Characteristics of China's Urban Entrepreneurs

Michael Troilo[1]

University of Michigan

Abstract. What are the characteristics of China's urban entrepreneurs? This paper seeks to address that question using survey data from five cities in China: Shanghai, Wuhan, Shenyang, Fuzhou, and Xian for the period 1996-2001. I first estimate the likelihood of being an entrepreneur at year-end 2001 using multinomial logit with a full model of demographic and psychological variables. I then employ a parametric survival model to analyze who becomes an entrepreneur over time from 1996 to 2001. I find support for some of the "traditional" relationships between entrepreneurship and age, gender, marital status, self-employed parents, and health. More importantly, I find that education and rising unemployment have persistently negative effects on the probability of self-employment in urban China. These surprising results indicate that context does indeed matter. The literature discussing entrepreneurial characteristics has focused almost exclusively on Western nations, particularly the United States and United Kingdom. The contribution of this paper is to examine the characteristics of entrepreneurs in a previously neglected yet prominent country, China.

Keywords: entrepreneurship, China, emerging markets, individual characteristics, self-employment.

1. Introduction

With millions of businesses starting every year worldwide, the phenomenon of entrepreneurship has received increased attention in management studies. What drives people to start a firm? Knight (1921) posited that it was the intelligence and foresight of entrepreneurs who grasped opportunities previously unseen that led to new firm formation. Schumpeter (1934) opined that the creativity of nascent entrepreneurs was the principal cause of new businesses. Both of these classical thinkers believed that economic, sociological, and psychological factors intermingled to produce a special breed of people called entrepreneurs.

Only recently has empirical analysis been brought to bear on the characteristics of entrepreneurs. The context of most of this research has been in Western countries, particularly the United Kingdom and the United States. The

---

seminal empirical work, such as Rees and Shah (1986), Evans and Leighton (1989), and Dolton and Makepeace (1990) was done in these two countries for the simple reason that data on self-employment and other demographic variables were both well-kept and readily available. This established a research trajectory of entrepreneurship in the West.

China, as the 3nd-largest economy in the world (World Bank 2004), is undeniably an important market. It also is the source of Confucian culture that has influenced entrepreneurship in other East Asian countries (Lee and Tsang 2001). It therefore merits more attention in the entrepreneurship literature.

This paper seeks to uncover the factors influencing the decision to become an entrepreneur in mainland China. Using the China Urban Labor Survey (CULS), a detailed questionnaire that covered a random sample of households and individual adults in five large cities, I first estimate the likelihood of becoming an entrepreneur with a probit function. This is done to compare my results with previous work. I then examine who becomes an entrepreneur over time from 1996 to 2001 utilizing a parametric survival model with a Weibull distribution. I also examine the correlations between the variables and perform other robustness checks.

I report two findings in this study. First, education has a negative impact on the probability of self-employment in urban China. Second, a higher rate of non-employment strongly decreases the likelihood of becoming an entrepreneur. These findings suggest that China's urban entrepreneurs are concentrated in low-tech industries, and that they have become self-employed as a matter of choice.

The paper is organized as follows. Section 2 details the relevant literature, with an emphasis on the variables that have been tested before. Section 3 describe the data and the methodology employed to test it. I present results in Section 4. Discussion, including avenues of future research, concludes the paper in Section 5.

2. Literature Review

I order the literature review in terms of the factors most commonly studied. These categories have some overlap among them. My primary objective is to instill structure into the review of the relevant literature.

2.1. Demographic Factors

Age: Given the energy needed to start a business, one would reasonably expect that entrepreneurship is primarily a young person's game. The empirical findings to date support this. Rees and Shah (1986), using as their dependent variable the probability of self-employment in their study of entrepreneurship in the U.K.,

determine that age has a non-linear effect on the probability of self-employment, rising quickly in the early years of employment and declining in the final years. Reynolds (1997), in his investigation of new firm gestation in the U.S., discovers that age is the most important determinant of the decision to start a new venture, with 17% of his sample in the 25-34 age bracket accounting for 69% of new firms. He also notes that age has a curvilinear effect on the probability of becoming an entrepreneur. His dependent variable is the probability of nascent entrepreneurship as defined by certain gestational behaviors, e.g. "writing a business plan". Schiller and Crewson (1997) find that age is positive and significant for both male and female entrepreneurs, though their data set is limited to people aged 26-35. Their context is the U.S., and their dependent variable is the probability of ever being self-employed.

Gender: Some of the seminal work, including Rees and Shah (1986) and Evans and Leighton (1989) ignore gender, since their datasets only include males. Dolton and Makepeace (1990) find that men are more likely than women to be self-employed. They estimate the probability of employment using data from the U.K. Delmar and Davidson (2000) conclude that gender is the strongest factor in the decision to become a nascent entrepreneur, with men more likely than women to start a new venture. Like Reynolds (1997) they estimate the probability of nascent entrepreneurship, but they compare the U.S., Norway, and Sweden. Blanchflower and Oswald (1998) discover a negative effect of being female on the probability of self-employment. They situate their study in the U.K. Bates (1995) establishes that women rely more on education and experience in choosing to enter self-employment, whereas men rely more on wealth. He predicts self-employment entry across industries using U.S. Census data.

Marital status: Schiller and Crewson (1997) suggest that marriage is a positive influence on the supply of female entrepreneurs but a negative influence on the supply of male ones. In contrast, Delmar and Davidson (2000) find that marital status has insignificant effect on the probability of becoming an entrepreneur. Reynolds (1997) adds nuance by establishing that being married is characteristic of those 25-34 year olds who pursue entrepreneurship as a "hobby", while those 25-34 year olds who are "committed to the entrepreneurial lifestyle" are usually single.

