.[1]

Finally, we discuss the role of personal preferences (such as risk attitude and over-confidence) that have been studied by existing literature in economics and management.[2] Even though we cannot specify which preference matters, our findings imply that such preferences play a role for the level differences in entrepreneurship across gender and other personal characteristics. For instance, we find that women have a significantly lower chance of firm creation, but this is not due to their higher wages. In fact, given the same score, they also have lower first-job wages. Thus, the difference can

[1] One interpretation is that wage premium to higher academic ability increases with economic growth. This interpretation is related to a literature on how skill premium changes with business cycles (e.g, Katz and Murphy 1992; Autor, Katz and Kearney 2008). However, we are not aware of systematic studies on how wage premium changes with business cycles in China. We let the data tell us whether the negative talent-firm creation link varies with business cycle.

[2] Among many others, Iyigun and Owen (1998) and Hayward, Shepherd and Griffin (2006) are examples in this line of research.

be due to personal tastes or liquidity constraints that shape one's preference in entrepreneurship choice.[3] However, unless one assumes that such preferences change systemically with scores and more so for certain gender (and other personal characteristics), one cannot explain our main findings on the academic talent-firm creation link and how this link varies by personal background and business cycle. As additional suggestive evidence, we show that risk attitude measures in our College Graduate Survey data confirm that women are more risk averse than men; in contrast, these risk measures appear not significantly correlated with *Gaokao* scores.

Overall, our findings imply that academic talent measured by *Gaokao* scores – the most important human capital measure in the Chinese education system – is positively associated with both entrepreneurial ability and wage ability. When it comes to talent allocation across occupations, we find more evidence for the argument that talents are attracted away by non-entrepreneurial sectors (e.g., Murphy, Shleifer and Vishny 1991; Baumol, Litan and Schramm 2007) than evidence for a simplified positive association between academic talent and entrepreneurship.

Section 2 describes the background and data. Section 3 presents the main finding on the native link between academic talent and firm creation within colleges, and Section 4 presents a conceptual framework that accommodates different interpretations for the negative link. Finally, Section 5 brings the implications of different interpretations to data, and Section 6 concludes.

2 Background and Data

2.1 College Entrance Exam and College Admission

All 2000 or so colleges in China admit students via a centralized college entrance exam system (i.e., *Gaokao*). *Gaokao* scores determine which college, if any, students will attend. The exam is written and graded by provincial education authorities. High school students were divided into natural sciences and social sciences tracks in their first or second year and took different subjects in the exam. Thus, *Gaokao* scores are only comparable within a give province-year-track (social or natural sciences). Therefore, we always control for province-year-track fixed effects when comparing the scores in our analysis.

We employ data covering all high school graduates participating in the National College Entrance Exam during 1999-2003 and their college admission outcomes across 2,056 colleges. The

[3]There exists a large literature on how liquidity constraints matter for entrepreneurship (e.g., Blanchflower and Oswald 1998).

median birth year for these cohorts is 1982, i.e., the median age is 33 in the year of 2015 (the final year of the firm registration records we examine).

This data provides detailed information on their performance in the exam including total scores and scores by subject. It also provides information on their colleges (if admitted) and majors. In addition, it includes biographical information of individuals like gender, *Hukou* (i.e., rural or urban identity), birth year, and their high schools, which we use to measure one's personal background.

It should be noted that the college dropout rate is very low in the Chinese context. In fact, the college education system in China is also known for being "strict entrance, easy out". The overall graduation rate for Chinese colleges is above 95%.[4] Thus, we do not study dropbox in this context.

2.2 Firm Registration and Updates

We are interested in firm creation and firm performance. As mentioned above, a recent literature has pointed out the importance to differ small business from big firms. We are able to contribute to this literature by using information on the universe of firms that had been registered in China by February 2015. For small business, an individual may face a trade-off between staying self-employed (not studied in this paper) and registering a firm.[5] Such considerations matter little for firms that create jobs for other people, as they have to be registered. The data allows us to examine both small and large firms.

This data includes 28 million firms, 11 million of which had been de-registered by 2015. The administrative registration records provide information on the owners (known as shareholders) and basic firm information including industry, firm location and registered capital size. The owners can be individuals and firms. Based on individual information, we know whether a college graduate i created a firm or not. We also know if a firm created by i invested in another firm as an owner, which we call "expansion" in our latter analysis.

The firm registration data does not provide employment information. The variable that proxies the scale of the firm is the size of registered capital, which indicates the maximum liability a firm has and hence matters for doing business. After a firm got registered, the owners can choose to update their registered capital size. For instance, according to the registration data, Huawei Technologies Co., Ltd. updated its registered capital size 17 times between 1995 (its founding year) and 2015,

[4]The Beijing-based Mycos Institute estimated in 2011 that China's college dropout rate is 3%, while the Ministry of Education said that year that the rate is 0.75%.

[5]Putting aside that fact that certain industries require registration for doing business, an important benefit of registering a firm is limited liability. On the other hand, staying self-employed without registering a firm may have implications on tax evasion.

raising the registered capital over 2,500 folds – from RMB15.81 million to RMB39.9 billion.

In our analysis, we use registered capital size to indicate a firm's scale and success. For this purpose, we focus information on registered capital in 2015 throughout the paper.[6]

2.3 Data Linking, Sampling and Summary Statistics

We start from the college data admission data and link it with firm owner information based on unique and encrypted national identity number.[7] The original college admission data includes 14.2 million records (out of 23 million records including those who failed to go to any college) accepted by a certain college. We face two challenges in linking the data.

The first is missing or invalid identity number. Because identity number is not used for college admission, provinces exhibit different record keeping patterns that generally improve over time. Out of the 14.2 million records, 3.9 million did not report identity number and 0.8 million reported ids that are invalid. We focus on the records with valid id in our main analysis, which includes 9.5 million exam records for 9.1 million unique individuals – the extra records belong to the repeated takers (see the next challenge below). When examining the pattern of missing personal identity number, we find it varies greatly across province-years. For instance, in 1999, Beijing and Guangxi had missing rates (including invalid cases) of 0.7% and 0 whereas Shanghai and Yunnan had missing rates of 100% and 56%. To check whether missing identity numbers creates any concern for our findings, we conduct additional analyses by restricting our data to the province-years with missing probabilities lower than 5% or even 1%.

The second is the presence of repeated takers. Due to the high stake of the exam, repeated takers are not rare. For a exam taker, we use the latest exam record for him or her because it determined which college he or she finally entered. To check whether repeated takers are important for our findings, we also report the results after excluding repeated exam takers.

Together, this linked data provides information for 9.1 million college students for whom we know their firm creation outcomes. To further help de-identification, we randomly sampled 20% of the college students (i.e., 1.8 million individuals) in the linked administrative data for our analyses. By 2015, the 1.8 million college graduates had established 170,087 firms. Below, we describe the key variables and present their summary statistics in Table 1.

[6] If a firm exited before 2015, its registered capital size refers to the most recent information before exit, which we observe in 2015.

[7] While the 15 or 18-digit Chinese national identity number includes information on one's home county and birth dates, we cannot observe such information because of the encryption.

Firm Creation We vary the definition of a firm according to the registered capital size. As mentioned above, we use the information updated in 2015. 7.2% of the college graduates had created any firm, but only 1.6% of them created a firm of registered capital size no less than RMB 2 million. The probabilities becomes 0.78% and 0.4% when restricting registered capital size to no less than RMB 8 million and 15 million. In our analysis, we use these different definitions to check how our findings vary by firm size.

***Gaokao* Scores** As reported in Panel A of Table 1, the mean and standard deviation of the *Gaokao* scores are 443.8 and 101.5. But as explained above, the scores are only comparable within province-year-tracks. After isolating province-year-track fixed effects, the standard deviation of the *Gaokao* scores is 87.7.

Since our main interest is the within-college comparison, we also report the residualized scores after isolating college fixed effects: the standard deviation becomes 65 points. If we add 12 major fixed effects, the standard deviation changes to be 63.1 points. Appendix Figure A.1 plots the density distribution of these variables.

Personal Background As shown in Panel A of Table 1, 54% of the colleges students are male, and 53% have urban *Hukou*. While we do not observe family socioeconomic status, high school quality provides a reasonable proxy because richer families can afford more elite high schools. We measure high school quality by their college outcomes. Specifically, we calculate the share of students admitted by the top 100 colleges in China (i.e., those designated by the government as the Project-211 colleges – the top 100 colleges in the 21^{st} century) for each high school and use the share to measure high school quality. As reported in Table 1, the bottom 25% and top 25% high schools have 4.8% and 19.4% of their students admitted by the top 100 colleges, indicating wide variations across high schools.

Firm Outcomes We present firm-level variables in Panels B and C of Table 1. The median firm was established in 2010, or around six years after one graduated from colleges. The top five industries are wholesale and retail (30.9%), leasing and business services (20.2%), scientific research and technology services (14%), manufacturing (8.5%), information technology and services (6.8%). The other 15 industries together account for the remaining 20%.

The 25th and 75th percentiles of registered capital size are RMB 200,000 and 2 million. We will examine registered capital size (in 2015) as an outcome according to our conceptual framework. Registered capital size is likely to be a noisy measure of firm success. For instance, access to credit

can also vary by personal background. To address this concern, we consider three alternative proxies for firm success in addition to capital size.

The first proxy is whether the firm was registered out of one's home province. 39% of the firms were registered out of home's home province. The fact that the majority of firms were established within one's home province indicates that it is not easy to establish a firm beyond one's home province due to local protection.

The second is expansion. If a firm invests in another firm as an owner, we can observe such expansion behavior in the firm registration data. 5% of the firms are able to expand this way.

A third proxy is whether a firm became publicly listed. This is a rare event: only 3.6 per 10,000 firms became publicly listed in our sample. Thus, this measure can be considered as a proxy for extreme success.

2.4 Supplementary Data

College Graduates Survey 2010-15 We cannot link the data with any administrative data on wages. Instead, we employ large scale surveys of college graduates conducted in 2010-15, which cover around 14,600 students from 90 colleges who reported their first-job wages. We designed the surveys to evaluate elite college premium and intentionally asked their *Gaokao* scores (see Jia and Li (2019) for more description of these surveys where we also show that the first job has a strong persistent effect on future jobs).

These surveys cover differ student cohorts (who were admitted to colleges during 2006-2011), for a limited number of colleges, and only provide information on the first jobs. However, since our main interest is within-college comparison, we employ data from these surveys to provide information on the relationship between *Gaokao* scores and first-job wages within colleges.

