Received: 13 October 2020 | Revised: 19 April 2021 | Accepted: 20 April 2021

DOI: 10.1111/pirs.12608

FULL ARTICLE

Unpacking the U-shaped relationship between related variety and firm sales: Evidence from Japan

Ren Lu[1] | Qing Song[1] | Ting Xia[2] | Daguo Lv[3] | Torger Reve[4] | Ze Jian[1]

[1]School of Business Administration, Guangdong University of Finance and Economics, Guangzhou, China

[2]School of Finance, Guangdong University of Finance and Economics, Guangzhou, China

[3]Institute of Finance and Public Management, Anhui University of Finance and Economics, Bengbu, China

[4]Department of Strategy and Entrepreneurship, BI Norwegian Business School, Oslo, Norway

Correspondence
Ting Xia, School of Finance, Guangdong University of Finance and Economics. Guangzhou, China.
Email: xiating2019@gdufe.edu.cn

Funding information
Social Science Foundation of Guangdong Province, China, Grant/Award Number: GD19CGL05

Abstract

The purpose of this paper is to study how related variety influences firm sales. We apply an instrumental variable method (as well as the most recent plausible instrumental variable method in robustness tests) to analyze more than 600,000 firm observations in all of Japan's 47 prefectures. We find that related variety, as a kind of regional industrial structure, has a U-shaped relationship with firm sales. This finding enriches the related variety perspective by supplementing micro-level evidence, revealing that the "related variety–firm sales" relationship is not linear, as most prior studies have suggested.

KEYWORDS
firm sales, Japan, related variety, U-shape

JEL CLASSIFICATION
R10; L21

1 | INTRODUCTION

How related variety influences the regional economy is a critical research topic in economics (e.g., Choi & Park, 2020; Erkus-Ozturk, 2018; Naldi et al., 2020). As a concept that describes a type of regional industrial structure (Frenken et al., 2007), related variety sheds light on regional economy development in the viewpoint of analyzing regional industries' relatedness (Frenken et al., 2007). Related variety is popular among scholars and policy makers

(Boschma & Iammarino, 2009; Castaldi et al., 2015; Frenken et al., 2007), and it also refers to regional innovation policy concepts (Asheim et al., 2011; Balland et al., 2019; Boschma, 2014).

Originating from Jacobs' (1961, 1969) concept of urban economics, most related variety studies focus on analyzing regional level issues, such as how related variety influences regional entrepreneurial path dependency (Aarstad & Kvitastein, 2019), how related variety constructs regional advantages (Asheim et al., 2011), and how related variety helps form regional-level patenting (Castaldi et al., 2015). While such studies offer many insights, they also have a drawback, in that related variety perspective's micro foundation is not solid (Content & Frenken, 2016). Theoretically, without studying firm behaviours and performance, related variety perspective is not an integrated theory. Empirically, micro-level investigations are needed since firms are at the core of economic activities. If the related variety perspective cannot directly provide insights to guide entrepreneurs' behaviour, it has few practical implications at the firm level. This is probably why Content and Frenken (2016) issued a call to combine related variety with micro-level works.

Only a few papers (for example, Aarstad et al. (2016, 2019), Cainelli and Ganau (2019)) have attempted to respond to Content and Frenken's call. However, as most previous studies suggested that related variety has a linear relationship with regional development (Boschma et al., 2012), a research gap appears: does such a conclusion apply to the firm level? Inspired by Boschma's (2005) theoretical paper, which suggested an inverted U-shape relationship between proximity and firm innovation, and by Cainelli and Lupi's (2010) empirical study that found a U-shaped relationship between variety and firm employment growth, we have reason to hypothesize that a non-linear relationship may exist.

A firm's total annual sales (hereafter, firm sales) are an ideal starting point from which to fill the above research gap. Firm sales are critical: although other factors can influence firm's development, such as innovations (Arzubiaga et al., 2018) and venture capital (Park & Steensma, 2012), it is firm sales that directly reflect whether or not a firm's products/services are accepted by consumers (Chandler et al., 2009). Firm sales are both a driver and result of firm behaviours. On the one hand, increasing firm sales creates possibilities to conduct firm strategies (such as innovation strategy and investment strategy). On the other hand, one of the goals of a firm is to create business values by increasing firm sales (e.g., Coad & Srhoj, 2020; Erhardt, 2021). Therefore, studying firm sales provides an excellent angle from which to extend related variety perspective at the micro-level. Furthermore, when related variety influences regional economic development (e.g., Content & Frenken, 2016; Fritsch & Kublina, 2018), such an economic development process should be shown as firm sales grow. By tracing the trajectory changes of firm sales caused by related variety, we can directly observe whether the related variety and firm sales relationship is linear or non-linear.

The purpose of this paper is to examine the implications of related variety on firm sales. We develop a theoretical framework that notes that the "related variety—firm sales" relationship appears as a U-shape. Such a U-shape is determined by two forces: when related variety is low, resource competition is a dominating force; when related variety is high, externalities are a dominating force. Drawing on both firm-level and regional-level panel data from Japan's 47 prefectures, we apply the instrumental variable (IV) method to test our framework and combine the most recent plausibly exogenous IV method to ensure robustness (Clarke & Matta, 2018; Kippersluis & Rietveld, 2018). All of our results show that our framework makes sense.

The present paper contributes by stepping towards closing the above-mentioned research gap. More specifically, we attempt to contribute insights into the micro-level foundation for the related variety perspective. Furthermore, this study shows that the influence of related variety on firm sales is not monotonous. In other words, the linear relationship may be just a special case in the U-shape. Finally, we find that many related variety studies build on European data (e.g., Boschma et al., 2012; Lazzeretti et al., 2010), and it seems that we lack empirical evidence from Asian countries, not to mention Japan. Thus, our paper not only represents a step towards closing the above-mentioned research gap, but also enriches related variety perspective by providing Japan's evidence.

The remainder of this paper is organized as follows. Section 2 discusses related variety and develops the study's hypothesis. Section 3 introduces methodology issues and Section 4 presents regression results. Since we apply an instrumental variable method, in Section 5 we discuss that, even if our instrumental variable is not perfectly exogenous, our regression results are still stable. We also conduct other robustness tests in this section. The final section offers a discussion and concluding remarks.

2 | THEORETICAL BACKGROUND AND HYPOTHESIS

This section is dichotomized. The first subsection presents what related variety is. The second subsection develops a theoretical hypothesis, showing mechanisms of how related variety influence firm sales.

2.1 | Definition of related variety and relevant knowledge

The term "related variety" was coined by Frenken et al. (2007) to describe one kind of industrial structure at the regional level. Related variety can be defined as "regional industry structures having complementary and overlapping knowledge base" (Aarstad et al., 2019). Such a definition is somewhat abstract, so we provide more details about related variety below.

There is a long scholarly tradition of studying how regional industrial structure influences regional economy (Drucker & Feser, 2012; Thabet, 2016). Since Alfred Marshall (1890), a lot of scholars have embraced so-called Marshall–Arrow–Romer (Arrow, 1962; Romer, 1986) externality, indicating that it is industrial specialization that pushes regional economies up (Duranton & Puga, 2000; Glaeser et al., 1992). More specifically, industrial specialization benefits economic growth from four aspects: professional labor pool, intermediate system, effective transactions, and local tacit knowledge (Lu et al., 2018). However, from the viewpoint of urban planners, industrial specialization is not a panacea (Jacobs, 1969). Regions with high industrial specialization lack the flexibility to respond quickly to exogenous changes and shocks, which leads to negative effects on regional growth (Combes, 2000; Paci & Usai, 2006). According to Jacobs externality, the essence of regionally economic growth reflects a cooperation among all local industries. Therefore, regions with industrial variety should have more opportunities to achieve growth. Considering that empirical studies on the impact of industrial variety on regional economy have not yielded consistent conclusions, Frenken et al. (2007) extended Jacobs' viewpoint by classifying industrial variety into two types: related variety and unrelated variety. Compared with unrelated variety, related variety is more representative of Jacobs' externalities associated with urbanization economies (Cainelli et al., 2016; Frenken et al., 2007). For instance, Beaudry and Schiffauerova (2009) and Cainelli et al. (2016), among others, have shown that related variety contributes positively to regional growth, rather than industrial variety in its general meanings.

Related variety is mainly built on an industrial classification system. Although Frenken et al. (2007) argued that their related variety perspective is built on knowledge base theory, it is difficult to judge whether two industries share the same or a similar knowledge base. For example, it is possible that two industries share the same or a similar knowledge base in terms of business model, but do not have such a base in terms of technology (Cainelli & Ganau, 2019). Accordingly, Frenken et al., (2007) based their related variety perspective on industrial classification system, regardless of whether it is the North American Industrial Classification System (NAICS), Standard Industrial Classification (SIC), the NACE system, or another system. For example, Content et al. (2019) assumed that four-digit sectors (other studies have used five-digit sectors (e.g., Aarstad et al., 2016) or three-digit sectors (da Silva et al., 2020, but this is not overly important) belonging to the same two-digit industry share the same or similar knowledge base; therefore, such four-digit sectors have high related variety. Content et al.'s method fits people's intuition. For example, in the textile industry (two-digit), the cotton sector (four-digit) and the silk sector (four-digit) have many possibilities to share knowledge with each other. Meanwhile, it seems less likely that the cotton sector would have many opportunities to share knowledge with, say, the aerospace manufacturing sector. The industrial classification system is not a perfect method in terms of fully uncovering related variety among industries/sectors (Neffke & Henning, 2013), but such a method is easily understood, feasible, and acceptable.

The level of related variety is determined by two factors under the same two-digit industry: the number of four-digit sectors and the size of each four-digit sector. This can be explained by the following example: considering that there is a textile industry (two-digit industry), the large number of four-digit sectors means that the textile industry encompasses a lot of four-digit sectors, such as the cotton sector, the silk sector, the wool sector, the velvet sector, etc. In this situation, each four-digit sector can be related to its peers, so the related variety is high. However, if the textile industry only contains the cotton sector, then it is impossible to relate the cotton sector to other peers. The related variety then is low. Regarding the size of each four-digit sector, if some four-digit sectors have dominated size in a two-digit industry, then related variety should be low. For easy understanding, Figure 1 visualizes what related variety is.