Geographic mobility: Delmar and Davidson (2000) find that nascent entrepreneurs are more geographically mobile than the control group. They also note that the U.S. enjoys a higher rate of nascent entrepreneurship than either Sweden or Norway. Among many reasons why this is so may be the greater mobility of Americans compared to Norwegians and Swedes. Lerner and Hendeles (1993) surveyed Russian immigrants in Israel and discovered that those

immigrants who were more mobile within Israel were more likely to be self-employed than those who were more stationary.

Income/wealth: Rees and Shah (1986) resolve that the probability of self-employment depends positively on the difference between self-employed versus employed earnings, thus supporting rational choice. Evans and Leighton (1989), examining a cohort of men in the U.S., find insignificant effect for income on the probability of self-employment, but highlight the fact that both education and work experience influence income as well as the probability of self-employment. Bates (1995) concludes that wealth, as measured by household net worth, has a significant positive effect on the likelihood of men becoming self-employed.

Education: Given the speculation that Schumpeter, Knight, and other classical theorists had about entrepreneurs possessing above-average cognitive ability, it is not surprising to learn that empirical work generally has demonstrated a significant positive effect of educational attainment on the probability of starting a new venture. Dolton and Makepeace (1990) report that schooling increases the likelihood of self-employment. Bates (1995) demonstrates that education increases the probability of new ventures in skilled services industries, though it is not important for entrepreneurship in construction. Reynolds (1997) notes that those who view entrepreneurship as a "hobby" tend to be more highly educated. Delmar and Davidsson (2000) confirm the significance of education for nascent entrepreneurs.

Self-employed parents: Having a role model of entrepreneurial behavior may reasonably be supposed to have a positive impact on the likelihood of self-employment. Existing self-employment literature supports this, particularly with regard to parents. Roberts (1991) found that entrepreneurs were 20% more likely to have self-employed fathers than non-entrepreneurs. Dunn and Holtz-Eakin (1996) found that people whose parents were self-employed were themselves more likely to become self-employed. Further support for the correlation of self-employed parents and increased likelihood of self-employment has also been furnished by Reynolds (1997) and Delmar and Davidsson (2000).

Work experience: Skills learned over the course of one's career may enable one to become an entrepreneur. More work experience would therefore increase the likelihood of self-employment. This has been demonstrated from the research of Evans (1989), Evans and Leighton (1989), Bates (1995), and Schiller and Crewson (1997).

Unemployment: Unemployed people should on balance be more likely than employed people to become self-employed, since their opportunity cost is lower. Mesch and Czamanski (1997) studied recent immigrants to Israel and found that those out of work were more likely to start their own business than those with

employment opportunities. Ritsila and Tervo (2002) corroborated this result through their study of the Finnish unemployed.

2.2. Psychological Factors

Health: I anticipate that would-be entrepreneurs would in general be in better health than the overall population, since new ventures require energy. Rees and Shah (1986) show that poor health is negative and significant on the probability of self-employment. Evans and Leighton (1989) include a poor health variable in their analysis of self-employment. It is negative as expected in all specifications, and is almost significant at the 10% level in two regressions.

Agreeableness: According to Barrick and Mount (1991), the trait of agreeableness includes characteristics like friendliness, tolerance, flexibility and willingness to trust. Ceteris paribus, a person who is more trusting is less likely to be an entrepreneur. Blanchflower and Oswald (1998) tested this idea with a hostility score but did not find it to be significant in their full model of variables. Wooten et al. (1999), in contrast, found that former executives who left their companies to start their own firms tended to be more skeptical and less trusting than those outplaced executives who didn't become entrepreneurs.
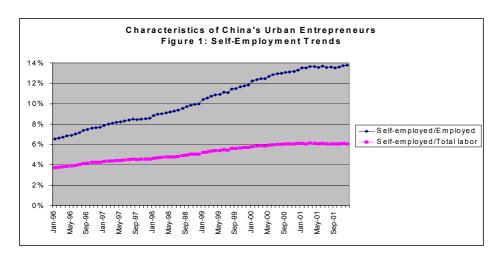
Optimism: Another personality trait where entrepreneurs may reasonably differ from non-entrepreneurs is optimism. Starting a new venture can often require a leap of faith, both in one's abilities and in what the future holds. Busenitz and Barney (1997) report that a sample of entrepreneurs consistently overestimated their ability to answer a set of questions correctly relative to a group of managers. Djankov et al. (2005) find a statistically significant difference in self-reported happiness between entrepreneurs and non-entrepreneurs, with entrepreneurs being higher.

Based on prior literature, I expect that the above demographic and psychological variables, with the exception of agreeableness, to have a positive impact. A negative result would be a surprise. I do not form any hypotheses about the relative strength of the variables to one another, as the literature is mixed regarding which variables are the most important for self-employment.

3. Data and Methodology

My analysis focuses on the determinants of self-employment at year-end 2001. I perform a multinomial logit estimation of the probability of being self-employed conditional on observed characteristics to compare the results with prior literature. After this static analysis, I study who becomes an entrepreneur over

time with a parametric survival model assuming a Weibull distribution. The reason for this is the increasing rate of self-employment over time, which can be seen in Figure 1 below. The Weibull distribution does not assume a constant rate of self-employment. I perform the parametric estimation by time-varying the independent variables where possible.

