doi:10.1111/pirs.12269

# Agglomeration, (un)-related variety and new firm survival in China: Do local subsidies matter?*

**Anthony Howell**[1], **Canfei He**[2], **Rudai Yang**[1], **C. Cindy Fan**[3]

[1] School of Economics, Peking University, No.5 Yiheyuan Road Haidian District, Beijing, 100871, People's Republic of China (e-mail: tonyjhowell@pku.edu.cn, rdyang@pku.edu.cn)
[2] College of Urban and Environmental Sciences, Peking University, No.5 Yiheyuan Road Haidian District, Beijing, 100871, People's Republic of China (e-mail: hecanfei@urban.pku.edu.cn)
[3] Department of Geography, UCLA, 1255 Bunche Hall, Los Angeles, CA 90095, USA (e-mail: fan@geog.ucla.edu)

Received: 18 February 2015 / Accepted: 1 September 2016

**Abstract.** We study empirically the effects of five different dimensions of agglomeration – specialization, diversity, related variety, unrelated variety, and city size – on the survival chances of new entrepreneurial firms in China. Consideration is further given to studying the mediating effects of local subsidies on new firm survival given different existing local industrial structures in those regions. In support of the 'regional branching' hypothesis, we find that increasing local related variety has a stronger positive effect on new firm survival than other types of agglomeration. We also find that receiving comparatively fewer subsidies motivates firms to seek out and benefit from local existing economies, which in turn, positively influence their chances of survival. By contrast, agglomerated firms that receive relatively more subsidies tend to be more likely to face financial distress leading to eventual market exit. The findings thus reveal that both the intensity and the location of state support matters in terms of optimizing positive agglomeration effects on firms' post-entry performance and survival.

**JEL classification:** R11, R12, C41, L10

**Key words:** Agglomeration, related variety, new firm survival, public subsidies, China

## 1 Introduction

New firms are often seen as the catalyst of innovation and employment growth and are acknowledged for their role in promoting regional competitiveness. Yet new firms face a number of factors, or liabilities, including a lack of sufficient resources, higher vulnerability to external shocks, and greater likelihood to operate farther from the industry's minimum efficient scale leading to cost disadvantages (Schutjens and Wever 2000). Due to such liabilities, less than half

* The authors would like to acknowledge funding support from the School of Economics at Peking University and the Natural Science Foundation of China No. 41425001, No. 41271130, and No. 71603009.

of all new start-ups are expected to survive more than five-years after entry, regardless of country contexts (Cefis and Marsili 2011). Since their premature exit results in a loss of economic growth opportunities, understanding the factors that influence the survival chances of new firms is essential to promoting a vibrant economy.

A well-developed body of literature rooted heavily in the tradition of industrial organization and strategic management attempts to reveal the key factors that may help mitigate new firms' initial liabilities and improve their survival chances (Dunne et al. 1989). In the traditional survival literature, however, the regional dimension has been largely excluded from empirical analyses in spite of the fact that new firms tend to locate to existing clusters in order to reduce costs associated with learning about the local business environment (Stuart and Sorenson 2003).

To incorporate a spatial perspective into the literature, economic geographers have increasingly focused their attention on studying the mechanisms of agglomeration and how externalities influence firm survival.[1] A growing number of empirical studies now exist that study the effects of spatial externalities on firm survival in countries ranging from Greece (Fotopoulos and Louri 2000), to the US (Acs et al. 2007), Germany (Fritsch et al. 2006; Falck 2007), UK (Boschma and Wenting 2007) and China (Howell 2015; He and Yang 2016). The findings from these studies, however, are mixed and sometimes directly contrast one another. Fotopoulos and Louri (2000), for instance, show that localization economies resulting from specialization positively influence new firm survival in Greece, while Folta et al. (2006) find that specialization reduces the survival chances of US firms.

It is also common to observe mixed outcomes within the same study on firm survival depending on the dimension of agglomeration under scrutiny. In the US context, Acs et al. (2007) show that city size and diversity are important determinants of firm survival, while specialization is negatively related to the firm's survival chances. Similarly in the German context, Staber (2001) studies the knitwear industry and finds that localization economies led to higher failure rates, while diversification reduced failure rates. In China, He and Yang (2016) find that specialization decreases the survival chances of Chinese firms, although the opposite effect is found for urbanization economies.

It is clear from the aforementioned studies that the effects of agglomeration are country specific and depend on the type of agglomeration. While existing studies have contributed a great deal of insight into the nature of the relationship between agglomeration and firm survival, comparatively few studies exist that systematically take into account multiple dimensions of agglomeration on new firm survival in transitioning economy contexts. Moreover, while many existing studies show that the effect of agglomeration depends on firms' characteristics, most studies focus on dimensions of the firm related to its age, size, productivity or whether it is a single- or multi-plant firm. By contrast, it is not very clear if access to state support, an important dimension of firm heterogeneity in China, influences the ability of new firms to benefit from different types of externalities.

The focus of this paper, therefore is to study the relationship between five different dimensions of agglomeration – specialization, city size, diversity, related variety and unrelated variety – and new firm survival. Considering the Chinese context, we also consider the effects of local state support (i.e., public subsidies) as a potential source of firm heterogeneity with respect to external economies (correcting for sample selection). The empirical strategy relies on a large sample of more than 135,000 new entrepreneurial firms in Chinese manufacturing from 1998–2007.

In the next section, we provide a brief overview of the relevant literature. Section 3 introduces the data and key variables. Section 4 discusses the model estimation strategy. The empirical findings are presented in Section 5, and Section 6 concludes.

[1] For a good overview of the literature, see Frenken et al. (2015).

## 2 Literature review

At least since Marshall (1890), the spatial concentration of economic activity is viewed as leading to performance-enhancing spillovers that are good for innovation, growth and economic development. At the micro-level, firms that co-locate together are expected to benefit from superior access to knowledge and cost-saving spillovers, which enables them to attain a comparative advantage in the market and survive longer (Tallman et al. 2004). However, it is recognized that if firms are unable to sufficiently exploit local externalities, they may become exposed to higher risks of failure due to extensive competition or having to pay higher rents typically associated with larger cities (Stuart and Sorenson 2003; Folta et al. 2006).

