RIETI Discussion Paper Series 12-E-002

Market Size and Entrepreneurship

SATO Yasuhiro
Osaka University

TABUCHI Takatoshi
RIETI

YAMAMOTO Kazuhiro
Osaka University

Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Paper Series 12-E-002
January 2012

Market Size and Entrepreneurship*

SATO Yasuhiro†
Osaka University and Research Institute of Economy, Trade and Industry

TABUCHI Takatoshi‡
University of Tokyo and Research Institute of Economy, Trade and Industry

and

YAMAMOTO Kazuhiro§
Osaka University and Research Institute of Economy, Trade and Industry

Abstract

In order to examine the impacts of market size on entrepreneurship, we estimate a monopolistic competition model that involves the workers' decisions to pursue entrepreneurship by using data on Japanese prefectures. Our results show that a larger market size in terms of population density leads to a higher incentive for individuals to become entrepreneurs. A 10% increase in the population density increases the share of people who wish to become entrepreneurs by approximately 1%. In contrast, such a positive effect on the self-employment rate is observed only for prefectures with very high or very low densities. The self-employment ratio is negatively associated with population density in prefectures with medium densities.

Keywords: market size, entrepreneurship, density economies, market expansion.
JEL classification: J62, L26, R12

---

* We thank the Center for Spatial Information Science at the University of Tokyo for providing us data on employment (Employment Status Survey). We are grateful to two anonymous referees, Md. Kamar Ali, Kristian Behrens, Kazumi Hitomi, Koji Karato, Miki Kohara, Noriko Mizutani, Se-il Mun, Ryo Okui, Ping Wang, Atsushi Yoshida, and participants of various seminars and conferences for their helpful comments and suggestions. Of course, we are responsible for any remaining errors. We acknowledge the financial support by the JSPS Grants-in-Aid for Scientific Research (S, A, B and C), the MEXT Grant-in-Aid for Young Scientists (B), and RIETI.
† Graduate School of Economics, Osaka University (Japan); E-mail:ysato@econ.osaka-u.ac.jp
‡ Faculty of Economics, University of Tokyo (Japan) and Research Institute of Economy, Trade and Industry (Japan); E-mail: ttabuchi@e.u-tokyo.ac.jp
§ Graduate School of Economics, Osaka University, (Japan); E-mail: yamamoto@econ.osaka-u.ac.jp

1 Introduction

We generally observe that the distribution of entrepreneurial activities is not even across different locations, and big cities such as New York, London, and Tokyo, apparently exhibit higher entrepreneurship than other cities. Is this view supported once we examine the data on entrepreneurship and what causes the difference in entrepreneurship across locations? To answer this question, we focus on the relationship between regional market size and the entrepreneurship.

Specifically, this paper aims to empirically investigate the association between the size of a region (the size of a home market) and entrepreneurship with regard to the determination of the spatial distribution of entrepreneurship in Japan. For this purpose, building on new economic geography by Ottaviano, Tabuchi, and Thisse [28],[1] we first develop a monopolistic competition model by introducing the workers' decision to pursue entrepreneurship and the technological externalities of agglomeration economies. Therefore, this model involves both pecuniary and technological externalities. We then estimate the relationship.

An increase in the home market size has positive and negative effects on the profits of firms and the decision of (potential) entrepreneurs regarding whether or not to start a new business. Positive effects arise from technological externalities of agglomeration economies due to density economies as evidenced by Ciccone and Hall [7] and Ciccone [6]. Such density economies are generated by technological spillovers and knowledge exchange with heterogenous entrepreneurs and workers (Jacobs [20], Carlino, Chatterjee and Hunt [5]), and by better matching in labor market pooling (Helsley and Strange [19]). Positive effects arise also from pecuniary externalities of market expansion due to the home market effect as shown by Krugman [21]. On the other hand, a negative effect stems from pecuniary externalities of competition among firms and from technological externalities of congestion diseconomies and increases in land rents and wage rates. If the positive effects dominate the negative one, entrepreneurs will establish a new firm in a large market.

In order to determine the dominant effect, we estimate the relationship between the market size and entrepreneurship that is derived from the developed model. In estimation, we employ the share of self-employment as the measure of entrepreneurship. This is the share of self-employed workers among the total employed workers, and is often used in the empirical literature on entrepreneurship. We also use data on the share of people who wish to start a new business among those who are either employed but want to change careers or unemployed but want to find a job. We consider this index to represent *potential*, *ex ante* entrepreneurship and indicate the degree of incentive of people to start a new business,

[1] Alternatively, we could use another new economic geography model such as Krugman [22] and Forslid and Ottaviano [11]. However, because these wage equations are highly nonlinear, we cannot derive equations suitable for empirical estimation of the relationship.

which is in contrast to the share of self-employment that represents *observed, ex post* entrepreneurship.

Following Ciccone and Hall [7] and Ciccone [6], we employ the population density as the index of market size, and investigate its relationship with observed/potential entrepreneurship. Our results show that the population density can be negatively related to the observed entrepreneurship. More specifically, although we obtain evidence of the positive effects of market expansion and density economies in the case of service industry, the negative effect of competition dominates the positive effects except in cities with very high or very low density. In the case of manufacturing industry, we cannot detect evidence of the density economies. In contrast to the observed entrepreneurship, potential entrepreneurship is positively related to the population density: a ten percent increase in the population density raises the share of ex ante entrepreneurs by approximately one percent. This result may give support to the view of "nursery cities" a la Duranton and Puga [10], where large cities are places that specialize in creating new firms and succeed through innovation. Thus, the market size has opposite effects on entrepreneurship in different phases.

Several studies have presented the relationship between the market size and entrepreneurship.[2] According to Berry and Reiss [4], empirical studies based on micro data showed that there are thresholds of regional population size enabling firms to establish themselves as going concerns, and that firm entry is high in regions with large population size. Similar results on the relationship between regional population size and firm entry are obtained by Reynolds et al. [30], who showed that higher growth of regional population leads to higher birthrate of firms. Harada [17] used data on workers who wish to start a business and indicated that workers are more likely to become entrepreneurs in regions with a larger gross prefectural domestic product. Okamuro and Kobayashi [25] and Okamuro [26] showed that the regional growth rate and firm density positively affect the entrepreneurship. Rosenthal and Strange [31] showed that entrepreneurship is higher in districts with higher employment density using data on new establishments in the New York Metropolitan Area. Glaeser and Kerr [14] measured entrepreneurship by new establishments that are independent from existing firms and presented that new entrants are likely to be attracted to areas with high overall levels of customers and suppliers, and strongly drawn to areas with many smaller suppliers. However, in the results of Glaeser [13], who adopted the self-employment rate and average firm size as the indices of entrepreneurship, the effect of employment density on entrepreneurship is insignificant. Glaeser, Kerr and Ponzetto [16] found little evidence of large sales on entrepreneurship. It suggests that lower fixed costs and a larger pool of entrepreneurial people lead to higher entrepreneurship. To sum up, existing studies on the relationship between market size and entrepreneurship show a highly mixed picture.

[2] The detailed discussion and survey on the research on entrepreneurship in the context of urban economics are given in Glaeser, Rosenthal and Strange [15].

The novelty of our paper can be illustrated by the following three features. First, we build a model to deal with the entrepreneurial decision in the monopolistic competitive market in a regional economy, and derive a reduced form of an equation for estimation from this model with microfoundation. Second, we measure entrepreneurship by the share of potential entrepreneurs as well as the share of observed entrepreneurs. This enables us to see if the market size is associated with the entrepreneurship in different stages of entrepreneurship. Third, we consider the possibility of non-linearity and non-monotonicity in the relationship between market size and entrepreneurship that has theoretical background. This allows flexibility in capturing this relationship and is in contrast to the existing studies mentioned above, most of which assume linearity and monotonicity.

The remainder of the paper is organized as follows. Section 2 presents a model of new economic geography with entrepreneurship and characterizes entrepreneur formation. Section 3 analyzes the entrepreneur formation and investigates the positive effects arising from density economies and market expansion. It provides a set of empirical results by using Japanese prefectural-base data. Section 4 checks the robustness of these results by different methods of estimations. Section 5 concludes and suggests future research directions.

2 Basic model

In this section, we present a simple new economic geography model involving entrepreneurship in order to examine how the market size affects entrepreneurship.

There are two goods in a region. The first good is homogeneous. Consumers have a positive initial endowment of this good that is also produced using labor as the only input under constant returns to scale and perfect competition. This good can be traded freely and is chosen as the numéraire. The other good is a horizontally differentiated product that is supplied by using labor under increasing returns to scale and monopolistic competition.[3]

There is a mass L of individuals who move between the two sectors according to the wage differential provided later. As in Ottaviano, Tabuchi, and Thisse [28], consumers' preferences are identical across individuals and are described by the following utility function of the quasi-linear form:

$$U = \alpha \int_0^n q(v)\mathrm{d}v - \frac{\beta}{2}\int_0^n (q(v))^2\mathrm{d}v - \frac{\gamma}{2}\left(\int_0^n q(v)\mathrm{d}v\right)^2 + q_0,$$

where $q(v)$ is the quantity of variety v, n is the mass of varieties, q_0 is the quantity of the numéraire, and α, β, and γ are positive parameters. We omit subscripts denoting regions until Section 3.3 for readability.

[3] This differentiated good may be services that comprise a continuum of varieties without being exported to the rest of the world. See Appendix 1 in Sato, Tabuchi and Yamamoto [32] when the good is freely traded with the rest of the world.

The budget constraint of an individual is given by

$$\int_0^n p(v)q(v)\mathrm{d}v + q_0 = \overline{q}_0 + I,$$

where $p(v)$ is the price of variety v, I is the individual income, and $\overline{q}_0$ is the initial endowment, which is supposed to be sufficiently large to ensure positive demand for the numéraire. As we see below, I is the wage income $\overline{w}$ for a worker and the rewards w to run a firm as an entrepreneur.

Maximizing the utility subject to the budget constraint yields the individual demand

$$q(v) = \frac{\alpha\beta + \gamma P}{\beta\left(\beta + \gamma n\right)} - \frac{1}{\beta}p(v),$$

where $P = \int_0^n p(v)\mathrm{d}v$ is the price index.

Turning to the supply side, the technology in the homogeneous good sector requires one unit of labor in order to produce $\overline{w}$ units of output. Because this good is a numéraire with costless trade, its price must be equalized to one across regions. Then, marginal cost pricing implies that the equilibrium wage of the sector is always equal to $\overline{w}$, which differs across regions due to exogenous differences in labor productivity.

In the differentiated good sector, each firm supplies a single variety under increasing returns to scale so that n is also regarded as the number of firms. The fixed requirement of entrepreneurial labor is one and the marginal requirement of labor is normalized to zero for mathematical tractability.