Figure 1: Self-Employment Trends

3.1. Data

I use the China Urban Labor Survey (CULS) compiled by Giles and Park (2005) in coordination with scholars at the Institute for Population and Labor Economics at the Chinese Academy of Social Services as well as provincial and municipal offices of the National Bureau of Statistics. The survey was administered at the end of 2001 in five cities each with populations exceeding 1 million. These cities are, in order of population from largest to smallest, Shanghai, Wuhan, Shenyang, Xian, and Fuzhou.

According to Giles and Park, the cities were selected to provide both regional diversity as well as variation in the size of the public versus private sectors. Both Fuzhou and Shanghai are relatively wealthy cities on the coast that have profited tremendously since the inception of economic reform in the late 1970s. In contrast, the other three cities are interior and have lagged in their economic development. The order of cities from richest to poorest in terms of mean wages and per capita GDP are: Shanghai, Fuzhou, Shenyang, Wuhan, and Xian (Giles and Park 2005). Shenyang is in the northeast, Wuhan is in central China, Xian is northwest, Shanghai is east, and Fuzhou is southeast.

Within each city, a proportional population sampling approach was used to sample an average of 15 registered urban households in each of 70 neighborhood clusters. Each household head was asked questions about the family, and then all

family members above age 16 who were no longer in school were interviewed separately. Surveys were conducted in 3,499 households (700 per city), with 8,109 adults over age 16 completing the survey. At the time of the survey, 5,787 adults were under mandatory retirement age and 4,238 were currently employed. The survey non-response rate was 16.5% (Giles and Park 2005).

The CULS includes individual calendar-based work histories with detailed questions about job changes, transitions to unemployment or retirement, changes in pension, health care, and housing benefits, and access to government programs since January 1996. Like many multi-purpose household surveys, it also includes a battery of questions on work status, income, expenditure, housing and consumer durables, productive assets and wealth, health, and household demographics (ibid). We focus on measuring the extent to which these various factors, demographic, social, and personal, predict the likelihood of self-employment.

The dataset includes the relevant household responses as well as the personal characteristics for each of the 8,109 individuals. Their work history is arranged by month; hence, there are 72 months available for analysis.

Table 1 below shows the proportion of self-employed to the total labor force (self-employment rate) and the proportion of total employed to the total labor force (employment rate) by city for the years 1996, 1999, and 2001 calculated from the CULS data. Overall, we see that the employment rate is dropping for all five cities and the self-employment rate is increasing. The trends among the cities are mixed. Shanghai has the lowest self-employment rate and the second-lowest employment rate in 2001, while Shenyang has the highest self-employment rate and the lowest employment rate for 2001. The employment rates for both cities fell steadily, but the self-employment rate for Shanghai stagnated while Shenyang's self-employment rate enjoyed healthy growth. Fuzhou seems to fare best among the cities over the 6-year time period when both rates are considered.

Table 1: Self-Employment and Total Employment by City

	1996		1999		2001	
	SE Rate	**Emp. Rate**	**SE Rate**	**Emp. Rate**	**SE Rate**	**Emp. Rate**
Total	4.24%	65.09%	5.70%	56.23%	6.06%	51.62%
Shanghai	1.77%	67.09%	2.34%	56.57%	2.29%	49.71%
Wuhan	4.83%	67.07%	6.34%	56.20%	6.69%	51.31%
Shenyang	5.10%	60.07%	7.58%	52.55%	8.56%	48.95%
Fuzhou	6.60%	68.17%	7.70%	61.45%	8.28%	56.88%
Xian	3.24%	62.59%	4.95%	54.41%	4.95%	51.62%

Notes:
"SE Rate" is the percentage of self-employed to the total labor force.
"Emp. Rate" is the percentage of employed (both self and employed by others) to total labor force.
Those not working may be unemployed, retired, on sick leave, etc.

3.2.   Methodology-Multinomial Logit Analysis

The dependent variable for the first part of the study is self-employment at year-end 2001. This is a variable equal to 0 if the individual indicated he or she was unemployed at December 2001, 1 if he or she was self-employed, and 2 if he or she was employed by a firm. The unemployed category forms the basis of comparison with the other two in the multinomial regression.

   Concerning demographic factors, the age variable is the age in 1996 or 2001. I also include age^2 as a control variable. The variable "male" is a dummy equal to 1 if a person self-reports as male on the survey.   "Married" is a dummy equaling 1 if the person responds as married and 0 otherwise. The variable "residencelength" measures the amount of time in years that the individual has lived in his/her city. "Housesize is the area in squared meters of an individual's domicile and serves as a proxy for wealth. "College" is a dummy equal to 1 if a person has had more than 12 years (high school) of education and 0 otherwise. Note that this variable includes not just those that attended university but also those who attended finishing schools and/or technical schools after high school. "Fatherselfemp" is a dummy equaling 1 if the person's father was self-employed and 0 otherwise. "Totalwork" measures the person's work experience in years. "Emprate01" is the employment rate calculated by dividing the total number of employed (self and by firm) by the total labor force for each city for year-end 2001.

   To capture psychological factors, I utilized three dummy variables. "Goodhealth" equals 1 if the respondent identified his/her physical health as either "excellent" or "good" on a 5-point scale with 1 being "excellent" and 5 being "very poor". "Satisfied" is a dummy variable equaling 1 if the person described him/herself as being either "very satisfied" or "satisfied" with his/her current standard of living on a 4-point scale. This question was included in the attitudinal portion of the survey, and the premise is that a more satisfied person finds their situation to be more agreeable. Another attitudinal question asked respondents to predict their expected economic condition five years hence versus today. People answering "much better" or "better" were given a value of 1 for our variable "optimistic", while those answering "no change" or "worse" received a 0.