When applying related variety to explain firm sales, the following issues are worth highlighting. Firstly, in management theory, firm sales are influenced by internal and external factors (Robbins & Coulter, 2018). Internal factors include firm behaviours and strategies such as marketing strategy (Martin et al., 2017). External factors are those that influence firm sales but are out of firm control, such as macro-economic environment and political institution (Robbins & Coulter, 2018). In sum, external factors cannot be ignored. Second, related variety is an indicator of regionally industrial structure. The size of related variety shows the degree of diversification among the four-digit sectors in a region. In microeconomics, the most straightforward way to study firm sales from the industry perspective is to consider two forces. One force is resource competition between regional industries and inter-industry spillovers. In general, the following logic appears "The related variety is low and therefore the region is monopolized by a

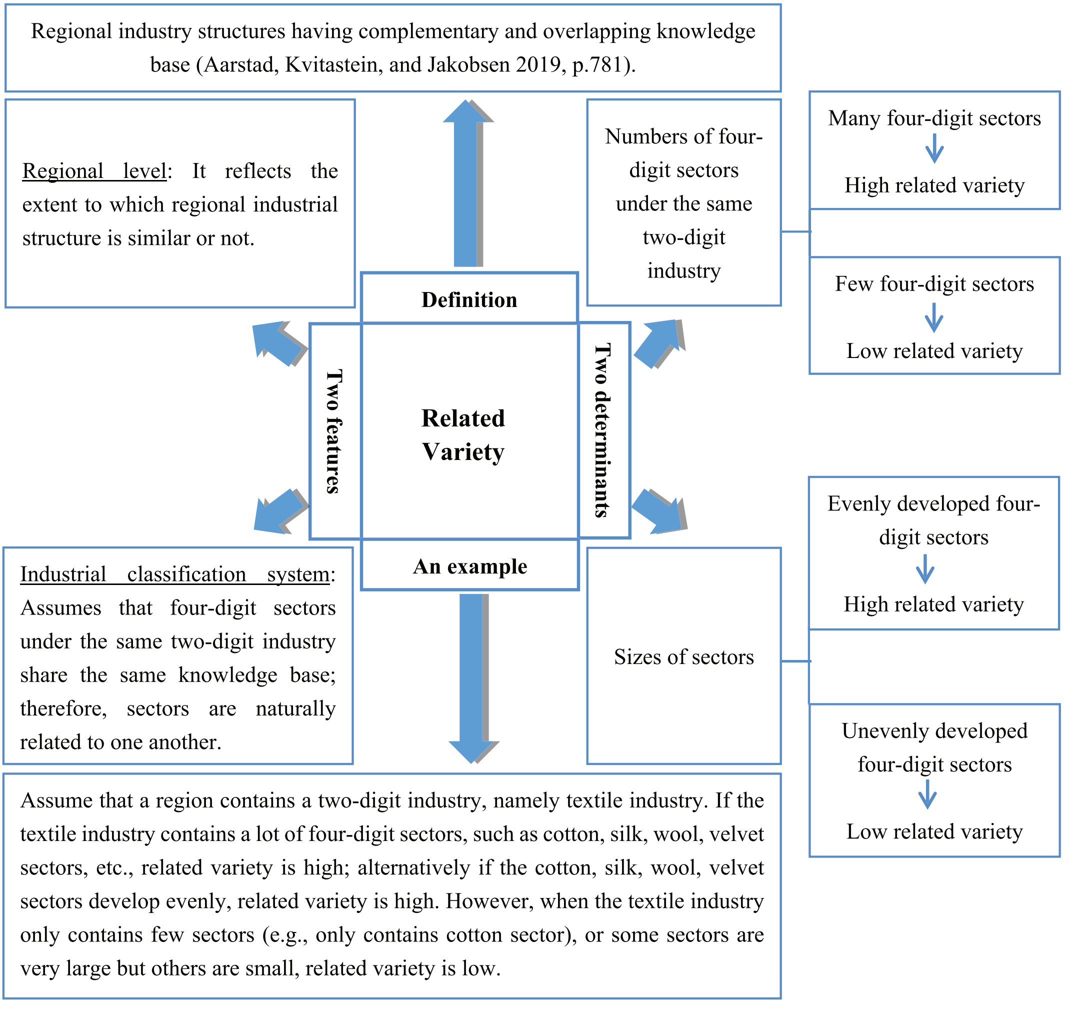

FIGURE 1 concept of related variety

few sectors, and thus the competition for resources seen in these sectors is modest." The other force is externalities: when the related variety is high, the diversity of industries is high, which means that the spillover is strong, and firms have a high probability of getting knowledge spillover from other sectors. Because of the importance of resource competition and externalities, our hypothesis development revolves around these two factors.

2.2 | Hypothesis development

As there is no empirical evidence on the relationship between related variety and firm sales, we develop a framework showing that the relationship of "related variety and firm sales" appears as a U-shape. Inspired by suggestions of Haans et al. (2016), we deconstruct the U-shape into two linear forces: resource competition and externalities. When related variety is low, resource competition is the main power that determines the relationship. When related variety is high, externalities would replace resource competition and dominate the relationship. Figure 2 visualizes how such two forces form the U-shape.

The first force is resource competition. In the resource-based view (e.g., Barney, 2001; Hoskisson et al., 2018), if a firm intends to increase its sales, its resources must have all four characteristics—valuableness, rareness, inimitability, and non-substitutability—at the same time. Drawing on the resource-based view, we illustrate the force of resource competition from two aspects: why firm sales would be high when related variety is low; and why increasing related variety from low to medium would cause firm sales to decrease.

We simply recall the example given in Figure 1: low related variety means that existing four-digit sectors (hereafter, sectors) have few or no peers under the same two-digit industry (hereafter, industry). With few peers, existing sectors naturally have monopoly power for controlling valuable and rare resources in the same industry. In addition, since only few incumbent sectors exist or the sizes of incumbent sectors develop unevenly, the reality is that either no peer sectors exist for imitating (let alone providing substitutable) resources owned by incumbent sectors, or small sectors have no capability to imitate and substitute big sectors' resources. In brief, if sectors' sizes are not evenly developed, small sectors would embed their business into that of large sectors (Wang et al., 2018).

When related variety increases from low to medium level, firm sales would decrease. Regarding valuableness, new sectors begin to be involved in competition for valuable resources with incumbent sectors. Accordingly, both incumbent sectors and new sectors are in a difficult situation. On one hand, the cost of obtaining valuable resources increases as there are more "players in the game." On the other hand, when firms pay more to sustain valuable resources, firm sales would decrease because firms have to decrease their expenditure on issues for benefiting sales, such as conducting innovations (e.g., Klement & Strambach, 2019) and spending on advertisements. Therefore, firm sales will go down. Regarding rareness, as new sectors flock in, incumbent sectors will gradually lose the monopoly power to control

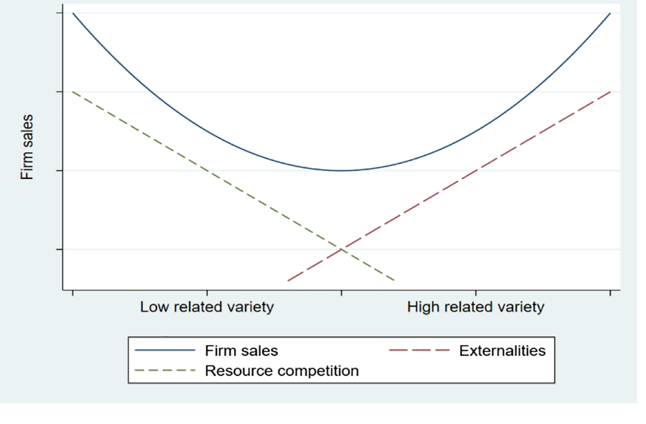

FIGURE 2 the dynamic change of firm sales

rare resources. The worst thing is that some resources may change from being non-rare to rare. For example, the number of all regional salespeople, which was previously sufficient for the cotton sector, may become insufficient when the silk sector enlarges and needs to hire salespeople. Because cultivating professional salespeople requires a long period of time (Tung & Worm, 2001), neither the cotton or the silk sectors get enough salespeople, and both may suffer from decreases in sales. Regarding inimitability and non-substitutability, increasing related variety implies that sectors begin to compete with one another. It is difficult for companies to keep their unique resources.

In conclusion, resource competition is a force that heavily influences the relationship between related variety and firm sales. The first force actually reveals a monopoly power and the process of diminishing such power: when related variety is low, firms in the incumbent sector "monopolize" regionally valuable and rare resources and face little threat, as their resources would not be imitated and substituted by peer sectors. Such monopoly power leads to firms having the capability to make high sales. When related variety increases, more firms in peer sectors are involved in resource competition, which pushes up the cost of valuable and rare resources and decreases the possibility that firms will use imitable and non-substitutable resources to increase sales.

The second force is externalities. When the level of related variety is in the process of becoming high, this indicates that the local related variety of industrial structure has developed to a mature stage, which will make externalities that were previously insignificant significant (Frenken et al., 2007). At this time, firms in the region will know how to make full use of externalities through "learning-by-doing" (e.g., Desdoigts & Jaramillo, 2020). Thus, the force dominating the relationship between firms changes from resource competition to externalities. Drawing on Jacobs's classic viewpoint, we explain the second force from two aspects: diversification and urbanization (Beaudry & Schiffauerova, 2009).

Diversification is a pillar of Jacobs's viewpoint, which notes that one of the most important approaches for firms to expand sales is innovation (Klement & Strambach, 2019; Porter, 1998). In Jacobs's words, "it is adding new works to existing ones": when many firms gather in a certain region, spatial proximity offers the possibility of knowledge spillovers. When related variety increases from medium to high, this means that incumbent sectors now have more opportunities to exchange knowledge with other sectors that have the same or similar knowledge base because knowledge exchange occurs primarily when the cognitive proximity between sectors is neither distant nor close (Frenken et al., 2007). In other words, each sector has the opportunity to absorb the knowledge it can utilize. The free flow of knowledge among sectors allows firms to fully exploit the knowledge spillovers to generate innovations, increase productivity, and achieve high sales.