Business Cycle in the Graduation Year To measure business cycle that matters for one's career choice, we consider GDP growth for one's home province and one's college province at the time one graduated from college. This data comes from statistical yearbooks. In our study period, the provincial GDP growth rates vary from 8.2% to 23.8%, with a mean of 13.4%.

Our main interest is not how business cycle affects the level of firm creation but how it affects the link between academic talent and firm creation.

3 Academic Talent and Firm Creation

3.1 Baseline Results

To examine the link between academic talent and firm creation, we employ the following specification at the individual level:

$$Firm_{i,pyt,c} = \beta Score_{i,pyt,c} + \alpha X_i + \lambda_{pyt} + \theta_c + \epsilon_{i,pyt,c}, \quad (1)$$

where $Firm_{i,pyt,c}$ a dummy variable indicates whether an individual i of province-year-track (pyt) in college (c) created a firm. We can vary the definition of a firm by its capital size: any firm, firms with registered capital no less than RMB 2 million, firms with registered capital no less than RMB 8 million, and firms with registered capital no less than RMB 15 million.

The key variable of our interest is the exam score ($Score_{i,pyt,c}$). In addition, X_i indicates one's personal characteristics that include gender, *Hukou* (rural vs. urban), high school quality and age fixed effects. While we do not have measures like parental income, it is reasonable to assume that male, urban, and those from better high schools enjoy more socioeconomic advantages than their counterpart. Since $Score_{i,pyt,c}$ is only comparable within province-year-track, we always control for province-year-track fixed effects (λ_{pyt}) in our analysis. We report standard errors are clustered at the college level.

We start examining the relationship between $Score_{i,pyt,c}$ and $Firm_{i,pyt,c}$ without considering colleges. In Appendix Figure A.2(a), we plot the correlations by various definitions of firms. As shown, the correlation is strongly negative when including all types of firms, becomes nonlinear and more positive as we restrict our definition to larger and larger firms. Because higher-score students go to better colleges, the patterns above mask both colleges and academic talent. As shown in Appendix Figure A.2(b), college selectivity (measured by median score of the admitted students) is generally positively associated with firm creation, especially for the large firms.

We are interested in β after controlling for college effects. To do so, we plot the patterns after isolating college fixed effects (θ_c). The patterns are clear (see Figure 1a): regardless of how we define firms, $Score_{i,pyt,c}$ is strongly and negatively associated with firm creation.[8] These strong and negative patterns hold if we further include major fixed effects (Figure 1b).

We present the estimation results in Table 2 where we report the coefficient for 1σ (standard deviation) difference in the exam score (i.e., 65 points in a college). Columns (1), (4) and (7) report the changing patterns along with firm size when not considering colleges. In contrast, columns (2),

[8]These negative patterns can be considered as the difference between Appendix Figure A.2(a) and Appendix Figure A.2(b). Controlling college fixed effects, however, is stricter than controlling for the median score of a college.

(5) and (8) show a consistent pattern within colleges: 1σ more in the exam score within a college decreases the probability of creating any firm by 0.77 percentage points or over 10% of the mean (7.12). The magnitudes are actually very similar if we define firms as those with registered capital no less than RMB 2 million or 15 million: 1σ more in the exam score within a college decreases the probabilities of creating these two type of firms by 0.17 and 0.04 percentage points, which are also around 10% of the mean (1.65 and 0.40). This finding remains similar if we control for 12-major fixed effects and personal characteristics (see columns (3), (6) and (9)).

In addition, gender difference is large: female is only half likely to create a firm than male; rural individuals and those from below-median high schools are also less likely to create firms than their counterparts, while the difference is smaller than gender difference. These findings suggest personal traits matter. For instance, female may be more risk averse or face discrimination in access to capital. We will allow for such differences in our conceptual framework.

3.2 Robustness Checks

Dealing with missing identity number We do not consider the exam records that miss personal identity number in our main analysis. To check whether this matters for our finding, we restrict our analysis to the province-years with missing probabilities lower than 5% and 1%. As reported in panels A and B of Appendix Table A.1, the patterns remain very similar to those in Table 2, suggesting that the missing in personal id is unlikely to be systematically correlated with the link between scores and firm creation.

Excluding repeated exam takers We include the college students who were repeated exam takers for completeness. If we exclude them from our analysis, again, we obtain estimates very similar to those from our baseline estimations (reported in Appendix Table A.2).

Patterns by colleges We wonder whether the the negative score-firm creation link differs greatly by college quality. We divide colleges into three groups: the top 10, those ranked 11-100, and the rest. As reported in Appendix Table A.3, the negative link between scores and firm creation hold for each group. The links also appear stronger within the top 100 than that within the rest. In our analysis in Section 5, we obtain similar patterns regarding high school quality. As discussed below, these patterns are consistent with differences in relative returns to scores for waged jobs.

Comparing students within college-major We focus on within-college comparison because major choice itself may reflect personal preferences in doing business. Nevertheless, the patterns also hold if we compare within a college-major by including college-by-major fixed effects (reported in Appendix Table A.4).

4 A Conceptual Framework

Here, we present a simple yet flexible conceptual framework built on earlier work (e.g., Evans and Jovanovic 1989; Levine and Rubinstein 2019). Specific to our context, we allow the mapping between exam scores to entrepreneurial ability to differ from the mapping from exam scores to wage ability. Our aim is to accommodate different interpretations of the negative link between scores and firm creation and discuss their testable implications. In Section 5, we confront each implication with data.

The first interpretation is that scores are negatively correlated with entrepreneurial ability, which we call the *"lower entrepreneurial ability"* hypothesis. The second is that scores are positively correlated with entrepreneurial ability and wage ability but provide a more precise signal for the latter. This hypothesis can also include the case where there is no significant correlation between scores and entrepreneurial ability. We term this interpretation the *"ability informativeness"* hypothesis. The third explanation is that scores are positively correlated with entrepreneurial and wage abilities, but the relative returns to scores are higher for waged jobs. We call this explanation the *"relative returns"* hypothesis.

In addition, we allow for personal preferences (such as risk attitudes) and unobserved individual-level idiosyncratic shocks to affect in occupation choice. For now, we do not assume that they are also systemically correlated with scores. Later, we will discuss whether they can explain our findings if they happen to be also correlated with scores.

Setup An individual i with score S_i and personal background X_i (e.g., gender and *Hukou*) in a college chooses to establish a firm or become a worker. We differ entrepreneurial ability ($A_{i,E}$) and wage ability ($A_{i,W}$) and allow for a flexible mapping between scores and these abilities using the following specifications:

$$A_{i,E} = e^{aS_i} \text{ and } A_{i,W} = e^{bS_i}, \tag{2}$$

where a can be positive or negative (or zero), reflecting different possible correlations between exam scores and entrepreneurial ability.

Similarly, we can allow b to have any sign. But for empirical relevance, we assume $b > 0$ (on which we later present evidence). If both b and a are positive, but $b > a$, then S_i is more informative for wage ability, and vice versa.

If individual i chooses to start a firm, he or she combines entrepreneurial ability ($A_{i,E}$) and physical capital K to produce Y:

$$Y_i = A_{i,E}^{\beta(X_i)} K^{\alpha},$$

where β captures the return to entrepreneurial ability, which is allowed to be affected by personal background (X_i) in a flexible way.

A firm owner's problem is to maximize the expected returns:

$$R_i = A_{i,E}^{\beta(X_i)} K^{\alpha} - rK_i.$$

Thus, we get the optimal amount of capital:

$$K_i^* = \left(\frac{\alpha A_{i,E}^{\beta(X_i)}}{r} \right)^{\frac{1}{1-\alpha}}. \tag{3}$$

Given that $A_{i,E} = e^{aS_i}$, we obtain the expected returns as follows:

$$\begin{aligned} \ln R_i^* &= \tfrac{1}{1-\alpha} \ln A_{i,E} \beta(X_i) - \tfrac{\alpha}{1-\alpha} \ln r + \ln \Gamma \\ &= \tfrac{1}{1-\alpha} a S_i \beta(X_i) - \tfrac{\alpha}{1-\alpha} \ln r + \ln \Gamma, \end{aligned} \tag{4}$$

where $\Gamma = \alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}$.

Instead, if individual i chooses to become a worker, his or her wage is a function of wage ability $A_{i,W}$ so that we have:

$$W_i = A_{i,W}^{\mu(X_i)},$$

where μ captures the return to wage ability, which is also allowed to be affected by personal background (X_i) in a flexible way.

Similar to existing literature, we capture personal preferences in occupation choice in a reduced-form way by assuming his or her utility includes an element $\delta(X_i)$ so that the utility of being a worker can be written as follows:

$$V_i = A_{i,W}^{\mu(X_i)} e^{\delta(X_i)}.$$

Thus, we have the expected returns from being waged worker:

$$\begin{aligned} \ln V_i &= \mu(X_i) \ln A_{i,W} + \delta(X_i) \\ &= b S_i \mu(X_i) + \delta(X_i) \end{aligned} \tag{5}$$

To be clear, the variation in $\delta(X_i)$ can stem from multiple sources. $\delta(X_i)$ can capture behavioral

factors: women may be more risk averse than men and hence are less likely to create firms. Underlying $\delta(X_i)$, however, can also be economic reasons. For instance, it women may find it more difficult to borrow money to start a firm and hence prefer to work as a waged employee. We do not attempt to enumerate each factor and consider them as a channel shaping an individual's occupation choice.

We also allow individual-level idiosyncratic shocks (ε_i) to affect occupation choice where ε_i follows a distribution of c.d.f $\mathbb{G}(\varepsilon)$.

An individual will choose to create a firm iff $lnR_i^* + \varepsilon_i \geq lnV_i$, which gives the following condition:

$$\frac{1}{1-\alpha} aS_i\beta(X_i) - \frac{\alpha}{1-\alpha} \ln r + \ln \Gamma + \varepsilon_i \geq bS_i\mu(X_i) + \delta(X_i). \tag{6}$$

Thus, the probability of creating a firm becomes:

$$\mathbb{P}(Firm_i) = \mathbb{G}\left\{ \left[\frac{1}{1-\alpha} a\beta(X_i) - b\mu(X_i) \right] S_i - \delta(X_i) - \frac{\alpha}{1-\alpha} \ln r + \ln \Gamma \right\} \tag{7}$$

From equation (7), we can see the negative link between scores and firm creation could be due to three different channels summarized by $\frac{1}{1-\alpha} a\beta(X_i) - b\mu(X_i)$:

(1) the *"lower entrepreneurial ability"* hypothesis assumes that $a < 0$;
(2) the *"ability informativeness"* hypothesis refers to $a \geq 0$, $b > 0$, and $b > \frac{1}{1-\alpha} a$;
(3) the *"relative returns"* hypothesis refers to $a \geq 0$, $b > 0$, and $\mu(X_i) > \beta(X_i)$.

