Advances in Applied Sociology, 2024, 14, 269-293
https://www.scirp.org/journal/aasoci
ISSN Online: 2165-4336
ISSN Print: 2165-4328

How do Related and Unrelated Varieties Impact Firm Economic Growth in China: An Empirical Analysis with Endogeneity

Yuting Jiang

School of Financial Mathematics and Statistics, Guangdong University of Finance, Guangzhou, China
Email: yuting_jiang@gduf.edu.cn

How to cite this paper: Jiang, Y. T. (2024). How do Related and Unrelated Varieties Impact Firm Economic Growth in China: An Empirical Analysis with Endogeneity. *Advances in Applied Sociology, 14,* 269-293.
https://doi.org/10.4236/aasoci.2024.146019

Received: March 31, 2024
Accepted: June 15, 2024
Published: June 18, 2024

Abstract

Using a sample of 84,868 Chinese manufacturing firms operating in the period 2006-2013, we analyze whether related and unrelated variety—measured by entropy method, affects firms' economic performance. Our results show that, with Heckman (1979) two-step sample selection model correcting only for sample-selection bias, unrelated variety has a negative and statistically significant impact. If we also correct the endogeneity of the main explanatory variables—related and unrelated variety, by employing Bartik's (1991) shift-share approach, we find that they become insignificant. We find a positive effect for related variety and negative effect for unrelated variety only if we consider highly-developed Chinese regions. Finally, we find that related variety has a positive effect only in the case of large firms.

Keywords

Agglomeration, Firm Economic Growth, Related and Unrelated Variety

1. Introduction

Spatial agglomeration is one of the main determinants of firms' economic performance and therefore of regional/local economic growth. Since the early 1990s, the focus in the literature has been on the benefits that firms derive from location in an agglomerated area such as an industrial district or a technological cluster. This stream of work generally shows that the geographic concentration of production tends to generate positive returns in terms of firm productivity (Cainelli et al., 2016; Cainelli & Ganau, 2019), firm innovation propensity (Cainelli & De Liso, 2005; Cainelli, 2008), and internationalization choices (Cainelli et al., 2014).

The literature identifies two main types of local externalities: 1) localization and 2) diversification economies (Glaeser et al., 1992) which depend respectively on the Marshall-Arrow-Romer (MAR) model and Jacobs externalities (Glaeser et al., 1992). The MAR theory suggests that knowledge spillovers occur among firms that belong to the same industry (localization economies) whereas Jacobs externalities (Jacobs, 1969) refer to the knowledge spillovers that occur among firms belonging to different industries (diversification economies).

A more recent strand of work has introduced the notion of related variety into the concept of Jacobs externalities (Frenken et al., 2007). In this case what matters is not diversification *per se* but related variety i.e. knowledge spillovers among firms operating in "different but related" sectors. In other words, inter-industry knowledge spillovers, i.e. the cross-fertilization of ideas, knowledge and technologies across industries, occur among sectors with common knowledge and technology bases.

This paper investigates the effect of related and unrelated variety on economic growth of a sample of 84,868 Chinese manufacturing firms in the period 2006-2013. We estimate a firm economic proportional growth equation *à la* Gibrat which includes these two measures of spatial agglomeration i.e. related and unrelated variety. According to a commonly accepted interpretation of Gibrat's Law, the growth rate of a firm is independent of its initial size. Thus, Gibrat's Law allows us to test whether the firm's initial size has any effect on its subsequent economic growth. We estimate an augmented version of the equation by adding two explanatory variables for agglomeration effects: related and unrelated variety. This allows us to test whether Gibrat's Law holds, and to investigate the role played by related and unrelated variety in firms' economic growth.

This paper contributes to the empirical literature on related variety in three ways. First, it is one of the first works to study an emerging economy such as China. During the years considered—2006-2013—China experienced a deep productive and technological transformation, and accelerated economic growth. Second, it is the first analysis of Gibrat's Law applied to the Chinese economy which also includes agglomeration measures and investigates the effect of related variety on the economic performance of Chinese firms using a large firm level dataset. Third, it takes account of the endogeneity of the main explanatory variables—related and unrelated variety—by employing Bartik's (1991) shift-share approach. To our knowledge, no previous work on related and unrelated variety addresses the potential endogeneity of these variables.

The paper is organized as follows. In Section 2 we review the literature on Gibrat's Law and the concept of related variety. Section 3 describes the data and the econometric methodology adopted. Section 4 presents and discusses the main results of our analysis and Section 5 concludes the work.

2. Related Literature

2.1. Gibrat's Law

Gibrat's Law or Gibrat's rule of proportionate growth was proposed first by Gibrat (1931). The original version hypothesized that, a skewed distribution consisting in a large number of additive and independent variables could be converted into a normal distribution by transforming the initial variables with a logarithmic function. Following work by Mansfield (1962) and Chesher (1979), Gibrat's Law began to be used to test proportional growth theory which is the basis of many empirical analyses. The current interpretation of Gibrat's Law differs slightly from its original version. Today, Gibrat's Law is understood as stating that a firm's proportional growth rate is independent of its absolute size, in other words, all firms, small and large, should grow at the same rate. This understanding of Gibrat's Law was tested first by Mansfield (1962) on the cases of three different industries. Many subsequent studies (Wagner, 1992; Geroski, 1995; Caves, 1998) provide estimates of this equation. However, in many cases the evidence does not support the Law (Reid, 1995; Audretsch, 1995; Harhoff et al., 1998; Weiss, 1998; Audretsch et al., 1999; Calvo, 2006).

Some authors hypothesize that rejection of the Law is due to the fact that small firms generally have a high probability of dying. Lotti et al. (2003) use quantile regression to show that Gibrat's Law holds for new entrants; in other words, estimates based on surviving firms might be affected by sample selection bias which tends to amplify the rapid growth of smaller firms.

For clarity, this paper uses Gibrat's Law as the starting point for an investigation of firm economic growth in China. Based on the common understanding of Gibrat's Law, we test whether the rate of firms' economic growth is independent of its initial income level. Gibrat's Law generally holds if firm growth is independent of determinants such as firm age and size. It assumes that the estimated coefficients of the firm's initial income level and firm age are not different from zero (Maine et al., 2010).

Although Gibrat's Law provides a useful framework to test proportionate growth, it is not able to capture all the determinants of firm income growth. Therefore, in our specifications we incorporate measures for agglomeration externalities to investigate their effects on firm income growth, and whether Gibrat's Law holds in estimates that include more variables.

2.2. Agglomeration and Related Variety

Analysis of agglomeration economies goes back to Alfred Marshall (1920) and since that time the idea that location within a bounded geographic area has a positive effect on firms' economic performance has been acknowledged by all economists. In a study of the cutlery and knitwear districts in Sheffield, and the knitwear district in Northampton, Marshall showed that firms located in these agglomerated areas benefitted from advantages, also called externalities, compared to firms operating in the same industries but located in non-agglomerated areas. Specifically, Marshall identified three mechanisms underlying agglomeration-related advantages: 1) concentration of a large number of highly specialized suppliers (input sharing); 2) availability of highly specialized workers (labor matching); and 3) existence of knowledge spillovers among local actors.

These three mechanisms explain why firms located in an agglomerated area, and benefitting from lower production costs, tend to achieve higher productivity with respect to their non-agglomerated counterparts. More recent empirical contributions generally confirmed this positive relationship between agglomeration and firms' economic performance (e.g. Henderson, 2003; Martin et al., 2011; Cainelli et al., 2016), propensity for innovation (Cainelli & De Liso, 2005; Cainelli, 2008), and internationalization choices (Cainelli et al., 2014).

