International Review of Entrepreneurship, Article #1692, 21(1): pp. 35-66.

Tracking the Long-Term Trajectory of International New Ventures' Innovation: The Moderating Role of Regional Multi-Cluster Diversity

Xiangcai Peng

Business School, China University of Political Science and Law, Beijing, China

Ren Lu[1]

School of Business Administration, Guangdong University of Finance & Economics, Guangzhou, China

Yinglin Liu

Institute of Applied Economics, Shanghai Academy of Social Sciences, Shanghai, China

Abstract. The purpose of this article is to study how the innovation performance of international new ventures (INVs) evolves over time, and how multi-cluster diversity as an external environment moderates INVs' innovation performance. We employed the latent growth curve modeling (LGCM) approach to analyze 5,744 INVs from 21 cities in the Guangdong Province (China). We found that INVs' innovation follows a positive linear trend over time, and that multi-cluster diversity plays a positive moderating role. Our findings enrich international entrepreneurship by showing the long-term effect of early internationalization on INVs' innovation under a multi-cluster environment.

Keywords: international new ventures, internationalization duration, innovation, multi-cluster diversity.

1. Introduction

International new ventures (INVs), which are firms that enter international markets shortly after their inception, have attracted widespread attention and in-depth discussions in the field of international entrepreneurship (IE) since the pioneering work of McDougall et al. (1994), Oviatt and McDougall (1994) and Knight and Cavusgil (2004). Innovation is one of the most important factors that influences an INV's performance: innovation shapes superior international performance of INVs (Knight and Cavusgil 2004), and INVs also benefit from

1. Corresponding author: Ren Lu, School of Business Administration, Guangdong University of Finance & Economics, Guangzhou, China. Email: biluren@qq.com

organizational and technological innovation by exporting (Azar and Ciabuschi 2017).

While many studies have investigated INVs' innovation from various perspectives, two main questions remain: (1) How does the long-term trajectory of INVs in terms of innovation evolve over time? (2) How does regional multi-cluster diversity, as a specific type of regional industrial context, impact INVs' innovation? To be more specific, although scholars have recently started to realize that INVs' internationalization is essentially an entrepreneurial behavior process that occurs over time (Cavusgil and Knight 2015), only limited attention has been paid to the long-term trajectory of INVs in terms of innovation (Acedo et al. 2021; Coviello 2015; Romanello and Chiarvesio 2017); in other words, how INVs evolve over time is still "embryonic" (Knight and Liesch 2016) and insufficiently understood (Coviello 2015). On the other hand, INVs' innovations are influenced by its regional environment, particularly local clusters (Cavusgil and Knight 2015). Although scholars have explored how a single industrial cluster influences INVs, they have overlooked the fact that many cities and metropolitans generally contain two or more clusters, which form different regional multi-cluster diversities (Bathelt and Zhao 2016; Li and Bathelt 2018). Merely considering a single cluster's effect on INV but ignoring multiple clusters' effects encounters the missing variable bias, which is a serious problem that causes research results to become imprecise and inconvincible.

In this paper, we aim to fill the two above-mentioned research gaps. By integrating the dynamic state approach and resource-based view, we develop a theoretical framework to investigate the dynamics of an INV's innovation since the implementation of its fast and rapid internationalization strategy and the moderating role of regional multi-cluster diversity. We applied latent growth curve models (LGCM) to analyze 33,881 firm-level longitudinal observations from 21 cities in the Guangdong Province (China). We found that INVs' innovation follows a positive linear trend over time, and that regional multi-cluster diversity positively moderates INVs' innovation performance.

Our paper makes both theoretical and practical contributions. First, building on Acedo et al. (2021), we expand the research frontier on how INVs evolve over time in terms of innovation. Our study observes frequent changes in INVs' innovation as the internationalization duration increases. Secondly, both Cavusgil and Knight (2015) and Zander et al. (2015) pointed out that studying how clusters influence INVs is a critical research direction. We respond to their call by introducing the concept of multi-cluster diversity into INV research. Thirdly, from a practical perspective, our results not only help policymakers carry out effective cluster policies that support the internationalization and innovation of new ventures, but also caution founders/managers that approaching the industrial cluster may be the wisest strategy.

The remainder of the paper is organized as follows. Section 2 provides relevant knowledge about two important concepts, namely international new ventures and multi-cluster diversity. Section 3 develops hypotheses. Section 4 introduces the data and details the econometric method, which is the latent growth curve model (LGCM). Section 5 reports the estimation results and various robustness checks. Section 6 concludes.

2. Concepts and Relevant Knowledge

This section consists of three parts. In Sub-section 2.1, we define international new ventures. Because we draw on the resource-based view to develop our hypothesis, we briefly present relevant theoretical background in Sub-section 2.2. Knowledge about regional multi-cluster diversity, which is the moderating variable of the present paper, will be introduced in Sub-section 2.3.

2.1. International New Ventures

In the early 1990s, a rapid internationalization phenomenon emerged in the international economy; many firms entered the international market shortly after their inception, rather than following a gradual internationalization process (McDougall et al. 1994; Johanson and Vahlne 2015; Knight and Cavusgil 2004). This phenomenon has attracted intensive attention in the field of entrepreneurship and international business, and prompted the emergence of the field of international entrepreneurship (Cavusgil and Knight 2015). Accordingly, scholars proposed different labels to identify firms that internationalize from inception, such as "international new ventures (INVs)," "global start-ups," "instant internationals," "early or young internationalizing firms," and "born globals" (Rialp et al. 2005; Oviatt and McDougall 1994; Zahra 2005; Knight and Cavusgil 2004). Among these, INVs, which are defined as young and entrepreneurial firms with export sales that represent at least 20 percent of their total sales within three years of inception, are influential (Zhou et al. 2010).

Over the past three decades, INV scholars mainly focused on describing, understanding, and explaining various antecedents and driving factors that influence the emergence of INVs (Kim and Cavusgil 2020; Gerschewski et al. 2018), including the impact of entrepreneurs' characteristics (Lee et al. 2016), organizational features (Martin et al. 2020; Buccieri et al. 2020), external environment (Pangarkar and Yuan 2021; Amdam et al. 2020; Acedo et al. 2021), networks (Loane and Bell 2006), the form and process of INVs' learning (Huang et al. 2020; Tuomisalo and Leppaaho 2019; Gerschewski et al. 2018; Øyna et al. 2018), and INVs' internationalization models (Hashai and Almor 2004; Gerschewski et al. 2015).

Recently, some literature has started to investigate international entrepreneurship from a dynamic perspective, and prior studies have mainly focused on a specific internationalization stage, tending to divide the internationalization process of INVs into pre-entry and post-entry phases in terms of time (Cavusgil and Knight 2015; Liesch et al. 2007). For instance, Acedo et al. (2021), based on 485 observation data of 97 INVs from 1990 to 2015 in Spain, found that the initial level of export intensity and geographic dispersion of INVs would affect their long-term trajectories. Although prior studies have empirically investigated the impact of early events on later outcomes, they did not provide answers to the question of how the long-term trajectory of INVs' innovation has evolved over time since INVs entered the international market.

2.2. Integrating the Resource-based View and the Dynamic State Approach

This paper traces INVs' growth process of innovation. Compared to other business theories such as transaction cost theory, business network theory, activity-based theory, and so on, the dynamic state approach and the resource-based perspective provide additional detail for understanding the growth and development of startups (Levie and Lichtenstein 2010; Barney 2001; Peng 2001). Accordingly, we integrate the dynamic state approach and the resource-based view.

The dynamic state approach (DSA) is a cutting-edge methodology in the field of international entrepreneurship research, which has gained significant attention in recent years. Compared to the traditional stage theory, this approach offers a more comprehensive and adaptable perspective for understanding firm growth and entrepreneurship (Levie and Lichtenstein 2010). Rather than following predetermined stages of development, DSA views firms as dynamic systems. Through dynamic states of synergy between the entrepreneur, the organization, and the market, etc., firms are able to achieve the most efficient development (Salvador Federico 2016). In other words, DSA takes into account a range of complex factors that influence business growth, including sustainability, social entrepreneurship, and industry-wide dynamics, etc. By studying these factors, we can determine what sustains dynamic states, when and where they change, and which contextual variables are most important in the process. Therefore, DSA has a wide range of research applications in areas such as organizational learning, social enterprise, and sustainability entrepreneurship.

The resource-based view (RBV) suggests that a firm is a collection of resources, both tangible and intangible, that can be transformed into unique capabilities providing sustained competitive advantage (Wernerfelt 1984; Barney 2001). According to RBV, a firm's survival and sustainability usually depend on its resources and ability that contribute to creating competitive advantages. However, Zahra and Das (1993) emphasize that resources alone do not directly transform into competitive advantage. It is the capabilities that lie behind the resources to acquire, allocate, utilize, protect, and integrate them that are critical to successful transformation. Transforming resources into a competitive advantage involves an iterative process of absorption, utilization, and transformation (Qiu et al. 2020). This iterative process usually includes two basic activities: searching resources and integrating resources (Sirmon et al. 2011). Searching resource refers to the process by which firms find new and valuable ideas within a large and diverse set of internal and external innovation resources (Laursen and Salter 2014). When a firm cannot fully access the resources it requires from within, it should search the external environment for resources controlled by other organizations. Integrating resources is the process by which a company obtains innovation by recombining or introducing new features to previously existing elements (Guan and Yan 2016; Felin et al. 2023). Therefore, for survival and sustainable competitive advantage, firms must not only maximize their existing resources and capabilities, but also develop new resources and capabilities.

In the present paper, we propose a framework to analyze the long-term innovation performance of INVs by combining DSA with RBV: exploring the long-term innovation performance of INVs by assessing the dynamic state and matching of resources, capabilities, and the external environment.

2.3. Regional Multi-cluster Diversity

In the field of international entrepreneurship, the importance of resources for the internationalization of INVs has drawn the attention of scholars to the resources that INVs possess (Westhead et al. 2001). However, scholars have tended to focus on how INVs obtain resources from international markets to sustain their survival and growth as they internationalized, and limited attention has been paid to how the industrial cluster environment in which INVs operate allows them to obtain the resources they need to undertake internationalization (Cavusgil and Knight 2015). In fact, INVs can benefit from internationalization only if they have access to the resources necessary to support them in their international activities (Fernhaber et al. 2008; Dunning 1998). As an important external environment for firms, industrial clusters have been shown to have a significant impact on processes such as resource acquisition, technology transfer and knowledge spillover, which further influence firms' performance and internationalization (Wennberg and Lindqvist 2008; Bathelt et al. 2004; Bathelt and Zhao 2016; Fernhaber et al. 2008; Porter 1996). In a cluster – which is defined as "geographic concentrations of interconnected companies, specialized suppliers, service providers, firms in related industries, and associated institutions in a particular field (Porter 2000, p. 15)" – INVs have access to skilled workers, specialized inputs for operations, and knowledge of opportunities and competitor activity. This can not only help INVs overcome the constraints associated with a limited history and small size, but also contribute to their rapid internationalization.