   Table 2 below, Panel A presents the descriptive statistics for the variables I employed in the multinomial estimation. Note that missing values for housesize reduce the number of observations from 8,109 individuals to 8,073. The average worker is rather old at 47.57 years, and 82% of the respondents are married. There is wide variance in house size, and 28% of the individuals have more than a high school education. These people tend to reside in a given city for a long time (40 years). The responses to the psychological variables are about evenly divided.

3.3. Methodology-Parametric Survival Estimation:Weibull Distribution

In the second stage of the analysis, I estimate who becomes an entrepreneur for the 72 months during 1996-2001 with a parametric survival model assuming a Weibull distribution. The dependent variable is self-employment across time (SE) with the default condition equal to 0. When a person becomes self-employed the model identifies this event as failure and assigns a 1 to that individual for that month.

I created time-varying variables from the initial set of static variables where possible for our survival model. I designate these new variables with a "2": age2, married2, etc. "Age2" is a monthly calculation of age using the person's date of birth as reported to surveyors in 2001, and "age2^2" is the squared term. "Married2" is still a dummy variable, but instead of being only 0 or only 1 as in the multinomial estimation it becomes 1 during the month and year of marriage for those married during 1996-2001. "College2" also functions in this manner. "Residence2" counts by month and year the residency of an individual in their given city based on their 2001 response to number of years lived in that city. I calculated total work experience by year and month, designated "totalwork", as indicated by the respondent in 2001.[2] To test duration dependence I created the squared version as a control. "Emprate2" is the monthly ratio of those working to the total available labor force by city, and is motivated by the data we presented in Table 1.

---

2. Note that work experience is not a created variable like the more common "potential work experience", which equals age-years of schooling-6.

Table 2: Descriptive Statistics

A) Multinomial Logit Analysis

Variable	**Units**	**Obs**	**Mean**	**Std.Dev.**	**Min**	**Max**
age6	years	8,108	47.57	15.17	15.05	93.25
age6^2	years	8,108	2,493.41	1,525.00	227.51	8,695.56
male	dummy var.	8,109	0.48	0.50	0.00	1.00
married	dummy var.	8,109	0.82	0.38	0.00	1.00
residencelength	years	8,096	40.05	16.04	0.00	94.00
housesize	square meters	8,073	46.30	27.16	2.00	400.00
college	dummy var.	8,109	0.28	0.45	0.00	1.00
fatherselfemp	dummy var.	8,109	0.02	0.14	0.00	1.00
totalwork	years	8,109	11.87	11.80	0.00	51.00
emprate01	percent	8,109	0.52	0.03	0.49	0.57
goodhealth	dummy var.	8,109	0.43	0.50	0.00	1.00
satisfied	dummy var.	8,109	0.54	0.50	0.00	1.00
optimistic	dummy var.	8,109	0.58	0.49	0.00	1.00

B) Weibull Parametric Estimation

Variable		**Obs**	**Mean**	**Std.Dev.**	**Min**	**Max**
age2	years, months	583,776	44.62	15.22	9.16	93.25
age2^2	years, months	583,776	2,223.67	1,445.94	84.03	8,695.56
male	dummy var.	583,848	0.48	0.50	0.00	1.00
married2	dummy var.	583,848	0.85	0.36	0.00	1.00
residence2	years, months	582,912	37.11	16.10	0.00	94.00
housesize	square meters	582,192	46.30	27.16	2.00	400.00
college2	dummy var.	583,848	0.24	0.42	0.00	1.00
fatherselfemp	dummy var.	583,848	0.02	0.14	0.00	1.00
totalwork2	years, months	583,848	9.35	10.62	0.00	50.08
totalwork2^2	years, months	583,848	200.11	315.68	0.00	2,508.34
emprate2	percent	583,848	0.59	0.06	0.48	0.70
goodhealth	dummy var.	583,848	0.43	0.50	0.00	1.00
satisfied	dummy var.	583,848	0.54	0.50	0.00	1.00
optimistic	dummy var.	583,848	0.58	0.49	0.00	1.00
age	years	8,108	47.57	15.17	9.16	93.25
age^2	years	8,108	2,042.68	1,376.02	84.03	8,695.56

For the other initial variables, I had to assume constancy due to lack of data or other compelling reasons (e.g. change of gender is possible but rare). I noted that although data did exist for when individuals purchased and sold their home almost no one (<1%) in the dataset reported buying or renting a second dwelling within the 1996-2001 timeframe. For the personal characteristics goodhealth, satisfied, and optimistic, there is only the response furnished at the end of 2001. I also checked change of status from married to single and found that the number of such changes was insignificant (<5% for married2). Summary statistics for these variables appear on Table 2 Panel B.

The parametric survival model is a form of accelerated event-time regression. Accelerated event-time regression is appropriate for this longitudinal analysis for two reasons. First, this method accounts for the fact that the duration of self-employment is right-censored, i.e. we do not observe what occurs after 2001. Secondly, accelerated event-time regression permits flexibility with regard to parametric distributions (Polidoro 2005).