A firm produces variety v and chooses a price $p(v)$ that maximizes its profit:

$$\pi(v) = \left(p(v) - \tau\left(D\right)\right)q(v)L - w, \tag{1}$$

where $\tau\left(D\right)$ represents the marginal costs of production, which include distribution costs involving the retail and wholesale costs within the region.[4] The marginal costs $\tau\left(D\right)$ are subject to technological externalities discussed in the introduction. They can be related to population density in two possible ways. On the one hand, a higher density may lead to lower marginal costs because of greater Marshallian externalities such as knowledge spillovers and labor pooling or because of lower distribution costs. It is well known in urban economics (Alonso [1]) that the average population density of a city is positively associated with city size, and hence, density economies may be regarded as agglomeration economies accruing from the high density or large size of the city. This implies that population density is high in regions involving large cities, where agglomeration economies are large. On the other hand, higher population density leads to a higher land rent and heavier congestion, which raise the marginal costs. However, the impact of a high land rent on the entrepreneurial decision varies from industry to industry. In general, firms in the manufacturing industry are more land intensive than those in the service industry,

[4] The specification (1) does not fully capture the characteristics of firms trading between regions. Rather, this specification fits the service and manufacturing industries without exporting outside the region.

indicating that a high land rent discourages entrepreneurship in the manufacturing industry relative to the service industry.[5] If the positive effects of density economies dominate the negative effects of congestion, τ is decreasing in the population density D. If the opposite holds true, τ is increasing in D.

Because firms are symmetric in the same region with respect to production technologies, they charge the same price in equilibrium, and therefore, we drop v hereafter. This is given by

$$p = \frac{\alpha\beta + \tau(D)(\beta + \gamma n)}{2\beta + \gamma n},$$

which exhibits the so-called procompetitive effect $\partial p / \partial n < 0$. Plugging the price into (1), the profit made by a firm is rewritten as follows:

$$\pi = \frac{(\alpha - \tau(D))^2 \beta L}{(2\beta + \gamma n)^2} - w.$$

Under free entry of firms, this profit equals zero so that the equilibrium rewards w to be an entrepreneur in the differentiated good sector is determined as

$$w = \frac{(\alpha - \tau(D))^2 \beta L}{(2\beta + \gamma n)^2}. \tag{2}$$

We follow existing microeconomic studies on the formation of entrepreneurship by assuming that individuals choose to become either a *worker* who obtains wage income $\overline{w}$ or an *entrepreneur* who obtains rewards w. Because the equilibrium wage in the homogeneous good sector is $\overline{w}$, the arbitrage between the two sectors by individuals requires

$$w = \overline{w} \tag{3}$$

so that individuals are indifferent. As will be seen later, the wage $\overline{w}$ differs across regions due to idiosyncratic regional conditions, and hence, the factor prices are not equalized across regions.

Let e be the share of entrepreneurs in the labor force so that the number of entrepreneurs in the region is eL. Since the mass of varieties, n, is equal to the number of firms, the labor market clearing implies $eL = n$. Solving this equation with (2) and (3), we can uniquely determine the entrepreneur share as[6]

$$e = \frac{1}{\gamma L}\left[\frac{(\alpha - \tau(D))\sqrt{\beta L}}{\sqrt{\overline{w}}} - 2\beta\right]. \tag{4}$$

[5] Parts of the positive effects of density economies for firms are negated by an increase in the land rent because the density economies are capitalized by the land rent. See Arzaghi and Henderson [2] for an empirical evidence of such capitalization in the advertising agency industry in Manhattan.

[6] If entrepreneurs are mobile across regions as in new economic geography models, they would migrate to a region with a higher utility until a spatial equilibrium is attained. This makes the population density D and the wage $\overline{w}$ in the RHS of (4) endogenous through the change in the number of firms. The resulting expression is no longer amenable to empirical analysis. We instead assume that they are immobile since people engage in interprefectural migration only 2.3 times in their entire life in Japan (Nakajima and Tabuchi [24]). In this sense, it would be fair to say that our estimation is not "structural" in the standard sense. Possible endogeneity biases are examined in Section 4.2.

Following Ciccone and Hall [7] and Ciccone [6], we use the population density as the measure of market size. This controls for variations in the prefectural areas, which are the units of regions in the empirical analysis in later sections. Let h be the area of inhabitable land so that $L = hD$ holds. Then, we can rewrite (4) as a function of population density. Taking the logarithm of it, we have

$$\ln(e) = \ln\left[\frac{(\alpha - \tau(D))\sqrt{\beta h D}}{\sqrt{\overline{w}}} - 2\beta\right] - \ln(\gamma h D). \tag{5}$$

The first term of the RHS in (5) includes two effects of population density D on entrepreneurship e. One is the *density economies/congestion diseconomies* that are described by changes in $\tau(D)$. If the density economies dominate the congestion diseconomies, increasing density reduces the marginal costs, enhances firm profits, and thus promotes entrepreneurship. However, if the former is dominated by the latter, increasing density discourages entrepreneurship. The other is the effect of *market expansion* that is represented by $\sqrt{\beta h D}$. An increase in market size enlarges the amounts of sales for a given cost-price margin $\alpha - \tau(D)$, raises profits, and augments the number of entrepreneurs.

The second term of the RHS in (5) represents a negative effect of D on e. This term indicates the *competition among firms* because it involves the substitutability parameter γ. The concentration of firms reduces firm revenues and profits because of the substitutability between varieties, and thus, undermines the incentive for pursuing entrepreneurship in the differentiated good sector.

Which effect dominates the others is not clear a priori by theory. Therefore, we resort to empirical analysis in order to reveal the overall effect of market size on the formation of entrepreneurship.

As we see later in detail, we estimate (5) by using data on the self-employment rate. Equation (5) can be regarded as the long-run relationship between market size and entrepreneurship. In addition, we examine how an incentive to become entrepreneurs responds to deviations from the long-run relationship by using the data on willingness to start up a business.[7] More precisely, letting e_p, e_o and $\widehat{e}_o$ denote the index of willingness to be entrepreneurs, the observed share of entrepreneurs and the share of entrepreneurs predicted by estimated equation of (5), we assume the following relationship:

$$e_p = f\left(\frac{\widehat{e}_o}{e_o}, D\right). \tag{6}$$

This equation represents that the incentive to be entrepreneurs depends on the ratio of the predicted entrepreneurship $\widehat{e}_o$ estimated by (5) to the observed entrepreneurship e_o, and on the market size D. $\widehat{e}_o$ can be interpreted as the capacity of market given various conditions such as D and $\overline{w}$. If the observed entrepreneurship e_o is high relative to $\widehat{e}_o$, people would think of this market as "tight" and have less incentive to become entrepreneurs. If $\widehat{e}_o/e_o$ is high, such a market is likely to allow more entrepreneurs

[7] As shown by Behrens and Robert-Nicoud [3], observed entrepreneurship e_o is the results of competition among potential entrepreneurs e_p, only a part of which can succeed in establishing a new firm and survive competition among firms.

to survive and people are more willing to become entrepreneurs. Here, we explore whether we observe such responses of incentives of being entrepreneurs to market tightness. Furthermore, we consider the possibility that the market size D affects incentives to be entrepreneurs. This dependence may be ascribed to spillovers of information among people, availability of appropriate human resources, and so on.

3 Empirical analysis

3.1 Model specification

In order to obtain an explicit equation that can be estimated, we have to specify the functional form of the distribution costs. We linearly approximate the distribution costs $\tau(D)$ as

$$\tau(D) = t_0 - t_1 D,$$

where t_0 and t_1 are constants. We assume $\alpha > t_0 > 0$ so that the differentiated good is produced. The sign of t_1 is to be determined by empirical analysis. If t_1 is shown to be positive, the density economies dominate the congestion diseconomies, whereas if t_1 is negative, the congestion diseconomies outweigh the the density economies. Thus, t_1 captures the net effect of population density on the marginal costs. Substituting this $\tau(D)$ into (5), the equation to be estimated is as follows:

$$\ln(e) = a_0 - \frac{1}{2}\ln(\overline{w}) - \ln(D) - \frac{1}{2}\ln(h) + \ln\left[-2\left(\frac{\overline{w}}{h}\right)^{1/2} + a_1 D^{1/2} + a_2 D^{3/2}\right], \tag{7}$$

where a_0, a_1, and a_2 are defined as

$$a_0 \equiv \ln\left(\frac{\beta}{\gamma}\right), \qquad a_1 \equiv \frac{\alpha - t_0}{\sqrt{\beta}}, \qquad a_2 \equiv \frac{t_1}{\sqrt{\beta}}.$$

As shown in the previous section, the population density as a surrogate for home market size has positive and negative effects on the entrepreneurship. By estimating equation (7) using the data explained below, we can determine whether the positive effect or the negative effect is dominant.

For empirical purposes, we specify a log-linear form of (6) as:

$$\ln(e_p) = \ln(\lambda_0) + \lambda_1 \ln\left(\frac{\widehat{e}_o}{e_o}\right) + \lambda_2 \ln(D). \tag{8}$$

where λ_0, λ_1 and λ_2 are parameters to be estimated. If $\lambda_1 > 0$ ($\lambda_1 < 0$), a tighter market discourages (encourages) entrepreneurship. If $\lambda_2 > 0$ by empirical analysis, then people have stronger incentives to become entrepreneurs in a larger market. If $\lambda_2 < 0$, people have less incentives.

3.2 Data and preliminary results

We use the data for years 1992, 1997 and 2002 on Japanese prefectures in order to estimate (7) and (8). We choose prefectures as the spatial unit of analysis because of the following reasons. First, availability of data is much better for prefectures than for other types of spatial units. If we use a smaller unit such as cities, we can use a limited number of data. Second, the economy is more or less homogeneous within a prefecture since the feudal system in the Edo Period, which was established about three hundred years ago. We employ two indices for entrepreneurship data. One is the share e_p of people who wish to start up a new business among those who are either employed but want to change careers or unemployed but want to find a job. The other is the share e_o of self-employment, which is calculated by the ratio of the number of self-employed people to the number of people in all jobs. All of these data are taken from the Employment Status Survey, published by the Ministry of Internal Affairs and Communications of Japan. e_p is interpreted as the share of potential entrepreneurs and represents the incentive to be entrepreneurs, whereas e_o is the observed share of entrepreneurs. Because data on self-employment is available for each industry, we examine the difference in observed entrepreneurship among industries.[8] Here, we consider the self-employment share for all industries (e_o), that in the manufacturing industry (e_{om}), and that in the service industry (e_{os}).

It is well known that there is no perfect measure of entrepreneurship (Glaeser et al. [15]). Our measures also have both good and bad points. Although e_p represents the share of people who is willing to start up a new business, it has no information on how many of them really become entrepreneurs. It shows the incentive of people and hence, the potential entrepreneurship, which itself is interesting to look into though it does not imply the entrepreneurship in the standard sense. In contrast, e_o describes the share of people who already have started in business. This index reflects the result of a firm creation/destruction process through time whereas e_p is measured by the wishes of people at one point in time and there is no accumulation through time in it. Moreover, e_o also counts those who are not considered to be entrepreneurs, such as those who took over their business from their parents. We think it important to use various measures in order to unveil the characteristics of entrepreneurship.[9] The population density D_{pop} is computed by dividing the prefectural population L (Population Estimates, published by the Ministry of Internal Affairs and Communications) by the inhabitable area H (Social Indicators by Prefecture published by the Ministry of Internal Affairs and Communications).[10] For the wage income

[8] We specify each industry by using the large category of Japan Standardized Industrial Classification (version 10).