Many cities and metropolitans generally contain more than one cluster, which form different types of regional multi-cluster diversity (Bathelt and Zhao 2016; Li and Bathelt 2018). When multiple clusters are co-located in geographic proximity (Bathelt and Li 2014), they show various features that differ from a single cluster. For example, Henn and Bathelt (2018) argue that actors between different clusters can exchange knowledge through various channels when multiple clusters coexist in a city, such as meetings, networking events, informal meetings and collaboration, thus causing cross-cluster knowledge convergence (Bathelt and Li 2020). For firms, through learning and absorbing the diverse knowledge from different industry clusters, they can not only break their dependence on local (internal) search (Posen et al. 2018), but also discover more new ideas and opportunities that can help them implement innovation when integrating that knowledge (Grant 1996). A small but increasing number of studies have highlighted the importance of investigating multiple cluster phenomena (e.g., Turkina 2018). In response to the call for relevant studies, building on Delgado et al. (2016)'s algorithm for industrial clusters and Duranton and Puga (2000)'s industry diversity idea, we particularly explore a special characteristics of multiple clusters, namely regional multi-cluster diversity (MCD). MCD here refers to the presence and coexistence of multiple industrial clusters within a given region or geographical area (e.g. city). These clusters are characterized by their unique cultural, social, economic, or demographic attributes that differentiate them from one another (Delgado 2020; Bathelt and Zhao 2016). More specifically, the essence of MCD contains two aspects: how many clusters there are in a region, and whether or not all individual clusters are equally developed. When regional multi-cluster diversity is high, it implies that a region consists of a lot of clusters and the size of each cluster is similar or the same, and vice versa. Studying MCD rather than a single cluster would provide us with a systematic understanding of the impact of the external environment on INV performance.

3. Hypothesis Development

3.1. Long-term Trajectory of INV's Innovation

In the literature on international entrepreneurship, many studies have revealed that INVs would take similar steps when they conduct their rapid internationalization strategy (Paul and Rosado-Serrano 2019; McDougall et al. 1994). In the beginning, due to limited resources and a lack of knowledge of international markets, INVs enter international markets and establish trade relationships mainly by exporting products or services (Alon et al. 2013; Paul and Rosado-Serrano 2019). Then, as trade relationships stabilize and resources and relevant knowledge accumulate, the INVs consider finding agents or establishing branches in the local market (Hennart 2014; Knight and Cavusgil 2004). Finally, they gradually integrate into local networks and localize their operations through friction and cooperation with actors in the local market (Øyna et al. 2018; Vanninen et al. 2017). Following the above-mentioned studies, we attempt to analyzing the long-term innovation performance of international new ventures (INVs) based on the dynamic states approach and the resource-based view. For the sake of demonstration, we briefly divide the dynamic state into three specific phases and discuss how the evolution of resources, capabilities, and external environments affect INVs' long-term innovation performance.

The start-up phase means the first several years that INVs begin to implement their international strategy. New ventures frequently perceive themselves as highly competent and thus leap to internationalize in order to seek opportunities (McDougall et al. 1994). INVs face a variety of challenges, such as limited resources, financial constraints, and imperfect technologies and products in the start-up phase (Knight and Cavusgil 2004). At this stage, INVs, as new entrants in international markets, often do not have sufficient basic resources to ensure their survival and sustainability, and this situation will drive them to actively seek internal and external resources (Mudambi and Zahra 2007; Zeng et al. 2010). Prior studies have noted that INVs' flexible organizational characteristics and the learning advantages of newness (LAN) create the conditions for acquiring resources as well as transforming acquired resources into innovative performance and competitive advantage (Autio et al. 2000; Pangarkar and Yuan 2021). On one hand, INVs' flexibility not only enables them to respond quickly to market changes, upgrade their products and services in response to market demand and feedback, and even develop new products (D'Angelo et al. 2020; Salomon and Shaver 2005), but also to seize key opportunities in international markets and experience fewer risks during the international expansion phase (Pangarkar and Yuan 2021). On the other hand, LAN can bridge the gaps between INVs' initial endowments and the capabilities required for effective competition through learning vicariously from the knowledge of incumbents (Posen and Chen 2013). When they interact with local actors and network members in foreign markets, LAN enables INVs to quickly acquire, assimilate, transform, and exploit external new knowledge (Wu and Voss 2015), which enables them to apply new ideas to the products, processes, or any other aspect of a firm's activities and thus enhance their own innovation performance.

In the growth and expansion phase, INVs can access resources that are controlled by other organizations (firms) more effectively by leveraging networks and strategic alliances, due to scale expansion and accumulation of resources (Gerschewski et al. 2018). During this phase, the establishment of agencies and affiliates not only enables INVs to acquire and accumulate knowledge of international markets, but also allows them to rapidly acquire resources through interacting with network members, strategic alliances and local actors, and effectively avoid opportunism (Mudambi and Zahra 2007). As integration in local markets increases, INVs' social capital and market position improves. At that time, access to resources is no longer a major issue for INVs, but they will face the challenges that it is difficult for INVs to mechanically inject resources to drive innovation and foster competitive advantage as their technologies, products, and services gradually become mature. At this point, INVs need to identify the value of their resources and recombine them to obtain resources that are aligned with the firm's growth path. Integrating resources not only improves the efficiency of resource utilization and avoids the phenomenon of sunken resource value, but also enables INVs to identify new innovation opportunities and thus achieve better innovation performance. The organizational capabilities, particularly absorptive capacity, influence the recognition of the value of new external resources during the integration process (Cohen and Levinthal 1990). Absorptive capacity not only makes the exchange and learning of existing knowledge possible (Ritala and HurmelinnaÅ]Laukkanen 2013), but also connects new external knowledge to its prior experience (Cohen and Levinthal 1990), and facilitates the generation of new ideas.

In the mature phase, INVs have well-developed business routines, as well as dedicated technology and market development units and other resources that can effectively support capability learning (Johanson and Vahlne 2015). INVs can develop new resources and capabilities in order to maintain and grow their competitive advantage. As Lewin et al. (2011, p. 84) explained, "new superior routines, capabilities, and new knowledge emerge through a dynamic interaction of internal and external variation, selection, and replication processes involving knowledge creation and change over time." While certain capabilities will vanish over time, INVs may benefit from more well-developed capabilities, particularly absorptive capabilities (Blesa and Ripollés 2021). After analyzing INVs' innovation performance at three phases, we found that INV's long-term innovation performance is the result of its effective use of resources and capabilities at various stages of its development (Penrose 1959). Accordingly, we propose the following hypothesis:

Hypothesis 1: INVs' innovation performance follows a positive linear trend over time after implementing a rapid internationalization strategy.

3.2. Moderating Role of Regional Multi-cluster Diversity

Innovation is accompanied by a continuous accumulation of resources and knowledge, and in order to improve innovation performance, companies need access to sufficient knowledge resources (Nonaka and Takeuchi 1995; Kurokawa et al. 2007). For INVs, in addition to international markets, the local environment can provide a variety of sources of the resources that INVs need to survive and grow.

The coexistence of multiple clusters can provide abundant heterogeneous and substitutable resources compared to the impact of a single cluster (Li and Bathelt 2018), which reduces the level of competition for resources among firms and significantly increases the availability of resources (Fernhaber et al. 2008). This enables INVs to easily compete for high-quality resources (such as experienced talents). In other words, MCD provides INVs with the resources they need to establish and maintain its operations. By constantly combining and recombining these resources, INVs are able to increase their own productivity levels while gaining access to new knowledge and innovation, thus reaching export thresholds and establishing distribution networks abroad more quickly (Fernhaber and Li 2010; Fernandes and Tang 2014). Consequently, rapid internationalization allows INVs to take advantage of narrow windows of opportunity to develop new products in international markets earlier than their competitors (Fernhaber et al. 2008). In addition, when MCD is high, it implies that a region contains a lot of resources, which will not only affect the initial level of export intensity and geographic dispersion of INVs (Boehe and Jiménez 2016), but can also influence INVs' efforts to learn new knowledge from external environments and conduct innovative activities (Demirkan 2018), thereby leading to a positive impact on their innovation. For instance, Acedo et al. (2021) found that INVs' long-term international performance is affected by the initial level of export intensity and geographic dispersion, so the impact of INVs' early internationalization strategy on innovation is constrained by MCD.

External networks of companies are particularly important for innovation, and for INVs, they are an important source of access to international opportunities (Coviello and Munro 1995; Pittaway et al. 2004). Particularly successful clusters are able to establish and maintain a variety of channels for low-cost knowledge exchange with relevant global hotspots (Bathelt et al. 2004), which creates a favorable external networking environment for firms within the cluster. Firms in other clusters in the same city benefit from the presence of such channels due to multiple clusters being co-located in geographic proximity. Thus, INVs in high-MCD regions will have many opportunities and access to establish relationships with firms and institutions in relevant hotspots globally (Capone and Lazzeretti 2018). INVs will then be more inclined to use these global pipelines to enter international markets earlier. The higher the level of MCD and the more complex the local and global network relationships, the more likely it is that INVs will seek innovation from domestic and foreign partners (Sapienza et al. 2006). In other words, INVs will have more opportunities to combine what they have learned in international markets with local resources, which will have an impact on their innovation performance. Therefore, we propose the following hypothesis:

Hypothesis 2: Multi-cluster diversity has a positive moderating effect on INV's innovation performance.

4. Methodology

4.1. Research Context and Data Sources

We chose manufacturing firms in Guangdong Province (China) as our research objective. As an emerging market country, China's increasing influence in international trade has affected not only large multinational firms, but also SMEs that enter international markets through exporting as the initial foreign entry model (Alon et al. 2013), especially after China's accession to the WTO in 2001. Guangdong is not only ranked first among all of China's provinces in terms of total imports and exports, but is also one of the most favored provinces for foreign direct investment. These conditions provide favorable conditions for the emergence of international new ventures and a good context for us to test the hypotheses proposed above. Figure 1 and Figure 2 provides details about the importance of Guangdong Province in China's foreign trade.

Figure 1. China's total imports and exports for each province ($100 million)
Note: the line with triangle is Guangdong.