The recent empirical work on agglomeration-related advantages is the seminal paper by Glaeser et al. (1992), which investigates two different forms of agglomeration forces and their effects on urban employment in the USA: the agglomeration advantages associated to the productive specialization of the local industry; and agglomeration advantages associated to the productive diversification in an urban area. The first type, generally measured by the level of productive specialization of an industry in the locality, captures knowledge spillovers among firms operating in the same industry. The idea is that physical proximity among firms belonging to the same industry facilitates the transmission of knowledge, ideas, information, and technologies among economic agents and promotes both knowledge spillovers among firms, and incremental and process innovations. These local externalities are usually described as localization economies.

The second type of agglomeration advantages, also called Jacobs externalities, consist of the knowledge spillovers across different industries located in the same geographic area. The notion of Jacobs externalities was proposed by Jane Jacobs (1969) in the context of American cities. Jacobs (1969) identified urban variety as one of the key mechanisms supporting and promoting the transfer of ideas, information, and knowledge among the different industries in a city/urban area. The cross-fertilization of information, knowledge, and innovation among firms in different industries generates advantages for individual firms and thus positive effects on the local system's aggregate economic performance. It has been shown that almost 70% of the innovations developed by one industry are used in another sector (Glaeser et al., 1992). It follows then that an industry located in an area characterized by a high degree of diversification and variety of the productive structure should grow faster thanks to these spillovers. In fact, the industry structure diversity/variety at the regional level promotes exchange and cross-fertilization of information, ideas, and technologies, which in turn, promote radical and product innovation. These local externalities described as diversification economies.

Debate over which of these two agglomeration forces, localization or diversification, contributes most to the economic development of a region/local system has been ongoing for a long time. Despite the large number of papers on this topic, the empirical literature is not conclusive.

Our understanding about the mechanisms underlying knowledge spillovers arising from local industrial diversification has been improved recently. It has been suggested that what matters is not diversification/variety *per se* but rather related variety. This assumes that knowledge spillovers within a region/local system occur among firms operating in "different but related" sectors. The literature on related variety (Frenken et al., 2007) shows that industrial variety in a locality is not a sufficient condition to guarantee cross-fertilization and the transfer of information and knowledge among different local industries. Technological/productive similarity among the industries located in a bounded geographic area is required in order to materialize these local externalities (Frenken et al., 2007). The authors suggest that the transfer of information, knowledge, and innovations can occur only among industries with the same or similar technology and knowledge bases. It is suggested also that the differentiated industrial mix in a local system/region can enhance opportunities to interact, copy, modify, and recombine ideas, practices, and technologies across sectors. Geographic proximity among firms makes this process of recombining existing pieces of knowledge in totally new ways more likely to occur. Recombination leads to new products and services. Traditional industries such as footwear are unlikely to transfer knowledge or technologies to high-tech sectors such as the biomedical industry. Thus, transfer and transmission processes are activated only if the cognitive distance among the firms operating in these different local industries is not too large (Nooteboom, 2000; Boschma & Iammarino, 2009). In their pioneering contribution, Frenken et al. (2007) operationalized the concept of local industrial variety using the entropy measure, and employed a standard statistical classification of industries to identify relatedness among the sectors within an industry. Specifically, they computed related variety as the weighted sum of entropy at the five-digit sector level for each two-digit level industry in a locality, while unrelated variety was operationalized as entropy at the two-digit industry level.

Many empirical studies focus on the relationship between related variety and regional economic growth. The first is Frenken et al. (2007) which analyzes the Dutch case, and shows positive returns from related variety on employment growth, but not productivity or unemployment growth, at the sub-national geographic NUTS3 level. Brachert et al. (2011) show that related variety was one of the main sources of German regional employment growth during the period 2003-2008. They found that related sectors foster regional economic growth. Firgo and Mayerhofer (2018) studied related variety and employment growth using highly disaggregated sub-regional data for Austria level and found that unrelated variety positively affects employment growth. Mameli et al. (2012) take account of sectoral heterogeneity and show that related variety seems to have a stronger influence on knowledge intensive service sectors compared to manufacturing industries. In an analysis of the impact of related variety in the case of Finland, Hartog et al. (2012) found no evidence of employment growth being affected by related variety. However, after decomposing high-tech and low/medium-tech sectors, they found that related variety had a positive impact on employment growth in high-tech sectors. Bishop and Gripaios (2010) argue that in investigating the effect of related variety on regional employment growth, distinguishing between manufacturing and services might be an oversimplification. Since sectors are heterogeneous, the mechanisms behind knowledge spillovers may differ among sectors. A more recent study by Aarstad et al. (2016) proposes a multi-level analysis of Norwegian data and shows that related industrial variety has a positive effect on firm innovation, and that unrelated variety has a negative effect on productivity.

Studies of China's economy investigating the effects of related and unrelated variety on firm level performance are scarce although Howell et al. (2018) examine the effects of related and unrelated variety on the survival chances of new entrepreneurial firms in China.

3. The Dataset and the Econometric Methodology

3.1. The Dataset

Our study focuses on firm level economic growth of a sample of Chinese manufacturing firms during the period 2006-2013. Our main source of data is the non-listed enterprise database, which is a micro-database at the enterprise level, and is also the most comprehensive enterprise database. The database is covering the basic information for enterprise with an income (in the main business activity) above 5 million of yuan in China, such as enterprise name, address, contact information, etc; financial information for enterprise, including revenue, profit, assets, liabilities and other financial indicators; production information, such as production capacity, output, capacity utilization, etc; labor supply information, such as number of employees and job distribution; industry category information, geographic location information (which region the enterprise is belonging to), ownership information (state-owned, privately owned, foreign-funded, etc), export status information, historical business data and so on. The database is compiled according to the standards published by the Chinese National Bureau of Statistics. It covers a wide range of economic activities such as the extractive and electricity industry, all manufacturing sectors, and gas and water production.

To construct our dataset we selected two years 2006 and 2013. We excluded observations with missing or invalid information on location (district level), industry category (2-digit level, according to the national industrial classification-2003, National industry classification and code (GB/T 4754—2002) are showed in **Appendix Table A1**), production, sales, and employment. We are interested in firms in operation both in 2006 and 2013. We dropped firms that started operations after 2006 and firms that changed location during the time period.

This left a final sample of 84,868 observations covering almost all 2-digit manufacturing industries from sector 13 to 43 (such as food manufacturing industry, textile industry, furniture manufacturing industry, medical manufacturing industry, paper and paper products industry, etc), except sectors 16 (Tobacco) and 25 (Oil processing, coking and nuclear fuel processing). The 84,868 sample observations are at firm level, with the variables including legal person code for identifying the exact firm, industry code at 2-digit level and industry code at 4-digit level for identifying the industry category of the firm, number of employees, main business income at year 2006, main business income at year 2013, starting operation year, industrial intermediate input, gross output value, value-added of industry, total fixed assets, total current assets, sales costs, sales revenue, province code, city code, district code (for geographic information of the firm) and so on.

