Ph. D. Dissertation in Economics

Exploring the Heterogeneity of Agglomeration Effects on Innovation: A Multifaceted Persepctive

혁신에 대한 집적 효과의 이질성에 대한 연구
:다면적인 관점으로

August 2023

Graduate School of Seoul National University
Technology Management, Economics, and Policy Program

Hyoji Choi

Exploring the Heterogeneity of Agglomeration Effects on Innovation: A Multifaceted Persepctive

지도교수 김연배

이 논문을 경제학박사학위 논문으로 제출함

2023 년 8 월

서울대학교 대학원
협동과정 기술경영경제정책 전공
최효지

최효지의 경제학박사학위 논문을 인준함

2023 년 8 월

위 원 장 황 준 석 (인)

부위원장 김 연 배 (인)

위 원 홍 아 름 (인)

위 원 전 보 강 (인)

위 원 장 필 성 (인)

Abstract

Exploring the Heterogeneity of Agglomeration Effects on Innovation: A Multifaceted Perspective

Hyoji Choi
Technology, Management, Economics, and Policy Program
The Graduate School
Seoul National University

Innovation has been considered an important driver for economic growth and industrial development. The cluster proposed by Porter has attracted the attention of many scholars as a strategy to promote innovation in the region and promote national competitiveness. However, many scholars question the effectiveness due to the lack of a systematic and comprehensive analysis of the externalities of clusters that encompass cluster strategies. In addition, there is a growing voice that it is necessary to understand heterogeneity due to the externalities of agglomeration and the multifaceted characteristics of innovation as well as whether it is conducive to innovation. Therefore, in this thesis, a quantitative and comprehensive analysis of the relationship between innovation and the external nature of agglomeration was performed through the multilevel analysis using Korean firm level data

(Chapter 3), and meta-analysis for selected 26 studies (Chapter 4).

Chapter 3 delves into a firm-level multilevel analysis model to explore the correlation between agglomeration-related metrics (region-industry units and regional unit) and innovation, employing data from Korean firms. Embedded within a theoretical framework predominantly rooted in the concepts of specialization and diversity, commonly adopted in existing literature, the study gauges agglomeration externalities by utilizing indicators largely associated with specialization and diversity. Furthermore, the analysis scrutinizes the scale effects of agglomeration, examining population density and a unique characteristic dummy variable for the Korean capital region. Notably, the study introduces a contextual factor linked to technological regimes, capturing the technological attributes that influence the agglomeration-innovation relationship, contributing to an investigation of the factors causing heterogeneity. As the analysis unfolds, the chapter examines the foundational presumptions of the existing theoretical framework and underscores the limitations of conventional methodologies used for analyzing the agglomeration-innovation relationship.

Chapter 4 rests on an extensive literature review, selecting 26 studies focused on innovation and agglomeration externalities as the basis for conducting a meta-regression analysis. Drawing insights from Chapter 2's literature review and the empirical analysis findings of Chapter 3, this section compiles and categorizes theoretical and empirical factors inducing heterogeneity. Of particular note is the examination of the heterogeneity inherent in innovation indicators, a commonly employed dependent variable in primary studies, as well as proxy measures capturing agglomeration externalities. Building on this categorization, the chapter proceeds to code moderator variables aligned with these factors, with the aim of quantitatively discerning how each factor's influence on the innovation-agglomeration relationship is manifested in research outcomes. The chapter's objective extends to investigating whether consistent patterns or shared insights can be extracted from the diverse outcomes generated by the meta-regression analysis of studies centered on the relationship between innovation and agglomeration, while accounting for varied theoretical and empirical factors.

Chapter 5 encapsulates the central findings of the thesis, spotlighting the heterogeneity present in literature concerning the agglomeration-innovation relationship, as illuminated in Chapter 2's literature review and analyzed in Chapter 3. The chapter aligns these insights with the quantified analysis results of heterogeneity unveiled through the meta-regression analysis in Chapter 4. It emphasizes that the variance in indicators employed across studies addressing similar research inquiries significantly contributes to the mixed results observed within the field. Furthermore, this section conducts an in-depth reassessment of the outcomes from Chapter 3's analysis, shedding light on the inherent limitations and deficiencies of the prevailing theoretical and methodological frameworks when addressing the contemporary landscape of innovation and agglomeration externalities within the literature. It provides insights into these shortcomings and proposes potential enhancements and revisions to these frameworks. Additionally, the chapter introduces recent advancements in the literature stream aimed at surmounting these constraints, illustrating the field's evolution in tackling the challenges at hand.

Moreover, this chapter underscores the contributions of the thesis by presenting a valuable perspective on the observed heterogeneity, accentuating the limitations of existing literature, and suggesting avenues for refining research methodologies. Through these insights and contributions, the chapter seeks to offer a comprehensive and thought-provoking conclusion to the study, providing a nuanced viewpoint on the intricate interplay between agglomeration and innovation within the evolving dynamics of the economic landscape and knowledge production processes.

Keywords: agglomeration, externliaites, spillovers, meta-regression, multilevel analysis, technological regime.

Student Number: 2014-30281

Contents

List of Tables

List of Figures

Chapter 1. **Introduction**

1.1 Research background and objective of the study

Since Schumpeter (1934), innovation has been regarded as the most important engine for industrial development and economic growth. For the past 30 years, the economics of industrial clustering has received much attention to promote innovation and drive economic growth. Initially driven by the work of Porter (1990), growth and industrial economists revived the traditional agglomeration theory that urban and regional economists had long taken for granted. The observation that knowledge production and innovative activities are strongly agglomerated geographically, in Europe (Caniels, 1999; Breschi; 1999) as well as in the US (Jaffe, 1989; Feldman, 1994; Audretsch and Feldman, 1995) has led many researchers to investigate the possible causes and mechanisms of this phenomenon.