Whether a firm benefits from or is harmed by the effects of agglomeration depends, in part, on the the dimension of agglomeration. In the literature, agglomerations are generally distinguished into three different dimensions: localization economies, urbanization economies and Jacobs externalities. Building on the early work of Marshall, localization economies are thought to arise from the spatial concentration of firms within the same industry. Firms potentially derive positive benefits resulting from a large pool of specialized labour, supplier-buyer linkages, and knowledge spillovers.

Urbanization economies are associated with large cities, where all firms may potentially benefit from greater access to larger markets, higher quality local amenities, better infrastructure, and public institutions, such as universities and research institutes. Jacob's (1969) externalities are thought to arise as firms benefit from many different industries within a region. Local industrial diversity is expected to lead to the combination of new ideas and more radical types of innovation due to spillovers that take place between industries (Glaeser et al. 1992).

More recently, Jacobs externalities have been further decomposed into related variety and unrelated variety (Boschma and Wenting 2007). Viewed from the perspective of portfolio diversification strategy, as a region attracts a larger number of unrelated industries, it will become better protected from external demand shocks that may have a devastating impact on one or a few sectors (Frenken et al. 2007). The notion of related variety takes into account the role of cognitive or technological relatedness between industries (Boschma et al. 2012), whereby co-located firms in different but similar industries can more easily communicate with one another, which helps facilitate between-industry knowledge spillovers.

From an evolutionary perspective, regions that develop and evolve into new related areas of economic activity, a process termed 'regional branching,' are associated with higher economic growth (Neffke et al. 2011). Proponents of regional branching argue that policy support should be directed towards attracting and developing new industries that share a close technological proximity to the existing local industrial mix in order to facilitate knowledge spillovers, as opposed to supporting leading industries that are already doing well. At the firm-level Howell et al. (2016) find that relatedness positively influences firms' performance, although the effects are not the same for all firms. The authors find that technological proximity has a stronger positive effect on higher productive firms in China, leading to asymmetrical benefits depending on the firm's own productive capabilities.

As in other agglomeration studies, the findings in Howell et al. (2016) point to the importance of taking into account the role of firm heterogeneity in studying the effects of related variety on firm performance outcomes. In China, state subsidies are an important source of heterogeneity that may influence the incentives and ability of firms to interact with their local environment. The next section turns to a discussion of agglomeration and state interventionist policies in the Chinese context.

### *2.1 Agglomeration and the role of state support in China*

In China, the central and local governments play a primary role in steering the direction, intensity and location of economic activities. For instance, central and local governments have implemented various location-based policies to encourage firms to concentrate together in special economic zones (SEZs), industrial districts and high-tech zones. Location policies have been relatively successful in China, playing an important role in accelerating its economic growth (Fan and Scott 2003). Hu et al. (2015), for instance, estimates that there have been more than 100 clusters designated by the state in over 60 cities since 1995, contributing 14 per cent to China's productivity growth between 1998 and 2007.

Coinciding with the rise of clustering policies in China, the opening-up of its economy to multinationals has led to intense local competition between domestic Chinese firms and foreign enterprises. Most Chinese domestic firms are technological laggards and cannot compete with the new foreign entrants, leading to their early exit from the market. The premature exit of Chinese firms due to unfair competition effects harms the Chinese economy and stifles opportunities for technological catch-up and product upgrading. To help mitigate these negative selection effects, central and local governments have increasingly relied on place-based policy supports directed towards firms operating in SEZs, industrial districts and high-tech zones in order to protect local profits and preserve state revenues (Howell 2016). Such policies include tax incentives, public subsidies, free or low cost loans, subsidized energy, subsidized raw materials, and land and technology.

He and Yang (2016), for instance, find that state support programmes (proxied by a dummy variable for whether or not a firm received public subsidies) directly improve the chances of firm survival in China, indicating that subsidies are an effective tool to help buffer domestic Chinese firms from negative competition effects with foreign enterprises. Beyond their direct effects, however, the authors do not consider whether subsidies may also indirectly influence firms' survival chances and subsequent performance depending on the local industrial structure.

On the one hand, Chinese policy-makers channel public subsidies towards firms operating in agglomerated areas in the hopes of mitigating the negative selection effect so that domestic firms can stay in the market long enough to hopefully take advantage of positive externalities. Firms that otherwise would have exited the market may instead benefit from new sources of ideas and knowledge that are expected to spillover within the same industry or between (related) industries, and in turn, spur technological catch-up and product upgrading.

On the other hand, firms that receive public financial support may become over-reliant on the state for its survival, thereby lacking the incentives to pursue profit-maximizing strategies. In turn, subsidy-receiving firms may not have the appropriate incentives to seek out and benefit from place-based economies, and may instead be exposed to increased risks of exit due to negative selection effects, thus necessitating perpetual future support from the state or face inevitable exit.

## 3 Data and variable development

Our study comprises more than 135,000 entrepreneurial firms within the first five years of operation. Besides their obvious policy importance, we focus on new firms to avoid capturing lifecycle effects where firms after a certain age tend to be less likely to seek out external knowledge (Audretsch and Lehmann 2005). In addition, a clear benefit of focusing on new firms is that they are less constrained by previous decisions, that is, past capital installments, thereby reducing concerns of endogeneity.

Our data are obtained from the Annual Report of Industrial Enterprise Statistics compiled by the State Statistical Bureau of China for the years 1998 through 2007. The data include all

firms with sales above five million Renminbi (approximately US$600,000), and contains extensive information on productivity, sales, employment, geographic location, industry affiliation and so forth. In total, our data captures over 90 per cent of productivity in China. We build a panel by linking firms over time using the firm's name, industry, address, etc. to assign unique numerical IDs.

### *3.1 Dependent variable: Measuring and interpreting firm survival*

Firms' survival span is developed using information on firm exit, along with entry and duration, which is obtained based on the firm's unique numerical ID. The entry year of the firm is identified for the first year, $t$, that the firm is observed but not in any years prior to $t$. The exit year of the firm is defined as the last year, $t$, that the firm reported information but not in the year $t+1, t+2, \ldots,$ 2007. The duration of a firm is defined by counting the number of years the firm is in operation, excluding its initial year of operation.