In this simple framework, personal preferences $\delta(X_i)$ (e.g., risk preferences) cannot explain the roles of scores S_i. Naturally, they can play a role if one assume that they vary systematically with S_i. We return to these possibilities in Section 5.

Testable Implications This very stylized model is useful as it captures different interpretations while delivering testable implications.

First, to evaluate the sign of a, we can rely on the implication of equation (3), which tells us the relationship between K_i^* and scores:

$$\ln K_i^* = \frac{1}{1-\alpha} \left[aS_i\beta(X_i) + \ln \alpha - \ln r \right]. \tag{8}$$

Thus, if $a < 0$ ($a \geq 0$), we expect to see a higher score is negatively (weakly positively) correlated with firm capital size. We can further extend this argument by exploiting other measures on the success of the firms. In such investigations, we will address the challenge of estimation based on entry by exploiting an instrument to predict firm entry.

Second, the *"ability informativeness"* hypothesis ($b > \frac{1}{1-\alpha}a$) and the *"relative returns"* hypothesis ($\mu(X_i) > \beta(X_i)$) can work together. Nevertheless, it is useful to know which channel dominates. If the *"ability informativeness"* hypothesis dominates, we expect to see little heterogeneity with respect to personal background (X_i) such as gender unless we assume that exam scores are not only a better signal for wage ability but also more precise for male (or female), which seems to be a strong assumption.

In contrast, if the *"relative returns"* hypothesis dominates, we expect to see heterogeneity with respect to personal background if relative returns vary by gender and other personal characteristics X_i. This hypothesis further predicts patterns on relative returns: for instance, if we find that the negative link between scores and firm creation is stronger for X_i, it implies that the wage returns to scores should be higher for X_i. Next, we test these implications.

5 Testing Different Interpretations

5.1 Scores and Firm Success

Estimation Strategy To investigate the sign of a, one may start with estimation equation (8) using the following specification:

$$\ln K_{f,i,pyt,c} = \gamma Score_{f,i,pyt,c} + \alpha X_i + \lambda_{pyt} + \theta_c + \epsilon_{i,pyt,c}, \tag{9}$$

where $K_{f,i,pyt,c}$ refers to registered capital of a firm f created by individual i. Recall that firms can update their registered capital and we use the most updated amount observed in 2015 as we are interested in firm success.

However, we are concerned whether entry is affected by unobservables that lead to bias in estimations conditional on entry. For instance, we assumed that individual-level shocks in equation (6) are not correlated with academic talent in our conceptual framework but we cannot observe these shocks. If they very systematically with academic ability, the estimation of equation (9) is biased. To deal with this challenge, we employ a Heckman two-step estimation by predicting entry (Heckman 1976). In particular, we add a stage of selection into entry before estimating the impact on firm outcomes:

$$Pr(Firm = 1|Z_{i,pyt,c}) = \Phi(Z_{i,pyt,c}) \tag{10}$$

where $Z_{i,pyt,c}$ includes X_i, λ_{pyt}, θ_c in equation (9) and an instrument variable.

We argue that the propensity of firm creation by peers within the same college-cohort (indicated by $\text{peer}_{c,t}$) provides a reasonable instrument to predict firm entry. Intuitively, higher-score students

may select into colleges with higher or lower propensity to create firms. But within colleges, it is difficult for students to predict the firm creation propensity for a specific cohort. This is indeed the case: in Table 3, we show that across colleges, one's scores is positively correlated with firms per capita among college peers (peer$_c$ in column (1)); however, within colleges, the relationship between one's score and firms per capita among peers in the same cohort (peer$_{c,t}$) is insignificant and close to zero (columns (2)-(3)). Meanwhile, within a college, $peer_{c,t}$ is a strong predictor for observing firm outcomes (column (4)). In terms of marginal effect, 1σ increase in $peer_{c,t}$ (0.05) is associated with a 1.5 percentage points higher probability of firm creation, or 16% of the mean.

For $peer_{c,t}$ (conditional on college c) to be a valid instrument, we also need it not to affect firm access beyond then entry channel. Below, we present the results using the Heckman method and the estimates based on entry. Then, we discussion the exclusion restriction of using $peer_{c,t}$ (conditional on college c) as an instrument.

Results Column (5) of Table 3 presents the within-college link between exam scores and log registered capital. It is more reasonable to compare capital within industries, and we find a stronger pattern if we employ demeaned log capital size by industry mean as the outcome (columns (6)). In column (7), we report the estimate when using the level of registered capital size (demeaned by industry mean) as the outcome, which says that 1σ within-college increase in scores is associated with RMB 220,000 more in registered capital within an industry. These coefficients are not large (around 1-3% of the mean) but are all positive. As discussed above, registered capital size is likely to be a noisy measure of firm success because access to capital can vary greatly across individuals.

We further measure firm success using three other proxies in columns (8)-(10). The first is based on the location of the firm. In particular, if the firm is founded out of one's home province indicates that the firm is able to overcome typical barriers in local protection, which serves as a possible success measure. The second is expansion: if a firm invests in another firm as an owner, we are able to observe this behavior in the firm data. Finally, becoming a publicly listed firm is a rare event that indicates more extreme success. Across all measures, we find positive and sizable correlations between exam scores and firm success: 1σ more in the exam score within a college increases the probabilities of investing out of home province and expansion by around 10% (i.e., 0.039/0.394 and 0.005/0.049). When it comes to the rare event of becoming listed, the estimate is less precise but the increase is even larger in magnitude, around 16% (0.057/0.359) of the mean.

As a comparison, we also report the estimates conditional on entry in panel B of Table 3 and visualize these patterns in Appendix Figure A.3. As shown, the estimates are close to the results

from the Heckman two-step method, suggesting that the selection bias is not quantitatively large.

Exclusion Restriction One challenge of using $peer_{c,t}$ (conditional on colleges) as an instrument is that peer firms per capita may affect firm success beyond the channel of entry, if peer networks also help a firm to succeed conditional on entry. To gauge the importance of this exclusion restriction concern, we further separate $peer_{c,t}$ into two groups: firms created by peers from the same home province ($peer_{c,t,homeprov}$) and firms created by peers from other provinces ($peer_{c,t,otherprov}$), normalized by the same number of college-college peers. Since it is more likely for an individual to benefit from $peer_{c,t,homeprov}$ than $peer_{c,t,otherprov}$, the difference in Heckman two-step estimates from using both instruments shed light on the importance of this concern.

As shown in columns (1)-(2) Appendix Table A.5, both $peer_{c,t,homeprov}$ and $peer_{c,t,otherprov}$ are strong predictors for entry but the former's effect is over three times the latter. As a result, we find that the Heckman two-step estimates on firm success are generally more precise when using $peer_{c,t,homeprov}$ as an instrument to predict entry. Nevertheless, the magnitudes of the estimates are very similar when using either instrument (columns (3)-(8)). These results suggest that the omitted channels through which the instruments affect success, if they exist, are not quantitatively large; otherwise, we would see a larger impact when using $peer_{c,t,homeprov}$ to predict entry.

Together, these findings imply that it is difficult for the *"lower entrepreneurial ability"* hypothesis to explain the link between scores and firm creation. To the contrary, the firms created by higher-score individuals are larger and more successful.

5.2 Ability Informativeness and Relative Returns

Score-Firm Creation Links Across Personal Background As a first check for the *"ability informativeness"* hypothesis and the *"relative returns"* hypothesis, we examine the score-firm creation link across gender, *Hukou*, and high school quality using the following specification:

$$Firm_{i,pyt,c} = \beta_1 Score_{i,pyt,c} + \beta_2 Score_{i,pyt,c} \times \mathrm{X}_i + \lambda_{pyt} + \theta_c \times \mathrm{X}_i + \epsilon_{i,pyt,c}, \tag{11}$$

where X_i indicates being female, rural or from below-median high schools. We also college effects to differ for these groups by including $\theta_c \times \mathrm{X}_i$ as controls.

Similar to Figure 1, Figures 2(a)-(c) plot the within-college correlations between scores and firm creation by gender, *Hukou* and high schools based on college enrollment outcomes. As shown, the negative score-firm creation link is stronger for the male (vs. female), for urban (vs. rural), and for those from better high schools (vs. the other high schools). For simplicity, we focus on firms

with registered capital no less than RMB 15 million in these figures and report similar patterns by firm size in Table 4.

The differences across personal background are large. For instance, the estimates in column (1) of Table 4 show that 1σ more in the exam score within a college decreases the probability of creating any firm by 0.96 for male but 0.58 for female, implying that the score-firm slope for women is 60% that for men. The magnitude differences are less striking between urban and rural individuals (0.8 vs. 0.7 in column (2)) but still sizable. The differences across high schools are also large: 1σ more in the exam score within a college decreases the probability of creating any firm by 0.98 for those graduated from above-median high schools but 0.66 for the others, implying that the score-firm slope for the latter is 66% that for the former. These differences are similar when we restrict firms to be the larger ones (columns (4)-(9)).

These results suggest that it is difficult for the *"ability informativeness"* hypothesis for account for the link between scores and firm creation unless one further assumes that scores are not only more informative for wage ability but even more informative for the male, the urban and those from better high schools.

Thus, we are left with the *"relative returns"* hypothesis. However, it is challenging to directly measure relative returns. Below, we present two sets of evidence that appear to support the relative returns interpretation.

First-job Wage Returns to Scores by Personal Background As mentioned above, we do not have administrative data on wages. Nevertheless, we have conducted a large scale college graduates survey where we obtain information on their first-job wages. While these first-job wages may underestimate life-time income, it is useful to investigate whether we observe heterogeneities with respect to gender, *Hukou* and high school quality that are consistent with (i.e., opposite to) the patterns in Figures 2(a)-(c).

This is indeed the case. We use the same specifications as equation (11) for our survey data and plot the correlation between log wage and exam scores in Figures 3(a)-(b). Limited by observations, the patterns are nosier than the patterns obtained from administrative data. Nevertheless, we observe a positive score and wage link (which supports that $b > 0$ in our conceptual framework) and stronger positive links for the male, the urban and those from more elite high schools. In our survey, we do not complete information to estimate high school quality according to college enrollment outcomes. Instead, we ask the students whether they graduated from a key high school designated by the government, which is an indicator of the quality of the high school.