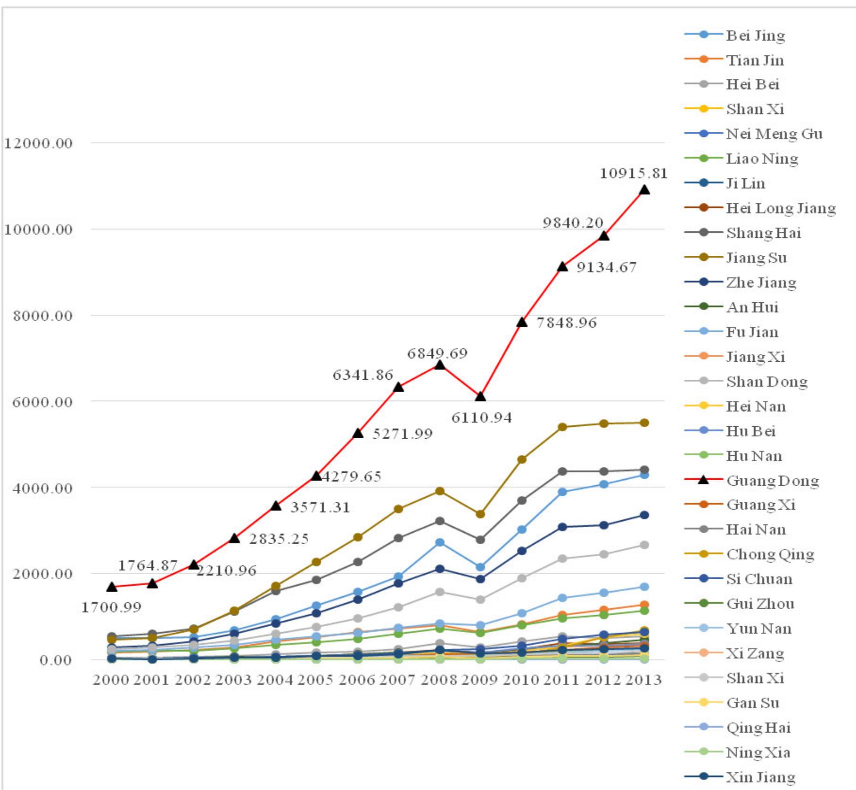
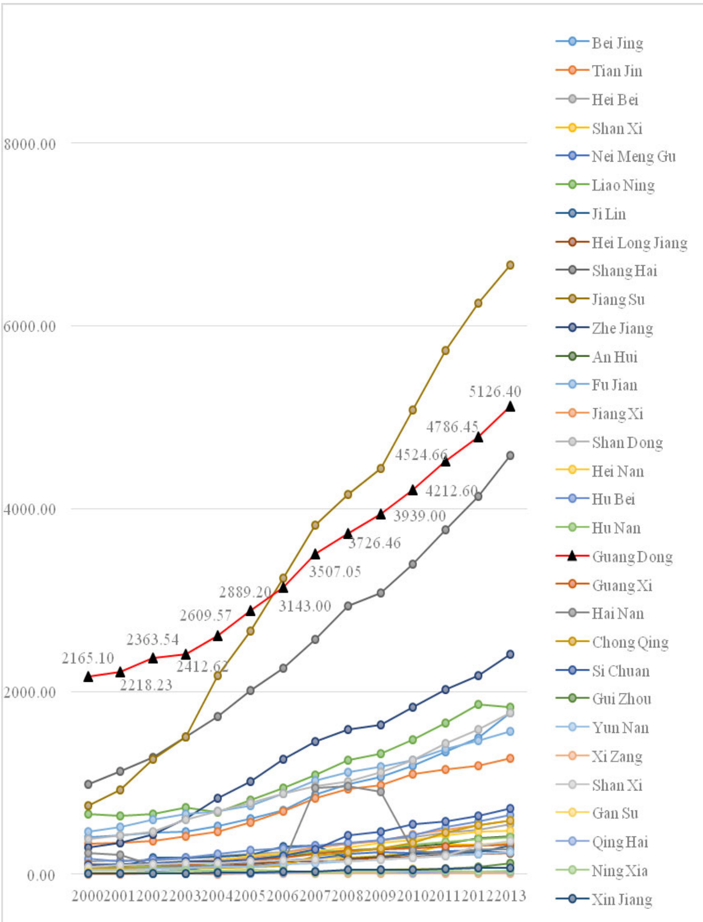


Figure 2. China's foreign direct investment for each province ($100 million)
Note: the red line with triangle is Guangdong; it was ranked as the first before 2006; after that, it was ranked as the second place.



Our data consist of two parts: firm-level data and city-level data. At the firm level, we first collected the firms' financial and export data from the RESSET Database (http://www.resset.cn/), which provides data on Chinese firms from 1998 to 2014 and customs data at the firm level from 2000 to 2016. We collected firms' patent data from the China National Intellectual Property Administration (https://www.cnipa.gov.cn/). At the city level, data are collected from the *China City Statistical Yearbook*, which provides us with detailed demographic, economic, and fiscal indicators at the city level. Finally, we obtained a data set containing the firm's financial information, patents, and export value. The time range selected for the sample data was from 2000–2013. We winsorized the top and bottom 1 percent of data to reduce the influence of outliers.

The initial sample includes 481,463 observations from all 21 cities in Guangdong Province. We quantified and selected INVs based on five principles: (1) firms with missing patent data were deleted; (2) firms with observation data of fewer than three years were deleted; (3) firms that do not export were deleted; (4) if the firm's age is older than three years when the first time exporting, it was deleted; (5) firms whose export intensity is less than 20 percent within three years of their establishment were deleted. The resulting sample consists of 33,881 observations, which includes 5,744 INVs.

4.2. Variables

4.2.1 Dependent Variable

Patents are generally regarded as a definite measure of a firm's innovation (Bronzini and Piselli 2016; Wang et al. 2021). Compared with other measurements (such as the number of new products launched by the firm or R&D expenditures), it is more objective to use patents to measure innovation. On the one hand, the products or processes that may be new for the firm may be outdated for the market (Bronzini and Piselli 2016). On the other hand, changes in the external environment may cause fluctuations in firms' R&D investment, which leads to the fact that R&D investment does not well reflect firms' true willingness to innovate. Unlike new products and R&D investment, patent applications are usually reviewed by experts for their novelty and applicability. Consequently, they can capture details about what is happening, making them a great way to measure innovation activity (Wang et al. 2021). We measure innovation by firm's annual patent applications (denoted by *A_Patents*), which is summed by the annual patent applications of invention patents and utility patents. We take the logarithm for normalizing the distribution:

$$A_{Patents_{tj}} = \ln\left(APat_{Invention_{t,j}} + APat_{Utility_{t,j}}\right)$$

where the subscripts $\boldsymbol{t}$ and $\boldsymbol{j}$ denote year and firms, respectively; $A_{Patents_{tj}}$ is the logarithm of the patents of firm $\boldsymbol{j}$ at year **t**. $APat_{Invention_{t,j}}$ is the number of invention patents applied for by firm $\boldsymbol{j}$ at year **t**, and $APat_{Utility_{t,j}}$ is the number of utility model patents applied for by enterprise $\boldsymbol{j}$ at year **t**.

4.2.2. Independent Variable

We are interested in the dynamics of INV's innovation performance over time since the implementation of its fast and rapid internationalization strategy. Thus, we use internationalization duration (*Duration*) as a proxy variable for time trends (Crespo et al. 2020). More specifically, *Duration* denoted the time (year) of INV's presence in international markets since the implementation of the fast and rapid internationalization strategy (*Duration* = 0,1, 2, 3, …).

4.2.3. Moderator

Regional multi-cluster diversity index (MCD) is a measurement of the external industrial environment, which reveals the degree to which a certain cluster industrial portfolio matches or deviates from the entire city portfolio (Duranton and Puga 2000; Delgado et al. 2016). We calculate the MCD in two steps.

In the first step, we calculated the number of clusters in a city. Delgado et al. (2016) coined a method to quantify industrial cluster, and we fully replicated and applied their method to quantify clusters in Guangdong Province. Specifically, Delgado et al.'s method consists of five steps: building a similarity matrix, choosing broad parameters, setting the clustering function, calculating performance scores, and assessing individual cluster's score (Delgado et al.'s method is complex, but they provided Stata code with which we apply it into our Guangdong sample). A significant merit of applying Delgado et al.'s method is that, when quantifying a cluster, such method takes industrial relatedness into account. It is the method that fully reveals Porter's cluster definition (please check Appendix 1 for more details).

As we have quantified each single cluster, in the second step, drawing on the method proposed by Duranton and Puga (2000), we measured regional multi-cluster diversity with three steps:

1. Calculating the share of cluster ***i*** in city ***k***'s employment (s_{ik}).
2. Calculating the share of cluster ***i*** in national employment (s_i).
3. Finally, calculating the absolute value of the difference between each single cluster's share in city's employment and its share in national employment, and the multi-cluster diversity index is given by:

$$MCD_k = {}^{1}\!/\!\textstyle\sum_{i=1}^{n} |s_{ik} - s_i|$$

where MCD_k denotes the level of multi-cluster diversity in city ***k***. According to the suggestions of Aiken et al. (1991), Cohen et al. (2003), and Hayes (2017), we centralized MCD to make the coefficients of the regression equation more explanatory.

4.2.4. Control Variables

We controlled for a variety of variables, at both the city level and firm level. Firstly, at the firm level, we controlled firm size (*Size*) measured by the total assets, debt-to-asset ratio (*Lev*) measured by dividing total assets by total liabilities and fixed assets (*lnFixed*) (Zheng et al. 2018; Zhang et al. 2020; Mukherjee et al. 2017). At the city level, we control the city's GDP (*lnGDP*), *lnFDI*, which is measured by the natural logarithm of the foreign direct investment in a city, and *RFIF*, which is measured by the number of registered foreign-invested firms (Nuruzzaman et al. 2019; Zhang and Song 2001). We also control industry capital intensity (*lnIndInt*), measured by dividing each industry's total fixed assets by its total employment (Zahra et al. 2018). Table 1 provides descriptive information on variables.

Table 1. Descriptive Statistics and Correlation matrix of Variables

	Mean	SD	N	A_Patents	Duration	Size	Lev	lnFixed	MCD	lnGDP	lnFDI	RFIF	lnIndInt
A_Patents	0.099	0.450	33,881	1.000									
Duration	4.155	2.728	33,881	0.148***	1.000								
Size	10.156	1.381	33,881	0.281***	0.314***	1.000							
Lev	0.576	0.277	33,881	0.010*	-0.070***	0.007	1.000						
lnFixed	8.488	1.783	33,881	0.216***	0.211***	0.821***	-0.140***	1.000					
MCD	1.417	0.914	33,881	-0.015***	0.085***	0.015***	0.022***	0.041***	1.000				
lnGDP	16.776	1.055	33,881	0.129***	0.254***	0.240***	0.141***	0.104***	0.308***	1.000			
lnFDI	11.467	1.248	33,881	0.103***	0.063***	0.210***	0.188***	0.102***	0.283***	0.835***	1.000		
RFIF	8.117	1.175	33,881	0.065***	-0.017***	0.140***	0.168***	0.046***	0.180***	0.727***	0.844***	1.000	
lnIndInt	3.698	0.807	33,881	0.114***	0.083***	0.370***	-0.040***	0.414***	0.149***	0.168***	0.193***	0.117***	1.000

4.3. Estimation Methodology

4.3.1. Latent Growth Curve Model

The latent growth curve model (LGCM) is a useful tool for analyzing longitudinal data to track the trajectories and changes of various phenomena over time. For a given measure (e.g., innovation, profitability), the LGCM assumes that each individual has a unique development path. In the current study, we argue that each INV's innovation has a distinct trajectory over time. When using LGCM, we must first fit a trajectory line for each INV's innovation. The fixed effects are then estimated based on the fitted trajectory line, which is represented by the mean intercept and mean slope. A random effect, which is represented by the variance of each line, will also be estimated. These variances represent differences in the trajectory lines, and if they are statistically significant, it indicates that the trajectory lines differ between individuals and do not follow the same pattern.