[9] We also used the number of establishments per capita for all industries, that for manufacturing industry, and that for service industry as alternative indices for e_o, e_{om}, and e_{os} in order to check the robustness of our results. All the results are similar to those obtained in this paper.

[10] Although the present study focuses on the population density, the employment density may be more important for

$\overline{w}$ of workers, we use the monthly total cash earnings (Monthly Labour Survey, published by Ministry of Health, Labour and Welfare). Table 1 lists the summary statistics and sources of all variables.[11]

[Insert Table 1 about here]

We first check the overall distributions of the shares of entrepreneurs across regions by using kernel density estimation. The estimated distributions of e_p, e_o, e_{om} and e_{os} are depicted in Figures 1-(a) to (d), respectively, where the horizontal axis represents the share of entrepreneurs and the vertical one indicates the density.[12]

[Insert Figure 1 about here]

Firstly, note that the values for entrepreneurship vary significantly across regions in each year: 5 to 11 percent with respect to potential entrepreneurs and 7 to 19 percent with respect to observed entrepreneurs. On average, the share of observed entrepreneurs is higher than that of potential entrepreneurs. The difference cannot be explained by whether accumulation through time is reflected or not, but can be explained at least partly by whether those who inherit business from their parents are included or not. That is, they are included in the observed share of entrepreneurs, but not in the potential share of entrepreneurs.

Secondly, both figures show that the distribution has shifted from year to year, which indicates that the entrepreneurship in all regions is affected by changes in the national economy as a whole. We observe the steady fall in the level of entrepreneurship, which may be due to the large trend in Japanese economy, such as declining Japanese economy during the 1990s and aging of population.

Before estimating equation (7), we provide some preliminary estimations of the relationship between population density and entrepreneurship. Table 2 shows the pooled, fixed-effects and random-effects estimation results in which the dependent variables are $\ln(e_p)$, $\ln(e_o)$, $\ln(e_{om})$ and $\ln(e_{os})$, and the independent variables are $\ln(D_{pop})$, $\ln(G_p)$ and the year dummy. G_p is the per capita gross prefectural domestic product (GPDP) (Prefectural Accounts, published by the Department of National Accounts,

entrepreneurs. However, results are unaltered even if we use the employment density. This is because the correlation coefficient between population density and employment density is extremely high ($\fallingdotseq 0.998$).

[11] The Hokkaido prefecture is the largest in terms of inhabitable area. The largest metropolitan area in Hokkaido is Sapporo with more than two million people, where many socioeconomic activities are concentrated. In contrast, there are many small cities and villages which lower the average population density of Hokkaido prefecture. This implies that the average population density of Hokkaido does not represent its regional characteristics, and hence, we have eliminated it from our sample, which leads to our present sample size of forty-six prefectures times three years.

[12] We used Silverman's default bandwidth in the kernel estimation. See Härdle [18] for more details.

Cabinet Office) and controls for the differences in overall economic conditions in prefectures. The year dummy controls for the year effects observed in Figure 1.

[Insert Table 2 about here]

In Table 2, the first, the second, and the third columns for each case represent the results estimated by OLS regression, the fixed effects model, and the random effects model, respectively. These estimated results indicate that potential entrepreneurship is positively related to population density, whereas observed entrepreneurship for all industries is negatively related to the population density. For the manufacturing industry, the observed entrepreneurship has no clear (significant) linear relationship to the population density. For the service industry, the estimated coefficients of the population density are positive but mostly insignificant. These results show the possibility that the relationship between entrepreneurship and market size, which is measured by the population density, can vary across the stages of entrepreneurship and across industries. However, these estimation results are far from conclusive because of the following reasons. First, although we expected unobserved prefectural effects, the fixed effects model turns out to be uninformative. This is because we have large standard errors due to little within variation in our sample. Put differently, while we have multiple years of data, the persistence of density means that we only observe cross-sectional patterns. Second, we cannot expect a random effects model to work. In the random effects model, the prefectural effect is thought as random. This is justified when we only sample a subset of the entire population of subjects. In our analysis, this does not hold true. Moreover, random effects methods do not control for unmeasured, stable characteristics of the prefectures, which we think highly relevant to our analysis. Note also that if there is a correlation between the prefectural effects and the covariates, the estimate of the random effects model is no longer consistent although the estimate of the fixed effects model is consistent. Hence, in the following analysis, we treat our data as pooled cross section and estimate (7) and (8) combined with relevant control variables in order to further analyze the relationship between population density and entrepreneurship.

3.3 Model for estimation

In estimating (7), we include the control variables that are considered to affect the entry decision of entrepreneurs according to the literature involving empirical studies of firm entry following the tradition of Orr [27].[13] It should be emphasized that the overall results remain qualitatively the same in the absence of the following control variables as will be shown in Table 3. First, we append the per capita GPDP (G_p) as before, the year-to-year comparison of the per capita GPDP ($g_p = G_{pt}/G_{pt-1}$), and

[13] See Geroski [12] for surveys on empirical studies of firm entry.

the price-cost margin (M), which describe the overall economic conditions of the corresponding region. G_p and g_p represent the current condition and future prospects, respectively. M is defined as $M =$ (GPDP$-$intermediate output$-$indirect tax less subsidies$-$compensation of employees)/GPDP, which is a proxy for the profitability of the business in the region. Second, in order to control the difference in industry structures, we add the shares of agriculture, manufacturing, and public sectors in the GPDP (S_{ha}, S_{hm}, and S_{hp}). All the data for G_p, g_p, M, S_{ha}, S_{hm}, and S_{hp} are sourced from Prefectural Accounts, published by the Department of National Accounts, Cabinet Office. Third, we also control the effective job opening to job applicant ratio (V), which is available in the Monthly Report of Public Employment Security Statistics, published by the Ministry of Health, Labour and Welfare. Table 1 provides the basic statistics and sources of these variables.

Introducing all of these variables into (7), the equation to be estimated for observed entrepreneurship is

$$\begin{aligned}\ln(e_{o,it}) = a_0 - \frac{1}{2}\ln(\overline{w}_{it}) - \ln(D_{pop,it}) - \frac{1}{2}\ln\left(h_{it}\right) + \ln\left[-2\left(\frac{\overline{w}_{it}}{h_{it}}\right)^{1/2} + a_1 D_{pop,it}^{1/2} + a_2 D_{pop,it}^{3/2}\right] \quad (9)\\ + b_1\ln(G_{p,it}) + b_2\ln(g_{p,it}) + b_3\ln(M_{it}) + b_4\ln(S_{ha,it})\\ + b_5\ln(S_{hm,it}) + b_6\ln(S_{hp,it}) + b_7\ln(V_{it}) + \text{year dummy} + u_{it},\end{aligned}$$

where u_{it} is the standard error term. We estimate this equation by the nonlinear least squares (NLS) method.

By estimating (9), we obtain data on $\widehat{e}_o$, with which we can also estimate the potential entrepreneurship (8):

$$\ln(e_{p,it}) = \ln(\lambda_0) + \lambda_1\ln\left(\frac{\widehat{e}_{o,it}}{e_{o,it}}\right) + \lambda_2\ln(D_{pop,it}) + \mu_{it}, \quad (10)$$

by the ordinary least squares (OLS) method, where μ_{it} is the standard error term.

3.4 Estimation results

Estimation results of (9) are reported in Tables 3 and 4. Table 3 displays the results using data for three years and Table 4 provides the results for each year. In Table 3, columns 1 and 2 are for all industries, columns 3 and 4 are for manufacturing industry, and columns 5 and 6 are for service industry. For each case, we present results with and without control variables.

[Insert Tables 3 and 4 about here]

We observe from these tables that the estimated value of a_2 is positive for all industries and service industry, and that it is less significant and sometimes negative for manufacturing industry. The sign of

a_2 represents the sign of t_1, which shows whether the density economies or congestion diseconomies in production are dominant. For all industries and service industry, the density economies prevail whereas it is ambiguous in manufacturing industry. This would reflect that the impacts of the land market are different in one industry to another. In general, production in manufacturing industry requires intensive land relative to that in service industry. This suggests the impacts of congestion in the manufacturing industry are likely to be stronger than those in the service industry.

Using the estimated regression equations (9) and (10), we simulate the elasticity $(D_{pop}/e)(\partial e/\partial D_{pop})$ of the entrepreneur share e_o with respect to the population density D_{pop}. Figure 2 provides the simulation results for the relevant range of D_{pop} when the elasticity is evaluated at the means of the inhabitable area h and $\overline{w}$. The horizontal axis represents the logarithm of D_{pop}, whereas the vertical axis describes the elasticity.

[Insert Figure 2 about here]

The first, second, and third graphs indicate the results of all industries (e_o), manufacturing industry (e_{om}), and service industry (e_{os}), respectively. We observe from this figure that population density is likely to have a negative impact on the observed entrepreneurship for the most part. This may indicate that due to intense competition between firms, *a higher population density eventually causes quite a few entrepreneurs to cease their business operations*. However, for very large metropolitan areas such as Tokyo and Osaka and for low density areas, the market size is positively related to the observed entrepreneurship. As shown below, this is caused by the strong density economies for very large metropolitan areas, whereas this is caused by the strong market expansion effect for low density areas.

The impact of population density on the entrepreneurship differs between industries. In Figure 2, the population density has a negative impact on e_{os} for intermediate population density and a positive impact on e_{os} for low and high population density in the service industry. On the other hand, it always has a negative impact on e_{om} in the manufacturing industry. These results indicate that agglomeration economies arising in urban areas are more important for the service industry than the manufacturing industry.

Using the estimation result of (9) for all industries, we next estimate (10). We first construct data on $\widehat{e}_o/e_o$ from the estimation result [2] in Table 3 for three years, and then regress it on $\ln(D_{pop})$. The OLS estimation results are shown in Table 5.

[Insert Table 5 about here]

Surprisingly, the estimated value of λ_1 is negative, implying that the potential entrepreneurship e_p rises as the market becomes tight, although it is insignificant. This may reflect (i) that a very tight market is regarded as a signal that firms can survive easily in that market, or (ii) that there are so many turnovers of firms in a very tight market that there are both much entry and exit of firms at the same time. On the other hand, the estimated value of λ_2 is positive, which implies that people in large cities with high D_{pop} have stronger incentives to become entrepreneurs for given market tightness. Since λ_2 is about 0.095, a ten percent increase in the population density increases the share of people who wish to become entrepreneurs by one percent. It may suggest that a high population density is considered to be an opportunity to start a business. This is because a high population density refers to large and diversified cities, where innovation is fostered and new products are developed (Duranton and Puga [10]).