I chose a Weibull distribution based on earlier survival analysis in the entrepreneurship literature (Audretsch and Lehmann 2004). The Weibull distribution suits my purposes in two ways. First, with the Weibull distribution we need not assume that the rate of becoming an entrepreneur is constant. There is nothing in the data that would lead us to assume a constant rate of self-employment a priori. Secondly, the Weibull distribution controls unobserved heterogeneity better than other distributions I might select (Audretsch and Lehmann 2004).

In using the Weibull distribution, I must make a strong assumption about left-censoring. Since I do not have the detailed work history of individuals prior to 1996, I must assume that no one in the dataset was self-employed before January 1996. I have tried to compensate for this by including as many variables as reasonably possible in the model that would impact becoming an entrepreneur, but in truth left-censoring is a difficult econometric problem to overcome. I think that the benefits of the model offset this strong assumption.

Following Woolridge (2002) and Audretsch and Lehmann (2004) I define the hazard function equations as follows. Let T be a non-negative random variable that describes the duration of an individual until he or she becomes an entrepreneur. The probability of becoming self-employed within an interval $(0,t)$ is given by the cumulative distribution function (cdf) $F(t)$. I assume for the sake of simplicity that $F(t)$ is both continuous and differentiable. The conditional probability of becoming self-employed within the interval $t+h$ given that the individual wasn't self-employed at time t, which is the hazard function, is:

$$\lambda(t) = \lim_{h \to 0} \frac{\Pr(t \leq T < t + h \mid t \leq T)}{h}$$

(1)

The survivor function is defined as S(t) = 1-F(t) = Pr(T>t). This is the probability that someone will not be self-employed past time t. Re-writing Equation 1 and taking limits, we find:

$$\lambda(t) = \lim_{h \to 0} \frac{F(t+h) - F(t)}{h} * \frac{1}{1 - F(t)} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$$

(2)

The cdf of the Weibull distribution is $F(t) = 1 - \exp(-\gamma t^\alpha)$ where gamma and alpha are non-negative parameters. The pdf is $f(t) = \gamma \alpha t^{\alpha-1} \exp(-\gamma t^\alpha)$ and is an exponential function when gamma equals 1. Using Equation 2, the hazard function of the Weibull distribution becomes:

$$\lambda(t) = \frac{f(t)}{S(t)} = \gamma \alpha t^{\alpha-1}$$

(3)

Note that when alpha equals 1 lambda equals gamma.

4. Results

4.1. Multinomial Logit Analysis

I first checked for multicollinearity of the variables prior to running the estimation. Table 3 below displays the correlation matrix. No significant degree of multicollinearity was noted.

Table 4 below presents the coefficients and z-scores from the multinomial estimation. The first column depicts the results for the self-employed, the second column shows the results for those employed by a firm, and the unemployed serves as the base.

Reviewing Table 4, we find that the impact of age is not significantly different for the self-employed vs. the employed. Both coefficients are statistically significant at the 1% and their value is approximately .2. Likewise, the age-squared term is almost identical in significance and magnitude.

Being male, on the other hand, increases the probability of self-employment, as does being married. Both of these variables are statistically significant at the 1% level for the self-employed, but only male is significant at the 1% level for the employed and married is insignificant. The male coefficient is 1.181 for self-employed versus only 0.951 for the employed. Furthermore, the married coefficient at 0.699 is relatively substantial for the probability of

entrepreneurship, being superseded only by the employment rate, gender, and having a self-employed father.

Table 3: Correlation Matrix of Independent Variables for Multinomial Logit Analysis

Independent variable	Age	Age^2	Male	Married	Residence length	House size	College	Father selfemp	Total work	Emprate01	Goodhealth	Satisfied	Optimistic
Age	1.0000												
Age^2	0.9840	1.0000											
Male	(0.0029)	(0.0045)	1.0000										
Married	0.1318	0.0396	0.0459	1.0000									
Residencelength	0.6840	0.6620	0.0178	0.0765	1.0000								
Housesize	0.0060	0.0100	0.0031	0.0510	(0.0653)	1.0000							
College	(0.1631)	(0.1399)	0.1012	(0.0618)	(0.1838)	0.1186	1.0000						
Fatherselfemp	(0.0469)	(0.0453)	0.0385	(0.0108)	(0.0293)	0.0188	0.0022	1.0000					
Totalwork	(0.0512)	(0.1395)	0.1866	0.2685	0.0400	(0.0191)	0.0763	0.0042	1.0000				
Emprate01	(0.0544)	(0.0525)	(0.0078)	0.0219	(0.0977)	0.3011	(0.0112)	0.0569	(0.0232)	1.0000			
Goodhealth	(0.3010)	(0.2770)	0.1086	(0.0763)	(0.2479)	0.0638	0.1444	0.0404	0.0138	0.0036	1.0000		
Satisfied	0.2015	0.2122	(0.0056)	0.0300	0.1040	0.1129	0.1278	0.0177	0.0216	0.0885	0.1023	1.0000	
Optimistic	(0.0704)	(0.0535)	0.0209	(0.0383)	(0.0728)	0.0594	0.1302	(0.0071)	(0.0126)	0.0188	0.1901	0.2765	1.0000

The first unusual discovery concerns education. Education, as measured by the college variable, has a negative effect on the probability of self-employment and is statistically significant at the 1% level, and has a highly positive effect on employment. I had expected education to have a positive effect on the likelihood of being an entrepreneur. I substituted another dummy variable, highschool (not shown on Table 4), for college to capture those with a high school education or above, and the result was still negative and significant at the 1% level.