4.3.2. Steps of Building Modeling

Following the existing literature on LGCM modeling (Raudenbush and Bryk 2002), we thus employ the two-level LGCM model to test the hypotheses proposed above. The first stage is the Level-1 model:

$$A_{Patents_{tj}} = \pi_{oj} + \pi_{1j} Duration_{tj} + \varepsilon_{tj} \quad (1)$$

where the indices *t* and *j* denote time (year) and firms, respectively; $A_{Patents_{tj}}$ is the patents of firm *j* at time t; $Duration_{tj}$ is the time variable vector that is used to capture the time trend in the model (e.g. *Duration* = 0,1, 2, 3, …); ε_{tj} is the level-1 random error, which represents variance across time, and it is assumed to be distributed normally, with a zero mean and variance of σ^2; The intercept π_{oj} and slope π_{1j} can vary between firms, so that each INV is allowed to has a unique developing trend line.

We allow each firm to have a separate intercept and slope. The firm-specific intercepts and slopes become (unobservable) outcomes in the Level-2 models:

$$\pi_{oj} = b_0 + b_1 Controls_j + \mu_{oj} \qquad (2)$$

$$\pi_{1j} = \gamma_1 + \gamma_2 Controls_j + \mu_{1j} \qquad (3)$$

where b_0 and γ_1 are the mean values of the intercept and slope, respectively; the error terms μ_{oj} and μ_{1j} define how the slopes and intercept vary between firms, and both of them are assumed to be normally distributed with a mean of zero and variance τ_π and τ_β, respectively; $Controls_j$ represents the control variables mentioned above.

Substituting the Level 2 into the Level 1 model, we obtain the reduced form of the model in Equation (4):

$$A_{Patents_{tj}} = b_0 + \gamma_1 Duration_{tj} + b_1 Controls_j + \gamma_2 Controls_j Duration_{tj} + \mu_{1j} Duration_{tj} + \mu_{oj} + \varepsilon_{tj} \qquad (4)$$

where the regression coefficients b_0 and γ_1 are fixed parameters having specific values, shared between all observations, and modeled as fixed effects. The terms included μ_{1j}, μ_{oj} and ε_{tj} are unobserved and modeled as random effects. Thus, the above LGCM model representation is divided into two parts: (1) the fixed part, which contains only fixed coefficients and provides the regression line, and (2) the random part, which contains only random effect terms that indicate how observations vary around the line.

In order to investigate how multi-cluster diversity moderates INVs' innovation performance, we then set the following model:

$$A_{Patents_{tj}} = b_0 + \gamma_1 Duration_{tj} + b_1 Controls_j + b_2 MCD_k + b_3 DMCD_k + \gamma_2 Controls_j Duration_{tj} + \mu_{1j} Duration_{tj} + \mu_{oj} + \varepsilon_{tj} \qquad (5)$$

where the index ***k*** denotes city; *MCD* represents the moderator (*MCD*), and *DMCD* represents the interaction term — namely *Duration* × *MCD*.

5. Regression Results And Robustness Tests

5.1. Regression Results

As shown in M1-M3 (Table 2), three different models were estimated: the null model, a model with random intercept, and a model with random intercept and slope. Although researchers usually ignore those models, analyzing the results of those models may help to accept or reject the research hypotheses. According to the model fit statistics, the information criterion (AIC, BIC) suggests that M3 is better than the other two models, and the LR test in M1 underpins the choice of multilevel modeling rather than applying OLS methods. However, it is not sufficient to verify the hypothesis based only on the results in the above model; it is also necessary to incorporate covariates and the interaction terms that we are interested in in the model, as shown in M4 and M5.

Regarding the fixed effects component in Table 2, the coefficients of *Duration* are positive and significant at the 1 percent level. In M4, for example, the coefficient of *Duration* is 0.015 and significant at 1 percent level, this suggests that the average level of INVs' innovation performance increases at a rate of 1.5 percent per year. This enables us to conclude that INVs' innovation performance follows a linear trend throughout time. Thus, Hypothesis 1 receives support. At the random effects component in Table 2, regarding the statistical significance of these variances, the estimations of σ^2, τ_π, and τ_β are higher than their respective standard errors, this suggests that there is significant variation in the annual innovation performance across time, and between firms. That is, different INVs have different innovation trajectories over time. Differences in firm characteristics (such as resources, capabilities, etc.) are important factors that lead to differences in its innovation performance.

Furthermore, the present study is also interested in the interaction effects, which are tested in M5, on firm innovation over time. As shown in M5 (Table 2), the coefficient of *DMCD* is 0.006 and significant at the 5 percent level, which denotes that the multi-cluster diversity plays an important role in launching and growing INVs, especially on its innovation.

Table 2. Multilevel regression analysis results

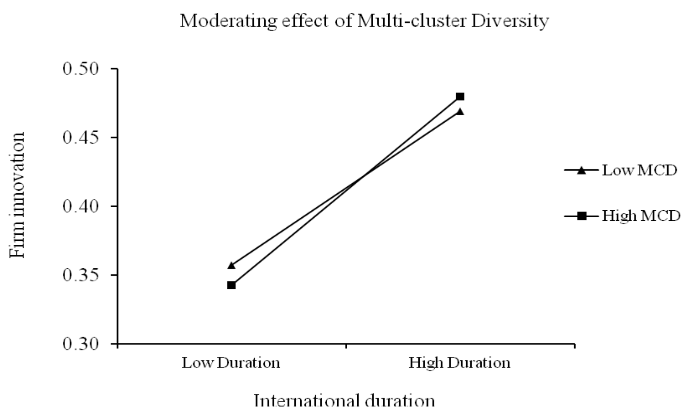
	M1	M2	M3	M4	M5
VARIABLES	A_Patents	A_Patents	A_Patents	A_Patents	A_Patents
Fixed effects	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Duration		0.024***	0.022***	0.015***	0.014***
		(0.001)	(0.001)	(0.001)	(0.001)
Size				0.033***	0.033***
				(0.003)	(0.003)
Lev				-0.005	-0.005
				(0.009)	(0.009)
lnFixed				0.002	0.002
				(0.002)	(0.002)
MCD				-0.037***	-0.027***
				(0.009)	(0.010)
lnGDP				0.023***	0.024***
				(0.005)	(0.005)
lnFDI				0.011**	0.010**
				(0.004)	(0.004)
RFIF				-0.017***	-0.017***
				(0.003)	(0.003)
lnIndInt				0.002	0.003
				(0.003)	(0.003)
Interaction Terms					

DMCD					0.006**
					(0.003)
Intercept	0.093***	0.001	0.006	-0.645***	-0.668***
	(0.004)	(0.005)	(0.004)	(0.056)	(0.057)
Random effects					
Firm-level Variance					
Intercept	0.289	0.288	0.189	0.189	0.189
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Duration			0.071	0.070	0.070
			(0.001)	(0.001)	(0.001)
% of total variance	46.31	44.68	49.06	48.96	48.96
Residual	0.335	0.329	0.270	0.270	0.270
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
% of total variance	53.69	55.32	49.94	51.04	51.04
Model fit statistics					
Log likelihood	-15,666.32	-15,118.27	-10,889.27	-10,699.92	-10,697.49
AIC	31,338.64	30,244.54	21790.54	21,427.84	21,424.98
BIC	31,363.93	30,278.26	21,841.12	21,545.87	21,551.44
LR test	10,654.53 *** (vs OLS)	1,096.11 *** (vs M1)	8,458*** (vs M2)	378.69*** (vs M3)	4.86** (vs M4)

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Figure 3 shows how the regional multi-cluster diversity moderate the relationship between international duration and INV's innovation. As internationalization duration increases, INVs located in cities with a high level of MCD will experience a higher level of innovation performance, compared with their peers who are located in cities with a low level of MCD. Therefore, Hypothesis 2 receives support.

Figure 3. Moderating effects of MCD on INV's innovation



5.2. Endogeneity Tests and Robustness Tests

5.2.1. Instrumental Variables Method

In this subsection, we consider the potential endogeneity caused by $Cov(Duration_{tj}, \varepsilon_{tj}) \neq 0$, which is also called Level 1 endogeneity. This endogeneity may be caused by the unobserved heterogeneity or omitted variable bias, we thus apply the instrumental variable (IV) method to solve this endogeneity problem.

We need to find a variable that is closely related to the *Duration* of the firm, but is independent of the firm's innovation as an instrumental variable. Following Angrist and Pischke (2014)'s suggestion, a good instrumental variable, in our paper, should satisfy the following two conditions (Clarke and Matta 2018):

1. There is a significant correlation between instrumental variable (IV) and independent variable (X). Mathematically, $\mathbf{cov}(X, IV) \neq \mathbf{0}$.

2. Instrumental variable (IV) should be uncorrelated with the error term, namely exogeneity. Mathematically, $\mathbf{cov}(IV, \varepsilon) = \mathbf{0}$.

The internationalization process of a firm is usually influenced by the geographical or/and economic factors of the region. In this paper, we take both kinds of those factors into consideration when we constructed the instrumental variables. Geographically, we used the reciprocal of the city's distance to Hong Kong as our first IV (*DisHK*). Condition (I) is easy to demonstrate. Since the implementation of China's reform and opening-up policy, Hong Kong, as one of the main gateways for China's foreign trade, has an important impact on the internationalization of firms in neighboring provinces, especially for Guangdong province. That is, the closer the city where the firm is located is to Hong Kong, the closer the firm will be to the foreign market. In this case, firms' willingness to export will be higher and they will be more likely to enter the international market earlier, compared with inland firms. Furthermore, geographic distance usually does not directly affect firm innovation, which means that condition (**II**) is satisfied.