3.5 Density economies or market expansion?

Population density potentially has a positive effect on entrepreneurship, though the range of a negative effect dominates that of a positive effect. As discussed at the end of section 2, this positive effect can arise from two sources: density economies and market expansion. We have seen in Table 3 that such a positive effect prevails in all industries and service industry. Focusing on the positive effect and determining the main source from the two yield some interesting findings.

The positive effect arises from changes in $[\alpha - \tau(D)]\sqrt{\beta h D}$ in (5). Therefore, we can determine the dominant source by computing the derivative of this term with respect to the population density D; that is, the positive effect can be decomposed as follows:

$$\frac{d}{dD}\left\{[\alpha - \tau(D)]\sqrt{\beta h D}\right\} = \underbrace{t_1\sqrt{\beta h D}}_{\substack{\text{density economies} \\ \equiv A(D) \\ \text{if } t_1 > 0}} + \underbrace{\frac{[\alpha - \tau(D)]\sqrt{\beta h}}{2\sqrt{D}}}_{\substack{\text{market expansion} \\ \equiv B(D)}}.$$

The relative strength of these two effects is given by

$$\frac{A(D)}{B(D)} = \frac{2D}{(\alpha - t_0)/t_1 + D}.$$

Because $(\alpha - t_0)/t_1 = a_1/a_2$ from (7), the estimates of this ratio can be obtained by using Table 3. We then depict the respective ratios of $A(D_{pop})/B(D_{pop})$ for all industries and service industry in Figure 3, respectively.

[Insert Figure 3 about here]

We observe that market expansion is more important than density economies for a wide range of population density. Density economies are more important than market expansion for entrepreneurship in regions with high population density like Tokyo and Osaka. This suggests that density economies are the dominant factor for entrepreneurship in regions involving large cities, whereas the market expansion is insignificant in such cities.

4 Robustness checks

We conduct two different estimations in order to check the robustness of our results as follows.

4.1 Nonparametric estimation

The analysis in the previous section may depend on the functional form of (7). While this restriction enables us to simulate the responses of entrepreneurship to the changes in population density, it may not be comprehensive enough to capture the actual relationship between entrepreneurship and population density. We therefore do not assume the explicit functional form in estimating the relationship between entrepreneurship and population density. This can be done by estimating the mean of the entrepreneur share conditioned on prefectural population density by Kernel estimation. Specifically, we estimate the values of $E(e_t|\ln(D_{pop}))$ for $t = p, o, om, os$.[14] The estimated conditional means are shown in Figures 4-(a) to (d). The results presented in Figures 4-(b) and (d) verify the findings regarding the observed entrepreneurship for all industries and for the service industry obtained in the previous section. However, there are slight differences between results here and those obtained in the previous section for the manufacturing industry. Figure 4-(c) shows a positive relationship between population density and entrepreneurship for small and very large cities. Such differences could come from not having the same set of control variables in the non-parametric estimation as in the NLS estimation. Finally, the result on the potential entrepreneurship in Figure 4-(a) confirms the results obtained in the previous section.

[Insert Figure 4 about here]

4.2 Endogeneity

Finally, we discuss the endogeneity issue. One may suspect that the differences in entrepreneurship among regions may affect the households' decision to migrate between regions, and hence change the population density. In order to check whether they make a significant impact on the estimation results or not, we conduct the instrumental variable (IV) estimation.

Since the pioneering work by Ciccone and Hall [7], it has been standard to use historical variables such as long lags of population density to instrument for the density of local population. This approach

[14] As earlier, we used Silverman's default bandwidth in the kernel estimation.

is thought to be appropriate "to the extent that (a) there is some persistence in the spatial distribution of population" (relevance: $Cov(D_{pop,i}, Z_i|\,.) \neq 0$, where Z denotes the set of instruments) and "(b) the local drivers of high productivity (entrepreneurship in our estimation) today differ from those of a long gone past" (exogeneity: $Cov(u_i, Z_i|\,.) = 0$) (Combes et al. [8], pp. 27). Here, we use the prefectural population density for years 1920 and 1940 (Japanese Imperial Statistical Yearbook), which are before the massive destruction by airborne attacks during World War II.[15]

As shown in Davis and Weinstein [9], Japanese regional population distribution has been stable before and after WWII despite the fact that most of the Japanese major cities are completely destroyed by the bombings. In fact, the correlation coefficient between the prefectural population density in 1920 or 1940 and that in 1992, 1997 or 2000 exceeds 0.9, and R^2 of univariate regression of the instrument on current population density exceeds 0.8 for both instruments. Moreover, we report in Tables 6-(a) and (b) the F-statistics relevant to the weak instrument tests (Stock and Yogo [33]). The F-statistics range from 120 to 620, which clearly exceed the rule-of-thumb critical values for the weak instrument test for the small number of instruments. Thus, it would be safe to say that our instruments are not weak and satisfy criterion (a).

We now turn to the exongeneity. First note that 1920 is the year before WWII (1939-1945), and Japan is in the midst of WWII in 1940. During this period, Japan experienced extremely rapid industrial changes. According to Miwa and Hara [23] (pages 7-29), the employment shares of the primary, secondary, and tertiary industries were 53.8, 20.5, and 23.7, respectively in 1920, then 44.3, 26.0, and 29.0 in 1940, and then 5.0, 29.5, and 64.3 in 2000.[16] The shares of the net domestic product (NDP) in the primary, secondary, and tertiary industries were 30.2, 29.1, and 40.7, respectively in 1920, and 18.8, 47.3, and 33.8 in 1940. The corresponding figures for 2000 (the share in the gross domestic product (GDP)) are 1.8, 28.6, and 69.6, respectively. These figures imply that the main industry has changed from agriculture in 1920 to manufacturing in 1940, and to service in 2000. Although the shares of the manufacturing sector do not differ much between these years, there have been significant changes in the industrial structure of the manufacturing sector during the past sixty years. Under the 2-digit classification, the textile mill products sector occupied the top share in the value of production (27.0 percent) of the manufacturing industry under the 1-digit classification in 1929 whereas the electrical machinery, equipment and supplies sector comes first (53.7 percent) in 2000. Moreover, in 1938, Japanese Imperial government enforced the National General Mobilization Act, which enabled the government to take control of many economic activities from 1938 to 1945. Finally, the air raids during WWII destroyed almost all major cities in Japan. In summary, Japan has experienced the drastic changes in industrial structure and economic

[15] These population densities are based on the total areas of prefectures, not on the inhabitable areas of prefectures.

[16] The unclassified category accounts for the remaining share.

system, and Japanese cities were almost "reset" physically. Hence, it is unlikely that entrepreneurship today is related to the population density in 1920 or in 1940, leading to satisfaction of criterion (b).

We estimate (9) by the nonlinear two-stage least squares (N2SLS) method and (10) by the two-stage least squares (2SLS) method. The results of IV estimation are shown in Table 6 and Figure 5.

[Insert Table 6 about here]

[Insert Figure 5 about here]

Observe that the estimation results are qualitatively very similar to those obtained in the previous section. We know from Table 6-(a) that the density economies dominate the congestion diseconomies (i.e., the estimated value of $a_2 = t_1/\sqrt{\beta}$ is significantly positive) for all industries and for the service industry, whereas the estimated value of a_2 is not significantly different from zero for the manufacturing industry. However, the estimated values of a_i are slightly different from those estimated by NLS (Table 3). Also, Table 6-(b) shows that a higher population density leads to higher potential entrepreneurship e_p. In Table 6, the coefficient of log(population density) is 0.085. This implies that a ten percent increase in the population density increases the share of potential entrepreneurs by about one percent, which is very similar to that obtained by OLS (Table 5).

Figure 5 displays how such changes in estimated values of a_i affect the elasticities of observed entrepreneurship with respect to population density and the relative importance of density economies to market expansion. In Figure 5-(a), the simulated elasticities are positive for all industries and for the service industry, and negative for the manufacturing industry but a lesser extent than in Figure 2. These results show that the difference in the impacts of population density on entrepreneurship is more prominent under IV estimation and indicate the existence of a negative bias due to endogeneity. Such a negative bias is thought to come from the negative correlation between the population density and the error term of (9). One possible explanation on this correlation may be that a positive shock on local economy, which stimulates the entrepreneurship there, raises the income level to increase the cost of living via rises in land rents. If such increases in the cost of living dominate the positive effects, this shock would lower the population density in the long run.

The results with the historical data of population density add slight modifications to those obtained in the previous sections. The modified results are summarized as follows. (i) For all industries and the service industry, the impact of population density on observed entrepreneurship is positive in small and large cities, and the impact is smaller (or even negative) in medium sized cities. (ii) The positive effect comes mainly from the density economies in large cities, whereas it stems from market expansion in small cities. (iii) The impact of density for the manufacturing industry is negative. (iv) A ten percent increase in the population density raises the share of potential entrepreneurs by about one percent.

5 Concluding remarks

This paper investigated the impact of market size on entrepreneurship. Our results show that a larger market stimulates potential entrepreneurship and provides the workers with stronger incentives to start a new business, but may lead to a lower ratio of self-employment, i.e., many potential entrepreneurs and few observed entrepreneurs. The analysis by industry indicates that the population density may have marginally positive impacts on the self-employment rate in the service industry, whereas it has significantly negative impacts on the self-employment rate in the manufacturing industry. Such difference may come from the fact that our model is for services rather than manufacturing because interregional trade is prohibited. Or it comes from the fact that the positive effect of density economies is observed in the service industry whereas it is not in the manufacturing industry. In the service industry, the positive effect of market expansion is larger than the positive effect of density economies in general and these two positive effects are smaller than the negative effect of congestion in absolute value. A possible reason for this is the ex post fierce competition among firms in the large market, which is made light of ex ante.

Several possible extensions are in order. First, the current analysis cannot provide the reason for the different impacts of market size on entrepreneurship in different stages. By utilizing firm level micro data, it may be possible to find the reason. In particular, it would be useful to analyze data regarding firm turnovers. Second, our analysis does not include insights into locational effects such that access to markets matters in starting up a business. It would be possible to take such an effect into consideration by including accessibility measures developed in new economic geography (Redding and Venables [29]). Third, the set-up costs may depend on the population density. Finally, it would be important to examine the effect of regional growth on entrepreneurship from a policy point of view. All these are important directions for future research.

References

[1] Alonso, W., 1964, *Location and Land Use*, Cambridge, MA, Harvard University Press.

[2] Arzaghi, M. and J. V. Henderson, 2008, Networking off Madison Avenue, *Review of Economic Studies* 75, 1011-1038.

[3] Behrens, K. and F. Robert-Nicoud, 2008, Survival of the fittest in cities: Agglomeration, selection, and polarisation, *CEPR Discussion Paper* 7018.

[4] Berry, S. and P. Reiss, 2007, Empirical models of entry and market structure, in: M. Armstrong and R. Porter (eds.), *Handbook of Industrial Organization, Volume III*, Amsterdam, North-Holland.

[5] Carlino, G.A., S. Chatterjee and R.M. Hunt, 2007, Urban density and the rate of invention, *Journal of Urban Economics* 61, 389-419.