3.2. Measuring Related and Unrelated Variety

We use entropy measures to define related and unrelated variety (Frenken et al., 2007). The entropy measures drawing from the construction of system entropy in information theory, are usually used to measure the disorder or randomness of a system, and typically used to describe the degree of disorder or uncertainty in a system. The calculation formulation is: $H(X)=\sum p(x)\log\frac{1}{p(x)}$, where X is the state of system, $p(x)$ is the probability of being in system state X. The higher probability of being in a specific state means the lower uncertainty of the overall system. The main advantage of entropy measures is that they can be decomposed at each digit sectoral level. The decomposable nature of entropy measure implies that variety can enter the regression without causing collinearity (Jacquemin & Barry, 1979; Attaran, 1986). Unrelated variety in each district is computed by entropy at the 2-digit level, related variety is computed by the weighted sum of entropy at the 4-digit level within each 2-digit sector.

More formally, let all four-digit sectors g located in district d at time t = 2006, fall exclusively into a two-digit sector j, where $j=1,\cdots,J$. The two-digit shares, P_{jdt}, are the sum of four-digit shares p_{gdt}:

$$P_{jdt}=\sum_{g\in j}p_{gdt} \tag{1}$$

The unrelated variety (UV), or entropy at the two-digit level is given by:

$$UV_{dt}=\sum_{j=1}^{J}P_{jdt}\log_2\left(\frac{1}{P_{jdt}}\right) \tag{2}$$

Related variety (RV), the weighted sum of entropy within each two-digit sector is given by:

$$RV_{dt}=\sum_{j=1}^{J}P_{jdt}\times\left[\sum_{g\in j}\frac{p_{gdt}}{P_{jdt}}\log_2\left(\frac{1}{p_{gdt}/P_{jdt}}\right)\right] \tag{3}$$

According to the clarification of how related and unrelated variety are defined and calculated above, we can say the higher value of unrelated variety means a higher degree irrelevant diversification, the higher value of related variety means a higher degree of relevant diversification.

3.3. Econometric Methodology

We investigate the effect of related and unrelated variety on firm level economic growth in China during the period 2006-2013. To our knowledge, this is the first analysis of Gibrat's Law in the context of the Chinese economy which also includes agglomeration measures. The firm level economic growth equation is defined as follows:

$$\begin{aligned} &\mathit{Income\ Growth}_{ijd} \\ &= \alpha + \beta \log\left(\mathit{Income}_{ijdt}\right) + \rho age_{ijdt} + vVD_{ijdt} + \delta \log\left(popdens_{ct}\right) \\ &\quad + \gamma RV_{dt} + \tau UV_{dt} + \epsilon_p + \theta_j + \vartheta_s + \varepsilon_{ijdt} \end{aligned} \tag{4}$$

The dependent variable is firm income growth, which is defined (in logs) as: (T = 2013, t = 2006): $\mathit{Income\ Growth}_{ijd} = \ln\left(income_{ijd,T}\right) - \ln\left(income_{ijd,t}\right)$. This variable measures the economic growth of firm i operating in the 2-digit sector j and located in district d between the year t = 2006 and T = 2013. We use the main business income of each firm to proxy for its income level.

In addition to related variety and unrelated variety, we include explanatory variables for: income level in the initial year 2006 ($Income_{ijdt}$), firm age (age_{ijdt}) defined as 2006 minus the firm's start year, and firm's level of vertical disintegration (VD_{ijdt}) defined as the ratio between industrial intermediate input and gross output value in 2006:

$$\mathit{Vertical\ Disintegration}_{ijdt} = \frac{\mathit{purchased\ intermediate\ input}_{ijdt}}{\mathit{gross\ output\ value}_{ijdt}} \tag{5}$$

To control for geographic heterogeneity in our main specifications we introduce a measure for population density ($popdens_{ct}$) calculated as the population of city c in 2006 per square-kilometer. This variable, which is in logarithm form, represents for urbanization economies.

ϵ_p denotes a set of geographic dummies defined at the provincial level to capture systematic differences across geographic areas in terms of natural resources, public infrastructures, social capital, industrialization and policy efficiency. In China administrative divisions include several levels. Provincial is the first level and includes provinces, autonomous regions, municipalities, and special administrative regions. The second level is the prefectural-city level which includes prefecture-level cities and prefectures. In 2019 there are 34 provincial units (23 provinces, 5 autonomous regions, 4 municipalities, and 2 special administrative regions) and 333 prefectural units. Cities are further sub-divided into districts. Central and local governments have different place-based policies such as tax incentives and public subsidies. Natural resources, technology, education, and health systems can vary significantly across regions.

θ_j is a set of industry dummies defined at the 2-digit level. They are introduced to control for productive, organizational, and technological differences. The scale theory points out that firm size influences firm economic growth. Therefore, we include ϑ_s which is a set of dummy variables, defined according to firm size (measured as number of employees). Specifically, we consider four size dummies: 1) small firms with 0 - 50 employees; 2) medium-small firms with 50 - 95 employees; 3) medium-large firms with 95 - 200 employees; and 4) large firms with over 200 employees.

Table A2 in Appendix shows Firms and employees distribution at province levels, and **Table A3** in Appendix show Firms and employees distribution at 2-digit level. Definition for main variables and Correlation matrix of selected explanatory variables are showed in **Tables A4-A6** respectively.

3.4. Identification Strategy

There are two main econometric problems related to our estimation procedure. The first concerns the sample selection and the second the endogeneity of our main explanatory variables: related and unrelated variety.

We estimate our firm level income growth equation adopting a Heckman (1979) two-step sample selection model. We observe income-growth only for the sub-sample firms that survived over the period 2006-2013, which means that if we estimate our main equation only for this sub-sample of firms, the OLS estimates would be biased. There is a clear sample-selection problem. To mitigate this problem, we adopt a two-step sample selection method. To capture the non-random survival of firms during years 2006 and 2013, first we estimate a probit regression for firm survival, with a binary dependent variable which takes value 1 if the firm is observed at both the beginning ($t = 2006$) and the end of the period ($T = 2013$) and is 0 otherwise. Then we estimate the augmented firm economic growth equation (4) including the inverse Mills ratio. The firm survival is modeled as an unknown non-linear function (Griffith et al., 2009) on firm size, fixed assets, output value, and total profit (all in logs) for the start year of 2006. These firm level characteristics are suitable excluded variables from the economic growth equation since they affect the probability of firm survival. Since the non-linear functional form determining the firm's exit decision is unknown, we follow Olley and Pakes (1996) and Pavcnik (2002) and adopt a semi-parametric specification which approximates the unknown function with a polynomial expansion in firm size, log fixed assets, log output value, and log total profit, and their interactions.

Summing up, the Heckman model is estimated as follows. First, the probit model is estimated for the whole sample; then the inverse Mills ratio (λ) obtained from the selection equation is added to the economic growth equation (4) as an additional regressor to correct for sample selection bias. Finally, the augmented version of the economic growth regression is estimated using OLS on the sub-sample of firms that survived over the 2006-2013 period. The Heckman two-step method provides unbiased and consistent estimates for the growth equation (4).

Although we corrected for sample selection bias, our estimates might still be biased by the (potential) presence of endogeneity of related and unrelated variety. Endogeneity occurs for several reasons including reverse causality. For example, related and unrelated variety might explain firm economic growth, but firm economic growth might induce a "leadership effect", which might attract other up-stream and down-stream firms to locate around this high-performing firm. This could lead to clustering/agglomeration processes along the supply chain which might generate related and unrelated variety. Also, exogenous shocks can affect firms' economic performance and regional industrial distribution simultaneously.