According to a WIPO report from 2019, the geography of innovation can seem to exhibit a paradoxical pattern. This is because scientific knowledge and innovation activities are progressively disseminating globally, yet they are also clustering in specific "hotspot" regions. Traditionally dominant economies are no longer exclusive contributors to scientific research and inventors, as new entrants, particularly Asian nations, are significantly contributing. The distribution of patenting activity among top economies has evolved, as depicted in **Figure 1-1.** Notably, China and Korea have played a substantial role in expanding the scope of knowledge production and innovation. Their combined patent registrations for 2015-2017 surpass 20%, a stark contrast to the less than 3%

recorded during 1990-1999. Australia, Canada, India, and Israel have also made notable contributions to the global diffusion of innovation.

Simultaneously, this heightened global dispersion of innovation has coincided with a heightened concentration of innovative activities within densely populated regions at the national level. In the United States, hotspots (or clusters) around New York, San Francisco, and Boston accounted for roughly 25% of all U.S. patents filed between 2011 and 2015. Similarly, in China, companies situated around Beijing, Shanghai and Shenzhen increased their share of all Chinese patents from 36% to 52% during the same time span (as indicated in **Table 1-1**).

Nevertheless, patent activity remains notably subdued in many middle-income nations and all low-income countries. A mere fraction, less than 19%, of the global pool of creative and scientific outputs originates from inventors or researchers located beyond the realms of hotspots and specialized clusters. Despite the transformative impact of technological advancements and digitalization on the global innovation landscape, over 160 countries continue to exhibit limited innovation activity, failing to establish domestic hotspots or niches clusters.

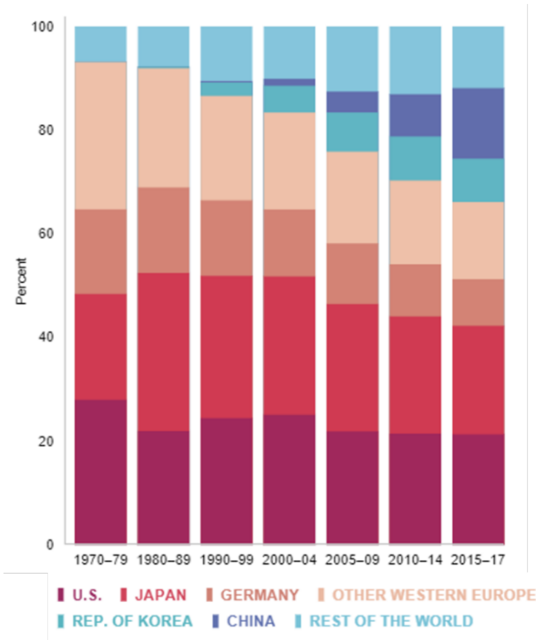


Figure 1-1 Evolution of patenting share by top economies: source (WIPO)

COUNTRY	PATENTS				PUBLICATION			
	1991-95	%	2011-15	%	2001-05	%	2011-15	%
CHINA	Beijing Shanghai Shenzhen- Hong Kong	36.5	Shenzhen- Hong Kong Beijing Shanhai	52.2	Beijing Shanghai Nanjing	43.9	Beijing Shanghai Nanjing	35.8
US	New York San Jose- San Francisco Boston	19.4	San Jose- San Fancisco New York Boston	23.4	New York Washington DC- Baltimore Boston	21.9	Boston New York Washington DC- Baltimore	21.4

Table 1-1 Top three Global Innovation Hotspot, patents and publications, US and China

Note: Source: Author's re-elaboration of WIPO (2019)

Confronted with these observations for economic realities, numerous scholars have been driven by a strong impetus to delve into the following inquiries: What are the distinguishing attributes of regions or countries contributing to knowledge production and innovation? What factors contribute to the concentration that propels certain regions or countries to flourish while causing others to lag behind? The two questions are closely interrelated, yet they distinctly emphasize different aspects. The initial question primarily revolves around the *emergence* of novel regions engaging in knowledge production and innovation, as exemplified by recent instances in Korea and China. The subsequent question closely pertains to the *intensifying* concentration of specific locales like New York, San Francisco, and Boston within the United States. *Both inquiries bear significant implications for the advancement and expansion of industries, regions, and nations. However, this study specifically directs its attention towards the second question*.

Building upon the foundational works of Marshall (1920), Weber (1929), Jacobs (1970), Romer (1986), and Porter (1990), scholars in economic geography, industrial economics, and innovation have directed their attention toward the externalities of concentration, also known as agglomeration, in addressing the second[1] question. In this context, two distinct types of externalities have gained recognition for their substantial influence on the process of knowledge generation and diffusion, as highlighted by Glaeser

[1] The traditional Theory associated with the first question is Alfred Weber (1909)'s Industrial Location Theory. As a first attempt to understand why firms locate in particular places (or regions) emerged, it deals with the least cost location that takes into account labor costs and transport costs between production factors and markets. The 'location' theory of knowledge production and innovation is not formalized and can be deeply related to industrial and R&D policies. This study aims to focuson the impact on innovation caused by internal and external economies of agglomeration, rather than the emergence of knowledge production and innovation in a specific place (or region).

et al. (1992). These externalities include specialization externalities, which predominantly operate within specific industries, and diversity externalities, which extend across various sectors as explored by Beaudry and Schiffauerova (2009).