We also consider the economic factors that facilitate the export and import trade of firms, namely trade intermediaries. Generally, trade intermediaries, also known as middlemen or intermediaries, refer to entities or individuals that facilitate the exchange of goods or services between buyers and sellers in a market (Ahn et al. 2011). In the present paper, we argue that trade intermediaries have a positive impact on the export and import behavior of firms. Specifically, trade intermediaries can not only influence firms' export decisions by reducing information frictions and export fixed costs, but can also help firms whose productivity has not yet reached the export threshold to enter international markets (Lu et al. 2017; Amiti and Weinstein 2011). Therefore, due to insufficient resources to establish distribution networks abroad, INVs often use the channels of trade intermediaries to explore international markets; that is, the extent to which trade intermediaries are active in the city's export and import trade can influence the export behavior of INV. Although trade intermediaries can reduce certain costs in the export process, they are not directly involved in the innovation process of the firm. In other words, trade intermediaries usually act as a platform. Through this platform, firms, especially starts-up, can enter the international market with lower information friction costs and export fixed costs, instead of conducting innovation (Lu et al. 2017; Akerman 2018).

Following Defever et al. (2020), we used the share of total annual exports of all trade intermediaries in the city's total annual exports and imports to measure the activeness of trade intermediaries in a city. Firstly, based on the firm-level customs data, we identify trade intermediaries by determining whether a firm's name contains a specific keyword. For example, the "trade," "economic and trade," "import and export," "commerce," etc.[2] Secondly, we identify the location of each trade intermediary through its address information and then count the total annual imports and exports of all trade intermediaries in a city. Finally, we use the share of total annual exports of all trade intermediaries in the city's total annual exports and imports as the second instrumental variable (*Intermediary*).

As both instrumental variables (*DisHK* and *intermediary*) are closely related to the firm's duration and not to the firm's innovation, we first simplify the multilevel model at level 1 as follows:

$$A_{Patents_{tj}} = \pi_{oj} + \pi_{1j} Duration_{tj} + \varepsilon_{tj} \qquad \textbf{(6)}$$

where the intercept π_{oj} and slope $\pi_{\mathbf{1}j}$ can vary between firms, as indicated by $\boldsymbol{j}$ subscript. Although we have controlled for firm-level and city-level variables, potential endogeneity caused by omitted variables may still exist. Ignoring such variables will make the error term in Level 1 correlated with the independent variable (*Duration*); that is, $\mathbf{cov}(\varepsilon_{tj}, Duration_{tj}) \neq \mathbf{0}$. The usual improvement method is to employ a two-stage least squares (2SLS) regression model to account for potential endogeneity (Hashai 2011).

The first-stage models are set as follows:

$$Duration_{tj} = \alpha_{oj} + \beta_{jk}\mathbf{IV}_{jk} + Controls_j + \delta_t + \varphi_j + \theta_{ij} \quad \textbf{(7)}$$

where IV_{jk} refers to the two instrumental variables mentioned above, namely *DisHK* and *Intermediary*; α_{oj} refers to constant term; δ_t and φ_j are the year fixed effect and industry fixed effect; and θ_{ij} is the error term. After obtaining the predicted value for endogeneity variables (*Duration*) by Equation (7), we

2. All keywords are originally in Chinese language. It is a traditional and common practice to identify firms whose names contain these keywords in customs data as trade intermediaries. Since we cannot observe the behavior of each firm importing and exporting products through trade intermediaries, we have kept the activity of trade intermediaries at the city level. In addition, since the research period of this paper is 2000–2013, which is inconsistent with the rise of new trade intermediaries such as cross-border e-commerce, this paper does not consider new trade intermediaries.

replace the *Duration* in Equation (6) with the predicted value and then perform the regression:

$$A_{Patents_{tj}} = \pi_{oj} + \pi_{1j}\widehat{Duration}_{tj} + \varepsilon_{tj} \qquad \textbf{(8)}$$

where $\widehat{Duration}_{tj}$ is the predicted value for endogeneity variables (*Duration*). Other variables have the same meaning as in Equation (6).

As shown in Table 3, for the first stage of the 2SLS model, we estimated Equation (7), and the regression results are shown in Model 6 (*DisHK*) and Model 8 (*Intermediary*). As predicted above, there are statistically significant correlations (p<0.01) between IVs (*DisHK* and *Intermediary*) and independent variable. The F-statistic of *DisHK* obtained from the first-stage regression is 77.457 and that of *Intermediary* is 43.895, which means that both *DisHK* and *Intermediary* are not weak instrumental variables in the present paper; that is, both *DisHK* and *Intermediary* are qualified IVs. As shown in M7 and M9, the coefficient of *Duration* is still positive and significant at the 1 percent level, and DMCD is also positive and significant. The estimation results of the instrumental variable method show that our findings obtained above remain robust.

Table 3.Endogeneity tests of multilevel regression analysis

	M6	M7	M8	M9	M10
VARIABLES	Duration	A_Patents	Duration	A_Patents	A_Patents
Fixed effects	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
DisHK	-1392.306***				
	(161.231)				
Intermediary			1.336***		
			(0.209)		
Duration		0.023***		0.022***	0.012***
		(0.002)		(0.002)	(0.002)
Interaction Terms					
DMCD		0.007*		0.008**	0.005*
		(0.004)		(0.004)	(0.003)
Intercept		-0.535***		-0.539***	-0.498***
		(0.055)		(0.055)	(0.099)
Controls	Yes	Yes	Yes	Yes	Yes
$\overline{\text{Controls}}$					Yes
Year fixed effect	Yes		Yes		
Industry fixed effect	Yes		Yes		
Random effects					
Firm-level Variance					
% of total variance		55.65		55.57	49.06
Residual					
% of total variance		44.35		44.43	50.94
Model fit statistics					
F-statistics	77.457		43.895		

Log likelihood			
AIC	-10,098.54	-10114.41	-10,670.15
BIC	20,227.09	20258.81	21,384.3
LR test	20,353.55	20385.27	21,569.78

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; In model 7 and 9, the *Duration* is the fitted value after instrumentalizing. Controls are the same as Table 2 has, which contains moderator (MCD). In model 6 and 8, *DisHK* and *Intermediary* are instrumental variables mentioned above. $\overline{\mathbf{Controls}}$ is the cluster mean of the Controls mentioned in Table 2.

5.2.2. Cross-level Endogeneity Tests: Correlated Random Effects (CRE) Modeling

Researchers who apply multilevel models in their works usually assume that the random effects (error terms) are uncorrelated with the regressors, but this assumption is often violated, leading to a potential endogeneity problem. Here, we call such endogeneity cross-level endogeneity, namely

$$Cov\left(Duration_{tj}, \gamma_{0j}\right) \neq 0$$

Antonakis et al. (2021) suggested that cross-level endogeneity can be relaxed by adding cluster means into the multilevel model. Mundlak (1978) and Snijders and Berkhof (2008) suggested that including the cluster means of a high level into the equation would be a straightforward solution to this endogeneity problem, and this method is also known as the correlated random effect (CRE) modeling approach. As Antonakis et al. (2021) explained, "the CRE model essentially unifies the RE and FE models." That is, compared with RE models and FE models, the CRE model has a more obvious advantage in dealing with the endogeneity problem mentioned above. Thus, we employ the CRE method to deal with potential cross-level endogeneity in our model (please see Antonakis et al. (2021) for detailed mathematical information about RE models and FE models).

Inspired by Antonakis et al. (2021), we set the CRE model as follows:

$A_{Patents_{tj}} = \pi_{oj} + \pi_{1j} Duration_{tj} + \varepsilon_{tj}$	$Level\ 1$
$\pi_{oj} = b_{00} + b_{o1} Controls_j + b_{o2} \overline{Controls}_j + \gamma_{oj}$ $\pi_{1j} = b_{10} + b_{11} \mathbf{Controls}_j + b_{12} \overline{Controls}_j + \gamma_{1j}$	$Level\ 2$

Substituting the Level 2 into the Level 1 model, we obtain the reduced form of the model:

$$A_{Patents_{tj}} = b_{00} + b_{o1} Controls_j + b_{o2} \overline{Controls}_j + b_{10} Duration_{tj} + b_{11} \mathbf{Controls}_j Duration_{tj} + b_{12} \overline{Controls}_j Duration_{tj} + \gamma_{oj} + \gamma_{1j} Duration_{tj} + \varepsilon_t$$

where $\overline{Controls}_j$ refers to the cluster mean of $Controls_j$; Other variables have the same meaning as in Equation (4). As shown in M10 (Table 3), the results are substantively similar to the baseline estimation and the bias is less pronounced, which shows that our findings still exist and remain robust.

5.2.3. Other Robustness Tests

The benchmark regression results may be affected by the measurement of key variables, so we changed the measurement of key variables for the sake of robustness.

First, changing the measurement of INVs. In this paper, an INV is defined as a firm with overseas sales that account for at least 20% of total sales within three years of its inception. To obtain robust results, we then redefined INV's measurement to include:

a. Firms with overseas sales that account for at least 30 percent of their total sales within three years of inception.

b. Firms with overseas sales that account for at least 25 percent of their total sales within two years of inception.

As shown in Table 4, M11 and M12 estimate definitions (a) and (b), respectively. The regression results show that the coefficient of *Duration* is significant and the direction is consistent with M5, which means that the previous conclusion is robust.

Second, changing the measurement of the dependent variable. Because a time lag exists between patent applications and formal authorization, we use the number of patents granted by the firm (*G_Patents*, the annual total number of granted patents for each firm) as the new independent variable. As shown in Table 4, M13 to M15 report the regression results successively, gradually applying the original INV's definition, definition (a), and definition (b) after changing the measurement of the dependent variable. The estimation results show that INVs' innovation performance follows a linear trend over time. The previous conclusions thus are robust.

Third, adding more control variables. we should also take into account the possible influence of individual clusters (IC) when we analyze MCD on the "duration-innovation" relationship. Ignoring this effect may affect the robustness of the estimation results. Therefore, we included individual cluster (IC) as control variables in the regression model to improve the robustness (Appendix 1 introduced the measurement steps for individual clusters). As shown in Table 4, M16 reports the regression results. The estimation results are substantively similar to the baseline estimation, which indicates that our findings still exist and remain robust.

Table 4. Other robustness tests of multilevel regression analysis

	M11	M12	M13	M14	M15	M16
VARIABLES	A_Patents	A_Patents	G_Patents	G_Patents	G_Patents	A_Patents
Fixed effects	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Duration	0.014***	0.011***	0.018***	0.017***	0.016***	0.015***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
DMCD	0.005**	0.008**	0.007**	0.007**	0.009**	0.006**
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
Intercept	-0.637***	-0.779***	-0.511***	-0.497***	-0.561***	-0.768***

	(0.056)	(0.081)	(0.051)	(0.051)	(0.071)	(0.236)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
IC						Yes
Random effects						
Firm-level Variance						
% of total variance	50.66	50.75	48.01	48.46	47.29	48.96
Residual						
% of total variance	49.34	49.25	51.99	51.54	52.71	51.04
Model fit statistics						
Log likelihood	-9285.323	-5167.093	-10811.92	-9433.694	-5007.403	-10686.573
AIC	18600.65	10364.19	21653.83	18897.39	10044.81	21433.15
BIC	18726.42	10480.82	21780.29	19023.16	10161.44	21686.06
Observations	32,359	17,600	33,881	32,359	17,600	33,881
Number of firms	5,487	2,922	5,744	5,487	2,922	5,744

Notes: Standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Controls are the same as Table 2 has, which contains moderator (MCD). IC denote the individual clusters mentioned in Appendix 1.