[6] Ciccone, A., 2002, Agglomeration effects in Europe, *European Economic Review* 46, 213-227.

[7] Ciccone, A. and R.E. Hall, 1996, Productivity and the density of economic activity, *American Economic Review* 86, 54-70.

[8] Combes, P.-P., G. Duranton, L. Gobillon and S. Roux, 2010, Estimating agglomeration economies with history, geology and worker effects, in E.L. Glaeser (eds.), *Agglomeration Economics*, Chicago, The University of Chicago Press.

[9] Davis, D.R. and D.E. Weinstein, 2002, Bones, bombs, and break points: The geography of economic activity, *American Economic Review* 92, 1269-1289.

[10] Duranton, G. and D. Puga, 2001, Nursery Cities: Urban diversity, process innovation, and the life cycle of products, *American Economic Review* 91, 1454-1477.

[11] Forslid, R. and G. Ottaviano, 2003, An analytically solvable core-periphery model, *Journal of Economic Geography* 3, 229-240.

[12] Geroski, P.A., 1995, What do we know about entry, *International Journal of Industrial Organization* 13, 421-440.

[13] Glaeser, E.L., 2007, Entrepreneurship and the city, NBER Working Paper 13551.

[14] Glaeser, E.L. and W.R. Kerr, 2009, Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain?, *Journal of Economics & Management Strategy* 18, 623-663.

[15] Glaeser, E.L., W.R. Kerr and G.A.M. Ponzetto, 2010, Clusters of entrepreneurship, *Journal of Urban Economics* 67, 150-168.

[16] Glaeser, E.L., S.S. Rosenthal, and W.C. Strange, 2010, Urban economics and entrepreneurship, *Journal of Urban Economics* 67, 1-14.

[17] Harada, N., 2005, Potential entrepreneurship in Japan, *Small Business Economics* 25, 293-304.

[18] Härdle, W., 1990, *Applied Nonparametric Regression*, Cambridge, Cambridge University Press.

[19] Helsley, R.W. and W.C. Strange, 1991, Agglomeration economies and urban capital markets, *Journal of Urban Economics* 29, 96-112.

[20] Jacobs, J., 1961, *The Death and Life of American Cities*, New York, Random House.

[21] Krugman, P.R., 1980, Scale economies, product differentiation, and the pattern of trade, *American Economic Review* 70, 950-959.

[22] Krugman, P.R., 1991, Increasing returns and economic geography, *Journal of Political Economy* 99, 483-499.

[23] Miwa, R. and A. Hara (eds.), 2007, *The Documents and Statistics of Japanese Economic History since the 1850s*, Tokyo: University of Tokyo Press.

[24] Nakajima, K. and T. Tabuchi, 2011 Estimating interregional utility differentials, *Journal of Regional Science* 51, 31-46.

[25] Okamuro, H. and N. Kobayashi, 2005, Determinants of regional variations in the business start-up ratio, RIETI Discussion Paper 05-J-014. (in Japanese)

[26] Okamuro, H., 2006, A comparative analysis of high-tech and low-tech industries, RIETI Discussion Paper 06-J-049. (in Japanese)

[27] Orr, D., 1974, The determinants of entry: A study of the Canadian manufacturing industries, *Review of Economics and Statistics* 56, 58-66.

[28] Ottaviano, G., T. Tabuchi and J.-F. Thisse, 2002, Agglomeration and trade revisited, *International Economic Review* 43, 409-436.

[29] Redding, S. and A.J. Venables, 2004, Economic geography and international inequality, *Journal of International Economics* 62, 53– 82.

[30] Reynolds, P.D., B. Miller and W.R. Maki, 1995, Explaining regional variation in business births and deaths: U.S. 1976-88, *Small Business Economics* 7, 389-407.

[31] Rosenthal, S.S. and W.C. Strange, 2005, The geography of entrepreneurship in the New York metropolitan area, *Federal Reserve Bank of New York Economic Policy Review* 11, 29-53.

[32] Sato, Y., T. Tabuchi and K. Yamamoto, 2009, Market size and entrepreneurship, *GCOE Discussion Paper* 36, Human Behavior and Socioeconomic Dynamics, Osaka University.

[33] Stock, J.H. and M. Yogo, 2005, Testing for weak instruments in linear IV regression, in: D.W.K. Andrews and J.H. Stock (eds.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, Cambridge: Cambridge University Press.

Appendix: Summary statistics for each year

[Insert Table 7 about here]

	Description	Mean	Std Dev	Minimum	Maximum	Source
e_p	share of people who wish to start up a business	0.0805	0.0141	0.0544	0.118	
e_o	share of self-employment (all industries)	0.133	0.0240	0.0710	0.191	
e_{om}	share of self-employment (manufacturing)	0.0776	0.0256	0.0250	0.171	Employment Status Survey
e_{os}	share of self-employment (service)	0.119	0.0137	0.0948	0.156	
E	number of employed workers (in thousands)	1371.1	1327.5	314.6	6677.0	
L	prefectural population (in thousands)	2616.0	2457.0	612.0	12219.0	Population Estimates
H	inhabitable area (km^2)	2147.6	981.2	832.6	4562.8	Social Indicators by Prefecture
D_{pop}	population density	1387.1	1622.6	372.7	8753.7	$= 1000 \times L/H$
D_{em}	employment density	725.8	875.8	181.0	4838.6	$= 1000 \times E/H$
$\overline{w}$	monthly total cash earnings (all industries, in thousand yen)	323.5	34.9	261.5	458.9	
$\overline{w}_m$	monthly total cash earnings (manufacturing, in thousand yen)	312.2	54.5	200.8	480.2	Monthly Labour Survey
$\overline{w}_s$	monthly total cash earnings (service, in thousand yen)	348.3	27.7	278.7	450.2	
G_p	per capita GPDP (in thousand yen)	3393.9	769.0	2165.9	7242.8	
g_p	year to year comparison of G_p	1.00	0.0182	0.960	1.06	
M	price-cost margin=(GPDP−intermediate output−indirect tax less subsidies−compensation of employees)/GPDP	0.241	0.0254	0.178	0.322	
S_{ha}	share of agriculture in the GPDP	0.0279	0.0200	0.000526	0.0898	Prefectural Accounts
S_{hm}	shares of manufacturing in GPDP	0.246	0.0855	0.0488	0.502	
S_{hp}	share of public sectors in the GPDP	0.109	0.0304	0.0580	0.203	
V	active job opening to job applicant ratio	0.839	0.374	0.230	1.91	Monthly Report of Public Employment Security Statistics
D_{pop20}	population density before the WWII (1920)	251.9	305.1	55.5	1726.8	Japanese Imperial Statistical Yearbook
D_{pop40}	population density during the WWII (1940)	356.4	605.0	71.9	3429.2	

Table 1. Summary statistics and data sources

Notes: Summary statistics are for pooled data across three years. See the Appendix for each year.

dependent variable	$\ln(e_p)$: ln(share of potential entrepreneurs)			$\ln(e_o)$:ln(share of entrepreneurs) (all industries)		
	[1]	[2]	[3]	[4]	[5]	[6]
$\ln(D_{pop})$	0.0761***	0.641	0.0708***	−0.131***	−0.0190	−0.128***
	(0.0149)	(0.627)	(0.0199)	(0.0185)	(0.255)	(0.0281)
$\ln(G_p)$	0.133**	0.710***	0.170**	−0.192***	−0.192**	−0.203***
	(0.0534)	(0.227)	(0.0694)	(0.0660)	(0.0922)	(0.0676)
const.	−4.32***		−4.59***	0.359		0.426
	(0.394)		(0.513)	(0.487)		(0.526)
fixed effect	no	yes	no	no	yes	no
random effect	no	no	yes	no	no	yes
year dummy	yes	yes	yes	yes	yes	yes
p-value, Hausman test			0.044			0.91
R^2	0.67	0.87	0.66	0.54	0.98	0.54
NOB	138	138	138	138	138	138

dependent variable	$\ln(e_{om})$: ln(share of entrepreneurs) (manufacturing)			$\ln(e_{os})$: ln(share of entrepreneurs) (service)		
	[7]	[8]	[9]	[10]	[11]	[12]
$\ln(D_{pop})$	−0.0272	0.0675	−0.0413	0.0252*	0.292	0.0202
	(0.0443)	(0.976)	(0.0682)	(0.0144)	(0.544)	(0.0355)
$\ln(G_p)$	−0.206	0.0662	−0.110	0.0798	−0.235	0.0397
	(0.3826)	(0.353)	(0.204)	(0.0514)	(0.197)	(0.0693)
const.	−0.603		−1.28	−2.88***		−2.60***
	(1.15)		(1.51)	(0.373)		(0.505)
fixed effect	no	yes	no	no	yes	no
random effect	no	no	yes	no	no	yes
year dummy	yes	yes	yes	yes	yes	yes
p-value, Hausman test			0.81			0.16
R^2	0.18	0.91	0.17	0.25	0.76	0.24
NOB	138	138	138	138	138	138

Table 2. Preliminary estimation results for entrepreneurship
Note: Robust standard errors clustered by prefecture are reported in parentheses. Significance levels: $* = 10\%, ** = 5\%, *** = 1\%$.

dependent variable	$\ln(e_o)$ (all industries)		$\ln(e_{om})$ (manufacturing)		$\ln(e_{os})$ (service)	
	[1]	[2]	[3]	[4]	[5]	[6]
estimated parameters						
$a_0(=\ln(\beta/\gamma))$	11.3^{***}	8.980^{***}	10.4^{***}	11.5^{***}	11.3^{***}	9.82^{***}
	(0.0810)	(0.859)	(0.269)	(3.64)	(0.0660)	(1.02)
$a_1(=(\alpha - t_0)/\sqrt{\beta})$	0.0663^{***}	0.0646^{***}	0.0690^{***}	0.101^{***}	0.0595^{***}	0.0610^{***}
	(0.00280)	(0.00440)	(0.00999)	(0.0339)	(0.00171)	(0.00472)
$a_2(=t_1/\sqrt{\beta})$	$0.406 \times 10^{-5***}$	$0.782 \times 10^{-5**}$	$0.141 \times 10^{-4**}$	-0.310×10^{-5}	$0.840 \times 10^{-5***}$	$0.906 \times 10^{-5***}$
	(0.134×10^{-5})	(0.345×10^{-5})	(0.719×10^{-5})	(0.494×10^{-5})	(0.114×10^{-5})	(0.327×10^{-5})
control variables						
$\ln(G_p)$		0.264^{*}		-0.495		0.137
		(0.139)		(0.465)		(0.188)
$\ln(g_p)$		-0.0966		-2.18		-0.729
		(0.518)		(1.71)		(0.766)
$\ln(M)$		-0.346^{***}		-0.127		-0.222
		(0.126)		(0.334)		(0.158)
$\ln(S_{ha})$		0.115^{**}		-0.287^{***}		0.0342
		(0.0456)		(0.0995)		(0.0574)
$\ln(S_{hm})$		-0.0250		-0.264		0.0263
		(0.0699)		(0.188)		(0.0774)
$\ln(S_{hp})$		-0.0179		-0.486		-0.0717
		(0.164)		(0.366)		(0.215)
$\ln(V)$		-0.0223		0.0907		-0.00667
		(0.0606)		(0.154)		(0.0794)
year dummy	yes	yes	yes	yes	yes	yes
R^2	0.66	0.73	0.31	0.40	0.12	0.19
NOB	138	138	138	138	138	138