Firms may exit the sample for several reasons, including foreclosure, restructuring or merger and acquisition (M&A), or the firm simply drops below the minimum sales threshold. Due to the minimum sales threshold we cannot make the strong claim that firm exit is equated with firm foreclosure. Nevertheless, we do define firm exit as an indicator of serious financial distress leading to firm exit.

We make this claim for three reasons. First, the minimum sales threshold is not strictly enforced. Over the time period of analysis, 5 per cent of privately-owned firms reported sales below the minimum threshold. This is good because it decreases the chances that the firm exits the sample solely due to dropping below the threshold. Second, we remove all firms that enter and exit the survey in the same year since these firms are most likely to be hovering around the sales threshold.

Third, we assign new IDs to firms that undergo restructuring, merger, or acquisition, when possible. The fraction of firms in a year that can be linked to a firm in the previous year ranges from 84.5 per cent in the years 1998–1999 up to 92.2 per cent in the final two years (2006–2007). Overall, 95.9 per cent of all year-to-year matches are constructed using firm IDs, and 4.1 per cent using other information on the firm.

Figure 1 plots the average 3-year survival probabilities (in quartiles) across 333 cities for all firms from 1998-2007. Regions with higher survival probabilities tend to be quite diffused along the coastal regions of China. The 3-year average survival probabilities in large cities, such as Beijing, Tianjin and Shenzhen are lower than in the areas in and around Shanghai and Chongqing. In the western parts of the country, the average survival rates of the firm tend to be located along the bottom quartile at under 80 per cent.

### *3.2 Agglomeration measures*

We develop several proxies to measure different types of agglomerations: (*i*) external economies that arise from the spatial concentration of firms in the same industry, the so-called localization economies (LOC) (Glaeser et al. 1992); (*ii*) external economies arising from the urban size, population and economic density, the so-called urbanization economies (city size); (*iii*) external economies that arise due to the spatial concentration of diverse industries (diversity) (Jacobs 1969); (*iv*) external economies that arise from the spatial concentration of different but related industries (related variety); and (*v*) external economies that arise from the spatial concentration of different and non-complimentary industries (unrelated variety) (Frenken et al. 2007).

Following Delgado et al. (2010), we use location quotients of employment at the 3-digit manufacturing sector level to proxy for localization economies. We use labour density to control for urbanization economies typically found in larger cities. To proxy for industrial diversity, we

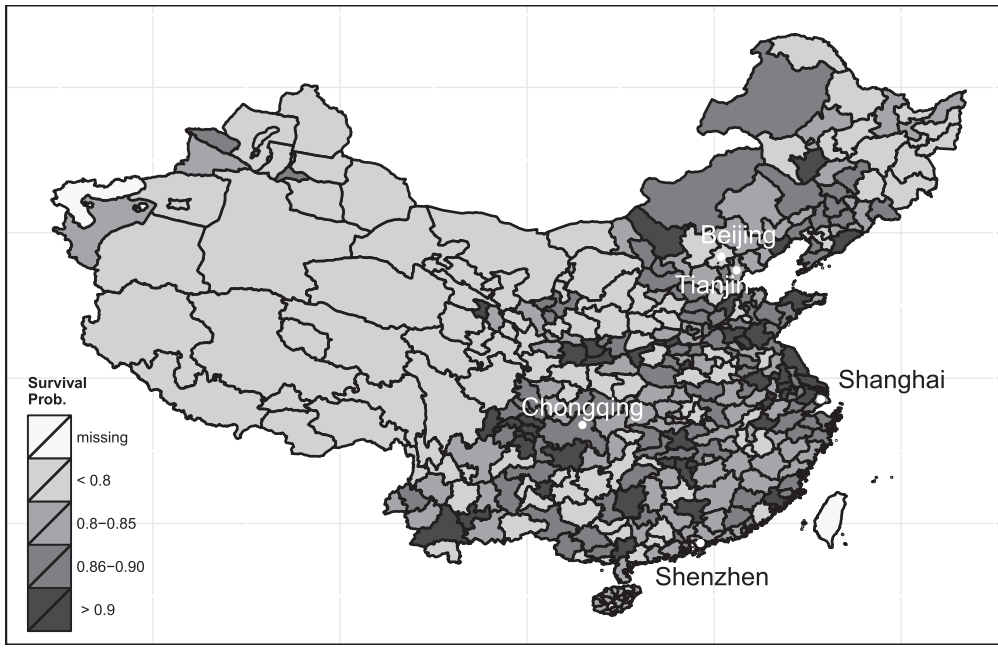


**Fig. 1.** Average 3-year Survival Rates, 1998-2007

use the total employment in all manufacturing sectors excluding the firm's own sector for each year across all cities.

We next distinguish between related variety and unrelated variety. Some earlier studies defined related industries as any subsectors that belong to the same sector (Frenken et al. 2007; Boschma and Iammarino 2009). This strategy, however, ignores the potential technological relatedness that may exist across subsectors (Essletzbichler 2013). Instead, we calculate similarity between subsectors on the basis of their shared use of input factors. The input mix reflects production technology implying that two subsectors that share similar input mixes also share close technological proximity (Frenken et al. 2007).

We use China's 2002 input-output tables with 122 sectors to capture similarity between two sector's input mixes. Following Los (2000), we use the Cosine Distance to measure technological relatedness, defined as the cosine between a pair of input coefficient vectors:

$$w_{m,j} = \frac{\sum_k \alpha_{mk} * \alpha_{jk}}{\sqrt{\left(\sum_k \alpha_{mk}^2 * \sum_k \alpha_{jk}^2\right)}}, \tag{1}$$

where $\alpha_m$ and $\alpha_j$ indicate the pair of input coefficient vectors, and $k$ denotes the $k_{th}$ input. $w_{m,j}$ ranges from 0 to 1, where a higher value indicates a higher degree of technological relatedness. As in Los (2000), we define two industries as being related if the value of $w_{m,j}$ is over 0.4, and unrelated otherwise.

We separately sum the total employment in industries that have a value of $w_{m,j}$ over 0.4 and less than or equal to 0.4 for each year across all cities to obtain the respective measures for related variety and unrelated variety. Note that the sum of these two measures adds up to our proxy for Jacobs externalities.