6. Discussion and Concluding Remarks

In the present article, we examined the dynamic changes of INV's innovation performance over time while testing for the role of regional multi-cluster diversity as a moderating mechanism. Based on 33,881 firm-year observations of 5,744 INVs that cover 21 cities in the Guangdong Province in China from 2000 to 2013, and by applying a latent growth curve model (LGCM), we found evidence that INVs' innovation follows a positive linear trend over time, and this relationship is positively moderated by regional multi-cluster diversity. Such results are stable and reliable after we conduct robustness tests.

Time, as the central issue of firm internationalization, is usually treated implicitly in the research on INVs' behaviors (Prashantham and Young 2011; Jones and Coviello 2005). Although prior INV studies have investigated various antecedents and driving factors that help explain the emergence of INVs and the performance of their internationalization, most of these studies were based on specific internationalization stages and ignored the subsequent dynamic changes of post-entry performance of INVs, especially in terms of innovation. In addition, INV's innovation will benefit from the external industrial cluster environment, but most studies have ignored the fact that clusters are often co-located. We fill the above research gaps by taking duration of internationalization and multiple cluster environments into consideration, as it can help us to understand the long-term performance of INVs in multiple cluster environments from a dynamic perspective since they entered the international market. Our findings provide theoretical and practical insights. For international entrepreneurship theory, our paper shows that INVs' innovation performance follows a positive linear tendency as internationalization duration increases. INVs will benefit from the positive microeconomic externalities derived from clusters. That is, for entrepreneurs, establishing their firms in multiple cluster environments will contribute to their early success when internationalized. In addition to being of academic interest, we argue that our study has relevance for policymakers. In a region where policymakers intend to promote international entrepreneurship and subsequent long-term success, they need to recognize that the impact of industrial clusters varies across time and firms due to the heterogeneity and life cycle of firms.

Meanwhile, there are certain limitations to this study that shed light on future research. The first limitation is that we have only examined what happens to manufacturing firms from China and we do not know whether our conclusions can be applied to other industries and countries. Secondly, although Delgado et al.'s novel clustering algorithm can effectively quantify industrial clusters, there may be alternative methods of doing this. Therefore, it is necessary to apply different industrial cluster indicators to test our findings again.

References:

Acedo, F. J., Coviello, N., and Agustí, M. (2021), "Caution ahead! The long-term effects of initial export intensity and geographic dispersion on INV development", *Journal of World Business*, 56(6): 101260.

Ahn, J., Khandelwal, A. K., and Wei, S.-J. (2011), "The role of intermediaries in facilitating trade", *Journal of International Economics*, 84(1): 73-85.

Aiken, L. S., West, S. G., and Reno, R. R. (1991), *Multiple Regression: Testing and Interpreting Interactions*. London: Sage Publications.

Akerman, A. (2018), "A theory on the role of wholesalers in international trade based on economies of scope", *Canadian Journal of Economics/Revue Canadienne d'Economique*, 51(1): 156-185.

Alon, I., Yeheskel, O., Lerner, M., and Zhang, W. (2013), "Internationalization of Chinese entrepreneurial firms", *Thunderbird International Business Review*, 55(5): 495-512.

Amdam, R. P., Lunnan, R., Bjarnar, O., and Halse, L. L. (2020), "Keeping up with the neighbors: The role of cluster identity in internationalization", *Journal of World Business*, 55(5): 101125.

Amiti, M. and Weinstein, D. E. (2011), "Exports and financial shocks", *Quarterly Journal of Economics*, 126(4): 1841-1877.

Angrist, J. D. and Pischke, J.-S. (2014), *Mastering 'Metrics: The Path from Cause to Effect*, Princeton, NJ: Princeton University Press.

Antonakis, J., Bastardoz, N., and Rönkkö, M. (2021), "On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations", *Organizational Research Methods*, 24(2): 443-483.

Autio, E., Sapienza, H. J., and Almeida, J. G. (2000), "Effects of age at entry, knowledge intensity, and imitability on international growth", *Academy of Management Journal*, 43(5): 909-924.

Azar, G. and Ciabuschi, F. (2017), "Organizational innovation, technological innovation, and export performance: The effects of innovation radicalness and extensiveness", *International Business Review*, 26(2): 324-336.

Barney, J. B. (2001), "Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view", *Journal of Management*, 27(6): 643-650.

Bathelt, H. and Li, P.-F. (2014), "Global cluster networks—foreign direct investment flows from Canada to China", *Journal of Economic Geography*, 14(1): 45-71.

Bathelt, H. and Li, P. (2020), "Processes of building cross-border knowledge pipelines", *Research Policy*, 49(3): 103928.

Bathelt, H., Malmberg, A., and Maskell, P. (2004), "Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation", *Progress in Human Geography*, 28(1): 31-56.

Bathelt, H. and Zhao, J. (2016), "Conceptualizing multiple clusters in mega-city regions: The case of the biomedical industry in Beijing", *Geoforum*, 75: 186-198.

Blesa, A. and Ripollés, M. (2021), "Critical capabilities for widening international new ventures' geographical scope", *European Management Journal*, 39(2): 201-213.

Boehe, D. M. and Jiménez, A. (2016), "How does the geographic export diversification–performance relationship vary at different levels of export intensity?", *International Business Review*, 25(6): 1262-1272.

Bronzini, R. and Piselli, P. (2016), "The impact of R&D subsidies on firm innovation", *Research Policy*, 45(2): 442-457.

Buccieri, D., Javalgi, R. G., and Cavusgil, E. (2020), "International new venture performance: Role of international entrepreneurial culture, ambidextrous innovation, and dynamic marketing capabilities", *International Business Review*, 29(2): 101639.

Capone, F. and Lazzeretti, L. (2018), "The different roles of proximity in multiple informal network relationships: Evidence from the cluster of high technology applied to cultural goods in Tuscany", *Industry and Innovation*, 25(9): 897-917.

Cavusgil, S. T. and Knight, G. (2015), "The born global firm: An entrepreneurial and capabilities perspective on early and rapid internationalization", *Journal of International Business Studies*, 46(1): 3-16.

Clarke, D. and Matta, B. (2018), "Practical considerations for questionable IVs", *The Stata Journal*, 18(3): 663-691.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003), *Applied Multiple Correlation/Regression Analysis for the Social Sciences*. Hillsdale, NJ: Erlbaum.

Cohen, W. M. and Levinthal, D. A. (1990), "Absorptive capacity: A new perspective on learning and innovation", *Administrative Science Quarterly*, 35(1): 128-152.

Coviello, N. (2015), "Re-thinking research on born globals", *Journal of International Business Studies*, 46(1): 17-26.

Coviello, N. E. and Munro, H. J. (1995), "Growing the entrepreneurial firm: Networking for international market development", *European Journal of Marketing*, 29(7): 49-61.

Crespo, N. F., Simoes, V. C., and Fontes, M. (2020), "Competitive strategies and international new ventures' performance: Exploring the moderating effects of internationalization duration and preparation", *Brq-Business Research Quarterly*, 23(2): 120-140.

D'Angelo, A., Ganotakis, P., and Love, J. H. (2020), "Learning by exporting under fast, short-term changes: The moderating role of absorptive capacity and foreign collaborative agreements", *International Business Review*, 29(3): 101687.

Defever, F., Imbruno, M., and Kneller, R. (2020), "Trade liberalization, input intermediaries and firm productivity: Evidence from China", *Journal of International Economics*, 126: 103329.

Delgado, M. (2020), "The co-location of innovation and production in clusters", *Industry and Innovation*, 27(8): 842-870.

Delgado, M., Porter, M. E., and Stern, S. (2016), "Defining clusters of related industries", *Journal of Economic Geography*, 16(1): 1-38.

Demirkan, I. (2018), "The impact of firm resources on innovation", *European Journal of Innovation Management*, 21(4): 672-694.

Dunning, J. H. (1998), "Location and the multinational enterprise: A neglected factor?", *Journal of International Business Studies*, 29(1): 45-66.

Duranton, G. and Puga, D. (2000), "Diversity and specialisation in cities: Why, where and when does it matter?", *Urban Studies*, 37(3): 533-555.

Felin, T., Kauffman, S., and Zenger, T. (2023), "Resource origins and search", *Strategic Management Journal*, 44(6): 1514-1533.

Fernandes, A. P. and Tang, H. (2014), "Learning to export from neighbors", *Journal of International Economics*, 94(1): 67-84.

Fernhaber, S. A., Gilbert, B. A., and McDougall, P. P. (2008), "International entrepreneurship and geographic location: An empirical examination of new venture internationalization", *Journal of International Business Studies*, 39(2): 267-290.

Fernhaber, S. A. and Li, D. (2010), "The impact of interorganizational imitation on new venture international entry and performance", *Entrepreneurship Theory and Practice*, 34(1): 1-30.

Gerschewski, S., Lew, Y. K., Khan, Z., and Park, B. I. (2018), "Post-entry performance of international new ventures: The mediating role of learning orientation", *International Small Business Journal*, 36(7): 807-828.

Gerschewski, S., Rose, E. L., and Lindsay, V. J. (2015), "Understanding the drivers of international performance for born global firms: An integrated perspective", *Journal of World Business*, 50(3): 558-575.

Grant, R. M. (1996), "Toward a knowledgebased theory of the firm", *Strategic Management Journal*, 17(S2): 109-122.

Guan, J. C. and Yan, Y. (2016), "Technological proximity and recombinative innovation in the alternative energy field", *Research Policy*, 45(7): 1460-1473.

Hashai, N. (2011), "Sequencing the expansion of geographic scope and foreign operations by "born global" firms", *Journal of International Business Studies*, 42(8): 995-1015.

Hashai, N. and Almor, T. (2004), "Gradually internationalizing 'born global' firms: an oxymoron?", *International Business Review*, 13(4): 465-483.

Hayes, A. F. (2017), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach*. New York: Guilford Publications.

Henn, S. and Bathelt, H. (2018), "Cross-local knowledge fertilization, cluster emergence, and the generation of buzz", *Industrial and Corporate Change*, 27(3): 449-466.

Hennart, J. F. (2014), "The accidental internationalists: A theory of born globals", *Entrepreneurship Theory and Practice*, 38(1): 117-135.

Huang, S., Ye, G. Y., and Shen, S. Q. (2020), "Initial entry speed and international performance of INVs: The moderating roles of executive's learning orientation and sub-national institutions", *Chinese Management Studies*, 14(1): 297-323.