Urbanization is the other pillar of Jacobs's viewpoint. Urbanization describes how certain areas develop from undeveloped to developed, and the process of urbanization goes hand in hand with the process of increasing related variety (Jacobs, 1969). The regional labour pool is an example. When related variety is high, this means that local urbanization is high, and a lot of talent would be clustered there. Because the relevance of different sectors is close, skilled employees can move freely among sectors. Cross-fertilization of knowledge and skills caused by job-hopping further reinforces the knowledge spillovers among sectors (Boschma et al., 2008), which will help firms increase efficiency and achieve high sales. Another example of urbanization is the supply chain system. Similar to labour pools, increasing related variety means that firms in similar sectors can share the same suppliers, which not only decreases costs, but also creates opportunities to sell more products because of a highly efficient supply chain system (Goe, 1991; Scott, 1986).

In brief, externalities are a force that helps explain why related variety and firm sales are positively related: when related variety develops from a medium level to a mature level, the positive effect of diversification brought by knowledge spillovers on firm sales becomes apparent. At the same time, urbanization causes more employee flow among sectors and the business supply system can realize economies of scale among sectors. All of these benefit firm sales.

Combining Force 1 (resource competition) and Force 2 (externalities), which play a dominant role at low and high related variety, respectively, we propose the following hypothesis:

Hypothesis: The relationship between related variety and firm sales is U-shaped. More specifically, firm sales are negatively influenced by related variety when it is below a certain tipping point and positively influenced by related variety when it is above that tipping point.

3 | METHODOLOGY

3.1 | Research context and data sources

In order to test our hypothesis, we chose Japan as our research objective. Japan has three advantages over other countries. First and foremost, as a country with a developed economy, Japan has established a mature industrial structure, thus quantifying related variety is feasible. Meanwhile, with Japan's rising economy after the Second World War, Japan's feature of related variety became distinct. For example, Aichi-ken has built high automobile industrial related variety given that Toyota's headquarters are there. Tokushima-ken, by contrast, is a less developed region with low related variety. Second, despite having the third largest economy in the world, Japan does not have special political authority to influence the world economy like the United States and China do. Last, but not least, we also found that Japanese economic statistics work is meticulous, which provides high-quality samples for our research.

We gathered data on all 47 of Japan's prefectures (prefecture refers to "ken" in Japanese language), which comprises 43 prefectures plus Hokkaido, Tokyo, Kyoto, and Osaka. All the data come from three sources. First, firm-level data come from the Orbis Global Enterprise Database, 2010–2014. Orbis is developed by the Bureau van Dijk company, a Moody's analytics company that captures and treats information about 300 million companies (Rocca et al., 2019). The second source is the Japan Meteorological Agency, which contains climatic data for every prefecture (http://www.jma.go.jp/jma/index.html). The Japan Meteorological Agency contributed the relevant data of the average annual precipitation of each prefecture as instrumental variable. The third source was e-Stat (https://www.e-stat.go.jp/en/), a portal site for Japanese Government Statistics, which provided regional demographics and economic data. The present study obtained an unbalanced panel dataset with 670,108 observations covering 331,515 firms for the period of 2010 to 2014. During this period, some firms may have gone bankrupt and disappeared, while new ventures came into the market. Therefore, the data are unbalanced; not every Japanese firm will remain in our observation for the entire period. No industry was excluded from the sample because related variety is an indicator showing the diversification of regional industrial structure. Any local industry is a part of constituting local industrial diversification. Based on the North American Industrial Classification System (NAICS), the calculation of annual related variety for 47 prefectures ultimately involves 274 four-digit sectors belonging to 24 two-digit industries.

3.2 | Variables

3.2.1 | Dependent variables

In line with Reuber and Fischer (1997) and Aspelund et al. (2005), firm sales – a critical indicator that measures performance of firms (Brannon et al., 2013) – is set as our dependent variable. When a firm achieves high sales, this usually implies that such a firm provides products/services that fit consumers' demands well (Delmar & Shane, 2006); high sales may also imply strong competitiveness (Delmar & Shane, 2006).

3.2.2 | Independent variables

Drawing on NAICS, we defined related variety as the weighted sum of the entropy across the number of employees at the four-digit sector within each two-digit industry (Frenken et al., 2007). Each four-digit sector *i* belong to only a two-digit industry S_g, where $g = 1, \ldots, G$. Then we sum the four-digit shares of employees p_i and obtain the two-digit shares:

$$P_g = \sum_{i \in S_g} p_i. \tag{1}$$

The entropy of related variety within each prefecture for a certain year is:

$$RV = \sum_{g=1}^{G} P_g H_g, \tag{2}$$

where:

$$H_g = \sum_{i \in S_g} \frac{p_i}{P_g} \ln\left(\frac{1}{p_i/P_g}\right). \tag{3}$$

As mentioned before, compared with the linear effects of related variety on the firm sales, we have paid more attention to the nonlinear effects of related variety on firm sales. Therefore, the quadratic terms of related variety are also important independent variables in this paper.

3.2.3 | Control variables

We control the following variables. At the firm level, we control firms' age and total assets. Coad et al. (2018) identified that, as an important corporate characteristic, age is conducive to high performance due to learning by doing. It is well documented that total assets, an important indicator of firm size, affect sales positively (Patin et al., 2020). At the regional level, we control regional population, the growth of income *per capita*, and the growth of gross prefectural product (Batabyal & Beladi, 2019; Mendoza-Abarca et al., 2015; Yang, 2000).

3.2.4 | Instrumental variables

We aim to discover how related variety influences firm sales. Although reverse causality may not be serious since related variety is at the regional level and sales are at the firm level, and although we have set several controls (as mentioned above), we still face the endogenous problem caused mainly by omitted variable bias (Gujarati & Porter, 2009). We have attempted to relieve the endogeneity by introducing an instrumental variable, which is each prefecture's annual precipitation.

We employed every year's average annual precipitation of 47 prefectures in Japan from 2010 to 2014. Precipitation generally has an influence on economy; for example, industries in desert climates differ from those of rainforest climates. While Waldman et al. (2006) and Waldman et al. (2008) have given good examples of applying precipitation as IVs, more details still need to be presented.

Jacobs's (1969) urban economics noted that abundant rain is a factor that causes people to live together and to form related industries. Precipitation is a factor that constitutes and helps form a business environment that attracts a certain sector to be located in geographic proximity. After that, the incumbent sector would attract more similar sectors to be located in geographic proximity. Therefore, a region's relative variety increases.

Furthermore, we find a consensus among scholars that agricultural production is affected by natural conditions, such as precipitation. This fact does not change much even if advanced agricultural technologies appear (Calzadilla et al., 2013). Byerlee et al. (2009) also pointed out that the development of non-agricultural industries within a region is highly correlated with local agricultural base. In other words, for a region with good precipitation condition, its agriculture would be strong, and its related variety might be low. Therefore, we hypothesize that annual prefectural precipitations and the level of related variety may show a U-shaped relationship. Prefectures with high/low precipitation are unfavourable to develop agriculture, which compels local economies to concentrate on industry, which then leads to high related variety, and vice versa.

In practice, regarding Japan's reality, two issues need to be highlighted. First, Japan's precipitation is unevenly distributed. Figure 3 shows five representative prefectures, ranked from the lowest precipitation to the highest precipitation. For example, Mie-ken received 3.315 litres of rainfall in 2011, but Miyagi-ken only received 1.108 litres in the same year. Secondly, taking the large population into account, Japan's precipitation *per capita* was not high. One of Japan's geological features is that most of its rivers are short and fast-flowing, and the amount of water provided by rivers is variable. Because of the small territory size and the difficulty of building large dams, precipitation constitutes one of the most important parts of Japan's water sources. Therefore, precipitation is still important for Japan's related variety. In sum, regardless of whether it is from theory or practice, we have reason to argue that Japan's regional precipitation can influence related variety.

Since a firm's sales mainly depend on its business strategies (e.g., Huang et al., 2020; Lu & Ma, 2011) and resources (e.g., Brinckmann et al., 2019), average precipitation should not be directly related to firm sales. Here we provide an extreme case. Japanese people frequently purchase umbrellas because precipitation often arrives unexpectedly. However, this does not mean that umbrella manufacturing firms can easily increase sales. One reason is that fierce competition pushes the price per unit to a low level. Thus, if an umbrella firm aims to increase sales, it would seem to be better to generate new styles, new business model, or high-quality products rather than expecting unpredictable and uncontrollable precipitation.

Table 1 provides definitions of all variables. Descriptive statistics and the correlation matrix are shown in Tables 2 and 3, respectively. We need to highlight one thing: although most firms in our observations were normal firms, with around US$1.5 million annual sales, some Japanese firms such as Sony, Panasonic, Hitachi, and Nissan made a hundred billion sales. In 2012, Toyota Motors even achieved 234.351 billion sales. Since such giants did contribute much to the regional industrial structure, we believe that it does not make sense if we deliberately "pretend" such giants only made small sales. Therefore, we do not simply Winsorize the data, which may lead to unusual estimators of interest. However, in the penultimate robustness test (see column (3) of Table 7), we prove that our results are valid even if we do not consider business giants and small firms.