Marshall (1890) observed that industries tend to specialize geographically due to the advantages of proximity within the same industry. This proximity facilitates intra-industry knowledge transfer, reduces transportation costs for inputs and outputs, and allows firms to leverage specialized concentrations for more efficient labor market pooling. In summary, Marshall’s arguments can be characterized as encompassing various pathways through which the effects of geographical proximity, specialization, and scale can manifest spatially. Jacobs (1969), on the other hand, asserted that diversity in cities, which stands in contrast to Marshall’s emphasis on specialization, serves as the driving force behind productive innovations. She contended that a higher sheer number and variety of divisions of labor enhance the economy’s inherent capacity to incorporate additional types of goods and services.

In a related vein, Porter (1990) engaged in the discourse of “competition externalities,” advocating the local competition, as opposed to monopoly, fosters industrial growth and the transfer of knowledge within geographically concentrated specialized industries. This perspective addresses strategic elements to ensure national competitiveness. Porter’s work led to a revive of Marshall and Jacobs’ arguments from the perspective of ‘*cluster.*’

In a seminal work on the industrial growth of cities, Glaeser et al. (1992) later formalized the ideas put forth by Marshall (1890), Arrow (1962), and Romer (1986), giving rise to the Marshall-Arrow-Romer (MAR) model. This model, rooted in endogenous growth theory, posits that industrial clustering in regions engenders knowledge spillover effects among firms, thereby promoting innovation within specific industries in those regions. This specialization encourages the exchange of knowledge, ideas, information, products, and processes through interactions among skilled workers, bypassing market transactions (Beaudry and Schiffauerova, 2009). Furthermore, this study aimed to identify the determinants of industrial development within cities by comparing the knowledge spillover in industrially specialized regions with the knowledge spillover in diversified regions in line with Jacobs' proposition, and the effects of competition. Through this study, traditional agglomeration theory and subsequent cluster discussions are integrated with endogenous growth theory from a knowledge economy perspective.

Hence, the primary distinction between the earlier argument concerning agglomeration externalities and Glaeser's framework lies in the reliance on "knowledge spillovers" within either specialized or diversified regions, as opposed to addressing the comprehensive externalities associated with concentration. Put succinctly, Glaeser et al. (1992) embarked on a distinct investigative path referred to as the MAR vs. Jacobs vs. Porter debate, which revolves around the concept of knowledge spillover. This initiation sparked a thoughtful pursuit to delve into the following inquiries: Does regional specialization or diversification assume a more significant role in shaping knowledge production and innovation? This question resonated throughout a wide spectrum of academic discourse, encompassing realms such as open innovation, economic geography, and endogenous growth, thus leading to extensive research endeavors to date. (Beaudry and Schiffauova, 2009; Enkel and Heil, 2014).

However, our understanding of the impact of agglomeration externalities, also known as cluster effects, on innovation remains intricate. From a theoretical standpoint, although the prior description mainly focused on positive effects, clusters can also have the negative effects from potential 'lock-in' or 'congestion' effects (Cooke, Uranga, and Etxebarria 1997; Boschma, 2005). In addition, increased conceptual ambiguity is also a hinderance to revealing our understanding as various paths to the effects of agglomeration covered in literature from various perspective such as growth theory, economic geography, and innovation literatures. In terms of analysis, since numerous empirical studies have delved into the relationship between clusters and innovation, utilizing data from diverse countries and time periods, but they have shown mixed results (Baptista and Swann, 1998; Beaudry and Breschi, 2003). The presence of mixed finding has largely hindered the formulation of definitive general conclusions.

Furthermore, the contention has emerged recently that the existing literature might be addressing the wrong question. This question constrains the answer into an oversimplified dichotomy, presenting an either-or scenario, while certain studies have firmly affirmed a positive correlation (Caragliu et al. 2016), the binary nature of the inquiry might obscure a more nuanced relationship. Audretsch and Belitski (2022) suggest that more pertinent and intriguing questions could center on the circumstances under which regional specialization enhances knowledge production and innovation, as opposed to the conditions favoring diversification for innovation.

In recent studies, conceptual frameworks and empirical studies have been conducted considering various *'contexts'* in consideration of the ***'multifaceted characteristics'*** of the relationship between innovation and agglomeration to explain ***'the heterogeneity of agglomeration effects on innovation.'*** However, the specifics of how, when, and which types of externalities are relevant for different forms of innovation remain shrouded in ambiguity, given the divergent nature of innovation activity across local, technological, and knowledge contexts. Despite an array of related research, the enigma surrounding agglomeration economies shows no signs of dissipating (Audretsch, 2022; Breschi and Lissoni, 2001; Greunz, 2004) For instance, concerning localized knowledge spillover, a mechanism for externalities within industrial clusters, Breschi (1992) asserts that this persistent ambiguity stems from two main sources of dissatisfaction: firstly, an excessive diversion of research efforts, both theoretical and empirical, from the examination of the role of geographical proximity in the economics of knowledge transfer, an area that remains somewhat contentious; secondly, the emergence of simplistic policy implications that evoke memories of past disappointments with initiatives such as science and clusters.

Based on the above discussion and the limitations in the existing literature, this study attempted *to identify the "ambiguity" results for the heterogeneous characteristics of the relationship between agglomeration and innovation* by conducting empirical regression analysis of Korean cases and meta-analysis on related literatures. Considering these circumstances, it is imperative to meticulously review the literature on agglomeration and innovation dynamics, which is extensively explored to bolster national competitiveness, effectively execute regional industrial strategies, and navigate the swiftly evolving terrain of knowledge production toward a sustainable socio-economic trajectory. Moreover, beyond the findings directly presented in the literature, there is a critical need to thoroughly scrutinize research methodologies.