Table 3. NLS estimation of (9) for pooled data (years 1992, 1997 and 2002)

Notes: Robust standard errors clustered by prefecture are reported in parentheses. Significance levels: $* = 10\%, ** = 5\%, *** = 1\%$. The dependent variable is the share of entrepreneurs.

dependent variable	$\ln(e_o)$ (all industries)			$\ln(e_{om})$ (manufacturing)			$\ln(e_{os})$ (service)		
year	1992	1997	2002	1992	1997	2002	1992	1997	2002
estimated parameters									
$a_0(=\ln(\beta/\gamma))$	10.4^{***}	8.22^{***}	7.57^{***}	13.3^{***}	7.59^{**}	9.68^{**}	11.3^{***}	8.52^{***}	8.95^{***}
	(1.25)	(1.14)	(1.17)	(4.39)	(3.71)	(4.80)	(1.59)	(1.57)	(1.04)
$a_1(=(\alpha-t_0)/\sqrt{\beta})$	0.0680^{***}	0.0651^{***}	0.0613^{***}	0.111^{***}	0.0525^{***}	0.0438^{***}	0.0582^{***}	0.0622^{***}	0.0618^{***}
	(0.00550)	(0.00448)	(0.00487)	(0.0394)	(0.0127)	(0.00695)	(0.00560)	(0.00701)	(0.00462)
$a_2(=t_1/\sqrt{\beta})$	0.294×10^{-5}	$0.965\times10^{-5**}$	$0.105\times10^{-4***}$	$-0.620\times10^{-5*}$	$0.358\times10^{-4**}$	$0.200\times10^{-4**}$	$0.860\times10^{-5*}$	$0.150\times10^{-4***}$	$0.588\times10^{-5**}$
	(0.323×10^{-5})	(0.349×10^{-5})	(0.341×10^{-5})	(0.374×10^{-5})	(0.144×10^{-4})	(0.811×10^{-5})	(0.463×10^{-5})	(0.422×10^{-5})	(0.268×10^{-5})
control variables	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.73	0.68	0.69	0.49	0.27	0.31	0.15	0.12	0.21
NOB	46	46	46	46	46	46	46	46	46

Table 4. NLS estimation of (9) for each year

Notes: Robust standard errors are reported in parentheses. Significance levels: $* = 10\%, ** = 5\%, *** = 1\%$. The dependent variable is the share of entrepreneurs.

dependent variable	$\ln(e_p)$
$\ln(\widehat{e}_o/e_o)$	-0.173
	(0.112)
$\ln(D_{pop})$	0.0950^{***}
	(0.0158)
const.	-3.06^{***}
	(0.113)
year dummy	yes
R^2	0.66
NOB	138

Table 5. OLS estimation of (10) for pooled data (years 1992, 1997 and 2002)

Notes: Bootstrapped robust standard errors clustered by prefecture are reported in parentheses. Significance levels: $* = 10\%, ** = 5\%, *** = 1\%$. $\widehat{e}_o$ is the predicted value of e_o by estimated equation [2] in Table 4.

dependent variable	$\ln(e_o)$ (all industries)			$\ln(e_{om})$ (manufacturing)			$\ln(e_{os})$ (service)		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
estimated parameters									
$a_0(=\ln(\beta/\gamma))$	7.27***	7.28***	7.40***	12.3***	12.3***	12.3***	5.41	5.63	6.34
	(1.60)	(1.61)	(1.57)	(2.90)	(2.93)	(2.97)	(6.98)	(5.96)	(5.28)
$a_1(=(\alpha-t_0)/\sqrt{\beta})$	0.0485***	0.0486***	0.0492***	0.0919***	0.0910***	0.0892***	0.0382**	0.0388**	0.0409**
	(0.00684)	(0.00706)	(0.00653)	(0.0280)	(0.0278)	(0.0273)	(0.0177)	(0.0157)	(0.0160)
$a_2(=t_1/\sqrt{\beta})$	0.159×10^{-4}***	0.159×10^{-4}***	0.153×10^{-4}***	0.389×10^{-5}	0.440×10^{-5}	0.502×10^{-5}	0.200×10^{-4}	0.196×10^{-4}*	0.177×10^{-4}
	(0.461×10^{-5})	(0.474×10^{-5})	(0.446×10^{-5})	(0.821×10^{-5})	(0.801×10^{-5})	(0.816×10^{-5})	(0.131×10^{-4})	(0.115×10^{-4})	(0.117×10^{-4})
instruments used:									
D_{pop20}: density 1920	yes	no	yes	yes	no	yes	yes	no	yes
D_{pop40}: density 1940	no	yes	yes	no	yes	yes	no	yes	yes
over-id test p-value	-	-	0.86	-	-	0.73	-	-	0.01
first-stagestatistics	141.7	206.0	121.8	141.7	206.0	121.8	141.7	206.0	121.8
R^2	0.67	0.67	0.68	0.38	0.38	0.38	0.1	0.11	0.11
NOB	138	138	138	138	138	138	138	138	138

(a) N2SLS estimation of (9) for pooled data (years 1992, 1997 and 2002)

dependent variable	$\ln(e_p)$		
	[1]	[2]	[3]
$\ln(\widehat{e}_o/e_o)$	-0.139	-0.140	-0.140
	(0.102)	(0.101)	(0.0999)
$\ln(D_{pop})$	0.0846***	0.0851***	0.0854***
	(0.0188)	(0.0201)	(0.0205)
const.	-2.99***	-2.99***	-3.00***
	(0.136)	(0.135)	(0.140)
instruments used:			
$\ln(D_{pop20})$: density 1920	yes	no	yes
$\ln(D_{pop40})$: density 1940	no	yes	yes
over-id test p-value			0.79
first-stagestatistics	500.9	622.0	317.5
R^2	0.66	0.66	0.65
NOB	138	138	138

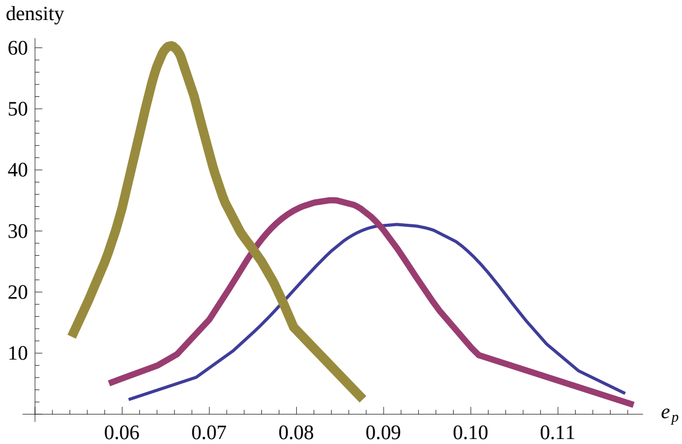
(b) 2SLS estimation of (10) for pooled data (years 1992, 1997 and 2002)

Table 6. Robustness check by IV estimation

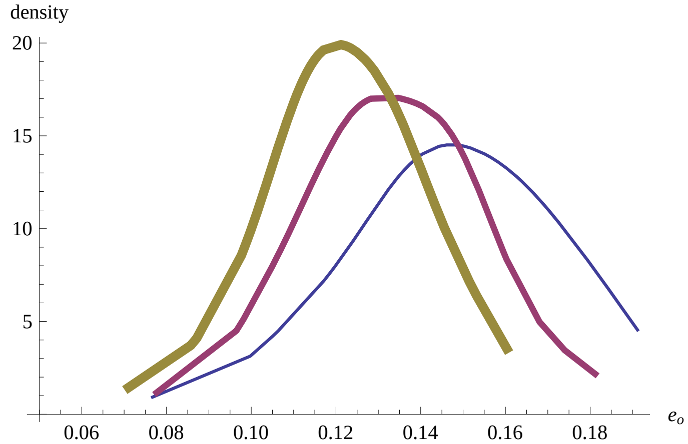
Notes: In (a), robust standard errors clustered by prefecture are reported in parentheses. In (b), bootstrapped robust standard errors clustered by prefecture are reported in parentheses. Significance levels: $* = 10\%, ** = 5\%, *** = 1\%$. The instrumental variables are the log lags of population density. In (b), $\widehat{e}_o$ is the predicted value of e_o by estimated equation [3] in (a).

	1992				1997				2002			
	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
e_p	0.091	0.012	0.061	0.12	0.083	0.011	0.059	0.12	0.067	0.0070	0.054	0.087
e_o	0.15	0.025	0.077	0.19	0.13	0.021	0.078	0.18	0.12	0.019	0.071	0.16
e_{om}	0.091	0.029	0.034	0.17	0.076	0.021	0.028	0.13	0.066	0.021	0.025	0.12
e_{os}	0.13	0.015	0.098	0.16	0.12	0.011	0.097	0.14	0.11	0.011	0.095	0.14
E	1367.5	1324.8	332.0	6635.0	1393.4	1350.2	332.0	6677.0	1352.4	1336.6	314.0	6653.0
L	2582.4	2429.2	615.0	11874.0	2618.8	2459.6	614.0	11808.0	2647.0	2535.1	612.0	12219.0
H	2140.1	986.2	832.6	4562.4	2140.6	986.0	833.3	4562.9	2162.1	993.0	850.5	4481.4
D_{pop}	1377.0	1622.3	386.4	8626.2	1393.2	1628.9	383.7	8556.9	1391.2	1652.7	372.8	8753.7
D_{em}	727.9	883.8	200.7	4820.2	740.3	893.9	196.6	4838.6	709.5	868.9	181.0	4766.8
$\overline{w}$	309.5	35.1	261.5	433.7	334.7	34.8	267.0	458.9	326.3	30.3	269.2	445.1
$\overline{w}_m$	291.9	51.3	200.8	436.4	320.1	54.3	211.1	472.9	324.7	53.2	233.2	480.2
$\overline{w}_s$	332.1	26.6	278.6	408.2	358.2	26.4	298.3	450.2	354.6	22.8	315.5	423.1
G_p	3201.6	770.7	2166.0	6731.5	3402.2	794.1	2230.2	7242.8	3578.0	709.6	2640.7	7218.6
g_p	1.0	0.019	0.98	1.1	1.0	0.015	0.96	1.0	1.0	0.014	0.96	1.0
M	0.25	0.027	0.19	0.31	0.24	0.025	0.18	0.29	0.24	0.024	0.20	0.32
S_{ha}	0.035	0.024	0.00083	0.090	0.027	0.019	0.00062	0.069	0.021	0.013	0.00053	0.054
S_{hm}	0.27	0.086	0.065	0.50	0.25	0.084	0.064	0.47	0.22	0.080	0.049	0.40
S_{hp}	0.10	0.030	0.058	0.18	0.11	0.033	0.062	0.20	0.11	0.027	0.060	0.17
V	1.2	0.40	0.30	1.9	0.79	0.25	0.23	1.3	0.57	0.14	0.30	0.88