To account for the endogeneity problem, we follow the strategy proposed by Autor and Duggan (2003), a modification of Bartik's (1991) shift-share approach. The main idea is that in absence of sector specific or local city level shocks each industrial sector at the local level (in our case the district level) will have experienced the same dynamics (in terms of employment) experienced at the national level over the period 2000-2005. In other words, the instrument variables should exclude any shock associated to China joining WTO in 2001 which would be specific to both the industrial sector and the local area. Accordingly, we construct instruments for related and unrelated variety. IV_{ru} accounts for sectoral variations at the four-digit level within the same two-digit sector, and is defined as:

$$IV_{rv} = \sum_{\substack{g=1 \\ g\in j}}^{G} \left\{ \left(\frac{n_{gd(t-6)}}{\sum_{\substack{g=1 \\ g\in j}}^{G} n_{gd(t-6)}} \right) \left[\log\left(n_{g(-d)(t-1)}\right) - \log\left(n_{g(-d)(t-6)}\right) \right] \right\} \tag{6}$$

where $n_{gd(t-6)}$ is the number of employees in the four-digit sector g within a two-digit sector j ($g \in j$), and located in the district d at time $t-6=2000$; the terms $n_{g(-d)(t-1)}$ and $n_{g(-d)(t-6)}$ denote the number of employees working in a four-digit sector g at the national level excluding the district d, at the times $t-1=2005$ and $t-6=2000$ respectively. We choose year 2000 (before China's accession to the WTO) to construct the "share" component in order to capture the original state before the shock (joining the WTO). While the "shift" component is defined for the period 2000-2005, considering a year lag (at 2005) to relax the endogeneity issue. The instrument for related variety specified in equation (6) for each four-digit sector g falling within a two-digit sector j, is used to calculate the employment shares in district d. This means that the share referring to year 2000 for each four-digit sector at the local level changes depending on the district considered. Then during the period 2000-2005 these four-digit sector shares multiplied by the change in employment in the same four-digit sector at national level but excluding the reference district d, are summed over the corresponding 4-digit sectors. Thus, the instrumental variable (IV) captures dynamics which are specific to the four-digit sectors within each two-digit sector for each district.

The instrumental variable (IV_{uv}) accounts for the variation at the unrelated two-digit sector level and is defined as follows:

$$IV_{uv} = \sum_{j=1}^{J} \left\{ \left(\frac{n_{jd(t-6)}}{\sum_{j=1}^{J} n_{jd(t-6)}} \right) \left[\log\left(n_{j(-d)(t-1)}\right) - \log\left(n_{j(-d)(t-6)}\right) \right] \right\} \tag{7}$$

where $j=1,2,\cdots,J$ denotes two-digit sectors. In this case, for each two-digit sector j located in district d, the IV is defined by calculating the shares in term of number of employees working for that specific two-digit sector j and located in district d among all two-digit sectors' employment within district d. This share of each two-digit sector, defined for the year 2000 ($t - 6 = 2000$) changes depending on the district d considered. Again, the "shift" term is calculated as the rate of change in terms of number of employees observed for the same two-digit sector and at the national level, but excluding the district d of reference, during period 2000-2005. Then the "shift" part multiplied by the "share" part we mentioned above, are summed over the corresponding 2-digit sectors, which constructs our instrument for unrelated variety, capturing the dynamics which are particular to each two-digit sector and each district.

Finally, we adopt the method proposed by Wooldridge (2010) to solve this estimation addressing sample selection and endogeneity simultaneously. First, we estimate a reduced-form selection equation using a probit model with the set of external instrumental variables (IV_{rv}, IV_{uv}) and the non-linear form exclusion restriction added to the exogenous variables entering Equation (4), and excluding the two endogenous variables (RV_{jct}, UV_{jct}). Second, we estimate firm economic growth equation via a Two-Stage Least Square (2SLS) regression with the inverse Mills ratio obtained from the first-stage selection model as an additional regressor. In addition, we also estimate equation (4) with a Generalized Method of Moments (GMM) approach for a comparison. Standard errors are clustered at the district level in all the specifications, which allows the error term to be correlated across the firms in each district (Bertrand et al., 2004).

The endogeneity of related and unrelated variety is tested using the Durbin χ^2 statistic and the Cragg-Donald Wald F statistic (Wooldridge, 2010). The null hypothesis is that the variables are exogenous. Our test rejects this hypothesis in the case of both specifications. This means that these two variables are endogenous in the equation. According to the weak identification test, the Wald F statistic is above the rule of thumb value of 10 in all specifications. Thus, in our case we have no weak instruments in our case.

To check whether these results hold for different sized firms, we split the full sample into four sub-samples according to the firm size (already defined). Also we focus on firms located in well-developed regions i.e. the top three Chinese regions based on the value of their gross regional product in the start year 2006.

4. Empirical Results and Findings

4.1. OLS Estimation Results with Correcting the Sample Selection

Table 1 reports the OLS estimates of our income growth equation correcting for the sample selection bias. In all specifications bootstrapped standard errors are clustered at the district level to allow the error term to be correlated across the firms in each district. Column (1) presents the results of our baseline specification without either related or unrelated variety. Columns (2) and (3) respectively show the results for the specifications with related and unrelated variety. Column (4) includes both variables.

Table 1. Sectoral variety and firm economic growth—OLS Method.

Estimation Method	OLS	OLS	OLS	OLS
$income\ growth_{ijd}$	(1)	(2)	(3)	(4)
$\log(income_{ijdt})$	−0.347***	−0.325***	−0.324***	−0.323***
	(0.011)	(0.011)	(0.011)	(0.011)
age_{ijdt}	−0.007***	−0.007***	−0.007***	−0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
VD_{ijdt}	−0.283***	−0.262***	−0.275***	−0.270***
	(0.044)	(0.041)	(0.040)	(0.040)
$\log(popdes_{ct})$	−0.132***	0.027	0.077*	0.082*
	(0.037)	(0.038)	(0.039)	(0.040)
Small size	Ref.	Ref.	Ref.	Ref.
Medium small size	−0.156***	−0.195***	−0.194***	−0.193***
	(0.019)	(0.019)	(0.019)	(0.018)
Medium large size	−0.243***	−0.294***	−0.288***	−0.288***
	(0.022)	(0.022)	(0.022)	(0.021)
Large size	−0.237***	−0.302***	−0.294***	−0.294***
	(0.027)	(0.026)	(0.026)	(0.025)
Geographic dummy	Yes	Yes	Yes	Yes
Industrial dummy	Yes	Yes	Yes	Yes
RV_{dt}		−0.394***		−0.002
		(0.038)		(0.056)
UV_{dt}			−0.026***	−0.025***
			(0.002)	(0.004)
lambda	−7.402***	−10.417***	−10.438***	−10.438***
	(0.713)	(0.729)	(0.739)	(0.720)
No. of Obs.	83,067	83,067	83,067	83,067
R-Squared	0.498	0.503	0.507	0.507
Log Likelihood	−119,115	−118,684	−118,369	−118,334
Selection Equation				
No. of Obs.	84,868	84,868	84,868	84,868
Pseudo-R-Squared	0.007	0.015	0.015	0.015
Log Pseudo-Likelihood	−8594.431	−8529.757	−8526.469	−8523.460
Wald Chi Square [p-value]	128.94 [0.000]	195.29 [0.000]	192.40 [0.000]	192.36 [0.000]

Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Bootstrapped Standard errors clustered at district level are shown in parentheses. All specifications include provincial geographic dummies and 2-digit industrial dummies and a constant term. lambda denotes the Inverse Mills Ratio from the selection equations.