In today's society, where information and knowledge are rapidly expanding, there is a growing emphasis on restraining futile knowledge production and researching knowledge production methods that adhere to proper approaches and directions. West et al. (2021) assert that misinformation disrupts societal scholarly learning and hinders the dissemination of knowledge. They also point out that the generation of misinformation occurs not only in mainstream media or social media but also within the realm of science, highlighting the need for a more serious consideration, especially given the substantial societal impact of scientific knowledge.

This thesis, while sharing this awareness of this issue, aims to enhance the efficiency of researchers' efforts in the realm of knowledge production and draw meaningful conclusions from the literature by quantitatively tracing the pathways through which the relationship between agglomeration and innovation can lead to the production of 'misinformation.' By doing so, it seeks to contribute to a deeper understanding of the dynamic between agglomeration and innovation, ultimately aiming to guide research efforts towards more effective knowledge production methods.

1.2 Outline of the study

This research is composed of five chapters, encompassing two primary sections that delve into the heterogeneity of agglomeration effects on innovation. These sections utilize multilevel analysis on firm-level data from Korea (Chap. 3) and meta-regression analysis (Chap. 4). The arrangement of this study is outlined in Figure 1-2, which encapsulates its motivation, theoretical background, key outcomes derived from empirical examination, and overall conclusion. The subsequent sections of this study are structured as outlined below.

Chapter 2 provides the theoretical background of this study by reviewing existing literature on the externalities of agglomeration and innovation. In Section 2.1, various pathways of agglomeration externalities are explored, shedding light on prominent types such as those represented by Marshall, Jacobs, and Porter. By tracing the trajectory of research literature, this section aims to uncover the conceptual ambiguities and heterogeneity in the relationship between agglomeration and innovation that may arise from a theoretical standpoint. Moving forward, Section 2.2 compiles relevant empirical studies and examines factors contributing to heterogeneity from an empirical perspective. Drawing from this literature, the study points out the absence of considerations for conceptual ambiguity, negative effects, and oversimplification of research questions. Moreover, it underscores the importance of numerous contextual factors that influence the relationship between agglomeration and innovation.

Chapter 1. Introduction

- Motivation and Objectives of study: Exploring the heterogeneity of agglomeration effects on innovation and estimating average values
- Outline of study

Chapter 2. Theoretical Background

- Three approaches to agglomeration externalities: MAR, Jacobs, Porter externalities
- The Heterogeneity of agglomeration effects on Innovation: A Multifaceted characteristics (Indicators of MAR and Jacobs externalities, Level of aggregation, Industry sectors, Geographical unit, Countries and regions, types of innovation – invention, innovation, commercialization)

Main Researches

Chapter 3. The Heterogeneity of agglomeration effects on Innovations: A Meta-Regression Analysis

- Estimates and identifies the factors of average effect and heterogeneity by type of agglomeration by collecting literature examining the relationship between innovation and agglomeration.
- 26 primary studies
- Methodology: Meta-regression analysis

+

Chapter 4. The effect of region-industry characteristics on firm innovation in different technological regime: A multilevel study

- Estimates and identifies the factors of average effect and heterogeneity by type of agglomeration by collecting literature examining the relationship between innovation and agglomeration.
- 3,664 Korean firm-level data
- Methodology: Multi-level analysis

Chapter 5. Conclusions

- Summary and main findings
- Limitation and opportunities for future research

Figure 1-2 Outline of this study

Chapter 3 employs a multilevel analysis model at the firm level to examine the relationship between agglomeration-related indicators (region-industry units and regional unit) and innovation using data from Korean firms. Within the theoretical framework primarily rooted in the concepts of specialization and diversity commonly adopted in existing literature, the study measures agglomeration externalities using indicators mainly associated with specialization and diversity. Additionally, an analysis focusing on the size effects of agglomeration is conducted by examining population density and a dummy variable representing the unique characteristics of the capital region in Korea. Notably, the study introduces a moderating factor related to technological regimes, capturing the technological characteristics as a contextual factor that influences the relationship between agglomeration and innovation. This aims to contribute to the investigation of factors contributing to heterogeneity. As the analysis results are interpreted, the study examines the underlying assumptions of the existing theoretical framework and highlights the limitations of conventional methodologies used in analyzing the relationship between agglomeration and innovation.

Chapter 4 is based on an extensive literature review and involves the selection of 26 studies focusing on innovation and agglomeration externalities. These selected studies serve as the foundation for conducting a meta-regression analysis. Drawing from the insights gained in Chapter 2's literature review and Chapter 3's empirical analysis results, this chapter compiles and categorizes the theoretical and empirical factors that induce heterogeneity. Notably, it scrutinizes the heterogeneity present in the indicators of innovation, the dependent variable commonly used in primary studies, as well as proxy measures capturing agglomeration externalities. Building upon this categorization, the chapter proceeds to code moderator variables that correspond to these factors, aiming to quantitatively ascertain the extent and direction in which each factor influences the relationship between innovation and agglomeration as revealed in the research results. The objective of this chapter is also to explore whether a systematic pattern or common insights can be extracted from the diverse outcomes obtained through the meta-regression analysis of studies focusing on the relationship between innovation and agglomeration, taking into account the varied theoretical and empirical factors.

In Chapter 5, the main findings of this thesis are summarized, focusing on the heterogeneity of literature regarding the relationship between agglomeration and innovation, as raised in Chapter 2's literature review and analyzed in Chapter 3. The chapter also aligns with the quantitative analysis results of heterogeneity identified through the meta-regression analysis in Chapter 4. It highlights that the variation in indicators used across studies addressing similar research questions contributes significantly to the mixed results observed in the field.

Moreover, this chapter undertakes a comprehensive reexamination of the results from Chapter 3's analysis, shedding light on the inherent limitations and shortcomings of the existing theoretical and methodological frameworks in addressing the current state of innovation and externalities of agglomeration within the literature stream. It offers insights into the deficiencies and proposes potential revisions and enhancements to these frameworks. Additionally, the chapter introduces recent developments in the literature stream aimed at overcoming these limitations, showcasing how the field is evolving to address the challenges presented.