Johanson, J. and Vahlne, J.-E. (2015), "The Uppsala internationalization process model revisited: From liability of foreignness to liability of outsidership", In: P. J. Buckley and P. Ghauri (Eds.), *International Business Strategy: Theory and Practice*, pp. 33-59. London: Routledge.

Jones, M. V. and Coviello, N. E. (2005), "Internationalisation: Conceptualising an entrepreneurial process of behaviour in time", *Journal of International Business Studies*, 36(3): 284-303.

Kim, D. and Cavusgil, E. (2020), "Antecedents and outcomes of digital platform risk for international new ventures' internationalization", *Journal of World Business*, 55(1): 101021.

Knight, G. A. and Cavusgil, S. T. (2004), "Innovation, organizational capabilities, and the born-global firm", *Journal of International Business Studies*, 35(2): 124-141.

Knight, G. A. and Liesch, P. W. (2016), "Internationalization: From incremental to born global", *Journal of World Business*, 51(1): 93-102.

Kurokawa, S., Iwata, S., and Roberts, E. B. (2007), "Global R&D activities of Japanese MNCs in the US: A triangulation approach", *Research Policy*, 36(1): 3-36.

Laursen, K. and Salter, A. J. (2014), "The paradox of openness: Appropriability, external search and collaboration", *Research Policy*, 43(5): 867-878.

Lee, I. H., Paik, Y., and Uygur, U. (2016), "Does gender matter in the export performance of international new ventures? Mediation effects of firm-specific and country-specific advantages", *Journal of International Management*, 22(4): 365-379.

Levie, J. and Lichtenstein, B. B. (2010), "A terminal assessment of stages theory: Introducing a dynamic states approach to entrepreneurship", *Entrepreneurship Theory and Practice*, 34(2): 317-350.

Lewin, A. Y., Massini, S., and Peeters, C. (2011), "Microfoundations of internal and external absorptive capacity routines", *Organization Science*, 22(1): 81-98.

Li, P.-F. and Bathelt, H. (2018), "Location strategy in cluster networks", *Journal of International Business Studies*, 49(8): 967-989.

Liesch, P., Weerawardena, J., Mort, G. S., Knight, G., and Kastelle, T. (2007), "Editors' Introduction to 'The Early and Rapid Internationalization of the Firm' Special Issue", *Journal of World Business*, 42(3): 232-235.

Loane, S. and Bell, J. (2006), "Rapid internationalisation among entrepreneurial firms in Australia, Canada, Ireland and New Zealand: An extension to the network approach", *International Marketing Review*, 23(5): 467-485.
Lu, J., Lu, Y., Sun, Y., and Tao, Z. (2017), "Intermediaries, firm heterogeneity and exporting behaviour", *The World Economy*, 40(7): 1381-1404.
Martin, S. L., Javalgi, R. G., and Ciravegna, L. (2020), "Marketing capabilities and international new venture performance: The mediation role of marketing communication and the moderation effect of technological turbulence", *Journal of Business Research*, 107: 25-37.
McDougall, P. P., Shane, S., and Oviatt, B. M. (1994), "Explaining the formation of international new ventures: The limits of theories from international business research", *Journal of Business Venturing*, 9(6): 469-487.
Mudambi, R. and Zahra, S. A. (2007), "The survival of international new ventures", *Journal of International Business Studies*, 38(2): 333-352.
Mukherjee, A., Singh, M., and aldokas, A. (2017), "Do corporate taxes hinder innovation?", *Journal of Financial Economics*, 124(1): 195-221.
Mundlak, Y. (1978), "On the pooling of time series and cross section data", *Econometrica*, 46(1): 69-85.
Nonaka, I. and Takeuchi, H. (1995), *The Knowledge-Creating Company*, New York: Oxford University Press.
Nuruzzaman, N., Singh, D., and Pattnaik, C. (2019), "Competing to be innovative: Foreign competition and imitative innovation of emerging economy firms", *International Business Review*, 28(5): 101490.
Oviatt, B. M. and McDougall, P. P. (1994), "Toward a theory of International New Ventures", *Journal of International Business Studies*, 25(1): 45-64.
Øyna, S., Almor, T., Elango, B., and Tarba, S. Y. (2018), "Maturing born globals and their acquisitive behaviour", *International Business Review*, 27(3): 714-725.
Pangarkar, N. and Yuan, L. (2021), "Industry life cycle, geographic diversification and performance of international new ventures", *Multinational Business Review*, 30(1): 62-80.
Paul, J. and Rosado-Serrano, A. (2019), "Gradual internationalization vs born-global/international new venture models: A review and research agenda", *International Marketing Review*, 36(6): 830-858.
Peng, M. W. (2001), "The resource-based view and international business", *Journal of Management*, 27(6): 803-829.
Penrose, E. T. (1959), *The Theory of the Growth of the Firm*, New York: Oxford University Press.
Pittaway, L., Robertson, M., Munir, K., Denyer, D., and Neely, A. (2004), "Networking and innovation: A systematic review of the evidence", *International Journal of Management Reviews*, 5(34): 137-168.
Porter, M. E. (1996), "Competitive advantage, agglomeration economies, and regional policy", *International Regional Science Review*, 19(1-2): 85-90.
Porter, M. E. (2000), "Location, competition, and economic development: Local clusters in a global economy", *Economic Development Quarterly*, 14(1): 15-34.
Posen, H. E. and Chen, J. S. (2013), "An advantage of newness: Vicarious learning despite limited absorptive capacity", *Organization Science*, 24(6): 1701-1716.
Posen, H. E., Keil, T., Kim, S., and Meissner, F. D. (2018), "Renewing research on problemistic search—A review and research agenda", *Academy of Management Annals*, 12(1): 208-251.
Prashantham, S. and Young, S. (2011), "Post-entry speed of International New Ventures", *Entrepreneurship Theory and Practice*, 35(2): 275-292.
Qiu, L., Jie, X., Wang, Y., and Zhao, M. (2020), "Green product innovation, green dynamic capability, and competitive advantage: Evidence from Chinese manufacturing enterprises", *Corporate Social Responsibility and Environmental Management*, 27(1): 146-165.
Raudenbush, S. W. and Bryk, A. S. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*. London: Sage.
Rialp, A., Rialp, J., and Knight, G. A. (2005), "The phenomenon of early internationalizing firms: What do we know after a decade (1993–2003) of scientific inquiry?", *International Business Review*, 14(2): 147-166.

Ritala, P. and HurmelinnaLaukkanen, P. (2013), "Incremental and radical innovation in coopetition—The role of absorptive capacity and appropriability", *Journal of Product Innovation Management*, 30(1): 154-169.

Romanello, R. and Chiarvesio, M. (2017), "Turning point: When born globals enter post-entry stage", *Journal of International Entrepreneurship*, 15(2): 177-206.

Salomon, R. M. and Shaver, J. M. (2005), "Learning by exporting: New insights from examining firm innovation", *Journal of Economics & Management Strategy*, 14(2): 431-460.

Salvador Federico, J. (2016), *Advances in the Research on Young Business Growth: Persistence, Impact on Profits and the Moderating Role of Firm Age*. Barcelona: Universitat Autònoma de Barcelona.

Sapienza, H. J., Autio, E., George, G., and Zahra, S. A. (2006), "A capabilities perspective on the effects of early internationalization on firm survival and growth", *Academy of Management Review*, 31(4): 914-933.

Sirmon, D. G., Hitt, M. A., Ireland, R. D., and Gilbert, B. A. (2011), "Resource orchestration to create competitive advantage: Breadth, depth, and life cycle effects", *Journal of Management*, 37(5): 1390-1412.

Snijders, T. A. B. and Berkhof, J. (2008), "Diagnostic checks for multilevel models", In: J. de Leeuw and E. Meijer (Eds.), *Handbook of Multilevel Analysis*, pp. 141-175. New York: Springer.

Tuomisalo, T. and Leppaaho, T. (2019), "Learning in international new ventures: A systematic review", *International Business Review*, 28(3): 463-481.

Turkina, E. (2018), "The importance of networking to entrepreneurship: Montreal's artificial intelligence cluster and its born-global firm element AI", *Journal of Small Business and Entrepreneurship*, 30(1): 1-8.

Vanninen, H., Kuivalainen, O., and Ciravegna, L. (2017), "Rapid multinationalization: Propositions for studying born micromultinationals", *International Business Review*, 26(2): 365-379.

Wang, Q.-J., Feng, G.-F., Wang, H.-J., and Chang, C.-P. (2021), "The impacts of democracy on innovation: Revisited evidence", *Technovation*, 108: 102333.

Wennberg, K. and Lindqvist, G. (2008), "The effect of clusters on the survival and performance of new firms", *Small Business Economics*, 34(3): 221-241.

Wernerfelt, B. (1984), "A resourcebased view of the firm", *Strategic Management Journal*, 5(2): 171-180.

Westhead, P., Wright, M., and Ucbasaran, D. (2001), "The internationalization of new and small firms: A resource-based view", *Journal of Business Venturing*, 16(4): 333-358.

Wu, A. and Voss, H. (2015), "When does absorptive capacity matter for international performance of firms? Evidence from China", *International Business Review*, 24(2): 344-351.

Zahra, S. A. (2005), "A theory of international new ventures: A decade of research", *Journal of International Business Studies*, 36(1): 20-28.

Zahra, S. A. and Das, S. R. (1993), "Building competitive advantage on manufacturing resources", *Long Range Planning*, 26(2): 90-100.

Zahra, S. A., Zheng, C., and Yu, J. (2018), "Learning advantages of newness: A reconceptualization and contingent framework", *Journal of International Entrepreneurship*, 16(1): 12-37.

Zander, I., McDougall-Covin, P., and Rose, E. L. (2015), "Born globals and international business: Evolution of a field of research", *Journal of International Business Studies*, 46(1): 27-35.

Zeng, S. X., Xie, X. M., and Tam, C. M. (2010), "Relationship between cooperation networks and innovation performance of SMEs", *Technovation*, 30(3): 181-194.

Zhang, K. H. and Song, S. (2001), "Promoting exports: The role of inward FDI in China", *China Economic Review*, 11(4): 385-396.

Zhang, X., Wu, W., Zhou, Z., and Yuan, L. (2020), "Geographic proximity, information flows and corporate innovation: Evidence from the high-speed rail construction in China", *Pacific-Basin Finance Journal*, 61: 101342.

Zheng, G., Wang, S., and Xu, Y. (2018), "Monetary stimulation, bank relationship and innovation: Evidence from China", *Journal of Banking & Finance*, 89: 237-248.

Zhou, L. X., Barnes, B. R., and Lu, Y. A. (2010), "Entrepreneurial proclivity, capability upgrading and performance advantage of newness among international new ventures", *Journal of International Business Studies*, 41(5): 882-905.

Appendix 1: Quantifying Clusters

We provide this appendix for those scholars who are interested in how to quantify clusters.