### *3.3 Firm-level covariates*

The main firm covariate of interest is the amount of subsidies it receives from central and local governments. We calculate the firm's subsidy intensity as the ratio of production-related subsidies to the total number of firm employees. In addition, we also include as firm level controls, the firm's current size and its square, sales growth and total factor productivity (TFP). Controlling for firm size and size squared addresses non-linearities in the survival-size relationship found in the existing literature. Including firm sales growth helps to avoid capturing the 'desperate' firm effects, where low growth firms may pursue a more risky survival strategy by switching products as a desperate measure to avoid imminent closure. TFP captures the firms' productivity and is derived using the 3-step procedure by Olley and Pakes (1996).

## 4 Model specification

The general hazard function represents the probability of failure of a firm during $t + \Delta t$ conditioned on the fact that the firm survives up to the time $t$. The hazard function is expressed as:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t <= T < t + \Delta t | T >= t)}{\Delta t} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}, \tag{2}$$

where $f(t)$ is the density function, $F(t)$ is the distribution function and $S(t)$ is the survival function. The survival function is $S(t) = exp(-\Lambda(t))$ and $\Lambda(t) = \int_0^t h(u)du$ is the cumulative hazard function.

While semi-parametric Cox's hazard and discrete-time models are most often employed to study firm exit, they frequently violate the proportionality assumption when examining multiple cohorts. Moreover, it is not easy to take into account issues related to left-truncation and right-censoring. Instead, the accelerated failure time (AFT) model provides a more appropriate alternative as it has a time-scaling factor that allows us to avoid violating the proportionality assumption.

In a AFT model, the survivor function at time t, $S(t|\boldsymbol{x}, \alpha)$, is assumed to be of the following form:

$$S(t|\boldsymbol{x}, \alpha) = S_0(t/\psi), \tag{3}$$

where $S_0(t)$ is the baseline survival model associated with a set of time-varying firm covariates, $\boldsymbol{x}$, time-varying industry covariates and random effects $\alpha$. The scaling factor $\psi$ is expressed as follows:

$$\psi(\boldsymbol{x}, \alpha) = exp(\eta) = exp(w + \boldsymbol{\beta'X}+), \tag{4}$$

where $\alpha = exp(w)$ is assumed to have a gamma distribution with distribution function $G(\alpha)$, and $\eta$ is the linear component of the model. The term $\alpha$ represents a frailty term with the mean of the distribution set to the value unity.

Our data is left-truncated since firms are observed at different points in time during the observation period, and subject to right censoring since not all firms exit the sample by the end reporting year in 2007. The likelihood function with left-truncated and right-censored observations is given in general form as:

$$L = \prod_{i=1}^{g} \int_0^{\infty} \left\{ \prod_{j=1}^{N} h(T)^c \left( \frac{S(T)}{S(E)} \right) \right\} dG(\alpha), \tag{5}$$

where $E_i$ takes into account the left truncation, giving the first time a firm enters into the panel; $c_i$ takes into account right censoring and takes the value of 1 for firms that fail and 0 for firms that are still active at the end of observation time.

In order to estimate the hazard function, an appropriate underlying distribution must be specified. We consider the log-logistic distribution since it has a flexible form that allows for monotonous functional forms, and other shapes as well. The hazard function with a log- logistic distribution is:

$$h(t|\boldsymbol{x}, \alpha) = \frac{\psi^{1/\lambda}\, t^{(1/\lambda - 1)}}{\lambda \left[1 + (\psi\, t)^{1/\lambda}\right]}. \tag{6}$$

The shape of the function is determined by $\lambda$. For $\lambda >= 1$, the functional form is decreasing monotonously and $0 < \lambda < 1$ has a bell-shaped form.

To obtain the survival probabilities of the firm, the hazard model in Equation (6) can be equivalently expressed as a log linear model expressed as:

$$log(D_{it}) = \beta_1 Subs_{it-1} + \beta_2 Agg^n + \beta_3 (Subs_{it-1} \times Agg^n) + \beta_4 \boldsymbol{X}_{it} + \alpha + \mu + \sigma \epsilon_i, \tag{7}$$

where $D_{it}$ is a random variable based on the duration, in years, of the firm. $Subs_{it-1}$ is the amount of subsidies the firm receives in year $t-1$, $Agg^n$ is a vector that contains each measure of agglomeration (outlined in subsection 3.2), and $\beta_1$ and $\beta_2$ are the respective corresponding parameters. $\beta_3$, a key parameter of interest, captures the effects of local state support given different local existing industrial structures on firm survival by multiplying the amount of subsidies the firm received in year $t-1$ with each respective agglomeration measure, $Agg^n$.

The vector, $\boldsymbol{X}_{it}$, includes the following set of control variables: firm size and its square, firm sales growth, and a set of industry, region and year dummies. As before, the term $\alpha$ represents a frailty term with the mean of the distribution set to the value unity. Finally, $\mu$ and $\sigma$ are unknown location and scale parameters, and $\epsilon_i$ has a distribution that determines $t_i$.

### *4.1 Identification issues*

Two key issues arise when trying to estimate equation (7). First, it is difficult to interpret the parameter, $\beta_2$, since the observed geographic influences on firm survival may be subject to issues related to selection bias in the location choice of the firm (Audretsch 1991). Second, the allocation of public subsidies is not likely to be randomly distributed to firms across different locations, thus making it difficult to interpret the parameters, $\beta_1$ and $\beta_3$. We will briefly discuss how we handle each of these estimation issues.

#### 4.1.1 Firm spatial sorting

It is known that larger firms may optimize on their location decision by moving into more agglomerated regions in order to take advantages of anticipated positive externalities. If selection occurs, the survival estimation results would lead to artificially high coefficients on the agglomeration measures. While this is certainly a theoretical possibility, as a practical matter, such bias is likely to be negligent for new entrepreneurial firms in China for the following reasons.

First, entrepreneurial firms do not typically engage in a multi-regional site selection process; rather, they tend to be wherever the owner's place of residence is located (Stam 2007). Second, the factors that influence firms' entry do not necessarily have the same effect on their chances of survival (Stuart and Sorenson 2003; Renski 2011), which we would expect to be the case if firms engaged in self-selection behaviour. Lastly, in the Chinese context, Au and Henderson (2006) argue that restrictions on migration (e.g., the Hukou system), obstruct individuals and firms from migrating into certain areas, thereby limiting self-sorting behaviour.