3.3 | Regressions: Panel model and 2SLS model

To assess how related variety influences firm sales, we ran regressions in three steps. Firstly, we applied panel model as our baseline regression. Second, we applied the two-stage least square (2SLS) method with IV to deal with

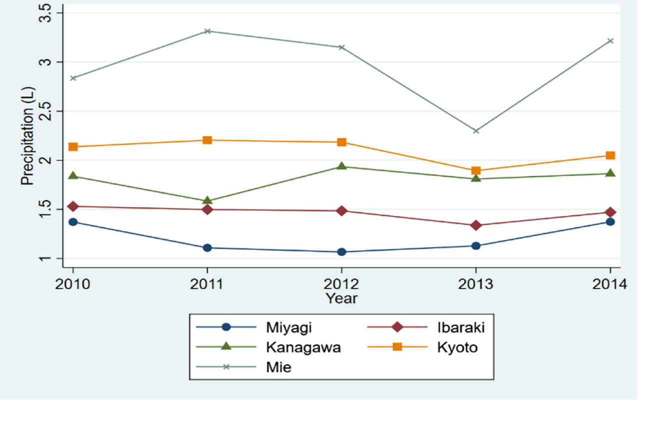

FIGURE 3 average annual precipitations of five representative prefectures

TABLE 1 Definition of variables

Variables	Definition	Notes
Dependent variables		
Sales	Firm's annual sales	Unit: million US$
Independent variables		
RV	Related variety calculated by entropy	See Equation 2
RV2	The square of related variety	
Control variables		
lnAge	Natural logarithm of (1 + [year-year of corperation+1])	At firm level
lnAsset	Natural logarithm of (1 + total assets)	At firm level
lnPop	Natural logarithm of population	At regional level
Income_growth	Growth rate of income per person	At regional level
GPP_growth	Growth rate of gross prefectural product	At regional level
Instrumental variables		
Precip	Prefectural average annual precipitation	Unit: liter

TABLE 2 Descriptive statistics

Variable	N	Min	P25	P50	P75	Max
Sales	670,108	0.000	0.621	1.573	6.572	234351.484
RV	670,108	0.783	1.385	1.671	2.000	2.284
RV2	670,108	0.613	1.918	2.792	3.999	5.217
Precip	670,108	0.998	1.485	1.717	1.948	3.315
Precip2	670,108	0.997	2.207	2.948	3.794	10.991
lnAge	670,108	0.693	2.708	3.258	3.689	5.844
lnAsset	670,108	0.000	5.790	6.966	8.486	21.703
lnPop	670,108	13.266	14.502	15.504	15.826	16.411
Income_growth	670,108	−7.900	−0.500	1.600	3.500	9.300
GPP_growth	670,108	−9.900	−0.600	0.500	1.600	10.000

endogenous problems. Third, we discussed the validity of IV in robustness tests in subsection 5.1. We provide more details below.

We employed the panel model with fixed effect, including the time fixed effect, industrial fixed effect, as well as regional fixed effect. The time fixed effect model is used to eliminate the common impact of different years on all firms. The industrial and regional fixed effect model can control impacts that change with the heterogeneities of industry and region, which alleviates endogeneity caused by omitted variables (Bell et al., 2019).

Our fixed effect model is set as follows:

$$Sales_{it} = \alpha_{1it} + \beta_{1it} RV_{it} + \beta_{2it} RV2_{it} + \gamma_{1it} Z_{it} + \mu_t + \theta_j + \nu_r + \varepsilon_{1it}, \tag{4}$$

where RV_{it} refers to the level of related variety in year t of region r, where firm i is located; $RV2_{it}$ is the square of RV_{it}; α refers to constant term. Z_{it} refers to the matrix of control variables shown in Table 1; μ_t is the time fixed effect; θ_j and ν_r are the industrial fixed effect and the regional fixed effect, respectively. ε is the error term.

TABLE 3 Correlation matrix

	Sales	RV	RV2	Precip	lnage	lnAsset	lnPop	Income_growth
Sales	1.000							
RV	0.040	1.000						
RV2	0.044	0.994	1.000					
Precip	0.007	−0.066	−0.043	1.000				
lnAge	0.039	−0.049	−0.050	0.009	1.000			
lnAsset	0.170	0.082	0.098	0.071	0.439	1.000		
lnPop	0.040	0.702	0.714	−0.199	−0.057	0.118	1.000	
Income	0.002	−0.238	−0.215	0.003	−0.016	−0.005	−0.001	1.000
GPP_growth	−0.004	−0.001	−0.008	−0.084	−0.043	−0.095	0.007	0.718

Although Equation 4 has controlled the fixed effects, potential endogeneity caused by omitted variables may still exist. In order to obtain regression results that are unbiased, efficient, and consistent, we introduce precipitation as related variety's IV in 2SLS regression (Feng et al., 2019). Furthermore, we also introduce quadratic precipitation as quadratic related variety's IV, which is necessary when running 2SLS regression for nonlinear relationship (for more mathematic details, please read Bun & Harrison, 2019; Ebbes et al., 2016). Only when the results of the first-stage regressions are significant can we run the second stage of regression since we need to obtain fitting values of endogenous variables from first-stage regressions to run second-stage regression (Gujarati & Porter, 2009).

The first-stage models are set as follows:

$$RV_{it} = \alpha_{2it} + \pi_{1it} Precip_{it} + \rho_{1it} Precip2_{it} + \gamma_{2it} Z_{it} + \mu_t + \theta_j + \upsilon_r + \varepsilon_{2it}, \quad (5)$$

$$RV2_{it} = \alpha_{3it} + \pi_{2it} Precip_{it} + \rho_{2it} Precip2_{it} + \gamma_{3it} Z_{it} + \mu_t + \theta_j + \upsilon_r + \varepsilon_{3it}, \quad (6)$$

where $Precip_{it}$ refers to the precipitation in year t of region r, where firm i is located in; $Precip2_{it}$ is the square of $Precip_{it}$; other variables have the same meaning as in Equation 4.

The second-stage model of 2SLS regression is the same with the panel model, as shown in Equation 4.

4 | REGRESSION RESULTS

4.1 | Panel model

We first estimate Equation 4 using panel model with fixed effects to test our hypothesis, and column (1) in Table 4 shows the result of regression. The quadratic coefficient of related variety is positive and significant $(\hat{\beta}_2 = 3.828;\ p < 0.1)$, which initially confirms the U-shaped effect of related variety on firm sales. Although the primary coefficient of related variety is not significant, it does not have an impact on whether the U-shaped relationship holds. The primary coefficient determines where the U-shaped relationship should be positioned; we discuss this in detail in subsection 5.2.1.

All the coefficients of control variables are significant statistically, except for age, which shows a negative relationship with sales here, but eventually becomes significantly positive in 2SLS. The signs of both total assets and population are positive, which meets our expectation. However, the negative signs of population and the growth of income per person maybe is a result of the aging Japanese population. Although social wealth enlarges, Japanese consumption motivation does not increase accordingly (Goh et al., 2020). Similarly, due to the low consumption

TABLE 4 Panel and 2SLS regression results

Variables	(1) Panel Sales	(2) 2SLS First stage RV	(3) 2SLS First stage RV2	(4) 2SLS Second stage Sales
RV	−5.927(8.082)			−305.006*(180.685)
RV2	3.828*(2.312)			112.484*(64.560)
Precip		−0.222***(0.055)	−0.779***(0.187)	
Precip2		0.068***(0.012)	0.223***(0.041)	
lnAge	−0.636(2.000)	0.001**(0.000)	0.005***(0.002)	3.470*(1.943)
lnAsset	27.988***(3.063)	0.002***(0.000)	0.009***(0.001)	23.054***(2.551)
lnPop	−230.132***(55.721)	−4.589***(1.054)	−10.540***(3.573)	−609.808**(254.238)
Income_growth	−0.913***(0.296)	−0.019***(0.002)	−0.043***(0.008)	−2.033**(0.935)
GPP_growth	0.975***(0.291)	0.015***(0.002)	0.033***(0.008)	2.043**(0.884)
Constant	3,022.204***(764.745)	64.953***(14.628)	148.001***(49.592)	8,526.983**(3,636.427)
Time fixed effect	Yes	Yes	Yes	Yes
Regional fixed effect	Yes	Yes	Yes	Yes
Industrial fixed effect	Yes	Yes	Yes	Yes
N	670,108	670,108	670,108	670,108

Notes: heteroscedasticity-adjusted standard errors reported in parentheses are clustered at region-industry level; *, **, and ***represent statistical significance at the 10%, 5% and 1% level, respectively; *N* refers to observations.

motivation arisen from the increased population dominated by the elderly people, the total population is also negatively related to firm sales.

4.2 | 2SLS model

For the first stage of the 2SLS model, we estimate Equations 5 and 6, and the regression results are shown in columns (2) and (3) of Table 4. As predicted in subsection 3.2.4, there are statistically significant correlations ($p < 0.01$) between IVs and independent variables, whether between precipitation and related variety or between the quadratic terms of them, which means that precipitation is not a weak instrumental variable in the present paper. For the second stage of the 2SLS model, we estimate Equation 4 again using the fitting values of related variety and its quadratic term. The regression result of 2SLS, presented in column (4) of Table 4, is consistent with that of panel regression. After adding IV into the estimations, the coefficient $\hat{\beta}_2 = 112.484$ $(p < 0.1)$ indicates that there is a significant U-shaped relationship between related variety and firm sales. Thus, our hypothesis is verified again. Such a conclusion has common ground with previous studies. Lucio et al. (2002) identified a U-shaped relationship when investigating the impact of industrial structure on productivity growth in Spanish industry, and they also used the instrumental variable method to relieve endogeneity. More recently, when considering the impact of related variety on firm performance, Cainelli and Ganau (2019) revealed that related variety is a positive predictor of firms' employment growth. The discrepancy between our nonlinear conclusion and their linear conclusion may be due to the fact that we use sales to measure firm performance.

5 | ROBUSTNESS TESTS

Our robustness tests are divided into two parts. Our first part particularly discusses the validity of our 2SLS method by applying the plausibly exogenous IV method (Kippersluis & Rietveld, 2018). This part is important since we are going to show that even if our IV is not perfect, our 2SLS results are still convincing. Our second part conducts several normal robustness tests such as U-shaped test, adding more control variables and changing sample, etc.