We report the estimates in Table 5. Despite the limited size of survey data, the differences in the links between score and first-job wages are large across gender and high school types: the score-wage slope for women is 22% that for men and the slope for non-key high school graduates is 39% that for key high school graduates. The difference is less striking between urban and rural individuals but still sizable: the slope for the rural is 59% that for the urban.

Thus, the wage patterns across personal background provide mirror images for the score-firm creation links. It should be noted the level differences across gender, *Hukou* and high schools in Table 2 remain unexplained. For instance, given the same score, women are half likely to create a firm as men, but they also have lower first-job wages, implying that the choice of starting a firm is shaped by factors beyond outside option. We return to this point in terms of $\delta(X_i)$ in Section 5.3.

Business Cycles in the Graduation Year The findings on the first-job wages suggest that first jobs matter in shaping occupation choice. Inspired by related literature on wage premium in business cycles, we further test the role of relative returns by exploiting business cycle in the graduation year. In principle, better economic growth can make it easier for higher-score students to get better wage jobs and start business. If the former (latter) dominates, we should observe a stronger (weaker) negative link between scores and firm creation. In either case, the evidence reveals the impact of relative returns.

We first consider GDP growth for one's home province and one's college province and divide the provinces based on the median growth rate and plot the links between scores and firm creation in Figure 4(a). For simplicity, we define firms in the figure as those with registered capital no less than RMB 15 million (i.e., the large firms) and report results for other definitions in the table below. As shown, when one's home province experienced better growth in the year he or she graduates, the score-firm creation link is even stronger. The likely explanation is that better economic growth alters the wage premium to academic talent (more than its impact on firm creation).

Instead of two groups, we further divide the sample into 5 quantiles based on home province GDP growth at the year of graduation and estimate equation (1) five times. We plot the coefficients in Figure 4(b), which again shows that the coefficients generally become more negative for better growth.

We report the estimates in Table 6 while considering different definitions of firms, which confirms the pattern in Figure 4: the negative score-firm slope for below-median home province growth is around 70% that for above-median home province growth. When comparing home province growth and college province growth in a horse race, we find that home province growth

generally matters more, suggesting that getting a job in one's home province is still important even if one completes college education in a different province.

This finding provide indirect evidence for the importance of relative returns. It also reassures that first jobs after college graduation matter, since the business cycle in the graduation year is closely related to first job outcomes.

5.3 The Role of Personal Tastes or Barriers $\delta(X_i)$

Existing studies have shown that personal preferences like risk attitude matter for entrepreneurship. While we do not study a specific personal preference in this paper and use a reduced-form way to capture their role (i.e., $\delta(X_i)$ in the conceptual framework), our findings do suggest that they play a role in explaining level differences across personal background.

In particular, *ceteris paribus*, we find that the female, the rural and those from lower-quality high schools are less likely to create a firm in columns (3), (6), (9) Table 2. Note that these differences cannot be explained by wage differences. If anything, we find that the female, the rural and low-quality high schools earn lower first-job wages in columns (1), (3) and (7) of Table 5. This suggests that $\delta(X_i)$ plays an important role in explaining the level differences across personal background. For instance, female and rural individuals may be more risk averse or may face disadvantages in accessing to credit to start a firm.

However, for personal tastes to explain our key findings on academic talents, we need to assume that (i) higher-score students with a college are more risk averse and (ii) the score-risk link differs systematically by personal background, both of which seem to be a strong assumption. In the 2011 waves of our College Graduate Surveys, we asked three questions on risk attitudes: (1) whether one prefers to get RMB 1,000 with certainty or get 0 and 2,000 with equal chance; (2) whether one prefers to lose RMB 1,000 with certainty or lose 0 and 2,000 with equal chance; (3) whether one agrees with the statement that ensuring certain returns to investment is more important than taking more risks to gain higher returns. Using answers to these three questions to measure risk attitude, we find that women are indeed more risk averse (when measured by question (1) and question (3)), suggesting that these measures are meaningful. In contrast, we find no significant correlations between *Gaokao* scores and these risk aversion measures, suggesting it difficult to use risk preferences to explain our key findings. We present these results in Appendix Table A.6.

In sum, personal tastes or barriers do play a role in entrepreneurial choices. However, unless one assumes that there exists a systematic association between certain trait and scores and that the score-trait links vary by personal background, this explanation alone cannot account for our key

findings on the negative links between *Gaokao* scores and firm creation by gender, *Hukou*, high school quality, and business cycle.

6 Conclusion

We investigate the link between academic talent and firm creation in this paper. We document that within colleges, individuals with higher college entrance exam scores are *less likely* to create firms. However, this is not because higher-score individuals have lower entrepreneurial ability. To the contrary, firms created by higher-score individuals are larger and more successful. The opposite patterns for *ex ante* firm creation and *ex post* firm success are more consistent with the hypothesis that scores are positively correlated with entrepreneurial ability but higher-score students are attracted away by non-entrepreneur sectors. Our study provides a depiction of the choice of academic talent in China. More broadly, it contributes to a general literature on the determinants of entrepreneurship and a literature on talent allocation.

Admittedly, academic talent is only one factor in determining firm creation. In the process of understanding personal outcomes, we also find that the female, the rural and those studied in lower-quality high schools are less likely to create firms. Moreover, our findings suggest that the entrepreneurial reluctance of these groups are unlikely to be due to better outside options. We believe it will be useful to uncover the factors underlying entrepreneurial reluctance by gender and other personal characteristics in the future work.

References

[1] Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008), "Trends in US Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics,* 90(2): 300-323.

[2] Baumol, William J., Robert E. Litan, and Carl J. Schramm (2007), *Good Capitalism, Bad Capitalism, and the Economics of Growth and Prosperity*. Yale University Press.

[3] Bertrand, Marianne, and Antoinette Schoar (2003), "Managing with Style: The Effect of Managers on Firm Policies," *Quarterly Journal of Economics,* 118(4): 1169-1208.

[4] Blanchflower, David G., and Andrew J. Oswald (1998), "What Makes an Entrepreneur?" *Journal of Labor Economics,* 16(1): 26-60.

[5] Bloom, Nicholas, and John Van Reenen (2007), "Measuring and Explaining Management Practices across Firms and Countries," *Quarterly Journal of Economics,* 122(4): 1351-1408.

[6] Caballero, Ricardo and Mohamad Hammour (1994), "The Cleansing Effect of Recessions," *American Economic Review,* 84(5): 1350-1368.

[7] Evans, David S., and Boyan Jovanovic (1989), "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *Journal of Political Economy,* 97 (4): 808-827.

[8] Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2013), "Human Capital and Regional Development," *Quarterly Journal of Economics,* 128: 105-164.

[9] Hayward, Mathew LA, Dean A. Shepherd, and Dale Griffin (2006), "A Hubris Theory of Entrepreneurship," *Management Science,* 52(2): 160-172.

[10] Heckman, James (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement,* 5(4): 475?492.

[11] Iyigun, Murat F., and Ann L. Owen (1998), "Risk, Entrepreneurship, and Human-Capital Accumulation," *American Economic Review,* 88(2): 454-457.

[12] Jia, Ruixue and Hongbin Li (2019), "Just Above the Cutoff: The Return to Elite Education in China," Working Paper.

[13] Katz, Lawrence F., and Kevin M. Murphy (1992), "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics,* 107(1): 35-78.

[14] Lazear, Edward (2004), "Balanced Skills and Entrepreneurship," *American Economic Review,* 94(2): 208-211.

[15] Levine, Ross, and Yona Rubinstein (2017), "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?" *Quarterly Journal of Economics,* 132(2): 963-1018.

[16] Levine, Ross, and Yona Rubinstein (2019), "Selection into Entrepreneurship and Self-Employment" NBER Working Paper No. w25350.

[17] Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny (1991), "The Allocation of Talent: Implications for Growth," *Quarterly Journal of Economics,* 106(2): 503-530.

[18] Oreopoulos, Philip, Till Von Wachter, and Andrew Heisz (2012), "The Short-and Long-term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics,* 4(1): 1-29.

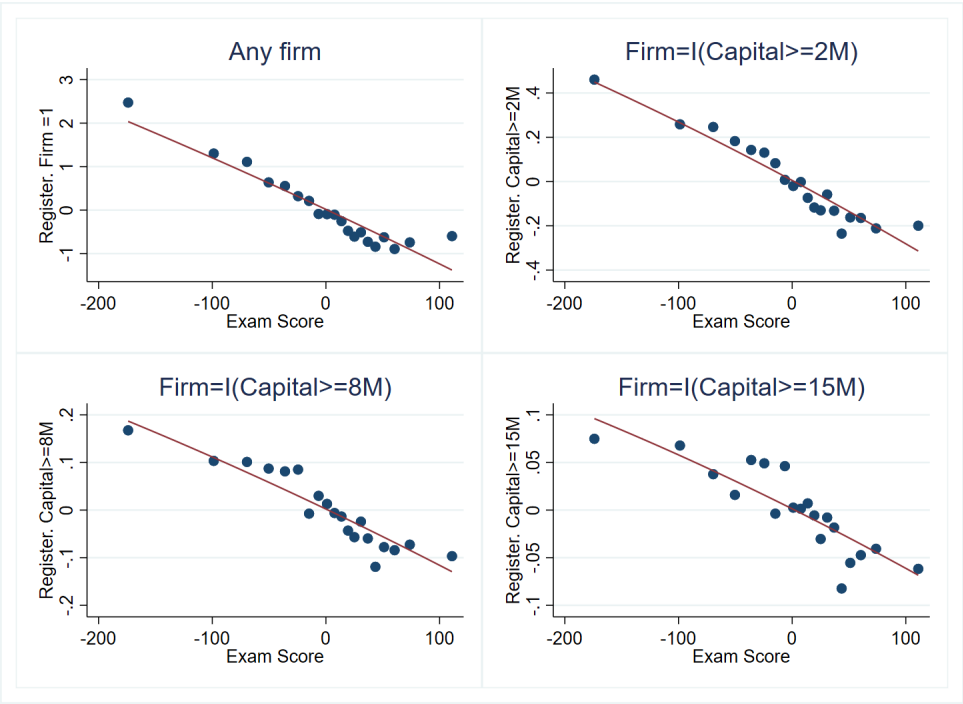
[19] Shleifer, Andrei (1986), "Implementation Cycles," *Journal of Political Economy,* 94(6)): 1163-1190.

[20] Shu, Pian (2016), "Innovating in Science and Engineering or 'Cashing In' on Wall Street? Evidence on Elite STEM Talent," Working Paper.