From the results in Table 1, the inverse Mills ratio (lambda) is negative and statistically significant indicating that we need to correct for sample selection.

Besides, column (2) and column (3) shows the strong negative effects of related variety and unrelated variety on firm's income growth when introducing them in the model respectively. Column (4) suggests only negative effects of unrelated variety on firm's income growth holds the significance when introducing the both variables, related variety and unrelated variety.

The negative impact of unrelated variety on firm's income growth, which is in line with previous work about the effects of related and unrelated variety on regional employment growth (Frenken et al., 2007; Saviotti & Frenken, 2008), could be considered as it may lead to resource fragmentation and can not promote effective knowledge exchange and collaborative innovation. The negative impacts of related variety on firm's economic growth is different with some previous works (Saviotti & Frenken, 2008; Boschma & Iammarino, 2009; Boschma & Frenken, 2012; Mameli et al., 2012) which related variety usually has positive effects on employment growth, value-added growth and labor productivity. The reason for this could be that related variety may lead to excessive consumption of resources (resource stress), such as land, water resources, energy, etc., which can increase production costs and affect firm's income growth.

Firm income level at the start year of the analysis has a negative effect on income growth; the estimated coefficients of the variable $\log(Income_{ijdt})$ are negative and statistically significant in all the specifications, and the estimated elasticities are similar at around 30%. Since the estimated coefficient of the firm income level in the initial year is different from 0, we can state that the Gibrat's Law does not hold for the Chinese case. Smaller firms grow at a different rate from larger firms. We find that the impacts of firm age and vertical disintegration are negative and significant. This confirms Gibrat's theory of unproportional increases in firms' income. Among firms that survived during the financial crisis, the larger the firm the worse their income performance: the estimated coefficients of medium, medium-large and large firms are all negative and highly statistically significant. Moreover, the magnitude of the coefficients increases with firm size. This finding is not surprising. Small firms which survived during the economic crisis generally experienced fewer shocks to their economic performance compared to larger firms. This might be that small firms tend to be less internationalized, and thus less exposed to external shocks than larger firms which are generally more involved in international activities. For this reason, they were less hit by the financial crisis and have had a better economic performance.

4.2. IV Estimation Results with Correcting the Endogeneity Issue

Results of the IV estimation correcting for endogeneity are reported in **Table 2**. Column (1) reports the results of the IV Two-Stage-Least-Square method, column (2) presents the results of the IV-GMM. Both methods provide quite similar results after 500 replications of the bootstrapped standard errors. The results suggest that after correcting for the endogeneity of related and unrelated variety these variables do not have a significant effect on firm level economic growth.

Table 2. Sectoral variety and firm economic growth—IV Method.

Estimation Method	IV-TSLS	IV-GMM
income growth$_{ijd}$	(1)	(2)
log(*income*$_{ijdt}$)	−0.346***	−0.346***
	(0.104)	(0.104)
age$_{ijdt}$	−0.007***	−0.007***
	(0.012)	(0.012)
VD$_{ijdt}$	−0.307	−0.307
	(1.395)	(1.395)
log(*popdes*$_{ct}$)	−0.045	−0.045
	(0.754)	(0.754)
Small size	Ref.	Ref.
Medium small size	−0.148***	−0.148***
	(0.327)	(0.327)
Medium large size	−0.226***	−0.226***
	(0.555)	(0.555)
Large size	−0.216***	−0.216***
	(0.619)	(0.619)
Geographic dummy	Yes	Yes
Industrial dummy	Yes	Yes
RV$_{dt}$	0.894	0.894
	(6.210)	(6.210)
UV$_{dt}$	−0.052	−0.052
	(1.154)	(1.154)
lambda	−7.582***	−7.582***
	(2.029)	(2.029)
Endogeneity Test(chi-square [*p*-value])	7.129 [0.028]	7.129 [0.028]
Cragg-Donald Wald F statistics	894.789	894.789
No. of Obs.	83,067	83,067
R-Squared	0.448	0.448
Log Likelihood	−120,235	−120,235
Chi Square	117954.10	117954.10
Selection Equation		
No. of Obs.	84,868	84,868
Pseudo-R-Squared	0.015	0.007
Log Pseudo-Likelihood	−8523.46	−8593.29
Wald Chi Square [*p*-value]	192.36 [0.000]	131.40 [0.000]

Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Bootstrapped standard errors clustered at district level show are in parentheses. All specifications include provincial geographic dummies and 2-digit industrial dummies and a constant term. Lambda denotes the Inverse Mills Ratio from the selection equations.

The estimated negative and significant coefficient for the inverse Mills ratio (lambda) suggests the need to correct for sample selection. We employ the Durbin χ^2 statistic to test for endogeneity of related and unrelated variety. Our null hypothesis is rejected: these variables are not exogenous. The values of the Cragg-Donald Wald F statistic are higher than the conservative cut-off value of 10 in both specifications. The sign of the estimated coefficients of unrelated variety remains negative (as in the OLS estimates) and confirms previous studies (Frenken et al., 2007). The magnitude of these coefficients is larger at around 0.05 compared to 0.025 in the OLS estimations. The negative sign in the former case shows that the OLS estimations are upward biased.

As mentioned before, exogenous shocks can affect firms' economic performance and regional industrial distribution simultaneously. Thus, after correcting endogeneity with Bartick shift-share approach, the impacts of related and unrelated variety on firm's income growth become insignificant. Besides, another aspect can be considered to induce this result. The two variables for agglomeration—related and unrelated variety—may be associated with the control variables—geographic dummies and firm size dummies—through the infrastructure level, talent supply, industry chain collaboration, market size and potential, international integration level, and policy supports—China's economy contains the tradition of government intervention. Industrial policy plays important roles in different stages of China's economic development, especially in developed regions. For further study, we do the sub-sample analysis for different firm size and for firms in highly developed regions. .

4.3. Robustness Check Results for Different Firm Size and Highly Developed Regions

The results in **Table 1** and **Table 2** allow us to distinguish between agglomeration externalities generated by different but related sectors and by totally different and unrelated sectors for the whole sample. This allows as to identify which is relatively more important for income growth across the firms in a localized industry. For further robustness check, we do the sub-sample analysis for different firm size and for firms in highly developed regions.

Table 3 reports the IV-TSLS estimates for different size firms. The estimated coefficients vary widely between groups. Inverse Mills ratios are positive for small and medium-small firms, but negative for medium-large and large firms. Related variety is significantly positive only for large firms which can include state-owned firms and subsidiaries of multinational companies. Nevertheless, we can say that these units often have stronger networking, R&D and resource integration capabilities, seem to be more able to absorb knowledge and be better benefit from technological advances from related sectors, as they are able to translate new technologies into innovative products or services more quickly, thus to transform these external flows into higher economic performance. Unrelated variety is not significant for all sub-groups which is in line with the results

Table 3. Sub-sample estimation for firm size heterogeneity (IV-TSLS).