Furthermore, this chapter emphasizes the contributions of this thesis by offering a valuable perspective on the observed heterogeneity, highlighting the limitations of the existing literature, and suggesting potential avenues for refining research approaches. By presenting these insights and contributions, the chapter aims to provide a comprehensive and insightful conclusion to the study, offering a thought-provoking perspective on the complex relationship between agglomeration and innovation in the context of evolving economic dynamics and knowledge production processes.

Chapter 2. **Theoretical background**

2.1 Three approaches to agglomeration externalities

Aforementioned before, the literature on industrial and regional economies distinguishes three theoretical approaches to understanding how industrial clustering generate externalities. The first is the Marshall-Arrow-Romer (MAR) approach (Marshall, 1920; Arrow, 1962; Romer, 1986), which argues that industrial agglomeration are more beneficial and stronger for firms within the same industry (intra-industry spillovers) and region. It means that knowledge externalities between firms only occur among firms of the same or similar industry, and thus can only be supported by regional concentration of the same or similar industries. These specialization externalities imply that it is likely to arise when the industry to which a firm's main activity belong is *relatively large*. (Frenken et al., 2005) Marshall said two other benefits of geographic concentration: labor market pooling[2] and transport cost savings. *Economies of scale* induced from shared inputs in the form of labor equipment and infrastructure between large concentrations of firms from the same industry are another critical source of Marshallian externalities (Krugman, 1991). These intra-industry spillovers are also known as localization (specialization) externalities, Marshall or MAR externalities. This study will use speicalization or MAR indistinctively.

Jacobs (1969), on the other hand, claims that the most important sources of knowledge spillovers are external to the industry within which the firm operates. Since the

---

[2] Workers are better protected from business uncertainty and demand shocks if located in a region with a large local labor market in their own industyr (Mukkala, 2004)

diversity of these kowledge sources is greatest in cities, she also argues that cities are the source of innovation. She emphasizes knowledge arising from a variety of geographically industries promotes innovation and productivity in the region. More recently, Caragliu et al. (2016:93) explained agglomeration externalities in regions that are the ciritical points of innovation, “the place where co-locating firms enjoy the presence of other creative companies, active in different industries and *cross-fertilizing* dieas through formal and informal exchanges of information.”

The third type of externality refers to Porter’s (1990) arguments, also associated with Jacobs, that compeition is better for growth. Strong competition in the same market provides significant incentives to innovate which in turn accelerate the rate of technical progress of hence of productivity growth. Combes (2000) emphasizes the fact that hifh competition acts as a strong incentive to R&D spending, since firms are forced to be innovate in order to survive (van Oort and Stam, 2006)

In summary, MAR externality emphasizes the spillover effect within the same industries based on the economic effect of scale, and Jacobs focuses on diversity emphasizing 'cross-fertilization' (a kind of “Economies of scope”). The critical conceptual distinction between Porter’s argument with Marshall’s and Jacobs’s one is that it emphasizes ‘competition,’ not patterns of agglomeration itself. In other words, agglomeration is discussed as one channel that can provide incentive through compeition. This means that the discussion may be more complicated than the other two cases. MAR externality has a physical dimension of 'scale', and Jacobs has a dimension of 'diversity', but 'competition' is difficult to simplify. Of course, both MAR and Jacobs's externalities are only relatively straightforward, and as mentioned above, the dichotomous discussion is unrealistic. One of the causes is the criterion of "the same industry."

Frenken et al. (2007) introduced the concept of related and unrelated variety. These concepts provide the more disentangled views of diversified industry structures, compared to previous conceptions. Frenken et al. (2007) argued that relatedness rather implies technological and cognitive proximity (Boschma, 2008) and is expected to be particularly conducive to Jacobs-type knowledge externalities (Frenken et al., 2007) On the other hand, Aarstad et al. (2016), which is grounded in the paradigm of evolutionary economic geography, insists that regional specialization is a two-dimensional construct; a low level of specialization can indicate a region with a high level of related or unrelated variety. In the end, discussion on the specialization and diversity are closely related to the question of how much the same or different and thus relatedness can be another dimensions describing patterns of agglomeration.

2.2 The Heterogeneity of Agglomeration Effects on Innovation: A Multifaceted Characteristics

As mentioned earlier, the three main dimensions that can directly describe the aspect of patterns of agglomeration are scale, diversity, and relatedness. This chapter summarizes other factors that may cause heterogeneity in agglomeration effects on innovation.

Indicators of MAR and Jacobs externalities

The most obvious differences among the studies are the one associated with the choice of independent and dependent variables. Frenken et al. (2005, p. 22) suggests that this “ambiguity in resultsis probably due, at least in part, to problems of [...] definitions of variety, economic performance, spatial scale and spatial and sec-toral linkages...”. Some studies, probably constrained by data availability, utilize the same index to measurethe impact of both specialization and diversity in the same variable (for example, the Hirschman–Herfindahl index in Loikkanen and Susiluoto, 2002). Authors then may interpret a positive sign (or high values) on the coefficient as evidence of prevailing Marshall externalities and a negative sign (or low values) as a proof of Jacobs economies. This methodology, however, may not be appropriate in some industries because both kinds of economies could be present simultaneously. The two externalities are obviously not mutually exclusive, since specialization is a particular characteristic of a certain sector within a local system, whereas diversity is a property characterizing the whole area. This implies that variables of other classification criteria may be more useful than variables for two externalities when evaluating the externalities of aggregation.