Table 4: Multinomial Logit Analysis of Self-Employment in 2001

	Dependent Variable	Self-employed selfemp2001	Employed by firm selfemp2001
	Independent variables		
Demographic	age	0.184***	0.235***
		(5.27)	(10.99)
	age^2	-0.003***	-0.004***
		(7.69)	(16.46)
	male	1.181***	0.951***
		(10.96)	(14.83)
	married	0.699***	0.146
		(3.57)	(1.38)
	residencelength	-0.003	-0.004
		(0.51)	(1.34)
	housesize	0.003*	-0.001
		(1.79)	(1.16)
	college	-0.405***	1.157***
		(2.59)	(15.19)
	fatherselfemp	0.961***	0.080
		(3.63)	(0.37)
	totalwork	0.001	0.057***
		(0.14)	(17.39)
	emprate01	4.802***	4.33***
		(2.53)	(3.57)
Psychological	goodhealth	0.359***	0.305***
		(3.29)	(4.67)
	satisfied	0.499***	0.672***
		(4.40)	(9.87)
	optimistic	0.119	0.273***
		(1.10)	(4.24)
	Constant	-7.868***	-6.311***
		(6.64)	(8.51)
	Observations	8,073	
	Pseudo R-squared	31.87%	

Absolute value of z statistics in parentheses
*** significant at 1%; ** significant at 5%; *significant at 10%

I then cross-tabulated the self-employed in 2001 with the 21 broad industries in which they were self-employed. I found that 48% of all the self-employed were working in the wholesale and retail industry, which includes food & beverage. The next largest percentages, 18%, work in the social services industry, which does not include either education or healthcare. The picture that emerges of China's entrepreneurs is one that mirrors the findings of Bates (1995): China's entrepreneurs are clustered in industries where education isn't important, just as Bates showed that education wasn't significant for entrepreneurs in industries like construction.

Having a self-employed father is statistically significant for the probability of entrepreneurship at the 1% level and is insignificant for the employed. I substituted self-employed mother and also found it to be statistically significant. This corroborates the importance of parental role models documented elsewhere in the literature.

Employment rate is significant at the 1% level for both the self-employed and the employed. It has more magnitude for the former (4.8) than the latter (4.3). This suggests that some of the unemployed may be seeking entrepreneurial opportunities. The parametric analysis should give a clear indication as to whether or not the unemployed seek to become entrepreneurs over time.

Among the psychological factors, good health and satisfied are significant at the 1% level for the likelihood of entrepreneurship. Two aspects of this result are striking. The first is that these attributes are also significant for the likelihood of employment, and the magnitude of the variables differs little between the self-employed and the employed. The second is that a negative value was expected for satisfied for self-employment, since dissatisfied people tend to have the skepticism necessary for entrepreneurship.

The result for good health supports the findings of both Rees and Shah (1986), Evans and Leighton (1989), and Djankov et al. (2005) that good health raises the probability of entrepreneurship, ceteris paribus. With regard to "satisfied", the result lends credence to the conclusions of Blanchflower and Oswald (1998) that attitudes like hostility have no bearing on the decision to become an entrepreneur. The insignificance of "optimistic" is more puzzling. It could be that my proxy doesn't capture that air of overconfidence and exuberance seemingly so common to entrepreneurs. Alternatively, perhaps the importance of attitudes pales in comparison to age, gender, marital status, etc. in the decision to be self-employed.

I also considered height, which Djankov et al. (2005) found to be a significant variable for the likelihood of entrepreneurship. I did not find it to be significant in our model. For this reason, height is not included on Table 4.

4.2. Weibull Parametric Estimation

I checked multicollinearity of the independent variables prior to running the parametric estimation.   The correlation matrix appears on Table 5 below. Naturally there is strong correlation between age and age squared and total work experience and total work experience squared.   Outside of these items multicollinearity doesn't seem to be an issue.

Table 5: Correlation Matrix of Independent Variables for Weibull Parametric Estimation

Independent variable	Age2	Age2^2	Male	Married2	Residence2	Housesize	College2	Fatherselfemp	Totalwork2	Totalwork2^2	Emprate2	Goodhealth	Satisfied	Optimistic
Age2	1.0000													
Age2^2	0.9815	1.0000												
Male	(0.0029)	(0.0046)	1.0000											
Married2	0.3805	0.2767	(0.0230)	1.0000										
Residence2	0.6887	0.6644	0.0176	0.2794	1.0000									
Housesize	0.0060	0.0101	0.0031	0.0190	(0.0650)	1.0000								
College2	(0.0484)	(0.0505)	0.1033	(0.0054)	(0.1044)	0.1313	1.0000							
Fatherselfemp	(0.0466)	(0.0448)	0.0385	(0.0277)	(0.0291)	0.0188	(0.0039)	1.0000						
Totalwork2	0.0229	(0.0704)	0.1778	0.2571	0.0973	(0.0224)	0.1169	0.0014	1.0000					
Totalwork2^2	0.1319	0.0574	0.1739	0.2076	0.1675	(0.0209)	0.0999	0.0066	0.9404	1.0000				
Emprate2	(0.1151)	(0.1078)	(0.0030)	(0.0093)	(0.1028)	0.0818	(0.0398)	0.0317	(0.0906)	(0.1018)	1.0000			
Goodhealth	(0.2990)	(0.2734)	0.1086	(0.1822)	(0.2467)	0.0638	0.0973	0.0404	(0.0090)	(0.0324)	(0.0287)	1.0000		
Satisfied	0.2002	0.2112	(0.0056)	0.0504	0.1037	0.1129	0.1338	0.0177	0.0299	0.0655	0.0567	0.1023	1.0000	
Optimistic	(0.0700)	(0.0521)	0.0209	(0.0828)	(0.0721)	0.0594	0.0929	(0.0071)	(0.0177)	(0.0084)	0.0032	0.1901	0.2765	1.0000

We next turn to Table 6 below to examine the probability of becoming an entrepreneur over time. The coefficients of the variables appear as hazard ratios. A value greater than one indicates increased likelihood of self-employment/ shorter duration until self-employment (failure), while a value less than one indicates decreased likelihood of self employment/longer duration to self-employment.   Z-scores indicate the significance of the effect as in the multinomial logit analysis.