Estimation Method	TSLS	TSLS	TSLS	TSLS
$income\ growth_{ijd}$	Small Size (1)	Medium Small (2)	Medium Large (3)	Large Size (4)
$\log(income_{ijdt})$	−0.723***	−0.588***	−0.431***	−0.241***
	(0.050)	(0.020)	(0.020)	(0.010)
age_{ijdt}	−0.013***	−0.013***	−0.016***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
VD_{ijdt}	−0.129	−0.081	−0.306***	−0.197**
	(0.100)	(0.060)	(0.090)	(0.080)
$\log(popdes_{ct})$	−0.162	−0.075	0.000	−0.252***
	(0.260)	(0.110)	(0.110)	(0.070)
Geographic dummy	Yes	Yes	Yes	Yes
Industrial dummy	Yes	Yes	Yes	Yes
RV_{dt}	0.806	0.9	0.791	0.619*
	(1.390)	(0.620)	(0.790)	(0.360)
UV_{dt}	−0.025	−0.04	−0.05	−0.024
	(0.100)	(0.040)	(0.050)	(0.020)
Lambda	4.631***	4.999***	−3.955***	−17.696***
	(1.560)	(1.480)	(1.340)	(2.320)
Endogeneity Test (chi-square [p-value])	7.393 [0.025]	9.362 [0.009]	3.652 [0.161]	18.283 [0.000]
Cragg-Donald Wald F statistics	27.033	180.622	158.923	453.833
No. of Obs.	14,376	19,765	21,665	27,261
Model F Statistic [p-value]	208.57 [0.000]	274.61 [0.000]	290.66 [0.000]	423.25 [0.000]
R-Squared	0.447	0.44	0.441	0.437
Log Likelihood	−20181.1	−28007.3	−30679.9	−39552.3

Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at district level are shown in parentheses. Industry dummy are constructed at 2-digit industrial level, Geographic dummy are constructed at province level.

for the whole sample. Firm level characteristics such as firm income level in the initial year, age, and vertical disintegration are significant for medium-large and large firms.

We ran a sub-sample analysis on the firms in top three most developed regions based on the regional gross product value in 2006. The results are presented in Table 4. The negative and significant coefficients for inverse Mills ratio indicate the need for sample selection. However, in contrast to the result for the whole sample, related variety has a positive and statistically significant impact on firm-level economic growth. The economic interpretation of this result is interesting. The industry related variety in a locality is not a sufficient condition to guarantee cross-fertilization and transfer of information, knowledge, and technologies among different local industries. For the mechanisms underlying

Table 4. Sub-sample estimation for top 3 developed regions (IV-TSLS).

Estimation Method	IV-TSLS
$income\ growth_{ijd}$	(1)
$\log(income_{ijdt})$	−0.344***
	(0.030)
age_{ijdt}	−0.009***
	(0.000)
VD_{ijdt}	−0.182*
	(0.100)
$\log(popdes_{ct})$	−0.068
	(0.150)
Small size	Ref.
Medium small size	−0.126***
	(0.040)
Medium large size	−0.203***
	(0.040)
Large size	−0.238***
	(0.060)
Geographic dummy	Yes
Industrial dummy	Yes
RV_{dt}	1.241*
	(0.710)
UV_{dt}	−0.060*
	(0.030)
lambda	−5.775***
	(1.580)
Endogeneity Test(chi-square [p-value])	8.424 [0.015]
Cragg-Donald Wald F statistics	769.899
No. of Obs.	33,669
R-Squared	0.427
F statistics [p-value]	323.848 [0.000]
Log Likelihood	−50320.77

Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at district level are shown in parentheses. Industry dummy are constructed at 2-digit industrial level, Geographic dummy are constructed at province level.

inter-industry spillovers to work requires that the region is characterized also by a high level of economic and technological development. This finding confirms previous empirical studies which show that related variety is more effective in a high-tech sector context. When we consider highly developed Chinese regions, unrelated variety also becomes significant. We find that unrelated variety has a negative and significant effect on firm economic growth. Finally, firm specific characteristics such as income level in the start year 2006, age, and vertical disintegration are all negative and significant.

5. Conclusion

We used a sample of 84,868 Chinese manufacturing firms during the period 2006-2013 to test Gibrat's Law based on the joint presence of related and unrelated variety (Frenken et al., 2007). Our results reject Gibrat's Law even if we include agglomeration externalities. We analyzed the roles played by related and unrelated variety in firm-level economic growth and found that unrelated variety has a significant and negative effect on firm economic growth if we correct for sample selection bias. After correcting for endogeneity, both related and unrelated varieties were insignificant. We consider this to be an interesting result raising questions about the findings of related variety in studies that do not correct for these econometric problems.

We obtained some interesting results when we disaggregate our dataset according to firm size and region of location. In the case of large firms, related variety seems to have a significant positive effect. This is also true for firms located in highly-developed Chinese regions where related variety significantly and positively influences firm economic performance, while unrelated variety has a negative impact. Confirming previous work we found that local knowledge spillovers are more easily absorbed by firms operating in similar and related sectors. Our results seem to confirm that these types of inter-industry knowledge spillovers are the main drivers of economic growth only among Chinese manufacturing firms in highly-developed regions and not in rural and less-industrialized regions. We showed that large firms, most likely state-owned firms or subsidiaries of multinational companies, seem to have more capacity to transform the opportunities in their local environment into higher economic performance.

Our research has some limitations. First, the analysis is based on a cross-section. Panel datasets would allow us to conduct a more sophisticated analysis of the dynamics of firms' economic growth by taking account of time lags. Second, our sample period is short, 2006-2013, and includes the period of the Great Recession which coincided also with significant changes of the Chinese economy.

To conclude, we believe that our study contributes to the empirical literature on the relationship between agglomeration and firms' economic performance in the context of a developing country China. We found no strong effects of related variety and unrelated variety on firm level economic growth after correcting endogeneity. According to our findings, it seems to rely mainly on firm and regional specific characteristics.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Aarstad, J., Kvitastein, O. A., & Jakobsen, S. E. (2016). Related and Unrelated Variety as Regional Drivers of Enterprise Productivity and Innovation: A Multilevel Study. *Research Policy, 45,* 844-856. https://doi.org/10.1016/j.respol.2016.01.013

Attaran, M. (1986). Industrial Diversity and Economic Performance in US Areas. *The Annals of Regional Science, 20,* 44-54. https://doi.org/10.1007/BF01287240

Audretsch, D. B. (1995). Innovation, Growth and Survival. *International Journal of Industrial Organization, 13,* 441-457. https://doi.org/10.1016/0167-7187(95)00499-8

Audretsch, D. B., Santarelli, E., & Vivarelli, M. (1999). Start-Up Size and Industrial Dynamics: Some Evidence from Italian Manufacturing. *International Journal of Industrial Organization, 17,* 965-983. https://doi.org/10.1016/S0167-7187(98)00002-2

Autor, D. H., & Duggan, M. G. (2003). The Rise in the Disability Rolls and the Decline in Unemployment. *Quarterly Journal of Economics, 118,* 157-206. https://doi.org/10.1162/00335530360535171

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research. https://doi.org/10.17848/9780585223940

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics, 119,* 249-275. https://doi.org/10.1162/003355304772839588

Bishop, P., & Gripaios, P. (2010). Spatial Externalities, Relatedness and Sector Employment Growth in Great Britain. *Regional Studies, 44,* 443-454. https://doi.org/10.1080/00343400802508810

Boschma, R., & Frenken, K. (2012). Technological Relatedness and Regional Branching. In *Beyond Territory* (pp. 64-81). Routledge.