Level of aggregation

An industry could appear as a statistically homogenous entity if a 1-digit or 2-digit industrial classification is used, whereas the same industry will present a wide variety of different activities if the analysis is based on a 6-digit breakdown. Frenken et al. (2005) expect diversity measured at the lowest level of aggregation (related variety) to be positively correlated with economic growth and employment growth. It can be said that the level of industrial agglomeration reflects relatedness. However, not all studies show which industry classification system was used.

Industrial sectors

An important difference in most studies lies in the selected industries. Analyzed data may come from only one industry (as in Beaudry, 2001 or Baptista and Swann, 1999). The analysis may also consider all the range of industries including non-manufacturing services such as wholesale and retail trade (as in Glaeser et al.,1992; Beaudry and Swann, 2001, 2007; Combes et al., 2004), but it is also common to completely exclude services andagriculture from the sample due to problems of data availability or productivity estimation in services. Furthermore, the methodology may involve an analysis of one manufacturing industry at a time (as inHenderson et al., 1995), which allows to distinguish the roles of either type of externalities in each industry. This approach, however, may not be applicable to all countries, especially in small countries with only are latively small number of locations where the selected industries can flourish (van Soest et al., 2002). An alternative approach here is to consider only a number of the largest industries of all types in each region (for example, the 6 largest industries in each city as in Glaeser et al., 1992, and the 5 largest industries as in King et al., 2002), which may *de facto* automatically increase concentration levels in each city. The selected range of industries used for the sample may yield further differences.

Geographical unit

The selected level of geographical aggregation and the division of the observed territory into regions for the study of geographical specificities is yet another source of possible discrepancy in the results. Baldwin and Brown (2004) argue that when testing for diversity, the geographic unit of analysis should not oversize labor market regions as it is on that level that product variety has an influence.

Countries and regions

As the economic environment and the dispersion of populationvary from one country to the next, we expect some differences to arise in the effect of agglomeration economies in various countries. Some authors have carried out simultaneous studies of several countries and found quite comparable results, as Henderson (1986) for the US and Brazil. Other researchers have encountered distincteffects of the two externalities for different countries, as Beaudry and Breschi(2000,2003) for the UK and Italy or Beaudry et al.(2001) for several European countries.

Invention, Innovation and commercialization

Innovation can take many forms. At one level, the results from innovative efffort can physical or abstract, such as computer programs or new business methods vs. a higher-definition TV. At a second and arguably more important level, innovation is a very broad term that includes multiple types of efforts. The types of advances sought by start-up can be quite different from those pursued by large universities or corporate R&D labs. The theoretical and empirical literature tends to lump everything together, but is is important to recognize to be one of sources of heterogeneities.

Chapter 3. **The Heterogeneity of Agglomeration Effects on Innovations: A Meta-Regression Analysis**

3.1 Introduction

Recent decades have witnessed a new wave of interests in clusters from researchers and policy makers, and supporting clusters have become a prevalent local strategy in promoting economic development (M.E. Porter 1990; P. R. Krugman1991; M. Feldman 2000; Storper and Scott 1995). Clusters are claimed to have positive effects on innovation, productivity, and resilience (Baptista 1998; Folta, Cooper, and Baik 2006; Treado and Giarratani 2008). This study addresses agglomeration' effects on innovation. It focuses on innovation for two major reasons. First, one important outcome of clusters is promoting innovative activities because clusters can foster the spillover of the elusive knowledge that is critical to innovation (M.P. Feldman 1994; Audretsch and Feldman 1996). Practically, some clusters do promote innovation and make the local economy prosperous, such as the Silicon Valley. Second, in modern economic growth theories, innovation is an important driving force of long-term economic success (Grossman and Helpman, 1990; Aghion, Harris, and Vickers 1997; Freeman and Soete1997). As a result, firms, regions and countries all try to improve their capacities of innovation in order to achieve better economic performances (Calantone, Cavusgil, and Zhao 2002; Morgan 2007; Mairesse and Mohnen 2001). Thus, understanding clusters' effects on innovation can yield important insights into the issues of regional economic development and provide policy implications to local authorities.

To date, our knowledge of agglomeration externlaites (or cluster's effect) on innovation is mixed. Theoretically, clusters may encourage innovation due to knowledge spillover effects but may also jeopardize innovation due to ''lock-in'' effects (Cooke, Uranga, and Etxebarria 1997; Boschma 2005). Many empirical studies have investigated therelationship between clusters and innovation with data from various countries and time periods, but the results are inconclusistent (Baptista and Swann 1998; Beaudry and breschi 2003). Mixed results largely prohibit us from reaching any general conclusions. Meta-analysis is suggested as a meaningfull way of comining empirical studies with contradicting and heterogeneous results (Rowenthal, 1991). Since individual studies inevitably suffer from problems such as meausrement artifacts, limited research range (relatively narrow geographical regions and time frames), and small sample size, combining and contrasticng results from multiple studies are necessary for the aim of rearching powerful robust general conclusion (Glass 1976) Yet to data little work like that has been done on the topic of clusters and innovation except the note worthy paper of de Groot, Poot and Smit (2010) and Fang (2015).