Table 6: Weibull Parametric Estimations of Becoming Self-Employed during 1996-2001

Weibull regression-log relative-hazard form Gamma frailty.
Dependent variable: SE-months until a person becomes self-employed

	Hazard Ratio
	1.33***
	(6.75)
age2^2	1.00***
	(8.13)
male	2.26***
	(6.32)
married2	2.09***
	(3.04)
residence2	0.99*
	(1.76)
housesize	1.01***
	(3.50)
college2	0.23***
	(7.50)
fatherselfemp	3.03***
	(3.20)
totalwork2	0.85***
	(7.09)
totalwork2^2	1.00***
	(4.99)
emprate2	0.32
	(0.79)
goodhealth	1.40**
	(2.44)
satisfied	1.02
	(0.18)
optimistic	0.88
	(0.98)
No. of subjects	8,074
No. of failures	595
Time at risk/obs.	548,859

Absolute value of z statistics in parentheses

Over time, the factors that increase the likelihood of becoming an entrepreneur are having a self-employed father, being male, being married, age, and good health. All of these variables are significant at the 1% level, with the exception of good health which is significant at the 5% level. This basically corroborates the results of multinomial analysis, and mirrors prior findings in the self-employment literature. Much of that literature focused on Western societies, so it could be said that there are some characteristics of entrepreneurs that appear to be universal.

On the other hand, we see again that having a college education significantly reduces the probability of becoming an entrepreneur over time. The hazard ratio is a scant .23 and this variable is significant at the 1% level. This impact reflects the fact that China is less developed than the West, where higher education among entrepreneurs is more prevalent.

Perhaps China's lagging development also explains why work experience plays little role in predicting the likelihood of entrepreneurship over time. While both total work experience and total work experience squared are statistically significant at the 1% level, the ratios are near to 1. In fact, hazard ratio for total work experience is less than 1, suggesting that increased work experience lowers the likelihood of becoming self-employed. Given the skewed distribution of China's urban entrepreneurs in the wholesale and retail industry prior work experience, like education, may have little value. A budding entrepreneur is less reliant on education or experience for opening a restaurant or clothing store.

Surprisingly, the employment rate is not only not significant to becoming an entrepreneur; it also has the opposite effect of what I predicted. The employment rates for all five cites declined during 1996-2001 while the self-employment rates generally rose. I had conjectured that those being laid off from restructuring would have sought opportunities in self-employment, but the results do not support this. Perhaps it is because those individuals being laid off tend to be older, and therefore they lack either the skills or motivation to become entrepreneurs.

Good health strongly increases the likelihood of becoming self-employed over time and is significant at the 5% level. Neither "satisfied" nor "optimistic are significant. In the multinomial analysis "satisfied" is significant while "optimistic" is not.

The results I have shown on Table 6 are a more robust version of a generic parametric estimation with a Weibull distribution, since I have checked for gamma frailty. As an additional robustness check I also ran the parametric estimation with a log-logistic function (not shown), both with and without the gamma frailty test. The results of these regressions corroborated my findings on Table 6. While the coefficients from the log-logistic regression differ from the hazard ratios of the Weibull regression, the direction was the same and the magnitude was approximately identical.

5. Discussion

There are two interesting findings I wish to highlight as a result of this study. The first is the persistent negative effect of education on the likelihood of self-employment. I would have supposed that more education would raise the probability of being an entrepreneur, which is what most of the literature finds. I discover, however, that most of China's urban entrepreneurs are in industries that do not require much education, namely wholesale/retail and social services. This finding corroborates Bates (1995), who concluded that in industries like construction, more education has a negative effect on the likelihood of becoming an entrepreneur.

The second interesting result concerns the employment rate. Based on the trends I observed in the data non-employment is increasing and self-employment is also increasing in each of the five cities during the 1996-2001 timeframe. It is natural to suppose that the newly unemployed become entrepreneurs, but the emprate2 variable tells another story. This trend variable strongly decreases the likelihood of becoming an entrepreneur and is insignificant. I speculate that those being laid off are older workers who are either unable or unwilling to become entrepreneurs.

In my opinion, these results underscore the importance of context for the study of entrepreneurship. While many variables, such as marital status and wealth, may have universal effects on the likelihood of entrepreneurship, others must be assessed in light of their setting. Urban China is at a point in the evolution of its market economy where most entrepreneurs are clustered in industries where education matters little, unlike the high-tech entrepreneurs of developed nations. Rising unemployment strongly decreases the likelihood of entrepreneurship and is insignificant contrary to expectations, because those workers who become unemployed are likely older and therefore unable or unwilling to transition to self-employment

I regard this paper as a first, modest attempt to characterize China's urban entrepreneurs, a topic about which there is a dearth of research. I have in mind several ways of improving our understanding of this phenomenon in future research. First, I have not examined the role institutions play in the shaping of entrepreneurial opportunities in China. Thus far I have focused only the individual, yet there is a growing literature on institutions and entrepreneurship that I'd like to address.