For these three reasons, the likelihood that Chinese firms optimize on their location decision is likely to be lower than in other country contexts and not be the main source driving the results. Nevertheless, as a simple sensitivity check, we remove all of the mobile firms that changed their industry or region during the time period from the sample. Then we re-estimate the survival model on only the subsample of stationary firms (See Table 4 below).

### 4.1.2 Subsidies selection bias

As an initial check for selection bias, Table 1 splits firms into respective groups according to whether or not they receive subsidies. The results show that subsidies tend to be channeled to firms that have higher local specialization, higher local related (unrelated) variety, and that are located in smaller cities. At the firm-level, the results further show that subsidies tend to go toward more successful firms, e.g. firms that tend to be larger in size, are more experienced, enjoy higher sales growth, and are more productive.

To help remove selection bias, we adopt the following two-stage approach. In the first stage we estimate the selection equation as follows:

$$SubsD_{it} = \begin{cases} 1 \; if \; Subs \; D^*_{it} = \beta_1 Agg^n + \beta_2 \mathbf{X}_{it} + \alpha_i + \epsilon_{it} > \overline{c} \\ 0 \; if \; Subs \; D^*_{it} = \beta_1 Agg^n + \beta_2 \mathbf{X}_{it} + \alpha_i + \epsilon_{it} \leq \overline{c}, \end{cases} \tag{8}$$

where $SubsD_{it}$ is an (observable) indicator function that equals 1 if firm $i$ receives subsidies, and 0 otherwise. $SubsD^*_{it}$ is a latent indicator function whereby the firm receives subsidies if these are above a given threshold $\overline{c}$. As before, $Agg^n$ includes each measure of agglomeration, and $\boldsymbol{X}_{it}$ includes a set of control variables. Lastly, $\alpha_i$ captures the unobserved firm heterogeneity and $\epsilon_{it}$ is an error term.

**Table 1.** Descriptive statistics

| | Whole sample of firms | Receive subsidies | | |
|---|---|---|---|---|
| | | No | Yes | Difference |
| Agglomeration measures | | | | |
| Location quotients | 2.571 | 2.566 | 2.643 | 0.077*** |
| Diversity (Ln) | 5.937 | 5.936 | 5.948 | 0.012* |
| – Related variety (Ln) | 7.485 | 7.474 | 7.633 | 0.160*** |
| – Unrelated variety (Ln) | 5.690 | 5.671 | 5.933 | 0.262*** |
| City size (Ln) | 5.334 | 5.339 | 5.269 | −0.070*** |
| Firm-level controls | | | | |
| Size (Ln) | 4.325 | 4.297 | 4.685 | 0.388*** |
| Age | 2.929 | 2.902 | 3.267 | 0.364*** |
| Sales growth (Ln) | 0.025 | 0.012 | 0.192 | 0.181*** |
| TFP (Ln) | 3.371 | 3.370 | 3.381 | 0.011 |

*Notes*: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

We estimate Equation (8) using a random effects probit model given the panel structure and the binary character of the dependent variable. The random effects structure assume that the errors are not correlated with the regressors, an unrealistic assumption in this case. To address this issue, the Mundlak (1978) specification is used by including a vector of means of the time-varying regressors as control variables to allow for some correlation between the random effects and the regressors.

Given the firm receives subsidies, in the the second stage we estimate the following specification:

$$Subs_{it} = \begin{cases} Subs^*_{it} = \beta_1 Agg^n + \beta_2 \mathbf{X}_{it} + \alpha_i + \epsilon_{it} & if\ Subs_{it} = 1, \\ Subs_{it} = 0 & if\ Subs_{it} = 0 \end{cases} \tag{9}$$

where $Subs^*_{it}$ is the unobserved latent variable representing the firm's subsidy intensity. The parameters and variables have their same interpretations as above. Correcting for selection effects, Equation (9) is estimated using the consistent estimator introduced in Wooldridge (1995). That is, a pooled OLS model is estimated including T inverse Mills ratios (interacted with time dummies), which are obtained from estimating T probit models (one for each year). Since inverse Mills ratios are used, we should find at least one exogenous variable that significantly affects the likelihood that the firm receives a subsidy, but is uncorrelated with the amount of subsidies given the subsidy allocation decision has already been made.

We recognize that it is difficult to find a satisfactory exclusion restriction, and in theory, the factors that influence the decision to subsidize a firm may also be related to the intensity of that subsidy. In our case, we take the firm's distance to nearest port as the identification instrument. We motivate this choice theoretically: subsidies are more likely to be given to local Chinese firms that are closer to ports to help them compete against foreign firms which tend to be concentrated nearby shipping ports to export goods more cheaply. The main determinants of the subsidy amount, however, are based on firms' productivity and overall performance and not their location. As an empirical check, the econometric results in Table 2 show that the distance measure, whose marginal effect is significant in the selection equation (column 1), does not affect firms' subsidy intensity in the outcome equation (column 2).

The results in Table 2 also show that firms tend to be more likely to receive subsidies and receive them more intensively if they are in areas with higher specialization, higher related variety and higher unrelated variety. By contrast, firms in smaller cities tend to be more likely to receive subsidies, although given the firm receives a subsidy, more subsidies tend to go to firms in larger cities. In terms of firm characteristics, the results show that firms that are larger, more experienced, experience higher growth, and more productive tend to be more likely to receive subsidies and receive them more intensively. Although in columns (2) and (3), there is a non-linear effect on firm size and the coefficient on firm sales growth is statistically not significant.

## 5 Effects of agglomeration on new firm survival

Table 3 provides the estimates of the effect of agglomeration on new firm survival using the AFT survival models log-logistic distribution with random effects. Column (1) includes only the spatial agglomeration measures, column (2) adds the firm-level covariates, column (3) decomposes the diversity measure into related variety and unrelated variety, and column (4) re-estimates the survival model on only the firms that received subsidies correcting for sample selection. Note that the agglomeration variables have been standardized for comparison purposes. The significance of the estimated effects is assessed by clustering standard errors at the city level to adjust for the potential correlation of errors among firms located within the same city.