This study provides a meta-analysis of relevant empirical studies on the relationship of clusters and innovation since the 1990s. It differs from de Groot, Poot, and Smit (2010) and Fang (2015) in four aspects. *First*, de Groot, Poot, and Smit focus on the regional level effects, while this article pays equal attention to the regional-level and the firm-level effects. Moreover, this article explicitly compares the results from firm-level, industry-level, and regional-level studies, providing us additional knowledge about whether clusters' effects on innovation are mostly captured by the individual firms, kept in the industry, or absorbed by the region. *Second*, de Groot, Poot, and Smit include studies with dependent variables varying from employment growth, productivity growth to innovation. Therefore, their research question is in fact much broader than that of this article. But large variations in the dependent variables prohibit a clear interpretation of the results and make the calculation of an average effect size inapplicable. This article restricts its concern to clusters' of regional agglomerations' effects on innovation, and papers that do not have an innovation-related dependent variable are excluded. By doing so, this article is able to calculate a relatively meaningful average effect size and arrives at results that are easier to interpret, at the expense of a smaller sample size. *Third,* Fang (2015) analyzed variables for all heterogeneity, but this approach may be meaningless in theory. Based on the previous literature review (Chap 2), this study attempted to distinguish the influence of the physical dimension explaining the pattern of cluster(agglomeration) by classifying the groups of variables representing the size, diversity, and relatedness of agglomeration.

This study attempts to address three questions. (1) What are the general conclusions of agglomeration externalities on innovation from previous studies? (2) Are previous studies homogeneous or heterogeneous in their estimated correlations between clusters and innovation? (3) If they are heterogeneous, what variables may serve as moderators? Namely, what variables may influence the direction and magnitude of clusters' effects on innovation?

Using the multilevel model, this article reveals that generally speaking, clusters have statistically positive effects on innovation. Using Cochrane's Q statistics and I^2 statistics, significant heterogeneity is found across individual studies, suggesting moderators may be at work in shaping the cluster–innovation relationship. Using the mulitlevel models, potential moderators such as how is cluster measured, which industry is primary in the cluster, firm size and so on, are identifed. This particle provides useful guidance for local authorities in the following ways. First, generally speaking, initiating a cluster strategy to promote innovation is promising. Second, for a specific cluster, the direction and magnitude of the relationship cannot be determined without considering a handful of important moderators, such as cluster characteristics (e.g., does it have high concentration/localization?), industries, and whether we care about the firm-level, industry-region level, or regional-level innovation performances. Third, based on the results of this article and relevant local data, we can form an expectation of the direction and magnitude of a specific cluster's effects on innovation. The cluster can either be an existing one or a hypothetical one.

3.2 Literature review

Researchers have long identified clusters' effects on innovation. They propose that clusters may benefit from innovation for several reasons. First, since at least part of the knowledge essential for innovation is elusive and uncodified, knowledge spillovers inside clusters are important for promoting innovation (Jacobs 1970, 1986; M. P. Feldman 1994; Audretsch and Fledman 1996). Second, the deepened specialization inside clusters enables firms to concentrate on limited processes of production, therefore, increases firms' chance of innovation in their specialty (Young 1928; Yang and Ng 1993; Maskell 2001). Third, colocating with rivalries exposes firms to great pressure and motivates them to innovate and maintain competitiveness (Burt 1987; Harrison, Kelley, and Gant 1996; M. E. Porter 1998). Fourth, informal social networks in clusters enable firms to cooperate more intensively and take more risk, which are important for innovation since innovative activities require a large amount of investment and the ability to deal with uncertainty (Gordon and Mccann 2000; Bathelt 2002; Feser and Luger 2003; M. Porter 2003). Fifth, clusters enhance creativity by attracting high-skilled labor and facilitating the communication and collaboration between them (Florida 2006; Florida, Mellander, and Stolarick 2008). Last but not the least, the lowered production costs due to transportation and information costs minimization, shared public intermediate inputs, labor pooling, and so on, enable firms to generate more profits and possibly increase their inputs into the innovative efforts (Marshall 1920; Lichtenberg 1960; Henderson 1986; Von Hippel 1988). All these forces lead to a striking concentration of innovation in the economic landscape (Breschi 1999; Paci and Usai 2000; Wang and Lin 2008). However, some researchers warn that clusters may also inhibit innovations. First, negative externalities such as congestion and overcompetition are common in clusters (Brezis and Krugman 1993; Baptista 1998). They may lower firms' profits and their inputs into the innovative activities. Second, knowledge spillovers, or by another name "knowledge leakage," may discourage a firm to innovate, since other firms can "free-ride" (Shaver and Flyer 2000; Baten et al. 2004). Third, the rigidity of relationships and repetitive information may lead to the "lock-in" effects, which limit firms' abilities to absorb outside knowledge (Boschma 2005; Moodysson and Jonsson 2007). Because of these forces, although innovation is spatially concentrated, it is not concentrated in a single location. Namely, the dispersing forces are at work (P. Krugman 1998; Beaudry and Breschi 2003). In addition to the theoretical debate, empirical results are mixed. Many empirical studies detect a positive relationship between clusters and innovation (Aharonson, Baum, and Feldman. 2004; Brenner and Greif 2006; Fornahl, Broekel, and Boschma 2011). Some reveal insignificant relationships (Beugelsdijk and Cornet 2002; Baten et al. 2007; Fitjar and Rodrı´guez-Pose 2011). Some even find out negative relationships (Acs and Audretsch 1988; Lee 2009). A few recent studies identify mixed results in their own regressions and they suggest some moderators may change the direction and the magnitude of the cluster–innovation relationship (Hamaguchi and Kameyama 2007; Hornych and Schwartz 2009; Fritsch and Slavtchev 2010).

To date, several moderators are identified by individual empirical studies, such as sectors/industries (Shefer and Frenkel 1998; Beaudry 2001; De Beule and Van Beveren 2012), whether the clusters are strong in firms' own specialization (Baptista and Swann 1998; Aharonson, Baum, and Feldman 2004), and the magnitude of the specialization/concentration (Hornych and Schwartz 2009; Fritsch and Slavtchev 2010).