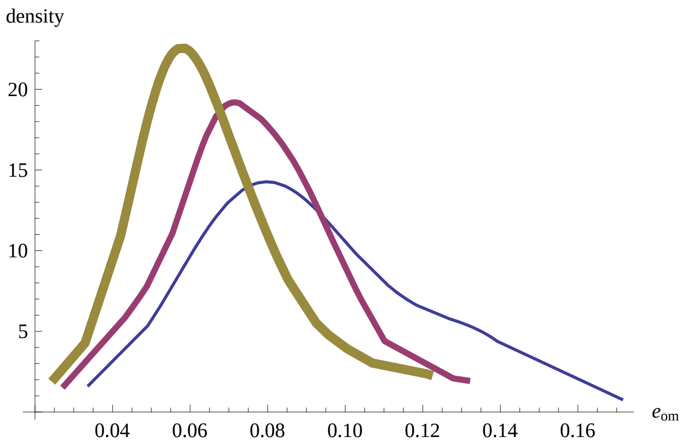
Table 7. Summary statistics for each year



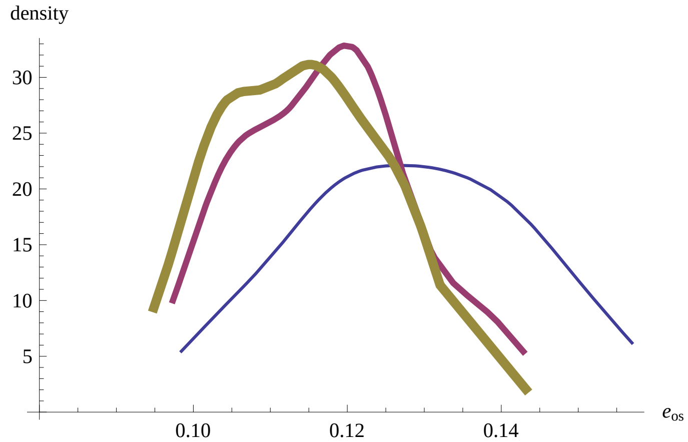
(a) e_p: potential (ex ante) entrepreneurs



(b) e_o: observed (ex post) entrepreneurs (all industries)



(c) e_{om}: observed entrepreneurs (manufacturing)



(d) e_{os}: observed entrepreneurs (service)

Figure 1. Kernel density estimations of the share of entrepreneurs

Notes: The thinnest lines are for year 1992, the second thinnest ones are for 1997, and the thickest ones are for 2002. The horizontal axis is the share of entrepreneurs and the vertical axis is the density.

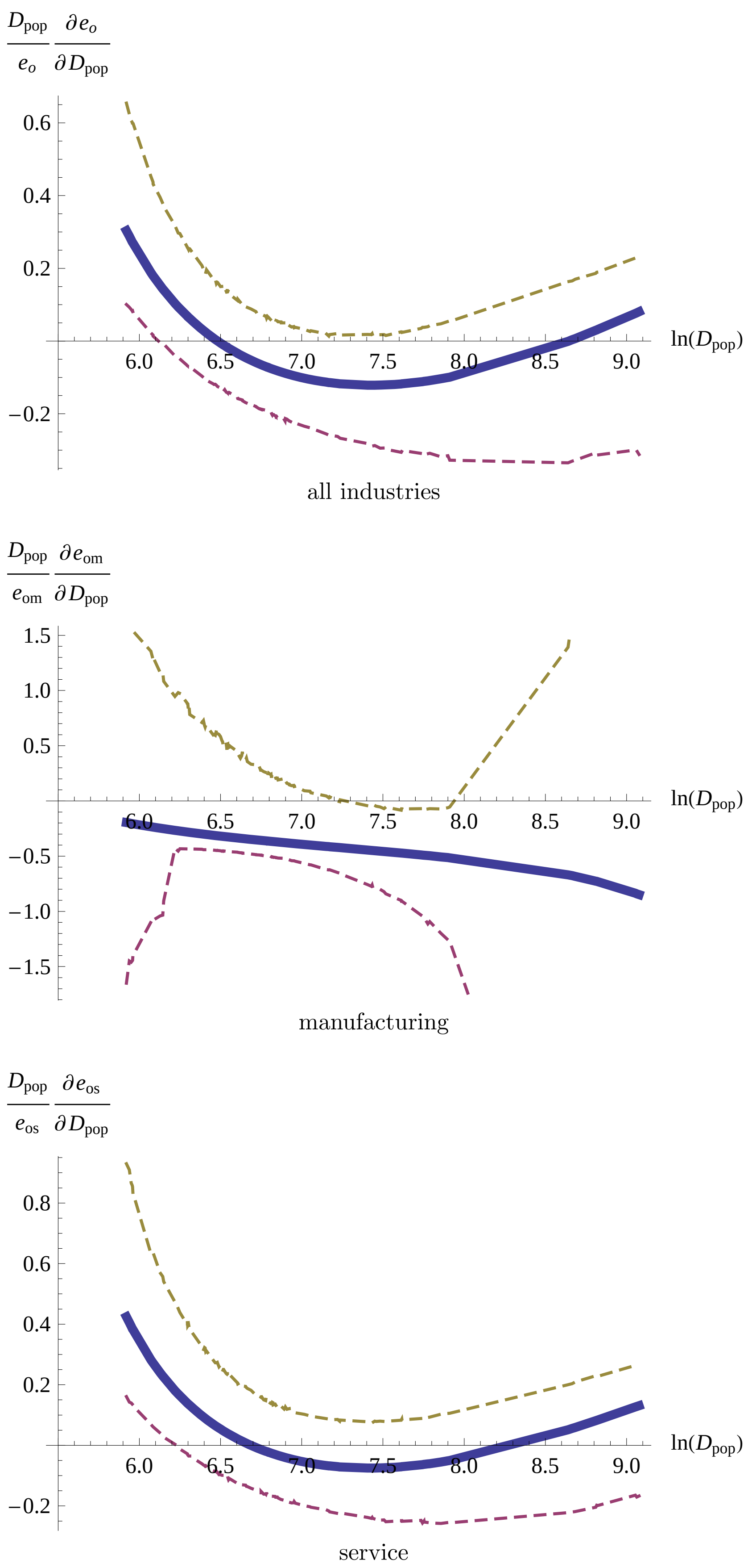


Figure 2. Simulated elasticities of entrepreneurship with respect to the population density

Notes: They are drawn by estimated equations [2], [4] and [6] in Table 3. Dotted lines describe the confidence intervals (95%).

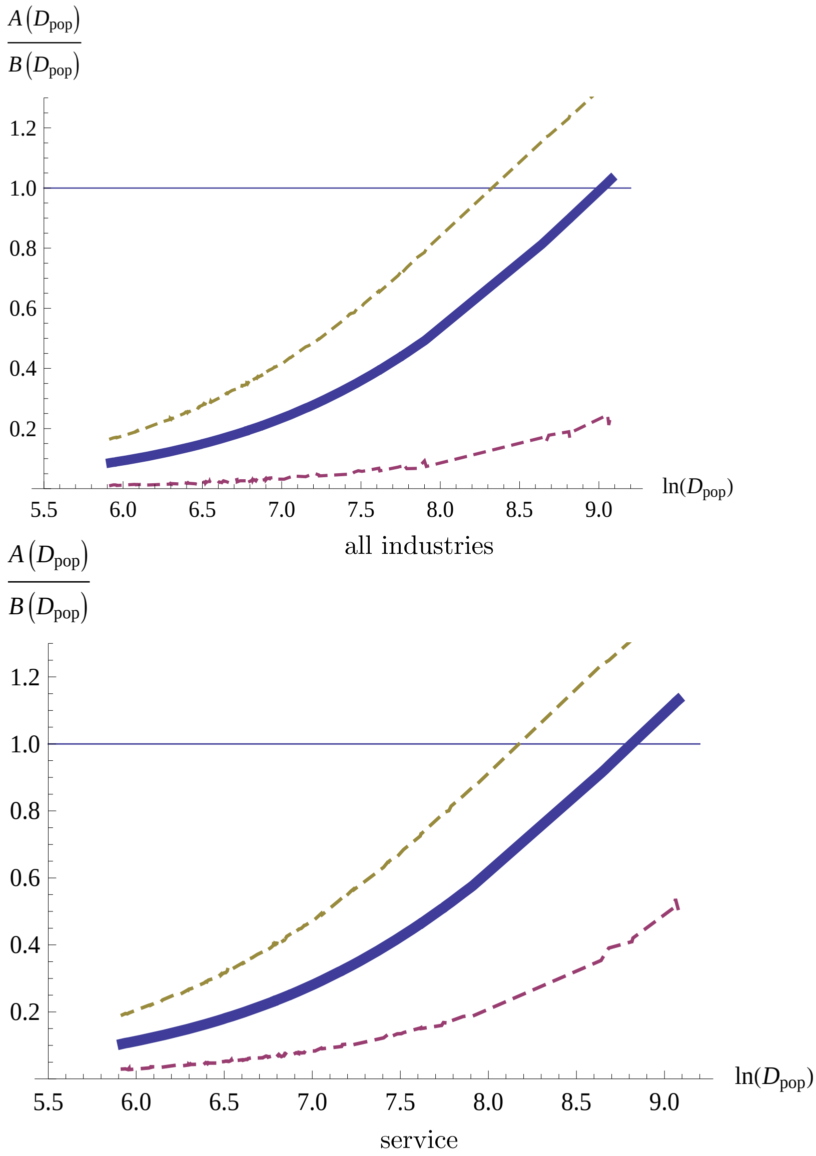
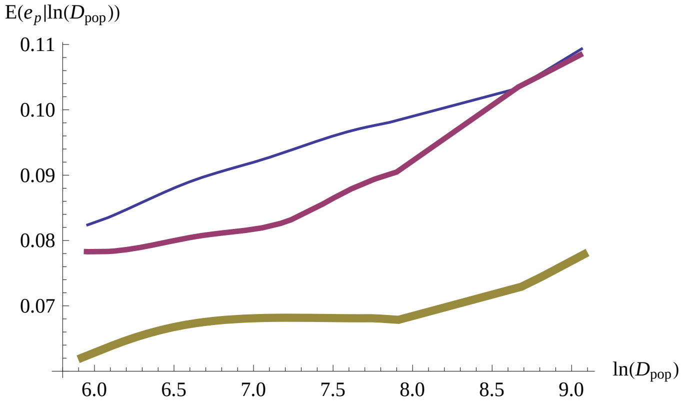
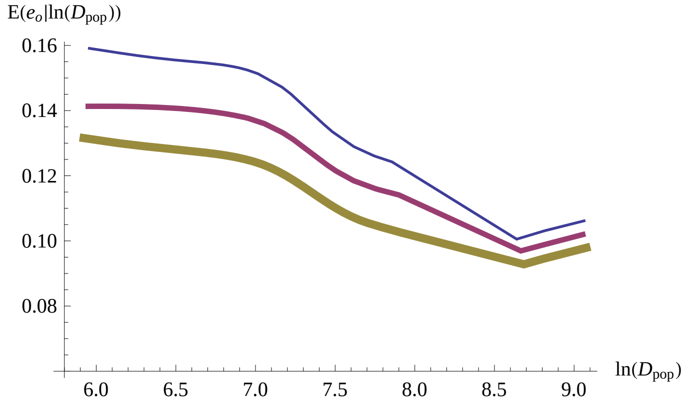