**Table 2.** Sample selection model

| | Receive subsidies (1=Yes, 0=No) | Subsidy-receiving firms only: subsidy intensity: | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Agglomeration measures | | | |
| LOC | 0.028*** | 0.030*** | 0.032*** |
| | (0.005) | (0.004) | (0.004) |
| Related Variety | 0.086*** | 0.018** | 0.004 |
| | (0.008) | (0.006) | (0.006) |
| Unrelated Variety | 0.089*** | 0.043*** | 0.075*** |
| | (0.006) | (0.005) | (0.005) |
| City Size | − 0.021*** | 0.046*** | 0.026*** |
| | (0.005) | (0.003) | (0.003) |
| Firm-level controls | | | |
| Size | 0.073* | 0.242*** | 0.264*** |
| | (0.032) | (0.024) | (0.024) |
| $Size^2$ | 0.032*** | −0.042*** | −0.043*** |
| | (0.003) | (0.003) | (0.003) |
| Age | 0.024*** | 0.044*** | 0.039*** |
| | (0.004) | (0.002) | (0.002) |
| Sales growth | 0.039*** | −0.005 | 0.007 |
| | (0.005) | (0.003) | (0.004) |
| TFP | 0.007* | 0.010*** | 0.022*** |
| | (0.003) | (0.003) | (0.003) |
| Distance to port | −0.027*** | 0.019 | ... |
| | (0.006) | (0.015) | |
| Time dummies | Yes*** | No | No |
| Industry dummies | No | Yes*** | Yes*** |
| Region dummies | No | Yes*** | Yes*** |
| IMR× time dummies | – | Yes*** | Yes*** |
| Means | Yes*** | Yes*** | Yes*** |
| Rho | 0.567 | – | – |
| Corrected predictions | 78.13% | – | – |
| Num. obs. | 337,074 | 25,382 | 25,382 |

*Notes*: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Bootstrapped standard errors in parentheses. Column (1) is estimated with a RE Probit model with the coefficients reported in marginal effects (at the means). Column (2) is estimated using the consistent estimator in Wooldridge (1995). Inverse Mills ratio (IMR) × time dummies is significant at 1% confirming the inclusion of a selection equation. Means of the time-varying regressors are included as control variables to allow for some correlation between the random effect and the regressors. Rho is the percentage of total variance contributed by the panel-level variance component (Rho > 0 indicates that the panel estimator is different from the pooled estimator).

The results show that the relationship between agglomeration and firm survival varies depending on the dimension of agglomeration under scrutiny. The coefficients on LOC are positive and highly significant, while the coefficients on diversity does not influence firm survival chances in any significant way. The coefficient on city size is negative and statistically significant in column (1), but becomes statistically not significant with the inclusion of firm controls in column (2). In column (3), the coefficient on related variety is positive and significant, while the one on unrelated variety is statistically not significant. In terms of economic impacts, related variety exhibits the largest positive effect on firm survival, increasing the probability of survival by a factor of 1.9.

The results on the firm-level covariates are largely consistent with the existing literature. The coefficient on subsidies is positive and statistically significant, even after correcting for

**Table 3.** Effects of agglomeration and subsidies on firm survival

| | Survival models: accelerated failure time with random effects | | | |
|---|---|---|---|---|
| | Agglomeration measures only (1) | Agglomeration measures and firm controls (2) | Related variety & unrelated variety (3) | Subsidy-receiving firms only (correcting for (sample selection) (4) |
| Agglomeration measures | | | | |
| LOC | 0.031*** | 0.013*** | 0.013*** | 0.016* |
| | (0.003) | (0.003) | (0.003) | (0.008) |
| Diversity | −0.002 | −0.005 | | |
| | (0.004) | (0.004) | | |
| – Related variety | | | 0.019* | 0.057* |
| | | | (0.006) | (0.019) |
| – Unrelated variety | | | −0.010* | −0.009* |
| | | | (0.005) | (0.004) |
| City size | −0.005* | −0.003 | −0.003 | −0.015 |
| | (0.002) | (0.002) | (0.002) | (0.013) |
| Firm-level controls | | | | |
| Subs_Int | | 0.006** | 0.005** | 0.008** |
| | | (0.002) | (0.002) | (0.003) |
| Size | | 0.340*** | 0.340*** | 0.531*** |
| | | (0.019) | (0.019) | (0.089) |
| $Size^2$ | | −0.020*** | −0.020*** | −0.037*** |
| | | (0.002) | (0.002) | (0.009) |
| Sales growth | | 0.009*** | 0.009*** | 0.038** |
| | | (0.003) | (0.003) | (0.015) |
| TFP | | 0.012*** | 0.011*** | 0.018*** |
| | | (0.002) | (0.002) | (0.004) |
| Region fixed effects | Yes | Yes | Yes | Yes |
| Industry fixed effects | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes |
| Log likelihood | –182,804 | –178,396 | –178,195 | –10524.6 |
| Num. obs. | 337,074 | 332,529 | 332,529 | 25,382 |

*Notes*: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. The t-statistics (in parentheses) are obtained using standard errors that are robust and clustered at the city level. The dependent variable is the duration period of the firm and the model estimated is a frailty accelerated failure time models using the log-logistic distribution. The coefficients are presented as survival probabilities and represent the conditional probability of competing a survival spell.

sample selection in column (4). The coefficients on firm size are positive and significant in each specification although the effects are non-linear. The coefficients on firms' sales growth and TFP are both both positive and significant.

### *5.1 Alternative model specifications and robustness checks*

Table 4 reports a series of alternative model specifications to check the validity of our initial findings between our agglomeration measures and new firm survival. Note that in each estimation, the individual terms remain included in all of the model estimations, but are not reported given their qualitatively unchanged values. Also, focus is placed on related variety and unrelated variety since the coefficient on Jacobs externality is statistically not significant in the baseline model. Note that a discrete hazard model is estimated in column (4) because it is more flexible in controlling for unobserved firm heterogeneity and addresses the issue of tied failure times.