Boschma, R., & Immarino, S. (2009). Related Variety, Trade Linkages, and Regional Growth in Italy. *Economic Geography, 85,* 289-311. https://doi.org/10.1111/j.1944-8287.2009.01034.x

Brachert, M., Kubis, A., & Titze, M. (2011). *Related Variety, Unrelated Variety and Regional Functions: Identifying Sources of Regional Employment Growth in Germany from 2003 to 2008 (No. 15/2011).* IWH Discussion Papers.

Cainelli, G. (2008). Industrial Districts: Theoretical and Empirical Insights. In C. Karlsson (Ed.), *Handbook of Research on Cluster Theory* (pp. 189-202). Edward Elgar Publishing. https://doi.org/10.4337/9781848442849.00019

Cainelli, G., & De Liso, N. (2005). Innovation in Industrial Districts: Evidence from Italy. *Industry and Innovation, 12,* 383-398. https://doi.org/10.1080/13662710500195991

Cainelli, G., & Ganau, R. (2019). Related Variety and Firm Heterogeneity. What Really Matters for Short-Run Firm Growth? *Entrepreneurship & Regional Development, 31,* 768-784. https://doi.org/10.1080/08985626.2019.1571636

Cainelli, G., Di Maria, E., & Ganau, R. (2014). An Explanation of Firms' Internationalisation Modes, Blending Firm Heterogeneity and Spatial Agglomeration: Microevidence from Italy. *Environment and Planning A, 46,* 943-962. https://doi.org/10.1068/a4690

Cainelli, G., Ganau, R., & Iacobucci, D. (2016). Do Geographic Concentration and Vertically Related Variety Foster Firm Productivity? Micro-Evidence from Italy. *Growth and Change, 47,* 197-217. https://doi.org/10.1111/grow.12112

Calvo, J. L. (2006). Testing Gibrat's Law for Small, Young and Innovating Firms. *Small Business Economics, 26,* 117-123. https://doi.org/10.1007/s11187-004-2135-5

Caves, R. E. (1998). Industrial Organization and New Findings on the Turnover and Mobility of Firms. *Journal of Economic Literature, 36,* 1947-1982.

Chesher, A. (1979). Testing the Law of Proportionate Effect. *Journal of Industrial Economics, 27,* 403-411. https://doi.org/10.2307/2097961

Firgo, M., & Mayerhofer, P. (2018). (Un)related Variety and Employment Growth at the Sub-Regional Level. *Papers in Regional Science, 97,* 519-547. https://doi.org/10.1111/pirs.12276

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies, 41,* 685-697. https://doi.org/10.1080/00343400601120296

Geroski, P. A. (1995). What Do We Know about Entry? *International Journal of Industrial Organization, 13,* 421-440. https://doi.org/10.1016/0167-7187(95)00498-X

Gibrat, R. (1931). *Les Inégalits économiques*. Sirey.

Glaeser E. L., Kallal H. D., Scheinkman J. A., & Shleifer A. (1992). Growth in Cities. *Journal of Political Economy, 100,* 1126-1152. https://doi.org/10.1086/261856

Griffith, R., Redding, S., & Simpson, H. (2009). Technological Catchup and Geographic Proximity. *Journal of Regional Science, 49,* 689-720. https://doi.org/10.1111/j.1467-9787.2009.00630.x

Harhoff, D., Stahl, K., & Woywode, M. (1998). Legal Form, Growth and Exit of West German Firms—Empirical Results for Manufacturing, Construction, Trade and Service Industries. *Journal of Industrial Economics, 46,* 453-488. https://doi.org/10.1111/1467-6451.00083

Hartog, M., Boschma, R., & Sotarauta, M. (2012). The Impact of Related Variety on Regional Employment Growth in Finland 1993-2006: High-Tech versus Medium/Low-Tech. *Industry and Innovation, 19,* 459-476. https://doi.org/10.1080/13662716.2012.718874

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47,* 153-161. https://doi.org/10.2307/1912352

Henderson, J. V. (2003). Marshall's Scale Economies. *Journal of Urban Economics, 53,* 1-28. https://doi.org/10.1016/S0094-1190(02)00505-3

Howell, A., He, C., Yang, R., & Fan, C. C. (2018). Agglomeration, (un)-Related Variety and New Firm Survival in China: Do Local Subsidies Matter? *Papers in Regional Science, 97,* 485-500. https://doi.org/10.1111/pirs.12269

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

Jacquemin, A. P., & Berry, C. H. (1979). Entropy Measure of Diversification and Corporate Growth. *The Journal of Industrial Economics, 27,* 359-369. https://doi.org/10.2307/2097958

Lotti, F., Santarelli, E., & Vivarelli, M. (2003). Does Gibrat's Law Hold among Young, Small Firms? *Journal of Evolutionary Economics, 13,* 213-235. https://doi.org/10.1007/s00191-003-0153-0

Maine, E. M., Shapiro, D. M., & Vining, A. R. (2010). The Role of Clustering in the Growth of New Technology-Based Firms. *Small Business Economics, 34,* 127-146. https://doi.org/10.1007/s11187-008-9104-3

Mameli, F., Iammarino, S., & Boschma, R. (2012). *Regional Variety and Employment Growth in Italian Labour Market Areas: Services versus Manufacturing Industries.* CIMR Research Working Paper.

Mansfield, E. (1962). Entry, Gibrat's Law, Innovation, and the Growth of Firms. *American Economic Review, 52,* 1023-1051.

Marshall, A. (1920). *The Economics of Industry.* Macmillan and Company.

Martin, P., Mayer, T., & Mayneris, F. (2011). Spatial Concentration and Plant-Level Productivity. *Journal of Urban Economics, 69,* 182-195. https://doi.org/10.1016/j.jue.2010.09.002

Nooteboom, B. (2000). *Learning and Innovation in Organizations and Economies.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199241002.001.0001

Olley, S., & Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica, 64,* 1263-1297. https://doi.org/10.2307/2171831

Pavcnik, N. (2002). Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants. *Review of Economic Studies, 69,* 245-276. https://doi.org/10.1111/1467-937X.00205

Reid, G. C. (1995). Early Life-Cycle Behaviour of Micro-Firms in Scotland. *Small Business Economics, 7,* 89-95. https://doi.org/10.1007/BF01108684

Saviotti, P. P., & Frenken, K. (2008). Export Variety and the Economic Performance of Countries. *Journal of Evolutionary Economics, 18,* 201-218. https://doi.org/10.1007/s00191-007-0081-5

Wagner, J. (1992). Firm Size, Firm Growth, and Persistence of Chance: Testing GIBRAT's Law with Establishment Data from Lower Saxony, 1978-1989. *Small Business Economics, 4,* 125-131. https://doi.org/10.1007/BF00389853

Weiss, C. R. (1998). Size, Growth, and Survival in the Upper Austrian Farm Sector. *Small Business Economics, 10,* 305-312. https://doi.org/10.1023/A:1007972518380

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

Appendix

Table A1. National industry classification and code (GB/T 4754—2002).