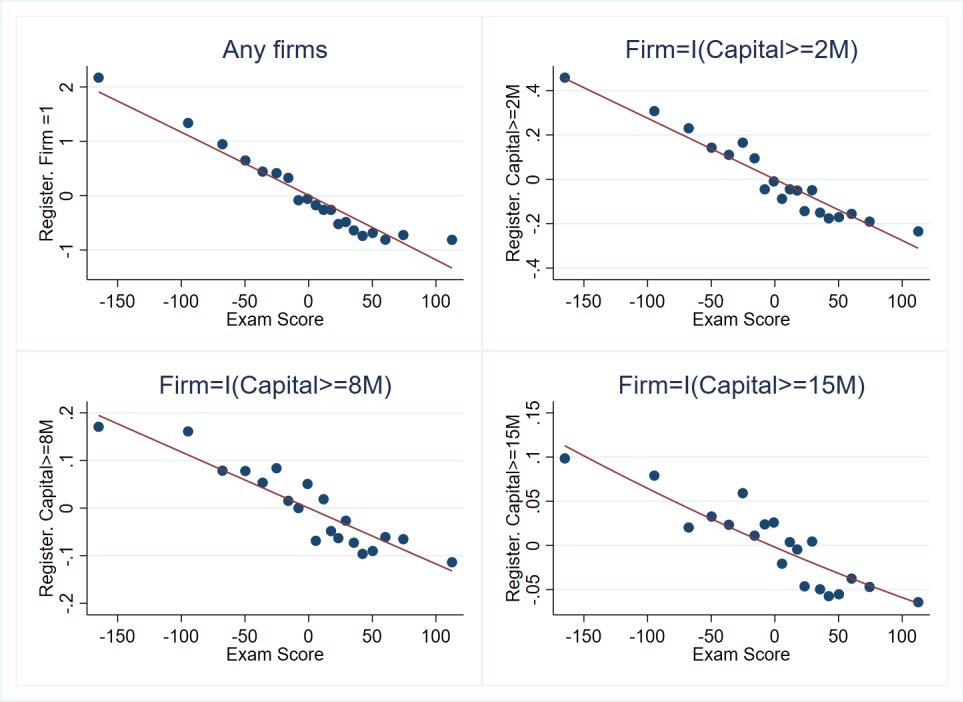
[21] Syverson Chad (2011), "What Determines Productivity?" *Journal of Economic Literature,* 49(2): 326-365.

Figure 1: Academic Talent and Firm Creation: Within-College Patterns

(a) College FEs



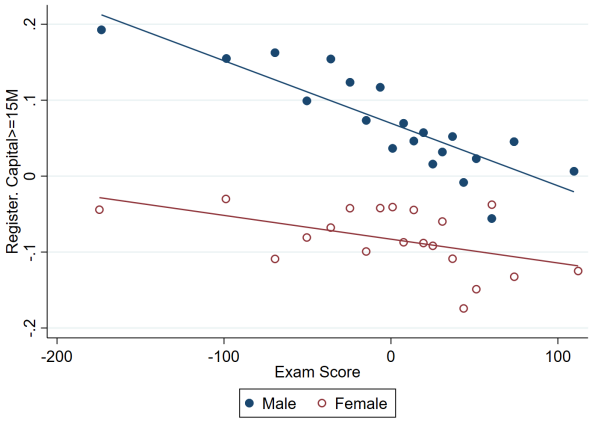
(b) College FEs + Major FEs



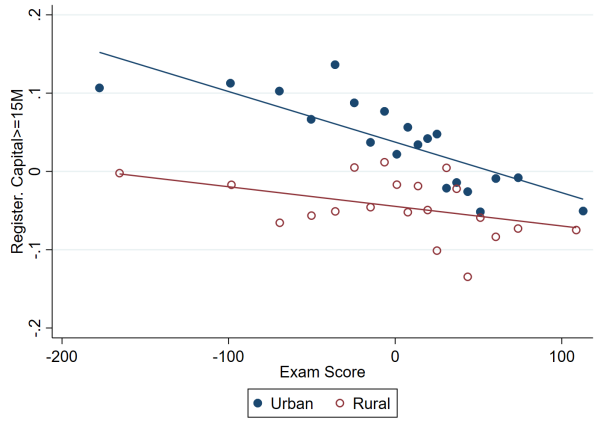
Notes: Firm=I(Capital≥2M/8M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/8 million/15 million. This figure shows that academic talent is negatively associated with firm creation within colleges. We regress exam scores and firm creation on province-year-track fixed effects and college fixed effects (and major fixed effects in (b)) and plot the correlations by bins.

Figure 2: Academic Talent and Firm Creation (Within Colleges), by Personal Background

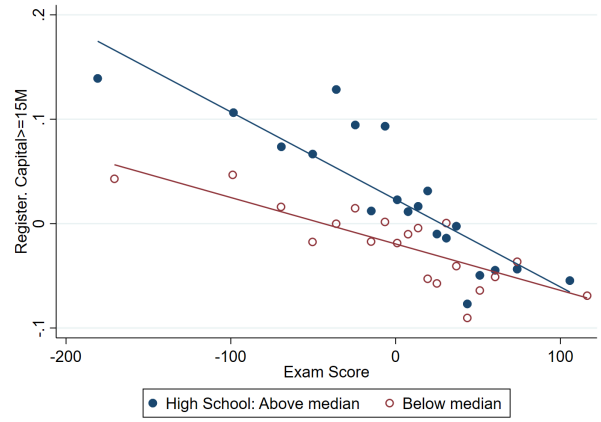
(a) Male vs. Female



(b) Urban vs. Rural



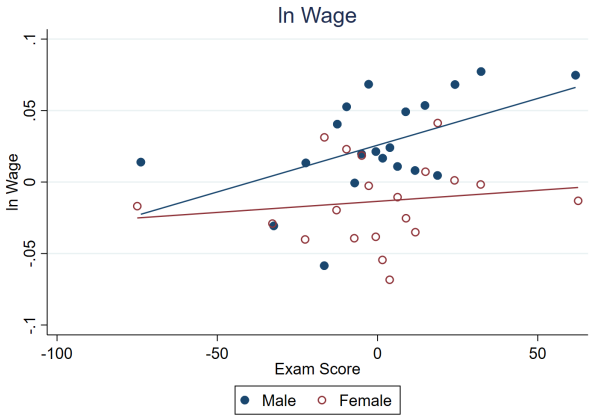
(c) Better vs. Worse High Schools



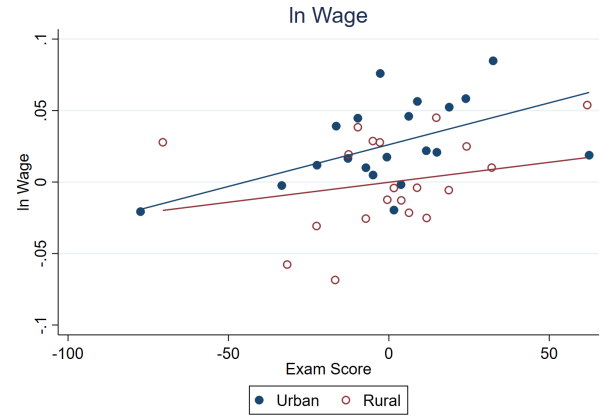
Notes: This figure shows that the negative score-firm creation link is stronger for male, urban individuals, and those from better high schools. We focus on large firms (those with registered capital no less than RMB 15 million) in these figures and show similar patterns when varying firm definition in Table 4.

Figure 3: Academic Talent and First-Job Wages (Within Colleges), by Personal Background