**Table 4.** Alternative model estimations

| | Survival | | | Exit | |
|---|---|---|---|---|---|
| | Stationary firms only (1) | Weibull distribution (2) | $1 million threshold (3) | Cox hazard (4) | Discrete time hazard (5) |
| Agglomeration measures | | | | | |
| LOC | 0.011* | 0.014* | 0.008** | −0.017** | −0.029* |
| | (0.005) | (0.007) | (0.003) | (0.006) | (0.012) |
| Related variety | 0.016** | 0.020*** | 0.013* | −0.024** | −0.004* |
| | (0.007) | (0.005) | (0.007) | (0.008) | (0.002) |
| Unrelated variety | −0.012** | −0.006*** | −0.006** | 0.015** | 0.013*** |
| | (0.004) | (0.002) | (0.003) | (0.003) | (0.006) |
| City size | −0.006 | −0.000 | −0.003 | 0.003 | −0.003 |
| | (0.004) | (0.003) | (0.003) | (0.003) | (0.006) |
| Region fixed effects | Yes | Yes | Yes | Yes | Yes |
| Industry fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes | Yes |
| Log likelihood | –94,241 | –179,793 | –103,150 | – | –198,118 |
| Num. obs. | 223,184 | 332,529 | 304,236 | 332,529 | 332,529 |

*Notes*: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. The t-statistics (in parentheses) are obtained using standard errors that are robust and clustered at the city level. The dependent variable is the duration period of the firm in columns (1), (2) and (5), and firm exit in columns (3) and (4). Each specification includes all controls from Table 3. Note that excluding column (1), each specification is estimated on the full sample of firms.

**Table 5.** Effects of local subsidies on firm survival

| | Survival models: accelerated failure time with random effects | |
|---|---|---|
| | All firms (without sample selection correction) (1) | Only firms that receive subsidie (correcting for sample selection) (2) |
| Interaction terms | | |
| Subs_Int x LOC | −0.006** | −0.010* |
| | (0.002) | (0.005) |
| Subs_Int x related variety | −0.009** | −0.008* |
| | (0.003) | (0.004) |
| Subs_Int x unrelated variety | 0.006** | 0.008** |
| | (0.002) | (0.003) |
| Subs_Int x city size | −0.009** | −0.006** |
| | (0.004) | (0.002) |
| Region fixed effects | Yes | Yes |
| Industry fixed effects | Yes | Yes |
| Year fixed effects | Yes | Yes |
| Log likelihood | –178,391 | –12,314 |
| Num. obs. | 332,529 | 25,382 |

*Notes*: $^{**}p < 0.01$, $^{*}p < 0.05$. The t-statistics (in parentheses) are obtained using standard errors that are robust and clustered at the city level. The dependent variable is the duration period of the firm and the model estimated is a frailty accelerated failure time models using the log-logistic distribution. The coefficients are presented as survival probabilities and represent the conditional probability of completing a survival spell. Each specification includes all controls from Table 3.

In column (1) of Table 4, all of the mobile firms that changed their industry or region during the time period from the sample are removed, and the model is re-estimated on only the sub-sample of stationary firms. The results reveal that the direction of the relationship between each agglomeration measure and firm survival remains the same. Albeit the size of the coefficients tend to be slightly smaller in magnitude, an indication that the initial results may be influenced by selection effects, but certainly not driven purely by them. The results in columns (2)–(5) further show that the initial findings are neither sensitive to the minimum sales threshold or to the estimation strategy.

### *5.2 Effects of local subsidies on firm survival*

Table 5 shows how increasing local state support influences new firms' survival likelihood given different types of local existing industrial structures exist in those regions. Interaction terms are created that multiply the amount of subsidies the firm received in the previous year with each measure of agglomeration. Column (1) re-estimates the model on the entire sample of firms. Column (2) re-estimates the survival model on only the sample of firms that received subsidies (correcting for selection bias).

The coefficients on the interaction terms between subsidies and LOC, related variety and city size are respectively negative and statistically significant. The results are largely consistent in both the uncorrected and corrected model, although, the size of the coefficients change. For instance, the size of the negative coefficient on the interaction term between subsidies and LOC is 40 per cent larger in the sample corrected model. One reason for this is that in the uncorrected model, a greater amount of subsidies gets distributed to better performing firms that are local industry leaders (e.g. have a higher LOC value), which artificially reduces the negative selection effects associated with specialization in the uncorrected model.

The main findings suggest that more heavily subsidized firms have a lower chance of survival than their less subsidized counterparts given higher values for LOC, related variety and city size, respectively. One reason for these negative interaction effects is because receiving too many subsidies reduces firms' incentives to pursue profit-maximizing strategies, since they incorporate an expectation that the state will intervene on their behalf in times of duress. By contrast, agglomerated firms that receive a fewer amount of subsidies are unable to rely completely on the state for their survival, and instead are motivated to exploit and benefit from localization economies and spillovers that are expected to take place within and between (related) industries.

There is an important caveat with the present interpretation of these findings, however. In general, more heavily subsidized firms are unlikely to declare bankruptcy given the nature of state support. Our results therefore reveal that more heavily subsidized firms versus their less subsidized counterparts are more likely to encounter financial distress (e.g., sales revenue drops well below the minimum sales threshold) when operating in more agglomerated areas (e.g., higher values for LOC, related variety and city size). In this regard, firms that become over-reliant on external state support while also facing intense competition effects or other negative externalities experience a greater threat to their sustainability in that they are less likely to maintain revenues above a certain threshold.

The only interaction term that returns a positive coefficient is the one between subsidies and unrelated variety. One reason why more heavily subsidized firms in areas with higher unrelated variety are more likely to survive than elsewhere may be because they do not have to compete against many local competitors. As a result, such firms can enjoy local monopolies that enable them to capitalize on profit-making activities, which in turn, helps them sustain their sales revenue above the minimum threshold enforced on our data.

## 6 Concluding remarks

Coinciding with China's opening up and market reforms, new firms in China face fierce competitive pressures that often lead to their premature exit from the market. The ability of firms to seek out and benefit from agglomeration economies matters for their survival prospects. In particular, understanding the factors that incentivize new firms to seek out and successfully exploit externalities are key to implementing successful place-based policies in China. This paper links five dimensions of agglomeration – specialization, diversity, related variety, unrelated variety, and city size – to new firms' survival prospects. Taking into account the Chinese context, consideration is given to studying how local state intervention influences the ability of firms to seek out and benefit from each dimension of agglomeration.