13	Agricultural and sideline food processing industry	28	Chemical fiber manufacturing industry
14	Food manufacturing industry	29	Rubber products industry
15	Wine, beverages and refined tea manufacturing	30	Plastic products industry
16	Tobacco Manufacturing industry	31	Non-metallic mineral products industry
17	Textile industry	32	Ferrous metal smelting and rolling processing industry
18	Textile and apparel industry	33	Non-ferrous metal smelting and rolling processing industry
19	Leather, fur, feathers and their products and footwear industry	34	Metal products industry
20	Wood processing and wood, bamboo, rattan, brown, grass product	35	General equipment manufacturing industry
21	Furniture manufacturing industry	36	Special equipment manufacturing industry
22	Paper and paper products	37	Automobile Manufacturing
23	Printing and recording media reproduction industry	39	Electrical machinery and equipment manufacturing
24	Culture, education, industry and art, sports and entertainment	40	Computer, communications and other electronic equipment manufa
25	Oil processing, coking and nuclear fuel processing	41	Instrumentation manufacturing industry
26	Chemical raw materials and chemical products manufacturing	42	Other manufacturing
27	Pharmaceutical manufacturing industry	43	Comprehensive utilization of waste resources

Table A2. Firms and employees distribution at provincial level.

provcd	Num. Firms	Percent (%)	Cum. (%)	provcd	Num. Employees	Percent (%)	Cum. (%)
130000	3882	4.01	4.01	130000	1,316,570	4.46	4.46
140000	777	0.8	4.82	140000	545,453	1.85	6.30
210000	6449	6.67	11.49	210000	1,600,686	5.42	11.72
230000	999	1.03	12.52	230000	388,082	1.31	13.04
320000	14,693	15.19	27.72	320000	4,238,566	14.35	27.39
330000	17,049	17.63	45.35	330000	3,957,860	13.40	40.79
340000	2517	2.6	47.95	340000	640,113	2.17	42.95

Continued

350000	6709	6.94	54.89	350000	2,128,293	7.21	50.16
360000	2181	2.26	57.14	360000	612,010	2.07	52.23
370000	14,625	15.12	72.27	370000	4,102,365	13.89	66.12
410000	4105	4.25	76.51	410000	1,319,083	4.47	70.59
420000	2619	2.71	79.22	420000	922,493	3.12	73.71
430000	4142	4.28	83.51	430000	875,858	2.97	76.68
440000	9071	9.38	92.89	440000	4,257,888	14.42	91.09
450000	1090	1.13	94.01	450000	319,555	1.08	92.17
510000	3020	3.12	97.14	510000	967,435	3.28	95.45
520000	406	0.42	97.56	520000	224,270	0.76	96.21
530000	610	0.63	98.19	530000	202,947	0.69	96.90
610000	1077	1.11	99.3	610000	544,508	1.84	98.74
620000	279	0.29	99.59	620000	216,444	0.73	99.47
630000	72	0.07	99.66	630000	35,317	0.12	99.59
640000	169	0.17	99.84	640000	49,158	0.17	99.76
650000	156	0.16	100	650000	71,604	0.24	100.00
Total	96,697	100		Total	29,536,558	100.00	

Note: This is firm distribution for original dataset.

Table A3. Firms and employees distribution at 2-digit level.

indcd2_2003	Num. Firms	Percent (%)	Cum. (%)	indcd2_2003	Num. Employees	Percent (%)	Cum. (%)
13	6397	6.62	6.62	13	1,167,541	3.95	3.95
14	2254	2.33	8.95	14	606,306	2.05	6.01
15	1482	1.53	10.48	15	469,599	1.59	7.60
16	69	0.07	10.55	16	67,821	0.23	7.83
17	8985	9.29	19.84	17	2,823,227	9.56	17.38
18	3644	3.77	23.61	18	1,432,771	4.85	22.23
19	2288	2.37	25.98	19	1,155,023	3.91	26.14
20	2058	2.13	28.11	20	350,344	1.19	27.33
21	1041	1.08	29.18	21	314,869	1.07	28.40
22	2519	2.61	31.79	22	583,204	1.97	30.37
23	1305	1.35	33.14	23	270,618	0.92	31.29
24	1036	1.07	34.21	24	407,719	1.38	32.67
25	695	0.72	34.93	25	369,000	1.25	33.92
26	7574	7.83	42.76	26	1,630,965	5.52	39.44
27	2309	2.39	45.15	27	705,169	2.39	41.83
28	605	0.63	45.77	28	207,121	0.70	42.53

Continued

29	1311	1.36	47.13	29	411,227	1.39	43.92
30	4215	4.36	51.49	30	844,833	2.86	46.78
31	7884	8.15	59.64	31	1,818,766	6.16	52.94
32	2266	2.34	61.98	32	1,631,851	5.52	58.46
33	2021	2.09	64.07	33	686,406	2.32	60.79
34	4711	4.87	68.95	34	1,044,448	3.54	64.32
35	8780	9.08	78.03	35	1,903,840	6.45	70.77
36	4205	4.35	82.37	36	1,149,000	3.89	74.66
37	4459	4.61	86.99	37	1,846,327	6.25	80.91
39	5967	6.17	93.16	39	1,854,056	6.28	87.19
40	3326	3.44	96.6	40	2,674,834	9.06	96.24
41	1356	1.4	98	41	478,999	1.62	97.86
42	1778	1.84	99.84	42	610,677	2.07	99.93
43	157	0.16	100	43	19,997	0.07	100.00
Total	96,697	100		Total	29,536,558	100	

Table A4. Definition for main variables.

Main varieble	Definition
$\log(income_{ijdt})$	$Income\ Growth_{ijd} = \ln\left(income_{ijd,T}\right) - \ln\left(income_{ijd,t}\right)$
age_{ijdt}	2006 minus the firm's start operation year
VD_{ijdt}	$Vertical\ Disintegration_{ijdt} = \frac{purchased\ intermediate\ input_{ijdt}}{gross\ output\ value_{ijdt}}$
$\log(popdes_{ct})$	the 2006 population in city c per square-kilometer
RV_{dt}	$RV_{dt} = \sum_{j=1}^{J} P_{jdt} \times \left[\sum_{g \in j} \frac{p_{gdt}}{P_{jdt}} \log_2\left(\frac{1}{p_{gdt}/P_{jdt}}\right)\right]$
UV_{dt}	$UV_{dt} = \sum_{j=1}^{J} P_{jdt} \log_2\left(\frac{1}{P_{jdt}}\right)$

Table A5. Correlation matrix of selected explanatory variables.

		[1]	[2]	[3]	[4]	[5]	[6]
$\log(income_{ijdt})$	[1]	1					
age_{ijdt}	[2]	0.2251	1				
VD_{ijdt}	[3]	0.0427	−0.0024	1			
$\log(popdes_{ct})$	[4]	0.0953	0.0264	0.0492	1		
RV_{dt}	[5]	0.1133	0.016	0.1042	0.4238	1	
UV_{dt}	[6]	0.1124	0.0163	0.0961	0.437	0.8966	1

Table A6. Descriptive statistics of dependent and continuous explanatory variables.

Variable	Mean	Std. Dev.	Min	Max
$income\ growth_{ijd}$	0.171	1.432	−12.622	8.039
$\log(income_{ijdt})$	10.599	1.270	5.328	18.872
age_{ijdt}	7.806	8.880	0.000	406.000
VD_{ijdt}	0.745	0.160	0.000	10.373
$\log(popdes_{ct})$	6.290	0.506	3.228	7.783
RV_{dt}	1.301	0.628	0.000	2.898
UV_{dt}	19.402	11.784	0.000	51.214
$\log(fixasset_{ijdt})$	10.063	1.437	5.257	18.390
$\log(profit_{ijdt})$	7.113	1.900	0.000	16.217
$\log(output_{ijdt})$	10.637	1.266	5.328	18.878