In order to quantify clusters comprehensively, Delgado et al. (2016) has proposed a novel approach for defining clusters by accounting for multi types of inter-industry linkages (skill, technology, supply, demand, and/or other linkages). We briefly repeated Delgado et al.'s cluster algorithm, which includes six steps: (1) creating the five fundamental similarity matrices M_{ij} to capture the relatedness between any two industries – those matrices are created respectively according to five essential indicators of the industry: employment (*LC-Emp*), establishments (*LC-Est*), the input-output (*IO*), labour occupation (*OCC*) links, and the co-agglomeration index (*COI*); (2) determining broad parameter choices β, namely suitable range for the number of clusters, according to the Chinese context; (3) based on previous similarity matrices M_{ij} and parameter choices , obtaining all cluster configurations *Cs* ($Cs=F(M_{ij},\beta)$) via clustering functions (*kmean*, *kmedian,* and *Hierarchical-Ward's)*; (4) calculating validation scores (VS) for each cluster configuration *C* and identifying the candidate configuration *C**, the one with the highest score; (5) assessing every cluster *c* in *C**; and (6) comparing all *c* in candidate configuration *C** to determine the finalized set of cluster definition *C***. In addition, Delgado et al. (2016) proposed expert judgment for some individual clusters, but we were unable to repeat this step because it has too much subjectivity and lacks criteria for selecting experts. Using the China Industrial Enterprises Database, we quantify clusters in Guangdong with six steps above, and more details are presented below.

Step 1: Similarity matrices

By grouping related industries into clusters, we defined a similarity matrix M_{ij} that captures the relatedness between any pair of industries i and j based on one indicator. In this step, we utilized five indicators: employment (*LC-Emp*$_i$), number of establishments (*LC-Est*$_{ij}$), the co-agglomeration index (*COI*$_{ij}$), measures of shared labor requirements (*OCC*$_{ij}$) and measures of buyers-supplier linkages (*IO*$_{ij}$). Table A1 shows the descriptive statistics of the five similarity matrices, plus one LC-IO-OCC$_{ij}$ matrix, which is the average value of matrixes *LC-Emp*$_i$, *LC-Est*$_{ij}$, *IO*$_{ij}$, and *OCC*$_{ij}$.

Table A1. Descriptive statistics for similarity matrices

	Mean	St.Dev	min	Max	Median	p99
LC-Emp$_{ij}$	.204	.204	-.267	.911	.171	.727
LC-Est$_{ij}$	.291	.239	-.438	.925	.277	.802
COI$_{ij}$	.013	.032	-.036	.860	.007	.1440
OCC$_{ij}$	.133	.290	-.038	1.000	.009	1.000
IO$_{ij}$	.011	.027	0.000	.597	.003	.151
LC-IO-OCC$_{ij}$	0.000	.704	-1.516	6.400	-.128	2.196

Step 2: Broad parameter choices β

The parameter choice β is set to decide the number of clusters, namely how many groups those investigated industries should be divided into. In the article of Delgado et al. (2016), 778 NAICS-6 industries in the United State are analyzed, and the range of β is 30-60. Due to practical constraints, we only studied 162 NAICS-4 industries, which is fewer than 778, and we thus set the value of 10-50 for β initially. After running the cluster function, we found that a value of more than 47 could have invalid results. Thus, we ultimately set the value for cluster function between 10 and 47.

Step 3: Clustering function

With the inputs of six similarity matrices in Step 1 and the parameter β, we can obtain all possibilities of cluster configurations via two main cluster functions ($Cs=F(M_{ij},\beta)$): the *hierarchical* function (with Ward's linkage) and centroid-based clustering functions (*kmean* and *kmedian*). Due to the limited data used to calculate the co-agglomeration index (COI_{ij}) and the shared labor requirements (OCC_{ij}), cluster functions produce less *Cs* than other similarity matrices.

We finally obtained 588 cluster configurations, and the results are presented in Table A2.

Table A2. Examples of cluster configurations generated

Similarity matrix M_{ij}	Parameter choices β		Clustering function C=F(M_{ij},β)	Number of *Cs*
	Number of clusters (numc)	Data		
LC-Employment (LC-Emp_{ij})	10-47	Raw	Hierarchical-Ward's	38
		Raw/Rst[a]	Kmean	76
		Raw/Rst	Kmedian	76
LC-Establishments (LC-Est_{ij})	10-47	Raw	Hierarchical-Ward's	38
		Raw/Rst	Kmean	76
		Raw/Rst	Kmedian	76
Labor occupation (Occ_{ij})	10-47	Raw	Hierarchical-Ward's	21
		Raw/Rst	Kmean	42
		Raw/Rst	Kmedian	42
IO (IO_{ij})	10-47	Raw	Hierarchical-Ward's	30
COI(COI_{ij})	10-47	Raw	Hierarchical-Ward's	38
LC-IO-Occ_{ij}	10-47	Raw	Hierarchical-Ward's	38

a. The underlaying data is either untransformed (Raw) or row-standardized (Rst).

Step 4: Performance scores for each C

After generating 588 Cs, we calculated validation scores (VS) for each *C* through capturing the extent to which individual clusters and industries have high within-cluster relatedness (WCR) compared to between-cluster relatedness (BCR) with other clusters. The VSs of each configuration *C* include two dimensions: *VS-Cluster*, indicating whether *individual clusters* in *C* are meaningfully different from each other, and *VS-Industry*, assessing the fit of i*ndividual industries* within their own clusters. See Table A3 for sub-scores and Table A4 for selected similarity matrices with the final score.

Table A3. Descriptive statistics: Validation scores for cluster configurations (Number of C_s=588)

	Description	Mean	St.Dev	min	max	p99
vs_clus	% of clusters with high WCRc (average of VS-Cluster sub-scores)	68.8	10.1	36.7	84.1	83.7
vs_ind	% of industries with high WCRic (average of VS-Industry sub-scores)	68.9	10.2	32.9	84.2	82.9
vs	Average VS-Cluster and VS-Industry	68.8	9.6	35.2	84.1	82.1
vs_clus_emp	% of clusters with high $WCR^{LC\text{-}Emp}$	86.6	7.2	65.2	100	100
vs_clus_est	% of clusters with high $WCR^{LC\text{-}Est}$	84.3	7.4	63.3	100	100
vs_clus_io	% of clusters with high $WCR^{LC\text{-}IO}$	80	17.1	40	100	100
vs_clus_occ	% of clusters with high $WCR^{LC\text{-}Occ}$	82.4	16.4	35.9	100	100
vs_ind_emp	% of industries with high $WCR^{LC\text{-}Emp}$	91.6	7.9	53.7	100	99.4
vs_ind_est	% of industries with high $WCR^{LC\text{-}Est}$	90.1	7.9	48.8	99.4	99.4
vs_ind_io	% of industries with high $WCR^{LC\text{-}IO}$	80.9	15.1	40.1	100	100
vs_ind_occ	% of industries with high $WCR^{LC\text{-}Occ}$	82.6	14.4	43.2	100	100

Table A4. Mean of validation scores (and sub-scores) by selected similarity matrices (M_{ij})

M_{ij}	Number of Cs，	Validation score VS (Avg sub-scores)	Validation sub-scores			
			$VS^{LC\text{-}Emp}$	$VS^{LC\text{-}Est}$	VS^{IO}	VS^{Occ}
LC-IO-Occ	38	79	65	65	93	93
LC-Emp	38	68	87	76	72	54
LC-Est	38	66	80	85	62	50
IO	30	74	58	60	87	87
Occ	21	79	65	64	90	98
COI	38	40	45	44	45	35

Step 5: Assessing individual clusters of candidate C*

In this step, we examined the clusters in *C** to assess whether there are industries that are more appropriately placed in other clusters and whether to combine or break individual clusters to promote the coherence of the clusters. We identified two types of possible outlier industries: systematic and marginal outliers. Systematic outliers are those with a low WCR_{ic} score, while marginal outliers are those with a high WCR_{ic} score that could be conceptually appropriately in another cluster.

Step 6: Comparing individual clusters of candidate C*

Based on the above work, we compared individual clusters in *C** with these criteria: the share of national cluster employment is supposed to be the 90th percentile and the share of national cluster establishment should be the 25th percentile. By applying the six steps above, we obtained the finalized set of cluster definition C**, which is generated using a hierarchical clustering function with 22 clusters and the multidimensional similarity matrix LC-IO-Occ. This configuration has the highest VS across all the Cs, with a score of 84.1 percent (see Table A5).

Table A5. Candidate cluster configurations C*s (top-10 rankings in VS)

C*s	Model choices			Validation scores					
				VS-Cluster		VS-Industry		VS	
	M_{ij}	Numc	Function	Rank	Score	Rank	Score	Rank	Score
C*	**LC-IO-Occ**	**22**	**Hiw**	**3**	**84.1**	**1**	**84.2**	**1**	**84.1**
C_2	LC-IO-Occ	21	Hiw	10	83.3	2	84.1	2	83.7
C_3	LC-IO-Occ	23	Hiw	8	83.7	4	83.6	3	83.7
C_4	LC-IO-Occ	24	Hiw	4	83.9	11	82.8	4	83.3
C_5	LC-IO-Occ	25	Hiw	28	82.5	13	82.7	5	82.6
C_6	EST-OCC	28	Hiw	21	82.6	18	82.3	6	82.4
C_7	EST-OCC	25	Hiw	24	82.5	21	82.3	7	82.3
C_8	LC-IO-Occ	26	Hiw	29	82.2	19	82.3	8	82.2
C_9	EST-OCC	27	Hiw	31	81.9	20	82.3	9	82.1
C_{10}	COI-IO-OCC	21	Hiw	54	81.0	14	82.7	10	81.8

Now, we have replicated Delgado et al.'s cluster algorithm and we have quantified all clusters in Guangdong during the investigation period. We must again emphasize that our results are transparent, rigorous, repeatable, and can fully represent Porter's cluster definition. Furthermore, although the distribution of 22 clusters varies from city to city due to the different industrial structure, 21 cities in Guangdong province rank at the top of the whole nation. In other words, Guangdong is the most economically developed province with the most diverse industrial structure. As shown in Table A6, Shenzhen has 22 clusters, and even Meizhou, located in a remote area, has 11 clusters.

Table A6. Distribution of 22 clusters in 21 cities in Guangdong province

city	Number of clusters	city	Number of clusters
Chaozhou	15	Shaoguan	17
Dongguan	21	Shantou	16
Foshan	20	Shanwei	14
Guangzhou	21	Shenzhen	22
Heyuan	14	Yangjiang	16
Huizhou	16	Yunfu	13
Jiangmen	21	Zhanjiang	15
Jieyang	19	Zhaoqing	19
Maoming	17	Zhongshan	20
Meizhou	11	Zhuhai	20
Qingyuan	17		