Our study produces the following two key results. First, compared to other types of agglomeration, related variety has the largest positive effect on firm survival. Second, firms that become over-reliant on external state support while also facing intense competition effects or other negative externalities experience a greater threat to their sustainability in that they are less likely to maintain revenues above a certain threshold.

The implications of these findings for policy are as follows. First, in order to help new firms to mitigate their initial liabilities, local policy-makers should adopt a 'regional branching' strategy (Neffke et al. 2011). That is, evolve the local industrial mix into new related industries in order to increase the supply of spillovers that are expected to take place between related industries. Second, the intensity and location of where subsidies get allocated matters. In order to promote regional competitiveness in China, allocating relatively fewer subsidies to new firms operating in regions with certain existing industrial structures may help motivate them to rely on agglomeration economies for their survival as opposed to becoming over-reliant on external finance from the state.

## References

Acs ZJ, Armington C, Zhang T (2007) The determinants of new-firm survival across regional economies: The role of human capital stock and knowledge spillover. *Papers in Regional Science* 86: 367–391

Au C, Henderson V (2006) Are Chinese cities too small? *Review of Economic Studies* 73: 549–576

Audretsch D (1991) New-firm survival and the technical regime. *Review of Economics and Statistics* 72: 441–450

Audretsch DB, Lehmann EE (2005) Does the knowledge spillover theory of entrepreneurship hold for regions? *Research Policy* 34: 1191–1202

Boschma R, Iammarino S (2009) Related variety, trade linkages, and regional growth in Italy. *Economic Geography* 85: 289–311

Boschma R, Minondo A, Navarro M (2012) Related variety and regional growth in Spain. *Papers in Regional Science* 91: 241–256

Boschma R, Wenting R (2007) The spatial evolution of the British automobile industry: Does location matter? *Industrial and Corporate Change* 16: 213–238

Cefis E, Marsili O (2011) Born to flip. Exit decisions of entrepreneurial firms in high-tech and low-tech industries. *Journal of Evolutionary Economics* 21: 473–498

Delgado M, Porter ME, Stern S (2010) Clusters and entrepreneurship. *Journal of Economic Geography* 10: 495–518

Dunne T, Roberts MJ, Samuelson L (1989) The growth and failure of US manufacturing plants. *The Quarterly Journal of Economics* 104: 671–698

Essletzbichler J (2013) Relatedness, industrial branching and technological cohesion in US metropolitan areas. *Regional Studies* 49: 752–766

Falck O (2007) Survival chances of new businesses: Do regional conditions matter? *Applied Economics* 39: 2039–2048

Fan CC, Scott AJ (2003) Industrial agglomeration and development: A survey of spatial economic issues in East Asia and a statistical analysis of Chinese regions. *Economic Geography* 79: 295–319

Folta TB, Cooper AC, Baik Y (2006) Geographic cluster size and firm performance. *Journal of Business Venturing* 21: 217–242

Fotopoulos G, Louri H (2000) Location and survival of new entry. *Small Business Economics* 14: 311–321

Frenken K, Cefis E, Stam E (2015) Industrial dynamics and clusters: A survey. *Regional Studies* 49: 10–27

Frenken K, Van Oort F, Verburg T (2007) Related variety, unrelated variety and regional economic growth. *Regional Studies* 41: 685–697

Fritsch M, Brixy U, Falck O (2006) The effect of industry, region, and time on new business survival: A multi-dimensional analysis. *Review of Industrial Organization* 28: 285–306

Glaeser E, Kallal H, Scheinkman J, Shleifer A (1992) Growth in cities. *Journal of Political Economy* 100: 1126–1152

He C, Yang R (2016) Determinants of firm failure: Empirical evidence from China. *Growth and Change* 47: 72–92

Howell A (2015) 'Indigenous' innovation with heterogeneous risk and new firm survival in a transitioning Chinese economy. *Research Policy* 44: 1866–1876

Howell A (2016) Firm R&D, innovation and easing financial constraints in China: Does corporate tax reform matter? *Research Policy (Forthcoming)* 45. doi:10.1016/j.respol.2016.07.002

Howell A, He C, Rudai Y, Fan C (2016) Technological relatedness and asymmetrical firm productivity gains under market reforms in China. *Cambridge Journal of Regions, Economy and Society (Forthcoming)* 9. doi:10.1093/cjres/rsw024

Hu C, Xu Z, Yashiro N (2015) Agglomeration and productivity in China: Firm level evidence. *China Economic Review* 33: 50–66

Jacobs J (1969) *The Economy of Cities*. Vintage, New York

Los B (2000) The empirical performance of a new interindustry technology spillover measure. In: Saviotti P, Nooteboom B, Elgar E (eds) *Technology and knowledge: From the firm to innovation systems*. Edward Elgar Publishing, Cheltenham

Marshall A (1890) *Principles of Economics*. Macmillan, London

Mundlak Y (1978) On the pooling of time series and cross-section data. *Econometrica* 46: 69–85

Neffke F, Henning M, Boschma R (2011) How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography* 87: 237–265

Olley GS, Pakes A (1996) The dynamics of productivity in the telecommunication equipment industry. *Econometrica* 64: 1263–1297

Renski H (2011) External economies of localization, urbanization and industrial diversity and new firm survival. *Papers in Regional Science* 90: 473–502

Schutjens VA, Wever E (2000) Determinants of new firm success. *Papers in Regional Science* 79: 135–159

Staber U (2001) Spatial proximity and firm survival in a declining industrial district: The case of knitwear firms in Baden-Wurttemberg. *Regional Studies* 35: 329–341

Stam E (2007) Why butterflies don't leave: Locational behavior of entrepreneurial firms. *Economic Geography* 83: 27–50

Stuart T, Sorenson O (2003) The geography of opportunity: Spatial heterogeneity in founding rates and the performance of biotechnology firms. *Research Policy* 32: 229–253

Tallman S, Jenkins M, Henry N, Pinch S (2004) Knowledge, clusters, and competitive advantage. *Academy of Management Review* 29: 258–271

Wooldridge J (1995) Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics* 68: 115–132