5.1 | Plausibly exogenous IV method

Although employing precipitation as IV seems sufficiently exogenous for related variety, we still face the question of whether "precipitation is a good IV." We address this challenge from a mathematical viewpoint. Referring to the practices of Conley et al. (2012), we set the following equation:

$$Sales_{it} = \alpha_{4it} + \beta_{1it} RV_{it} + \beta_{2it} RV2_{it} + \varphi_{1it} Precip_{it} + \varphi_{2it} Precip2_{it} + \gamma_{4it} Z_{it} + \mu_t + \theta_j + \upsilon_r + \varepsilon_{4it}, \tag{7}$$

where φ_1 and φ_2 refer to the direct effect of IV (precipitation) on the independent variable (firm sales). Following Angrist and Pischke's (2009) suggestion and Chen et al. (2018) practice, a good instrumental variable, in our paper, should satisfy the following three conditions:

1. There is a significant correlation between precipitation and related variety. Mathematically, π_1 and ρ_1 in Equation 5, π_2 and ρ_2 in Equation 6 are statistically not equal to 0.
2. Precipitation should be uncorrelated with error term.
3. The precipitation should not directly influence firm sales. Mathematically, this means that the φ_1 and φ_2 in Equation 7 are statistically insignificant.

The first condition is referred to as correlation, and the latter two are collectively referred to as exogeneity (Clarke & Matta, 2018). It is not difficult to prove that condition (I) regarding correlation is satisfied: precipitation is related to related variety, which is supported by the first stages of 2SLS (see columns (2) and (3) in Table 4). By contrast, conditions (II) and (III) regarding exogeneity are hard to verify. Fortunately, IV is exogenous as long as either condition (II) or condition (III) is satisfied (please see Clarke & Matta, 2018, for mathematic details). In our paper, testing condition (III) is feasible. According to Conley et al.'s (2012) plausibly exogenous IV method that allows for the existence of IV's endogeneity via taking the initiative to relax condition (III), What we should do is to prove that even when the direct effect of precipitation on firm sales (φ_1 or φ_2 in Equation 7) is statistically significant, our 2SLS result is still robust.

To facilitate the estimation of whether precipitation has a direct effect on firm sales, we can obtain a reduced-form Equation 8 according to Equations 5, 6 and 7:

$$Sales_{it} = \alpha_{it} + (\beta_{1it}\pi_{1it} + \beta_{2it}\pi_{2it} + \varphi_{1it}) Precip_{it} + (\beta_{1it}\rho_{1it} + \beta_{2it}\rho_{2it} + \varphi_{2it}) Precip2_{it} + \gamma_{it} Z_{it} + \mu_t + \theta_j + \upsilon_r + \varepsilon_{it}. \tag{8}$$

The presumptions that both φ_1 and φ_2 are statistically insignificant are strong. Therefore, following suggestions by Conley et al. (2012), we relax this hypothetical premise by allowing φ_1 and φ_2 fit two normal distributions with mean $\mu_{\varphi 1}$ and variance $\Omega_{\varphi 1}$, and mean $\mu_{\varphi 2}$ and variance $\Omega_{\varphi 2}$, respectively (mathematically: $\varphi_1 \sim N(\mu_{\varphi 1}, \Omega_{\varphi 1})$, $\varphi_2 \sim N(\mu_{\varphi 2}, \Omega_{\varphi 2})$). Kippersluis and Rietveld (2018) developed a method to determine $\mu_{\varphi 1}$ and $\mu_{\varphi 2}$. They argued that we should find a subsample from the full sample, where related variety is not influenced by precipitation (that is, π_1 and ρ_1 in

Equation 5, π_2 and ρ_2 in Equation 6 are statistically insignificant at the same time). Following this guideline, after estimating Equation 8 in the subsample, we can obtain estimators ($\beta_{1it}\pi_{1it} + \beta_{2it}\pi_{2it} + \varphi_{1it}$) and ($\beta_{1it}\rho_{1it} + \beta_{2it}\rho_{2it} + \varphi_{2it}$), namely the coefficients of IVs, which in Kippersluis and Rietveld's view are good estimators to replace $\mu_{\varphi 1}$ and $\mu_{\varphi 2}$ in full sample. Inspired by Kippersluis and Rietveld, we use the variation of precipitation as a standard to find the subsample. If the average annual precipitation in a region varies widely, this means that the local climate is hard to predict. However, city development requires a relatively stable infrastructure, such as drainage systems, to be built in order to decrease the damage from climate uncertainty. To some extent, for two regions with comparative levels of development, the more unstable the climate, the better the infrastructure (this is not hard to understand in Japan, which is a natural-disaster-prone country). Therefore, a better infrastructure means that the local industrial structure is impacted less by the unstable climate. Following this logic, we find that the subgroup contains 5,172 observations, in which the variation of precipitation is greater than or equal to 0.853 liters. After calculation, $\mu_{\varphi 1}$ and $\mu_{\varphi 2}$ are 3.308 and -0.593, respectively (see Kippersluis and Rietveld, 2018 for the concrete calculation methods).

Both Kippersluis and Rietveld (2018) and Clarke and Matta (2018) initially discussed $\Omega_{\varphi 1}$ and $\Omega_{\varphi 2}$ in linear regressions. In our case, after calculation we find that $\hat{\Omega}_{\varphi 1} = 0.136$ and $\hat{\Omega}_{\varphi 2} = 0.005$ for RV and RV2, respectively. However, given that applying $\hat{\Omega}_{\varphi 1}$ and $\hat{\Omega}_{\varphi 2}$ in the linear method to the quadratic method would lead to nonsymmetric and highly singular variance matrix in regression, and considering that $\hat{\Omega}_{\varphi 1}$ is 4.1% of $\hat{\mu}_{\varphi 1}$ and $\hat{\Omega}_{\varphi 2}$ is 0.8% of $\hat{\mu}_{\varphi 2}$, which means that variances of φ_1 and φ_2 are fairly small, we simply replace $\Omega_{\varphi 1}$ and $\Omega_{\varphi 2}$ as 0. Our method does not violate Kippersluis and Rietveld's theory.

As shown in columns (1) and (2) of Table 5, the insignificant coefficients $\hat{\pi}_1$ and $\hat{\rho}_1$ in Equation 5, $\hat{\pi}_2$ and $\hat{\rho}_2$ in Equation 6 imply that there is no relationship between precipitation and related variety in the subgroup. Columns (3) and (4) show that in the remaining sample, Precipitation and Precipitation2 are always statistically significant, which is consistent with columns (2) and (3) of Table 4 in the full sample. Columns (5) and (7) show that the statistical significance of IVs is consistent in the full sample and the remaining sample, respectively. The insignificant estimators of IVs (namely ($\beta_{1it}\pi_{1it} + \beta_{2it}\pi_{2it} + \varphi_{1it}$) and ($\beta_{1it}\rho_{1it} + \beta_{2it}\rho_{2it} + \varphi_{2it}$), see column (6)) suggest no direct impact of precipitation on firm sales (that is, the $\hat{\varphi}_1$ and $\hat{\varphi}_2$ in the subsample are not significant, because all of $\hat{\pi}_1$, $\hat{\pi}_2$, $\hat{\rho}_1$, and $\hat{\rho}_2$ are also not significant at this time), which means that precipitation do not violate condition (III) and is a perfect IV in the subsample. Column (8) shows that even if condition (III) is not satisfied in the full sample, the relation between regional related variety and firm sales is still U-shaped $\left(\hat{\beta}_2 = 157.478;\ p < 0.01\right)$. Therefore, the result of 2SLS regression is robust.

5.2 | Other robustness tests

We showed in subsection 5.1 that our regression results are robust even if condition (III) cannot fully be satisfied. In this subsection 5.2, we further conduct four types of robustness tests. First, we perform a U-shaped test by analysing the location of the extreme points and the slopes on either side of the extreme points. Second, we add more firm-level control variables. Third, we conduct regression after excluding super-huge and small firms from the full sample. Fourth, we regress mean-centred related variety against firm sales.

5.2.1 | U-Shaped test

We find that the signal of the related variety squared term is always positive in our regressions. However, such evidence alone is not enough to prove the U-shape exists because it is possible that, in a monotone interval, the quadratic term still can be statistically significant (Lind & Mehlum, 2010). Following Lind and Mehlum (2010) and Haans et al. (2016), we do the following things: we first calculate the extreme point of the U-shape and then check slopes in the left and right sides of the extreme point. The results are given in Table 6. The interval of related variety is

TABLE 5 Plausibly exogenous IV for RV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	First stage	**First stage**	**First stage**	**First stage**	**Reduced form**	**Reduced form**	**Reduced form**	**Plausibly exogenous IV**
Samples	**Subsample**	**Subsample**	**Remaining sample**	**Remaining sample**	**Full sample**	**Subsample**	**Remaining sample**	**Full sample**
Variables	**RV**	**RV2**	**RV**	**RV2**	**Sales**	**Sales**	**Sales**	**Sales**
Precip	0.045 (0.041)	−0.025 (0.122)	−0.180*** (0.003)	−0.561*** (0.009)	−12.767*** (1.770)	3.308 (2.388)	−12.873*** (1.740)	
Precip2	0.001 (0.008)	0.034 (0.024)	0.059*** (0.001)	0.173*** (0.002)	2.785*** (0.346)	−0.593 (0.465)	2.820*** (0.350)	
RV								−429.856*** (80.377)
RV2								157.478*** (28.874)
N	5,172	5,172	664,936	664,936	670,108	5,172	664,936	670,108

Notes: control variables are the same as Table 4 has; *, **, and ***represent statistical significance at the 10%, 5% and 1% level, respectively; *N* refers to observations.

TABLE 6 U-test for related variety and sales

	Lower bound	Upper bound
Interval	0.783	2.284
Slope	−128.882	208.859
t-value	−1.613	1.818
P > t	0.053	0.035

Notes: the extreme point is 1.356; overall test of presence of a U shape: t-value = 1.610; P > |t| = 0.053.