(a) Male vs. Female



(b) Urban vs. Rural



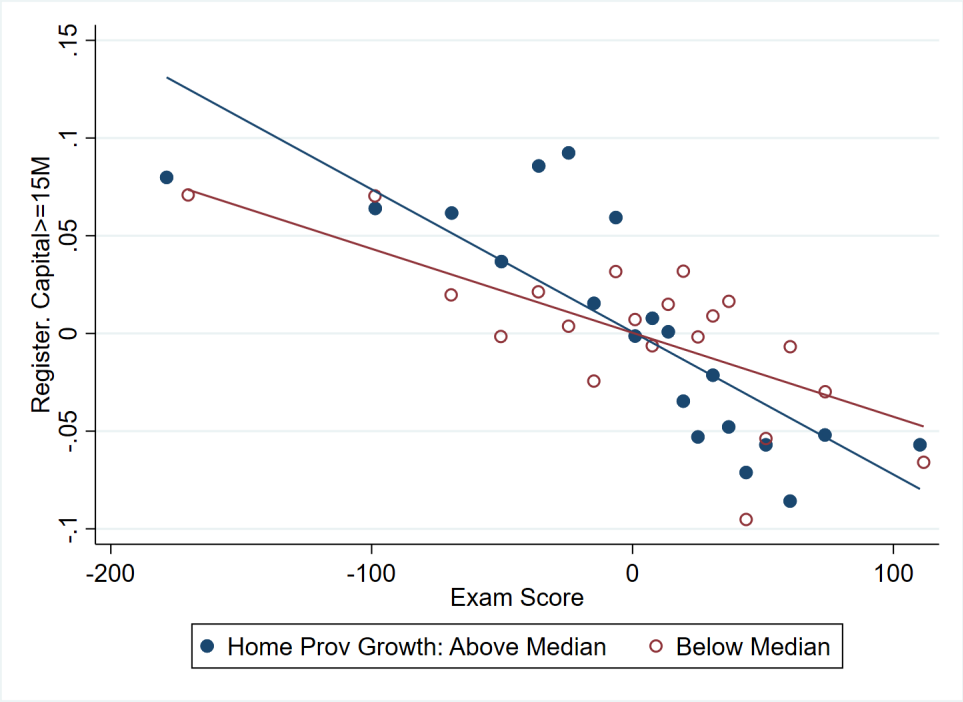
(c) Better vs. Worse High Schools



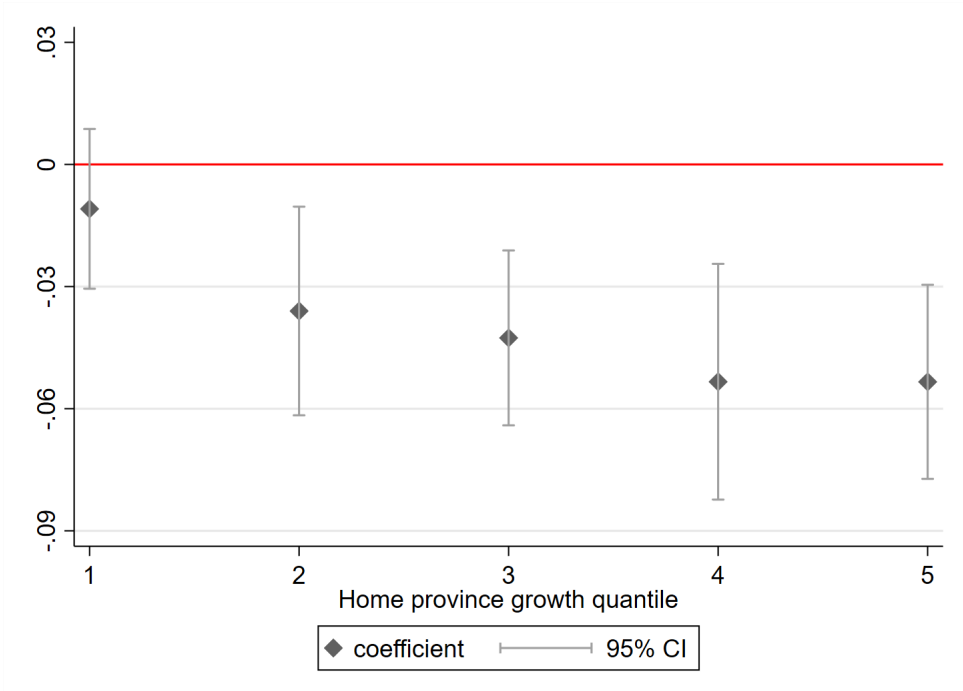
Notes: This figure shows that the positive score-wage link is stronger for male, urban individuals, and those from better high schools. The wage refers to the first job wages. This data comes from the College Student Survey we conducted during 2010-2015 (see Jia and Li 2019 for more discussion).

Figure 4: Academic Talent and Firm Creation (Within Colleges), by Business Cycle

(a) Above vs. Above Median



(b) Estimates by 5-Quantiles



Notes: This figure shows that the negative score-firm creation link is stronger when one's home province experienced better GDP growth in his or her graduation year. We focus on large firms (those with registered capital no less than RMB 15 million) in these figures and show similar patterns when varying firm definition in Table 6.

Table 1: Summary Statistics

A. Information on Individuals	Mean	Std Dev	p25	p75	Obs.
Prob of Creating a Firm (*100)					
Any Firm	7.21	25.87			1,814,547
Firms w. Capital>=RMB 2 Million	1.65	12.73			1,814,547
Firms w. Capital>=RMB 8 Million	0.78	8.82			1,814,547
Firms w. Capital>=RMB 15 Million	0.40	6.29			1,814,547
College Entrance Exam (*Gaokao*) Score	443.8	101.5	377	515	1,814,547
Gaokao Score (w. prov-yr-tr FEs)		87.7			
Gaokao Score (w. prov-yr-tr FEs+college FEs)		65.0			
Gaokao Score (w. prov-yr-tr FEs+college FEs + major FEs)		63.1			
Birth Year	1982.7	1.7	1982	1984	1,814,547
Male	0.54	0.50			1,814,547
Urban	0.53	0.50			1,814,547
High School (Sh. Stud. in Top 100 Colleges)	0.14	0.13	0.048	0.194	1,814,547
B. Information on Firms	Mean	Std Dev	p25	p75	Obs.
Registered Capital	632	6042	20	200	170,087
ln Registered Capital	4.28	1.74	3.04	5.30	170,087
Out of Home Province	0.39	0.49			170,087
Expanding (Investing in Other Firms)	0.05	0.22			170,087
Becoming Listed (*1000)	0.36	18.94			170,087
Establishment Year	2010.6	3.69	2009	2013	169,668
C. Firms by Industry	Percent				
Wholesale and Retail	30.91				
Leasing and business services	20.19				
Scientific research and tech services	13.96				
Manufacturing	8.54				
Info. Trans, software and i.t. services	6.84				
the other 15 industries	19.56				

Table 2: Academic Talent and Firm Creation in Administrative Data, Across- and Within-Colleges
Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Any firm (mean: 7.12)			Firm=I(capital≥2M) (mean: 1.65)			Firm=I(capital≥15M) (mean: 0.40)		
Gaokao score (sd)	-0.458	-0.774	-0.750	-0.042	-0.172	-0.171	0.007	-0.037	-0.040
	(0.032)	(0.033)	(0.033)	(0.011)	(0.011)	(0.012)	(0.004)	(0.005)	(0.005)
Female			-3.530			-0.862			-0.183
			(0.064)			(0.025)			(0.011)
Rural			-0.220			-0.238			-0.092
			(0.049)			(0.022)			(0.011)
High school (below median)			-0.435			-0.155			-0.044
			(0.047)			(0.022)			(0.011)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
college FE		Y	Y		Y	Y		Y	Y
major FE			Y			Y			Y
age FE			Y						Y
Effective Obs. (excl singletons)	1,814,547	1,814,534	1,799,656	1,814,547	1,814,534	1,799,656	1,814,547	1,814,534	1,799,656

Notes: Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million. Unconditional on colleges, the relationship between one's college entrance exam score and firm creation depends on firm size (columns (1), (4) and (7)). However, once conditional on colleges, one's is strongly and negatively associated with firm creation (columns (2), (5) and (8)). This finding is robust to defining firms by size and considering personal background and major fixed effects (columns (3), (6), (9)). Standard errors in the paraphrases are clustered at the college level.

Table 3: Academic Talent and Firm Success in Administrative Data, Within Colleges
Firm-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Validity of the instrument			**Selection**	**A. Heckman two-step estimates**					
Dependent Var.	$peer_c$*100	$peer_{c,t}$*100		marginal	ln capital	ln capital within ind	capital within ind	Out of home prov	Expand	Listed (*1000)
mean dependent var	9.40	9.43	9.43				632	0.394	0.049	0.359
Gaokao score (sd)	0.102	-0.007	-0.004		0.006	0.009	22.345	0.039	0.005	0.057
	(0.044)	(0.042)	(0.008)		(0.004)	(0.004)	(13.155)	(0.001)	(0.001)	-0.044
$peer_{c,t}$				0.301						
				(0.016)						
λ (inverse mills ratio)					-0.370	-0.386	-369.1	-0.104	-0.007	0.264
					(0.100)	(0.095)	(323.0)	(0.026)	(0.012)	(1.090)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
college FE		Y	Y	Y	Y	Y	Y	Y	Y	Y
age FE, female, rural, high sch type			Y	Y	Y	Y	Y	Y	Y	Y
Effective Obs. (excl singletons)	1,853,092	1,858,053	1,858,053	1,858,053	1,858,053	1,858,053	1,858,053	1,858,053	1,858,053	1,858,053
					B. Estimates conditional on entry					
					ln capital	ln capital within ind	capital within ind	Out of home prov	Expand	Listed (*1000)
					0.006	0.010	22.577	0.040	0.006	0.059
					(0.005)	(0.005)	(12.230)	(0.002)	(0.001)	(0.037)
prov-track-year FE					Y	Y	Y	Y	Y	Y
college FE					Y	Y	Y	Y	Y	Y
age FE, female, rural, high sch type					Y	Y	Y	Y	Y	Y
Effective Obs. (excl singletons)					169,358	169,341	169,341	169,358	169,358	169,358

Notes: The table shows that academic talent is positively associated with firm success. To address the concern of endogenous entry, we employ firm creation probability in the previous cohort within the same college as an instrument. Columns (1)-(3) show that while one's own score is correlated with peers' firm creation propensity across colleges, it is not the case any more within a college-cohort. Column (4) shows that the instrument is indeed a strong predictor for firm creation. Panel A reports the Heckman two-step estimates and Panel B the estimates conditional on entry. Standard errors in the paraphrases in Panel B are clustered at the college level.

Table 4: Academic Talent and Firm Creation in Administrative Data, by Personal Background
Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Any firm (mean: 7.12)			Firm=I(capital≥2M) (mean: 1.65)			Firm=I(capital≥15M) (mean: 0.40)		
Gaokao score (sd)	-0.957	-0.800	-0.982	-0.199	-0.181	-0.216	-0.048	-0.039	-0.051
	(0.044)	(0.036)	(0.045)	(0.017)	(0.014)	(0.018)	(0.007)	(0.007)	(0.008)
Female×Gaokao score (sd)	0.382			0.054			0.019		
	(0.039)			(0.018)			(0.008)		
Rural×Gaokao score (sd)		0.102			0.055			0.017	
		(0.042)			(0.020)			(0.009)	
Below-med Highsch×Gk score (sd)			0.321			0.059			0.018
			(0.045)			(0.021)			(0.009)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Female×college FE	Y			Y			Y		
Rural×college FE		Y			Y			Y	
Below-med Highsch×col FE			Y			Y			Y
Effective Obs. (excl singletons)	1,814,500	1,814,464	1,814,502	1,814,500	1,814,464	1,814,502	1,814,500	1,814,464	1,814,502

Notes: Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million. This table shows that the negative score-firm creation link is stronger for male, urban individuals, and those from better high schools. These patterns are robust to defining firms by capital size. Standard errors in the paraphrases are clustered at the college level.

Table 5: Academic Talent and First-job Wages in Survey Data, by Personal Background
Individual-level Analysis

	ln First-Job Wage					
	(1)	(2)	(3)	(4)	(5)	(6)
Gaokao score (sd)	0.045	0.046	0.034	0.034	0.053	0.054
	(0.013)	(0.013)	(0.013)	(0.012)	(0.015)	(0.015)
Female×Gaokao score (sd)	-0.037	-0.036				
	(0.021)	(0.021)				
Rural×Gaokao score (sd)			-0.016	-0.014		
			(0.020)	(0.022)		
Non-key Highsch×Gaokao score (sd)					-0.032	-0.033
					(0.019)	(0.020)
Female	-0.041					
	(0.013)					
Rural			-0.025			
			(0.008)			
Non-key High School					-0.017	
					(0.011)	
prov-track-year FE	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y
Female×college FE		Y				Y
Rural×college FE				Y		
Non-key Highsch×college FE						Y
Effective Obs. (excl singletons)	14,674	14,674	14,140	14,140	11,534	11,534

Notes: This table shows that the positive score-wage link is stronger for male, urban individuals, and those from better high schools. This data comes from the College Student Survey we conducted during 2010-2015 (see Jia and Li 2016 for more discussion). Standard errors in the paraphrases are clustered at the college level.