Figure 3. Relative importance of density economies to market expansion

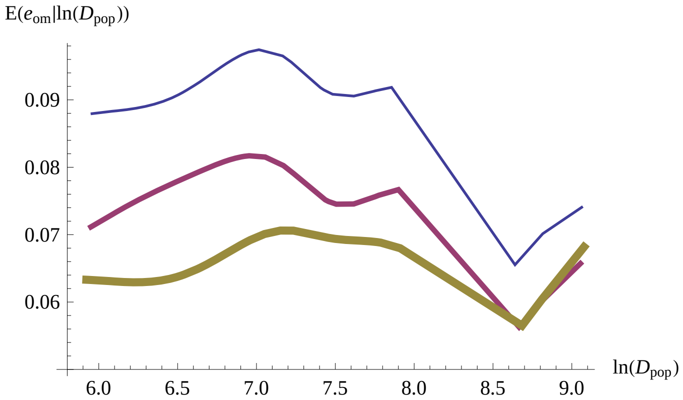
Notes: The are computed by estimated equations [2] and [6] in Table 3. Dotted lines describe the confidence intervals (95%)



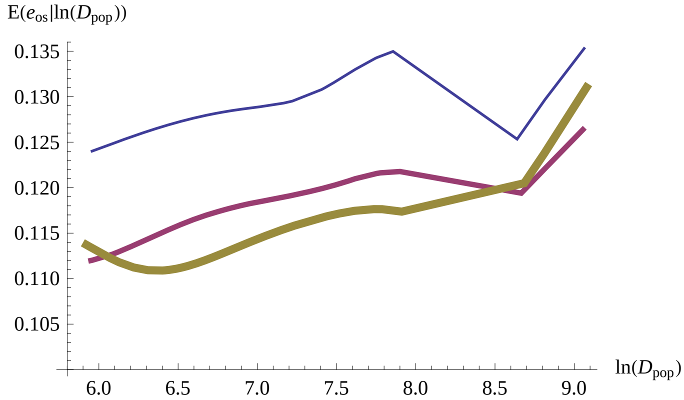
(a) $E(e_p|\ln(D_{pop}))$



(b) $E(e_o|\ln(D_{pop}))$



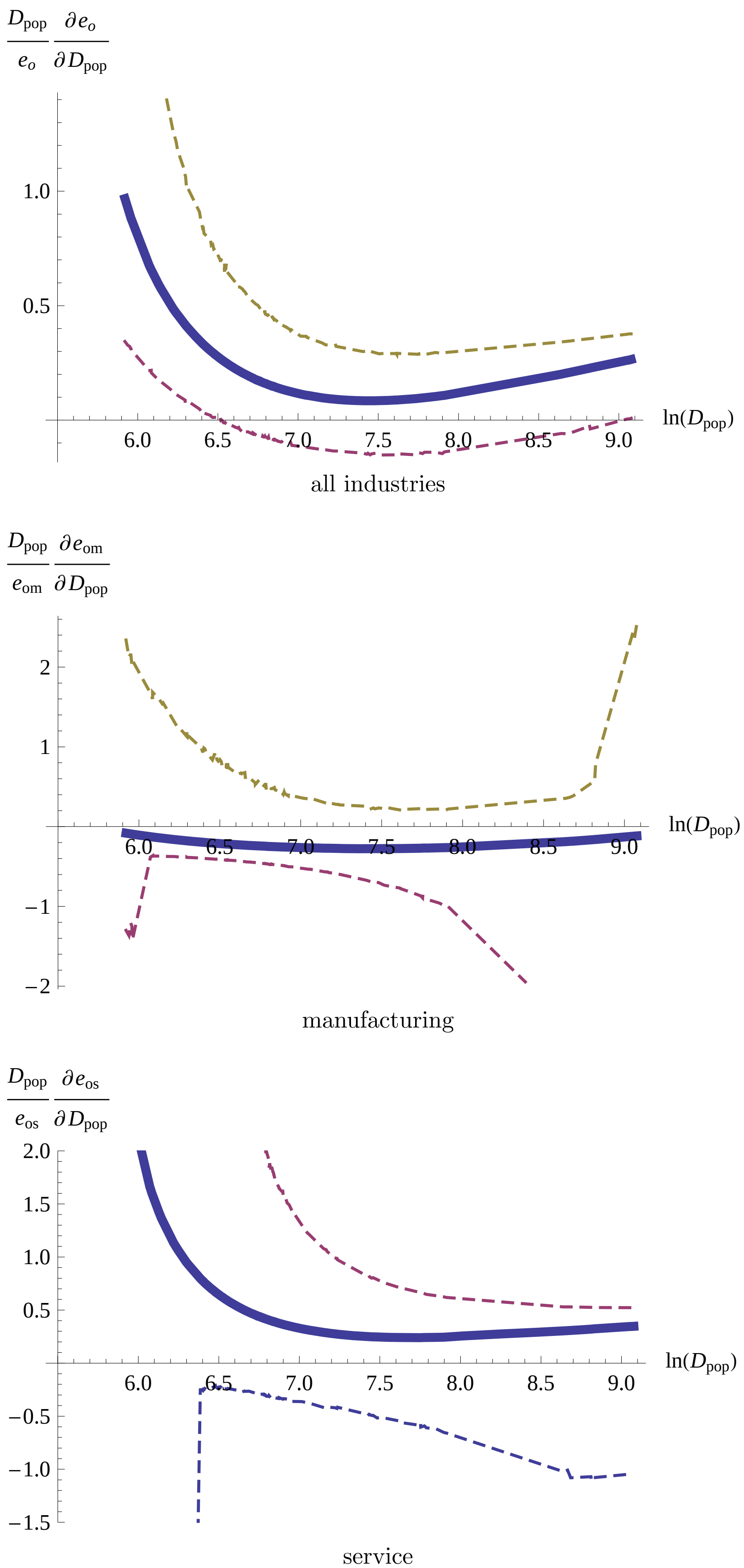
(c) $E(e_{om}|\ln(D_{pop}))$



(d) $E(e_{os}|\ln(D_{pop}))$

Figure 4. Kernel estimations of the conditional mean of e_o and e_p.

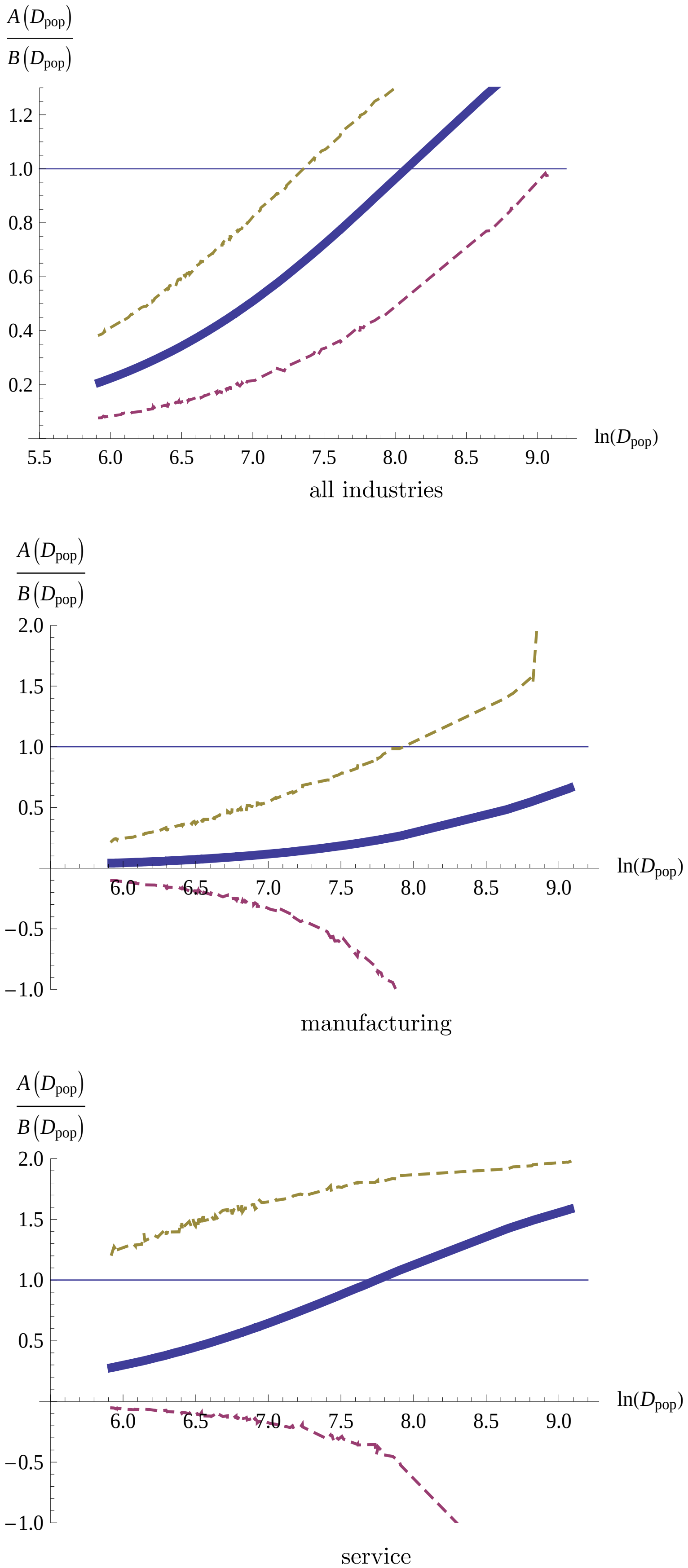
Notes: The thinnest lines are for year 1992, the second thinnest ones are for 1997, and the thickest ones are for 2002. The horizontal axis is a logarithm of the population density and the vertical axis is the conditional mean of e_o or e_p.



(a) Simulated elasticities of entrepreneurship with respect to population density.

Notes: They are drawn by estimated equations [3], [6] and [9] in Table 6-(a). Dotted lines describe the confidence intervals (95%)

Figure 5. Robustness check by IV estimation.



(b) Relative importance of density economies to market expansion.
Notes: We used estimated equations [3], [6] and [9] in Table 6-(a). Dotted lines describe the confidence intervals (95%)

Figure 5. Robustness check by IV estimation. (continued)