TABLE 7 Other robustness tests

	(1)	(2)	(3)	(4)
Variables	Sales	Sales	Sales	Sales
RV	−265.358*** (71.714)	−589.308*** (154.689)	−3.505*** (1.213)	
RV2	100.244*** (25.303)	226.531*** (57.382)	1.478*** (0.465)	
c_RV				71.309* (36.827)
c_RV2				112.547* (64.603)
lnCliability	0.942*** (0.192)	0.120 (0.424)		
ROE		0.000 (0.002)		
N	669,445	501,398	335,163	670,108

Notes: control variables are the same as Table 4 has; *, **, and ***represent statistical significance at the 10%, 5% and 1% level, respectively; *N* refers to observations.

[0.783, 2.284] with the extreme point 1.356. In the left interval [0.783, 1.356], the slope is −128.882 and statistically significant ($p < 0.1$); in the right interval [1.356, 2.284], the slope is 208.859 with statistical significance ($p < 0.05$). This test reconfirms the existence of the U-shape relationship. Figure 4 visualizes such a U-shape.

5.2.2 | Adding more control variables

Because of the data quality, we do not have many variables at the firm level. In this case, we have still added two more controls in our 2SLS regression. The purpose of this robustness test is to show that, even in the case of a

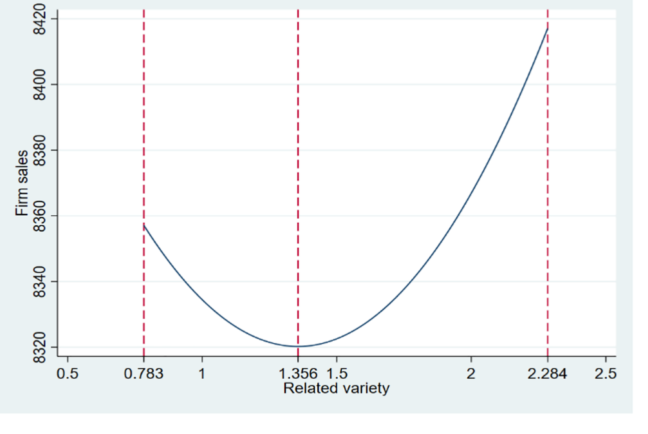

FIGURE 4 the U-shaped relationship between sales and related variety

decrease in observations, our results remain robust. We first control firms' the current liabilities and take its natural logarithm (*lnCliability*), because it approximately reflects the short-term solvency that is closely related to firm sales (Catao & Milesiferretti, 2014). As shown in column (1) of Table 7, the U-shaped relationship between related variety and firm sales remains robust after controlling for current liabilities $\left(\hat{\beta}_2 = 100.224;\ p < 0.01\right)$. Then, based on current liabilities, we add firms' return on equity (*ROE*) as another control variable. The higher the index, the stronger the profitability of the firm (Lemmon & Zender, 2010). The regression result remains in line with that of 2SLS in subsection 4.2, and the U-shape is even steeper with the addition of these two firm-level control variables ($\hat{\beta}_2 = 226.531$ and $p < 0.01$, see column (2) of Table 7).

5.2.3 | Excluding super-huge and small firms from full sample

One may question why all coefficients estimated by the panel model and the 2SLS model are so different in terms of magnitude. This question can be partly explained by the different regression methods. Larcker and Rusticus (2010) argued that 2SLS regression is generally better than panel regression due to more unbiased estimation. And, according to Jiang (2017), it is a normal phenomenon that applying 2SLS regression would get a larger coefficient than that of applying panel regression. In addition to the common explanation, perhaps the present study can provide more answers to this question: as mentioned in our descriptive statistics (Table 2, subsection 3.2.4), Japan has business giants such as Toyota and Honda, who pulled up the heterogeneous treatment effect across our full sample in the 2SLS regression (Jiang, 2017). In this robustness check, we exclude super-huge and small firms from the full sample by dropping observations in which firm sales are above 75% or below 25% of the total, and conduct 2SLS regression again. Our aims are to examine whether the regional industrial structure will have a different impact on the sales of local medium-sized firms and whether the high values of coefficients result from those business giants. The influence of related variety on firm sales is negative first and then positive; that is, there is still a U-shaped relationship between them (see column (3) of Table 7). Furthermore, the magnitudes of all coefficients drop sharply.

5.2.4 | 2SLS regression after mean centring

Although multicollinearity issue may not be important for our paper, we still provide extra demonstration by mean centring and comparing variance inflation factors (VIF). In baseline 2SLS regression, its VIF is up to 29.52 and there is high correlation between variables indeed, especially related variety and quadratic related variety. Here we adopt a mean centring approach to reduce multicollinearity (Paccagnella, 2006). We subtract *RV* from its mean value to obtain *c_RV*, based on which we obtain the squared term *c_RV2*. After 2SLS regression against centring related variety and its quadratic term, all coefficients are stable and, in particular, the quadratic coefficient of interest remains almost unchanged (see column (4) in Table 7, compared with column (4) in Table 4), but now the VIF decreases to 2.02. That means multicollinearity does not change our conclusion.

To sum up: all regression results show that our hypothesis receives supports; that is, the relationship between related variety and firm sales takes a U-shape.

6 | DISCUSSION AND CONCLUDING REMARKS

In this paper we have studied how related variety as a kind of regional industrial structure influences firm sales. Drawing on firm data in all 47 Japanese regions, we apply 2SLS with a relatively new method that can help discuss the validity of our instrumental variable (Clarke & Matta, 2018; Conley et al., 2012; Kippersluis & Rietveld, 2018). Our study shows that related variety has a U-shaped relation with firm sales. Such a conclusion is stable and robust.

Related variety studies have a tradition of concentrating on regional-level issues and, in most cases, related variety appears to have a linear and positive influence (such as) on regional employment (e.g., Firgo & Mayerhofer, 2018) and on regional innovation (e.g., Miguelez & Moreno, 2018). Ejdemo and Örtqvist (2020) took a step further. In their recent study on Swedish regions, the authors showed that related variety and regional innovation appeared to have an inverted U-shape relationship. Building on what Ejdemo and Örtqvist (2020) have done, our results showed that if the research focus changes from the regional level to the firm level and concentrates on firm business performance (in this paper, annual sales) rather than innovations, related variety and firm sales could be U-shaped relationship. Although our results differed from Ejdemo and Örtqvist's results, our findings do not violate business theory (Chanaet al., 2020; Younkin & Kashkooli, 2020). For example, Younkin and Kashkooli's (2020) study of the American music industry found that the most successful artists were those that either provided the most familiar music or the most distinctive music. Artists who were somewhere between familiarity and distinctiveness found it difficult to achieve high sales. By the same token, we argue that when a region's related variety is sufficiently low or high, the region will have a specific feature of regional industrial structure. To sum up, our findings supplement Ejdemo and Örtqvist's (2020) argument by replacing sales with innovation as the dependent variable. Our findings also reflect that firm performance differs from regional development.

Our results have two practical implications. One is that entrepreneurs should have different goals and behaviours from regional policy-makers. Policy-makers are responsible for the entire local economy; therefore, they have to balance regional industrial structure (Ejdemo & Örtqvist, 2020). A region with either very high related variety or very low related variety, however, benefits firms to make business strategies, because entrepreneurs can easily know what advantages they can take from the regional industrial structure. The second practical implication is that innovations and sales are different. Local externalities, especially knowledge spillover, benefits for generating innovations, but such benefit will decrease as related variety increases. However, sales are dominated by resource competition force and externality force.

The present study has certain limitations, three of which we list here. First, due to a lack of data, we only test how related variety influences firm sales. In the future, scholars may use other data to test how related variety influences other types of firm performance. The second limitation concerns a methodological issue. Our paper applies entropy to quantify related variety. Entropy is a popular indicator that has appeared in many studies on related variety (e.g., Content et al., 2019; da Silva et al., 2020; Fritsch & Kublina, 2018). However, Neffke and Henning (2013) noted that entropy is not a perfect indicator because it does not necessarily have relationships to those four-digit sectors in the same two-digit industry. Accordingly, Neffke and Henning suggested that we should observe that how people flow among different industries, and let such flow reflect industrial relatedness. We expect that, in the future, our results can be tested using Neffke and Henning's method. Lastly, as we employed statistics to reveal how related variety influences firm sales, such a method is static rather than dynamic. More specifically, our paper cannot reflect the concrete mechanism of how related variety causes firm sales change.

Given that most related variety studies at the firm level departed from innovation theory (e.g., Aarstad, Kvitastein, & Jakobsen, 2016; Ejdemo & Örtqvist, 2020), in the future scholars may combine related variety perspective with business theory. Considering that related variety per se is an indicator that reflects a type of regional industrial structure, related variety can be used to understand firm's business behaviours and performances. We have offered a few remarks about the U-shaped relation between the related variety and firm sales, hoping other business scholars and economic geographers will come up with more insights.

ACKNOWLEDGEMENTS

This paper was supported by Social Science Foundation of Guangdong Province, China (Grant No: GD19CGL05). We thank two professors Damian Clarke (Universidad de Santiago de Chile) and Hans van Kippersluis (Erasmus School of Economics) for providing details of applying the plausible instrumental variable method in the U-shape study. We owe a debt of gratitude to the editor of the *Papers in Regional Science*, Professor John Winters, and to our two anonymous reviewers for their suggestive and helpful comments.

ORCID

Ren Lu https://orcid.org/0000-0002-7923-6766

REFERENCES

Aarstad, J., & Kvitastein, O. A. (2019). Entrepreneurial path dependency in labour market regions: a longitudinal panel study of related and unrelated variety. *Small Enterprise Research*, *26*(3), 253–267. https://doi.org/10.1080/13215906.2019.1644192

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*(4), 844–856. https://doi.org/10.1016/j.respol.2016.01.013

Aarstad, J., Kvitastein, O. A., & Jakobsen, S.-E. (2019). Location decisions of enterprise R&D investments as a function of related and unrelated regional industry structures: A multilevel study. *Papers in Regional Science*, *98*(2), 779–797. https://doi.org/10.1111/pirs.12366

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. https://doi.org/10.1515/9781400829828

Arrow, K. J. (1962). The economic implications of learning by doing. *Review of Economic Studies*, *29*(3), 155–173. https://doi.org/10.2307/2295952

Arzubiaga, U., Kotlar, J., Massis, A. D., Maseda, A., & Iturralde, T. (2018). Entrepreneurial orientation and innovation in family SMEs: Unveiling the (actual) impact of the board of directors. *Journal of Business Venturing*, *33*(4), 455–469. https://doi.org/10.1016/j.jbusvent.2018.03.002

Asheim, B. T., Boschma, R., & Cooke, P. (2011). Constructing regional advantage: Platform policies based on related variety and differentiated knowledge bases. *Regional Studies*, *45*(7), 893–904. https://doi.org/10.1080/00343404.2010.543126

Aspelund, A., Bergutby, T., & Skjevdal, R. (2005). Initial resources' influence on new venture survival: A longitudinal study of new technology-based firms. *Technovation*, *25*(11), 1337–1347.