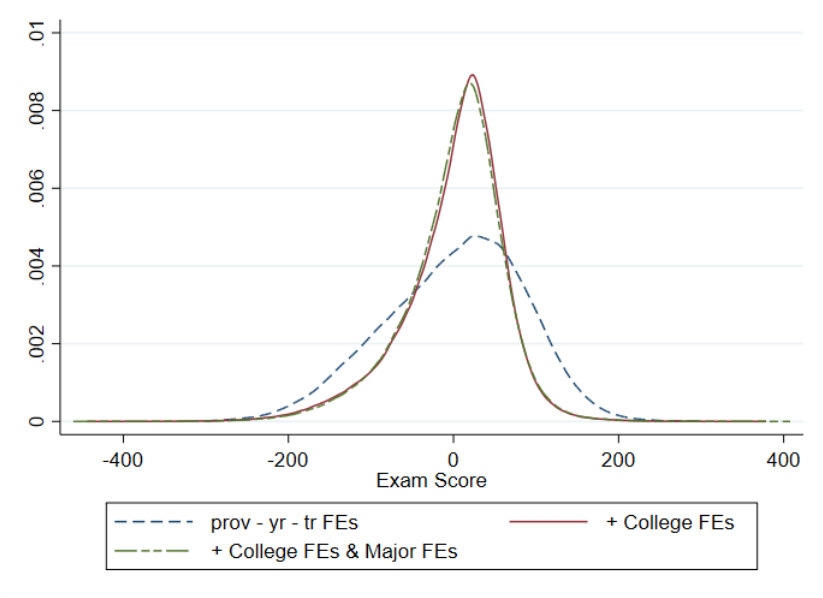
Table 6: Academic Talent and Firm Creation in Administrative Data, by Business Cycle in Graduation Year
Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Any firm (mean: 7.12)			Firm=I(capital≥2M) (mean: 1.65)			Firm=I(capital≥15M) (mean: 0.40)		
Gaokao score (sd)	-0.664	-0.710	-0.663	-0.142	-0.149	-0.140	-0.022	-0.020	-0.018
	(0.036)	(0.036)	(0.037)	(0.013)	(0.013)	(0.014)	(0.006)	(0.006)	(0.006)
Better Growth (birth prov)×Gaokao score	-0.278		-0.301	-0.063		-0.064	-0.030		-0.014
	(0.042)		(0.048)	(0.017)		(0.023)	(0.008)		(0.013)
Better Growth (college prov)×Gaokao score		-0.183	0.026		-0.047	-0.002		-0.034	-0.024
		(0.042)	(0.050)		(0.017)	(0.024)		(0.008)	(0.013)
Better Growth (birth prov)	1.905		2.297	0.407		0.512	0.206		0.111
	(0.294)		(0.362)	(0.122)		(0.181)	(0.060)		(0.099)
Better Growth (college prov)		1.142	-0.495		0.248	-0.121		0.223	0.144
		(0.310)	(0.386)		(0.129)	(0.190)		(0.061)	(0.100)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
age FE, female, rural, high school type	Y	Y	Y	Y	Y	Y	Y	Y	Y
Effective Obs. (excl singletons)	1,814,528	1,814,528	1,814,528	1,814,528	1,814,528	1,814,528	1,814,528	1,814,528	1,814,528

Notes: Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million. This table shows that the negative score-firm creation link is stronger when one's home (or college) province experienced better GDP growth in his or her graduation year. These patterns are robust to defining firms by capital size. Standard errors in the paraphrases are clustered at the college level.

Online Appendix

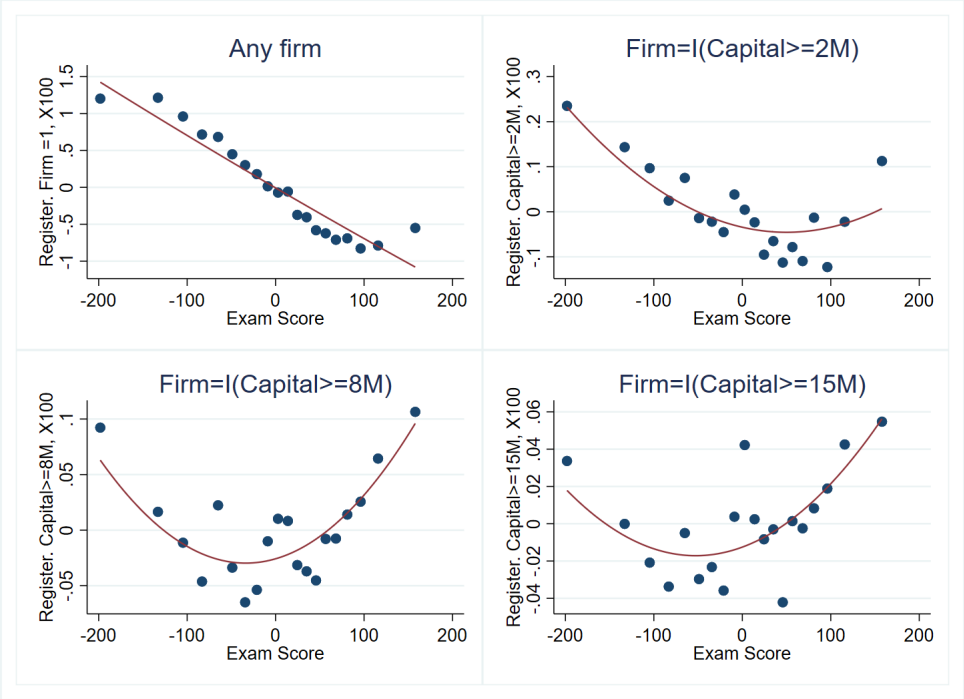
Figure A.1: Gaokao Score Distribution



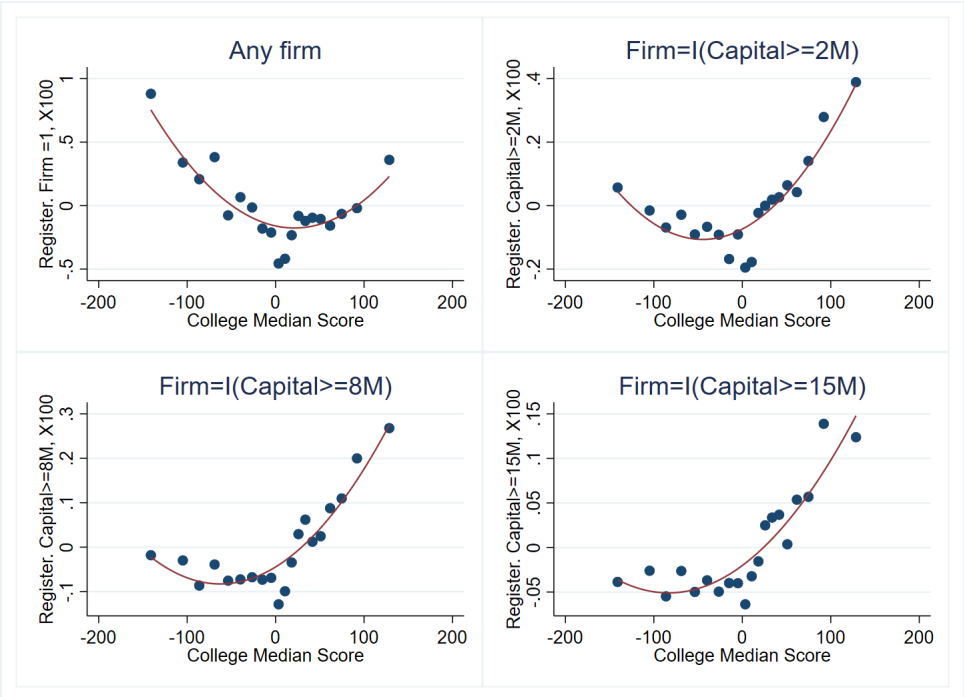
Notes: This figure presents the density of the exam scores. The standard deviation of the *Gaokao* scores after isolating province-year-track fixed effects is 87.7. It becomes 65 after controlling for college fixed effects, and 63 after controlling for college fixed effects and major fixed effects.

Figure A.2: Academic Talent and Firm Creation: Unconditional on Colleges & College Selectivity

(a) Scores vs. Firm Creation Unconditional on Colleges



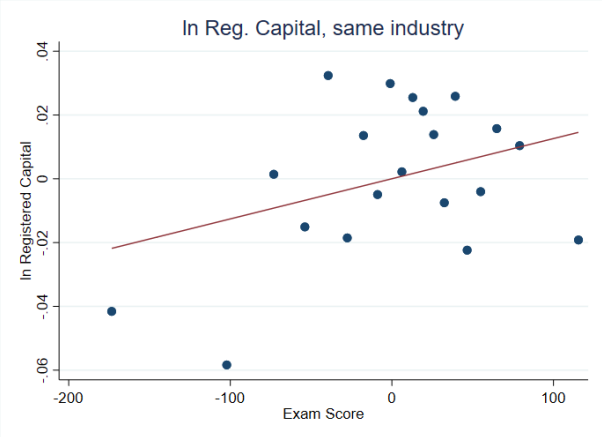
(b) College Median Score vs. Firm Creation



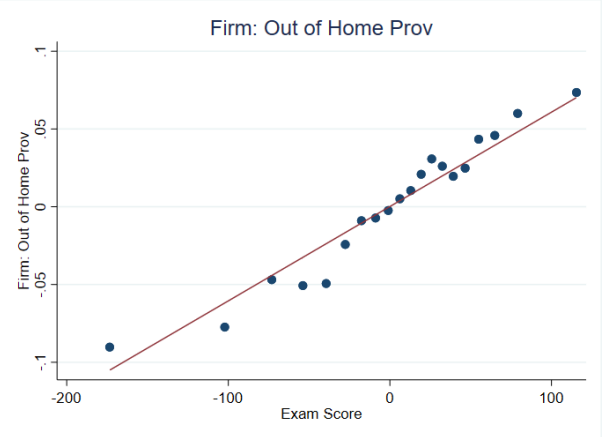
Notes: (a) shows that the relationship between academic talent and firm creation differs by firm size; (b) shows that college selectivity is positively correlated with firm creation, especially for large firms. The differences between (a) and (b) are similar to the within-college patterns in Figure 1.