Second, I plan to delve more deeply into the differences between self-employment and small business ownership. In subsequent work I will differentiate between those who are self-employed and those who actually own small firms and employ others, and I will investigate the sources of these differences. To this point I am content with identifying the self-employed as entrepreneurs.

Third, I will investigate regional variations in the rate of entrepreneurship more intensively. There is no reason to suppose, a priori, that entrepreneurship occurs at the same rate among the five large Chinese cities in our sample. Indeed, among Chinese the residents of Shanghai have a reputation for business savvy in much the same way that Americans regard Manhattanites. It would be worthwhile to investigate the differences among these cities.

Finally, I may attempt to obtain more and better proxies for some of the attitudinal/psychological factors that correlate with entrepreneurship. The CULS may be updated in the summer of 2005. This would allow me a chance to ask respondents more pertinent questions about their motivations and other psychological traits.

References:

Audretsch, D. and Lehmann, E. (2004), "The Effects of Experience, Ownership, and Knowledge on IPO Survival: Empirical Evidence from Germany", *Papers on Entrepreneurship, Growth, and Public Policy*, ISSN 1613-8333.

Barrick, M. and Mount, M. (1991), "The Big Five Personality Dimensions and Job Performance: A Meta-Analysis", *Personnel Psychology*, 44, 1-26.

Bates, T. (1995), "Self-Employment Entry across Industry Groups", *Journal of Business Venturing*, 10(2), 143-156.

Blanchflower, D. and Oswald, A. (1998), "What Makes an Entrepreneur?", *Journal of Labor Economics*, 16(1), 26-60.

Busenitz, L. and Barney, J. (1997), "Differences Between Entrepreneurs and Managers in Large Organizations: Biases and Heuristics in Strategic Decision-Making", *Journal of Business Venturing*, 12(1), 9-30.

Djankov, S., Miguel, E., Qian, Y., Roland, G. and Zhuravskaya, E. (2005), "Who Are Russia's Entrepreneurs?", *Journal of the European Economic Association,* forthcoming.

Delmar, F. and Davidsson, P. (2000), "Where Do They Come From? Prevalence and Characteristics of Nascent Entrepreneurs", *Entrepreneurship and Regional Development*, 12, 1-23.

Dolton, P. and Makepeace, G. (1990), "Self-Employment among Graduates", *Bulletin of Economic Research*, 42(1), 35-53.

Dunn, T. and Holtz-Eakin, D. (1996), "Financial Capital, Human Capital, and the Transition to Self-Employment: Evidence from Intergenerational Links", working paper, Syracuse University.

Evans, M. (1989), "Immigrant Entrepreneurship: Effects of Ethnic Market Size and Isolated Labor Pool", *American Sociological Review*, 54, 950-962.

Evans, D. and Leighton, L.S., "Some Empirical Aspects of Entrepreneurship", *The American Economic Review*, 1989, 79(3), 519-535.

Giles, J. and Park, A. (2005), "How Has Economic Restructuring Affected China's Urban Workers?", *China Quarterly*, forthcoming.

Knight, F. (1921), *Risk, Uncertainty, and Profit*, New York: Houghton Mifflin.

Lee, D. and Tsang, E. (2001), "The Effects of Entrepreneurial Personality, Background, and Network Activities on Venture Growth", *Journal of Management Studies*, 38(4), 583-602.

Lerner, M. and Hendeles, Y. (1993), "New Entrepreneurs and Entrepreneurial Aspirations among Immigrants from the Former U.S.S.R. in Israel", in N. Churchill, S. Birley, W. Bygrave, J. Doutriaux, E. Gatewood, F. Hoy, and W. Wetzel (eds), *Frontiers of Entrepreneurship Research*, Babson Park: Babson College, 562-575.

Mesch, G. and Czamanski, D. (1997), "Occupational Closure and Immigrant Entrepreneurship: Russian Jews in Israel", *Journal of Socio-Economics*, 26(6), 597-611.

Polidoro, F. (2005), unpublished dissertation, University of Michigan.

Rees, H. and Shah, A. (1986), "An Empirical Analysis of Self-Employment in the U.K.", *Journal of Applied Econometrics*, 1, 95-108.

Reynolds, P. (1997), "Who Starts New Firms? Preliminary Explorations of Firms-in-Gestation", *Small Business Economics*, 9, 449-462.

Ritsila, J. and Tervo, H. (2002), "Effects of Unemployment on New Firm Formation: Micro-level Panel Data Evidence from Finland", *Small Business Economics*, 19, 31-40.

Roberts, E. (1991), *Entrepreneurs in High Technology*, New York: Oxford University Press.

Schiller, B. and Crewson, P. (1997), "Entrepreneurial Origins: A Longitudinal Inquiry", *Economic Inquiry*, 35, 523-531.

Schumpeter, J. (1934), *The Theory of Economic Development: An Inquiry into Profits, Capital Credit, Interest, and the Business Cycle*, Cambridge: Harvard University Press.

Woolridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT University Press.

Wooten, K., Timmerman, T., and Folger, R. (1999), "The Use of Personality and the Five-Factor Model To Predict New Business Ventures: From Outplacement to Start-Up", *Journal of Vocational Behavior*, 54, 82-101.

World Bank website, www.worldbank.org/data