Balland, P.-A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialization policy in the European Union: Relatedness, knowledge complexity and regional diversification. *Regional Studies*, *53*(9), 1252–1268. https://doi.org/10.1080/00343404.2018.1437900

Barney, J. B. (2001). Is the resource-based "view" a useful perspective for strategic management research? Yes. *Academy of Management Review*, *26*(1), 41–56.

Batabyal, A. A., & Beladi, H. (2019). Preference matching, income, and population distribution in urban and adjacent rural regions. *Papers in Regional Science*, *98*(5), 2201–2208. https://doi.org/10.1111/pirs.12436

Beaudry, C., & Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, *38*(2), 318–337. https://doi.org/10.1016/j.respol.2008.11.010

Bell, A. J., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality & Quantity*, *53*(2), 1051–1074. https://doi.org/10.1007/s11135-018-0802-x

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, *39*(1), 61–74. https://doi.org/10.1080/0034340052000320887

Boschma, R. (2014). Constructing regional advantage and smart specialisation: Comparison of two European policy concepts. *Scienze Regionali*, *1*(1), 51–68.

Boschma, R., Eriksson, R., & Lindgren, U. (2008). How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity. *Journal of Economic Geography*, *9*(2), 169–190. https://doi.org/10.1093/jeg/lbn041

Boschma, R., & Iammarino, S. (2009). Related variety, trade linkages, and regional growth in Italy. *Economic Geography*, *85*(3), 289–311. https://doi.org/10.1111/j.1944-8287.2009.01034.x

Boschma, R., Minondo, A., & Navarro, M. (2012). Related variety and regional growth in Spain. *Papers in Regional Science*, *91*(2), 241–256.

Brannon, D. L., Wiklund, J., & Haynie, J. M. (2013). The varying effects of family relationships in entrepreneurial teams. *Entrepreneurship Theory and Practice*, *37*(1), 107–132. https://doi.org/10.1111/j.1540-6520.2012.00533.x

Brinckmann, J., Villanueva, J., Grichnik, D., & Singh, L. (2019). Sources of strategic flexibility in new ventures: An analysis of the role of resource leveraging practices. *Strategic Entrepreneurship Journal*, *13*(2), 154–178. https://doi.org/10.1002/sej.1313

Bun, M. J., & Harrison, T. D. (2019). OLS and IV estimation of regression models including endogenous interaction terms. *Econometric Reviews*, *38*(7), 814–827. https://doi.org/10.1080/07474938.2018.1427486

Byerlee, D., Janvry, A. D., & Sadoulet, E. (2009). Agriculture for development: Toward a new paradigm. *Annual Review of Resource Economics*, *1*, 15–31. https://doi.org/10.1146/annurev.resource.050708.144239

Cainelli, G., & Ganau, R. (2019). Related variety and firm heterogeneity: What really matters for short-run firm growth? *Entrepreneurship and Regional Development*, *31*(9–10), 768–784. https://doi.org/10.1080/08985626.2019.1571636

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*(2), 197–217. https://doi.org/10.1111/grow.12112

Cainelli, G., & Lupi, C. (2010). Does spatial proximity matter? Micro-evidence from Italy. In N. De Liso & R. Leoncini (Eds.), *Internationalization, Technological Change and the Theory of the Firm* (1st ed., pp. 163–186). London: Routledge.

Calzadilla, A., Rehdanz, K., Betts, R., Falloon, P., Wiltshire, A., & Tol, R. S. J. (2013). Climate change impacts on global agriculture. *Climatic Change*, *120*(1), 357–374. https://doi.org/10.1007/s10584-013-0822-4

Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: An analysis of US state-level patenting. *Regional Studies*, *49*(5), 767–781. https://doi.org/10.1080/00343404.2014.940305

Catao, L., & Milesiferretti, G. M. M. (2014). External liabilities and crises. *Journal of International Economics*, *94*(1), 18–32. https://doi.org/10.1016/j.jinteco.2014.05.003

Chana, C. S. R., Parhankangasb, A., Sahaymc, A., & Ood, P. (2020). Bellwether and the herd? Unpacking the U-shaped relationship between prior funding and subsequent contributions in rewardbased crowdfunding. *Journal of Business Venturing*, *35*(2), 105–121.

Chandler, G. N., McKelvie, A., & Davidsson, P. (2009). Asset specificity and behavioural uncertainty as moderators of the sales growth: Employment growth relationship in emerging ventures. *Journal of Business Venturing*, *24*(4), 373–387. https://doi.org/10.1016/j.jbusvent.2008.04.002

Chen, Y., Wang, L., & Zhang, M. (2018). Informal search, bad search? The effects of job search method on wages among rural migrants in urban China. *Journal of Population Economics*, *31*(3), 837–876. https://doi.org/10.1007/s00148-017-0672-x

Choi, T., & Park, J.-I. (2020). Related variety and productivity in the environmental industry: Estalishment-level evidence from South Korea. *Letters in Spatial and Resource Sciences*, *13*(3), 287–295. https://doi.org/10.1007/s12076-020-00259-w

Clarke, D., & Matta, B. (2018). Practical considerations for questionable IVs. *The Stata Journal*, *18*(3), 663–691. https://doi.org/10.1177/1536867X1801800308

Coad, A., Holm, J. R., Krafft, J., & Quatraro, F. (2018). Firm age and performance. *Journal of Evolutionary Economics*, *28*(1), 1–11. https://doi.org/10.1007/s00191-017-0532-6

Coad, A., & Srhoj, S. (2020). Catching gazelles with a lasso: Big data techniques for the prediction of high-growth firms. *Small Business Economics*, *55*, 541–565. https://doi.org/10.1007/s11187-019-00203-3

Combes, P. (2000). Economic structure and local growth: France, 1984–1993. *Journal of Urban Economics*, *47*(3), 329–355. https://doi.org/10.1006/juec.1999.2143

Conley, T. G., Hansen, C. B., & Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, *94*(1), 260–272. https://doi.org/10.1162/REST_a_00139

Content, J., & Frenken, K. (2016). Related variety and economic development: A literature review. *European Planning Studies*, *24*(12), 2097–2112. https://doi.org/10.1080/09654313.2016.1246517

Content, J., Frenken, K., & Jordaan, J. A. (2019). Does related variety foster regional entrepreneurship? Evidence from European regions. *Regional Studies*, *53*(11), 1531–1543. https://doi.org/10.1080/00343404.2019.1595565

da Silva, G. D., Gonçalves, E., & de Araújo Junior, I. F. (2020). Related variety and employment growth: A spatial dynamic model for Brazilian microregions. *Regional Science Policy & Practice*, *12*(1), 105–123.

Delmar, F., & Shane, S. (2006). Does experience matter? The effect of founding team experience on the survival and sales of newly founded ventures. *Strategic Organization*, *4*(3), 215–247. https://doi.org/10.1177/1476127006066596

Desdoigts, A., & Jaramillo, F. (2020). Bounded learning by doing, inequality, and multi-sector growth: A middle-class perspective. *Review of Economic Dynamics*, *36*, 198–219. https://doi.org/10.1016/j.red.2019.10.001

Drucker, J., & Feser, E. (2012). Regional industrial structure and agglomeration economies: An analysis of productivity in three manufacturing industries. *Regional Science and Urban Economics*, *42*(1–2), 1–14. https://doi.org/10.1016/j.regsciurbeco.2011.04.006

Duranton, G., & Puga, D. (2000). Diversity and specialisation in cities: Why, where and when does it matter? *Urban Studies*, *37*(3), 533–555. https://doi.org/10.1080/0042098002104

Ebbes, P., Papies, D., & Heerde, H. V. (2016). Dealing with endogeneity: A nontechnical guide for marketing researchers. In C. Homburg, M. Klarmann, & A. Vomberg (Eds.), *Handbook of Market Research* (pp. 1–37). Springer, Cham. https://doi.org/10.1007/978-3-319-05542-8_8-1

Ejdemo, T., & Örtqvist, D. (2020). Related variety as a driver of regional innovation and entrepreneurship: A moderated and mediated model with non-linear effects. *Research Policy*, *49*(7), 104073. https://doi.org/10.1016/j.respol.2020.104073

Erhardt, E. C. (2021). Measuring the persistence of high firm growth: Choices and consequences. *Small Business Economics*, *56*, 451–478. https://doi.org/10.1007/s11187-019-00229-7

Erkus-Ozturk, H. (2018). Related variety and innovation: Evidence from the tourism industry. *Tijdschrift voor Economische en Sociale Geografie*, *109*(2), 256–273. https://doi.org/10.1111/tesg.12275

Feng, L., Fu, T., Apergis, N., Tao, H., & Yan, W. (2019). The role of government intervention in financial development: Micro-evidence from China. *Accounting and Finance*, *59*(5), 2855–2878. https://doi.org/10.1111/acfi.12559

Firgo, M., & Mayerhofer, P. (2018). (Un)related variety and employment growth at the sub-regional level. *Papers in Regional Science*, *97*(3), 519–547. https://doi.org/10.1111/pirs.12276

Frenken, K., Oort, F. V., & Verburg, T. (2007). Related variety, unrelated variety, and regional economic growth. *Regional Studies*, *41*(5), 685–697. https://doi.org/10.1080/00343400601120296

Fritsch, M., & Kublina, S. (2018). Related variety, unrelated variety and regional growth: The role of absorptive capacity and entrepreneurship. *Regional Studies*, *52*(10), 1360–1371. https://doi.org/10.1080/00343404.2017.1388914

Glaeser, E. L., Kallal, H., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, *100*(6), 1126–1152. https://doi.org/10.1086/261856

Goe, W. R. (1991). The growth of producer services industries: Sorting through the externalization debate. *Growth and Change*, *22*(4), 118–141. https://doi.org/10.1111/j.1468-2257.1991.tb00565.x

Goh, S. K., McNown, R., & Wong, K. N. (2020). Macroeconomic implications of population aging: Evidence from Japan. *Journal of Asian Economics*, *68*, 101198. https://doi.org/10.1016/j.asieco.2020.101198

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). McGraw-Hill Irwin.