Figure A.3: Firm Capital Size and Other Success Measures (given entry)

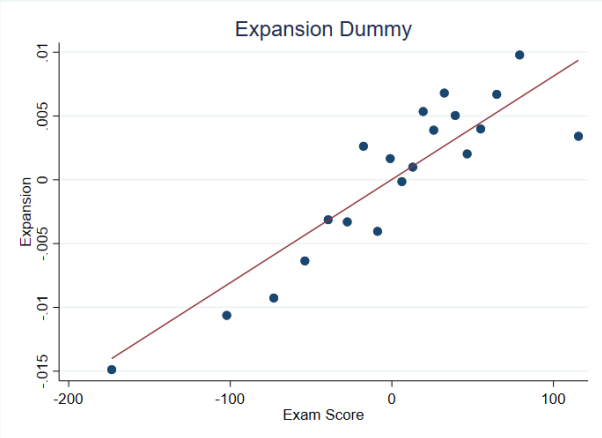
(a) log Registered Capital



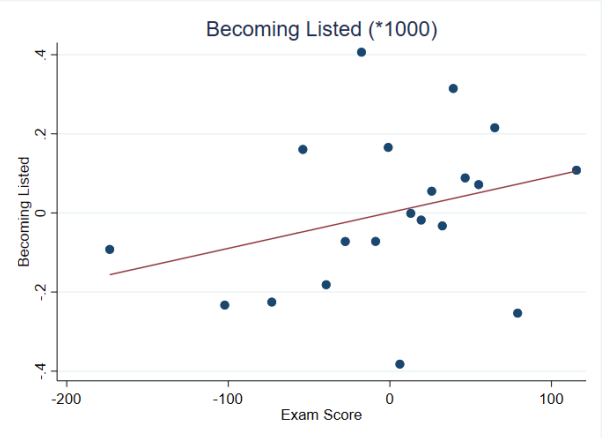
(b) Firms Created Out of Home Provinces



(c) Expansion



(d) Becoming Listed



Notes: This figure shows that given entry, academic talent is positively associated with firm capital size and other success measures including the propensity of (i) creating a firm outside one's home province, (ii) investing in other firms as a shareholder, and (iii) becoming publicly listed. We deal with the challenge of selection into entry using a Heckman two-step method in the main text.

Table A.1: Academic Talent and Firm Creation in Administrative Data, Provinces with Little Missing Info.
Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Any firm (mean: 7.12)			Firm=I(capital≥2M) (mean: 1.65)			Firm=I(capital≥15M) (mean: 0.40)		
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
college FE		Y	Y		Y	Y		Y	Y
major FE			Y			Y			Y
age FE			Y						Y
A. province-years with id missing rate < 5%									
Gaokao score (sd)	-0.500	-0.782	-0.774	-0.052	-0.167	-0.174	0.008	-0.034	-0.036
	(0.040)	(0.046)	(0.046)	(0.014)	(0.016)	(0.017)	(0.006)	(0.007)	(0.007)
Female			-3.729			-0.934			-0.180
			(0.084)			(0.034)			(0.014)
Rural			-0.138			-0.270			-0.104
			(0.068)			(0.031)			(0.016)
High school (below median)			-0.346			-0.171			-0.017
			(0.065)			(0.032)			(0.016)
Effective Obs. (excl singletons)	943,371	943,355	932,683	943,371	943,355	932,683	943,371	943,355	932,683
B. province-years with id missing rate < 1%									
Gaokao score (sd)	-0.620	-0.951	-0.917	-0.076	-0.219	-0.217	0.014	-0.041	-0.040
	(0.059)	(0.068)	(0.069)	(0.022)	(0.026)	(0.026)	(0.009)	(0.012)	(0.012)
Female			-3.955			-0.991			-0.207
			(0.125)			(0.050)			(0.021)
Rural			-0.071			-0.314			-0.131
			(0.102)			(0.049)			(0.026)
High school (below median)			-0.184			-0.179			-0.015
			(0.097)			(0.051)			(0.027)
Effective Obs. (excl singletons)	428,432	428,416	421,944	428,432	428,416	421,944	428,432	428,416	421,944

Notes: Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million. This table shows that our baseline finding holds if we restrict data to individuals in province-years whose id missing rates are lower than 5% (Panel A) or even 1% (Panel B). Standard errors in the paraphrases are clustered at the college level.

Table A.2: Academic Talent and Firm Creation in Administrative Data, Excluding Repeated Exam Takers
Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Any firm (mean: 7.12)			Firm=I(capital≥2M) (mean: 1.65)			Firm=I(capital≥15M) (mean: 0.40)		
Gaokao score (sd)	-0.432	-0.715	-0.685	-0.034	-0.164	-0.162	0.009	-0.035	-0.039
	(0.031)	(0.033)	(0.034)	(0.011)	(0.012)	(0.013)	(0.005)	(0.006)	(0.006)
Female			-3.587			-0.879			-0.185
			(0.069)			(0.028)			(0.013)
Rural			-0.241			-0.261			-0.100
			(0.058)			(0.025)			(0.013)
High school (below median)			-0.471			-0.168			-0.059
			(0.059)			(0.028)			(0.014)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
college FE		Y	Y		Y	Y		Y	Y
major FE			Y			Y			Y
age FE			Y						Y
Effective Obs. (excl singletons)	1,310,549	1,310,535	1,299,402	1,310,549	1,310,535	1,299,402	1,310,549	1,310,535	1,299,402

Notes: Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million. This table shows that our baseline finding holds after excluding exam takers. Standard errors in the paraphrases are clustered at the college level.

Table A.3: Academic Talent and Firm Creation in Administrative Data, By College Rank
Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Any firm (mean: 7.12)			Firm=I(capital≥2M) (mean: 1.65)			Firm=I(capital≥15M) (mean: 0.40)		
college rank	Top 10	11-100	100+	Top 10	11-100	100+	Top 10	11-100	100+
Gaokao score (sd)	-0.951	-1.331	-0.627	-0.267	-0.325	-0.141	-0.087	-0.076	-0.032
	(0.169)	(0.094)	(0.032)	(0.088)	(0.034)	(0.012)	(0.042)	(0.017)	(0.005)
Female	-3.819	-3.597	-3.525	-1.164	-1.013	-0.834	-0.231	-0.254	-0.171
	(0.541)	(0.154)	(0.071)	(0.327)	(0.076)	(0.026)	(0.145)	(0.038)	(0.011)
Rural	-0.186	-0.491	-0.176	-0.370	-0.368	-0.214	-0.192	-0.140	-0.083
	(0.326)	(0.125)	(0.054)	(0.133)	(0.064)	(0.024)	(0.070)	(0.033)	(0.011)
High school (below median)	-0.303	-0.286	-0.455	-0.380	-0.216	-0.145	0.056	-0.081	-0.040
	(0.452)	(0.149)	(0.050)	(0.229)	(0.079)	(0.023)	(0.105)	(0.033)	(0.012)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
major FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Effective Obs. (excl singletons)	34,250	265,026	1,500,371	34,250	265,026	1,500,371	34,250	265,026	1,500,371

Notes: Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million. This table shows that our baseline finding holds across college ranks. The links within the elite colleges appear stronger. Standard errors in the paraphrases are clustered at the college level.

Table A.4: Academic Talent and Firm Creation in Administrative Data, Within College-Major Comparison
Individual-level Analysis

	(1) Any firm	(2) Firm=I(capital≥2M)	(3) Firm=I(capital≥15M)
Gaokao score (sd)	-0.698	-0.166	-0.041
	(0.033)	(0.012)	(0.005)
Female	-3.543	-0.868	-0.184
	(0.064)	(0.025)	(0.011)
Rural	-0.206	-0.233	-0.090
	(0.049)	(0.022)	(0.011)
High school (below median)	-0.431	-0.152	-0.044
	(0.047)	(0.022)	(0.011)
college-by-major FE	Y	Y	Y
prov-track-year FE	Y	Y	Y
age FE	Y	Y	Y
Effective Obs. (excl singletons)	1,799,176	1,799,176	1,799,176

Notes: Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million. This table shows that our baseline finding holds within a college-major. Standard errors in the paraphrases are clustered at the college level.

Table A.5: Exclusion Restriction Discussion: Peers by Proximity and Firm Success
Firm-level Analysis

	(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Var.	Selection		Instruments	ln capital within ind		Out of Home Prov		Expand		Listed*1000	
				$peer_{c,t}$ from		$peer_{c,t}$ from		$peer_{c,t}$ from		$peer_{c,t}$ from	
				home prov	other prov	home prov	other prov	home prov	other prov	home prov	other prov
$peer_{c,t}$ home prov	0.161										
	(0.016)										
$peer_{c,t}$ other prov		0.044									
		(0.015)									
Gaokao score (sd)				0.011	0.010	0.035	0.034	0.005	0.005	0.059	0.061
				(0.004)	(0.005)	(0.001)	(0.034)	(0.001)	(0.001)	(0.045)	(0.045)
				-0.600	4.818	1.200	-35.053	-0.005	-0.053	-0.759	27.826
λ (inverse mills ratio)				(0.124)	(5.400)	(0.060)	(39.294)	(0.016)	(0.371)	(1.407)	(40.624)
prov-track-year FE	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y
college FE	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y
age FE, female, rural, high sch type	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y
Effective Obs. (excl singletons)	1,858,053	1,858,053		1,858,053	1,858,053	1,858,053	1,858,053	1,858,053	1,858,053	1,858,053	1,858,053

Notes: This table shows that using peers within the same college-cohort-home provinces to predict entry generates similar estimates on success as using the same college-cohort-other provinces to instrument for entry, suggesting that exclusion restriction is not a critical concern. Otherwise, we would see a larger impact on success when using the former as an instrument because it is more likely to help with firm success. Standard errors in the paraphrases are clustered at the college level.

Table A.6: Academic Talent and Risk Averse in Survey Data, Within College Comparison
Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	lottery1: rather gain w. certainty		lottery2: rather lose w. certainty		investment: ensuring certain return	
Gaokao score (sd)	0.035	0.055	0.032	0.046	-0.005	0.034
	(0.051)	(0.055)	(0.049)	(0.047)	(0.052)	(0.054)
Female		0.124		0.000		0.109
		(0.047)		(0.033)		(0.044)
Rural		-0.002		-0.020		-0.011
		(0.055)		(0.062)		(0.050)
High school (below median)		0.019		0.054		0.086
		(0.032)		(0.043)		(0.034)
prov-track-year FE	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y
Effective Obs. (excl singletons)	3,066	2,936	3,063	2,933	3,090	2,957

Notes: We define one being risk averse in three ways: (1) if he or she answers yes to "do you rather gain RMB 1000 with certainty than gain RMB 2000 and 0 with equal chance?"; (2) if he or she answers yes to "do you rather lose RMB 1000 with certainty than lose RMB 2000 and 0 with equal chance?"; (3) if he or she agrees or strongly agrees with the statement that "in investment, it is more important to ensure certain returns than take risks to pursue higher returns". This table shows that women are more risk averse measured by (1) and (3). However, there is no any strong correlation between scores and risk attitudes measured in all three ways. Standard errors in the paraphrases are clustered at the college level.