Haans, R. F. J., Pieters, C., & He, Z. (2016). Thinking about U: Theorizing and testing U- and inverted U-shaped relationships in strategy research. *Strategic Management Journal*, *37*(7), 1177–1195. https://doi.org/10.1002/smj.2399

Hoskisson, R. E., Gambeta, E., Green, C., & Li, T. X. (2018). Is my firm-specific investment protected? Overcoming the stakeholder investment dilemma in the resource-based view. *Academy of Management Review*, *43*(2), 284–306. https://doi.org/10.5465/amr.2015.0411

Huang, J., Liu, L., & Lu, R. (2020). Industry risk-taking and risk-taking strategy of born-global firms: An empirical study based on industrial variety perspective. *Asia-Pacific Journal of Business Administration*, *12*(1), 1–21.

Jacobs, J. (1961). *The death and life of great American cities*. Random House.

Jacobs, J. (1969). *The economy of cities*. Vintage.

Jiang, W. (2017). Have instrumental variables brought us closer to the truth. *The Review of Corporate Finance Studies*, *6*(2), 127–140. https://doi.org/10.1093/rcfs/cfx015

Kippersluis, H. V., & Rietveld, C. A. (2018). Beyond plausibly exogenous. *The Econometrics Journal*, *21*(3), 316–331. https://doi.org/10.1111/ectj.12113

Klement, B., & Strambach, S. (2019). Innovation in creative industries: Does (related) variety matter for the creativity of urban music scenes? *Economic Geography*, *95*(4), 385–417. https://doi.org/10.1080/00130095.2018.1549944

Larcker, D. F., & Rusticus, T. O. (2010). On the use of instrumental variables in accounting research. *Journal of Accounting and Economics*, *49*(3), 186–205. https://doi.org/10.1016/j.jacceco.2009.11.004

Lazzeretti, L., Capone, F., & Cinti, T. (2010). The regional development platform and "related variety": Some evidence from art and food in Tuscany. *European Planning Studies*, *18*(1), 27–45. https://doi.org/10.1080/09654310903343518

Lemmon, M. L., & Zender, J. F. (2010). Debt capacity and tests of capital structure theories. *Journal of Financial and Quantitative Analysis*, *45*(5), 1161–1187. https://doi.org/10.1017/S0022109010000499

Lind, J. T., & Mehlum, H. (2010). With or without U? The appropriate test for a U shaped relationship. *Oxford Bulletin of Economics and Statistics*, *72*(1), 109–118. https://doi.org/10.1111/j.1468-0084.2009.00569.x

Lu, R., & Ma, J. (2011). Can Western interorganizational governance mechanisms be applied to a guanxi-based market? *Asia-Pacific Journal of Business Administration*, *3*(2), 114–131. https://doi.org/10.1108/17574321111169812

Lu, R., Reve, T., Huang, J., Jian, Z., & Chen, M. (2018). A literature review of cluster theory: are relations among clusters important? *Journal of Economic Surveys*, *32*(4), 1201–1220. https://doi.org/10.1111/joes.12255

Lucio, J. D., Herce, J. A., & Goicolea, A. (2002). The effects of externalities on productivity growth in Spanish industry. *Regional Science and Urban Economics*, *32*(2), 241–258. https://doi.org/10.1016/S0166-0462(01)00081-3

Marshall, A. (1890). *Principles of Eeconomics*. MacMillan.

Martin, S. L., Javalgi, R. G., & Cavusgil, E. (2017). Marketing capabilities, positional advantage, and performance of born global firms: Contingent effect of ambidextrous innovation. *International Business Review*, *26*(3), 527–543. https://doi.org/10.1016/j.ibusrev.2016.11.006

Mendoza-Abarca, K. I., Anokhin, S., & Zamudio, C. (2015). Uncovering the influence of social venture creation on commercial venture creation: A population ecology perspective. *Journal of Business Venturing*, *30*(6), 793–807. https://doi.org/10.1016/j.jbusvent.2015.04.003

Miguelez, E., & Moreno, R. (2018). Relatedness, external linkages and regional innovation in Europe. *Regional Studies*, *52*(5), 688–701. https://doi.org/10.1080/00343404.2017.1360478

Naldi, L., Criaco, G., & Patel, P. C. (2020). Related and unrelated industry variety and the internationalization of start-ups. *Research Policy*, *49*(10), 104050. https://doi.org/10.1016/j.respol.2020.104050

Neffke, F., & Henning, M. (2013). Skill relatedness and firm diversification. *Strategic Management Journal*, *34*(3), 297–316. https://doi.org/10.1002/smj.2014

Paccagnella, O. (2006). Centering or not centering in multilevel models? the role of the group mean and the assessment of group effects. *Evaluation Review*, *30*(1), 66–85. https://doi.org/10.1177/0193841X05275649

Paci, R., & Usai, S. (2006). *Agglomeration economies and growth in Italian local labour systems 1991–2001*. Centre for North South Economic Research.

Park, H. D., & Steensma, H. K. (2012). When does corporate venture capital add value for new ventures? *Strategic Management Journal*, *33*(1), 1–22. https://doi.org/10.1002/smj.937

Patin, J.-C., Rahman, M., & Mustafa, M. (2020). Impact of total asset turnover ratios on equity returns: Dynamic panel data analyses. *Journal of Accounting, Busines & Management*, *27*(1), 19–29.

Porter, M. E. (1998). *On competition*. Harvard Business School Press.

Reuber, A. R., & Fischer, E. (1997). The influence of the management team's international experience on the internationalization behaviors of SMES. *Journal of International Business Studies*, *28*(4), 807–825.

Robbins, S. P., & Coulter, M. (2018). *Management* (14th ed.). Tsinghua University Press.

Rocca, M. L., Stagliano, R., Rocca, T. L., Cariola, A., & Skatova, E. (2019). Cash holdings and SME performance in Europe: The role of firm-specific and macroeconomic moderators. *Small Business Economics*, *53*(4), 1051–1078. https://doi.org/10.1007/s11187-018-0100-y

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, *94*(5), 1002–1037. https://doi.org/10.1086/261420

Scott, A. J. (1986). Industrial organization and location: Division of labor, the firm, and spatial process. *Economic Geography*, *62*(3), 215–231. https://doi.org/10.2307/144006

Thabet, K. (2016). Industrial structure and total factor productivity: The Tunisian manufacturing sector between 1998 and 2004. *Annals of Regional Science*, *56*(3), 639–662.

Tung, R. L., & Worm, V. (2001). Network capitalism: the role of human resources in penetrating the China market. *International Journal of Human Resource Management*, *12*(4), 517–534. https://doi.org/10.1080/713769653

Waldman, M., Nicholson, S., & Adilov, N. (2006). Does television cause autism? National Bureau of Economic Research Working Paper12632.

Waldman, M., Nicholson, S., Adilov, N., & Williams, J. (2008). Autism prevalence and precipitation rates in California, Oregon, and Washington counties. *JAMA Pediatrics*, *162*(11), 1026–1034.

Wang, Y., Assche, A. V., & Turkina, E. (2018). Antecedents of SMEs embeddedness in inter-organizational networks: evidence from China's aerospace industry. *Journal of Small Business and Entrepreneurship*, *30*(1), 53–75. https://doi.org/10.1080/08276331.2017.1391368

Yang, H. (2000). A note on the causal relationship between energy and GDP in Taiwan. *Energy Economics*, *22*(3), 309–317. https://doi.org/10.1016/S0140-9883(99)00044-4

Younkin, P., & Kashkooli, K. (2020). Stay true to your roots? Category distance, hierarchy, and the performance of new entrants in the music industry. *Management Science*, *31*(3), 604–627.

How to cite this article: Lu, R., Song, Q., Xia, T., Lv, D., Reve, T., & Jian, Z. (2021). Unpacking the U-shaped relationship between related variety and firm sales: Evidence from Japan. *Papers in Regional Science*, *100*(5), 1136–1157. https://doi.org/10.1111/pirs.12608

DOI: 10.1111/pirs.12608

Resumen. El objetivo de este artículo es estudiar cómo la variedad relacionada influye en las ventas de las empresas. Se aplicó un método de variables instrumentales (así como el método más reciente de variables instrumentales plausibles en las pruebas de robustez) para analizar más de 600.000 observaciones de empresas en las 47 prefecturas de Japón. Se encontró que la variedad relacionada, como una forma de estructura industrial regional, tiene una relación en forma de U con las ventas de las empresas. Este hallazgo enriquece la perspectiva de la variedad relacionada al complementar la evidencia a nivel micro, y revela que la relación entre "variedad relacionada-ventas de la empresa" no es lineal, como han sugerido la mayoría de los estudios anteriores.

抄録: 本稿では、関連ある多様性(related variety)が企業の売上にどのように影響するかを検討する。操作変数法(および頑健性の検証における最新の最適な操作変数法)を適用し、日本の全47都道府県における60万以上の企業の調査結果を分析する。地域産業の構造の一種として、関連ある多様性は、企業の売上とU字型の関連性を持つことが分かった。この知見は、ミクロレベルのエビデンスを補完することにより関連ある多様性の視点を豊かにし、「関連ある多様性と企業の売り上げ」の関連が、ほとんどの既存研究が示したような線形ではないことが明らかになった。