However, it is hard to generalize beyond individual studies to decide whether these moderators are effective in general. What's more, individual studies usually use data from a single country (even a single region), a single study level (firm, industry, or regional level), and the time frames are limited. Variables such as geographical region, time frame, and study level are hard, if possible, to be identified as moderators in individual studies, despite the fact that their moderating effects may be important.

This article combines the contradicting empirical results in a meaningful way and identifies moderators. By using results from previous empirical studies since the 1990s, this article arrives at a general conclusion based on a super large sample, which includes all the individual samples in the selected studies. Since the sample compasses different countries, industries, centuries, and data levels, moderators undetectable in individual studies can be identified.

3.3 Estimation framework

3.3.1 A meta-regression analysis

A meta-analysis is a statistical analysis that combines the results of multiple scientific studies. Meta-analyses can be performed when there are multiple scientific studies addressing the same question, with each individual study reporting measurements that are expected to have some degree of error. The aim then is to use approaches from statistics to derive a pooled estimate closest to the unknown common truth based on how this error is perceived. (Stanley and Doucouliagos, 2010) Like all statistical techniques, data fuels meta-analysis. However, "data" in the meta-analysis context are the complex products of the research process. Typically, meta-data will be comprised of estimates of some economic association (also known as "effect sizes") linked to key dimensions of the research process that produced these effects.

In the primary literature, there are several types to be coded coefficient extracted from regression (1) dummy-dummy form; (2) dummy-continuous; (3) patent or innovation count-cluster dummy; (4) count-countinous (5) countinous-cluster dummy dummy and (6) countinous-countinous.

In despite of the the heterogeneity in reported effect, to make these effects comparable so as to be able to combine them, I follow Doucouliagos and Stanley (2009) and Stanley and Doucouliagos (2012) by transforming these effects into parital correlation coefficient (PCCs). The PCC is a unit-free measure of the magnitude and direction of the association between two variables (innovation and agglomeration/cluster)

$$pcc_i = \frac{t_i}{\sqrt{t_i^2 + df_i}} \quad and \quad se\_pcc_i = \sqrt{\frac{(1 - pcc_i^2)}{df_i}} \tag{3.1}$$

where t stands for the t-statistics on the estimated agglomeration externalities and df for the degrees of greedom extracted from the repective estimate in the primary literature.

Doucouliagos (2011) suggests that a partial correlation that is less than ± 0.07 can be regarded as small, even if it is statistically significant. The partial correlation indicates strong association (large effect) it is greater than ± 0.33.

3.3.2 The search protocol and data

In order to select the studies to be included in the meta-analysis, we searched in the EBSCO (Business Source complete, EconLit with Full Text, Regional Business News) and Web of Science database for papers estimating a localized spillover effect, using the keywords "Mashall" or "Jacobs" or "proximity" or "agglomeration" or "cluster" and "innovation" or "innovative performance" or "firm performance". When searching in the Web of Science, this study restricted the analysis to fields of "Economics", "planning and Development", 'Economics, Econometric and Finance", "management", "Business", and "Geography" obtained 3795 articles in EBSCO and Web of Science. Reviewers read the titles and abstract of all studies captured in the above electronic search, using a range of first-stage inclusion criteria designed to ascertain if the study: (i) investigates the linear effect[3] of localized spillovers on innovations; (ii) has an empirical dimension as opposed to a theoretical focus only; (iii) is NOT a review only; and (iv) do NOT use R&D intensity or expenditure as an innovation proxy. More specifically, the model specification for the linear relationship between innovation and agglomeration can be written as follow:

---

[3] Studies with quaratic terms go beyond our scope of this meta-analysis because we want explore the linear relationship between innovation and agglomeration. In order to analyze a model including a quadratic term, it is necessary to collect only studies dealing with the quadratic term model and perform meta-regression analysis. In addition, studies for 'relatedness' can be seen as an attempt to interpret MAR and Jacobs externalities on one line rather than understanding them separately (Boschma, 2017; Frenken et al. 2007), so it was not included in this analysis because there are incompatible differences in the main explatory variables and analysis contexts.

$$
\begin{aligned}
&\text{Innovation} \\
&= \varphi\text{agglomeration externalities(specialization, diversity, urbanization etc.)} \\
&+ \beta_1\text{firm characteristics(size, age, R\&D, firm hunam capital, etc.)} \qquad (3.2) \\
&+ \beta_2\text{regional characteristics(human capital)} \\
&+ \text{Others(ex. other knowledge sources)}
\end{aligned}
$$

The latter are designed to ensure that the included study: (i) use dummy, patent/product/innovation count and share product innovation in sales as dependent variable; (ii) discusses and documents the data used, the estimation methodology in light of theoretical and econometric literature; and (iii) reports 'regression coefficient' estimates to together with standard errors or t-values and associated sample sizes. The process led to inclusion of 15 primary studies. The number eventually increased to 26 as a results of discovering new studies through snowballing and manual search. For our baseline analysis, database comprises 413 estimates of the effect-size – agglomeration externalities, from different 26 primary studies. Table A1 and Table A2 provides information about their main characteristics and summary of studies' indicators of agglomeration externalities respectively.

The estimates in the baseline sample range between a minimum of -0.2651 (PCC by Anokhin, 2019) and maximum of 0.7742 (PCC by Hornych and Schwarz), with a median of 0.00777. The innovation indicators used in 26 primary studies can be largely divided into three groups: stock for patent counts of patent counts ifself or the number of innovstions (Y_count), innovation dummy including product, process and radical innovation dummy (Y_Dummy), innovation measured by the share of total sales of new products developed through innovation (Y_Sale). Table 3-1 summarizes the primary studies using the innovation indicators of each group. 41.4% use innovation dummy measure as a dependent variable, with count (patent/innovation/product) for 39.7% and innovative sales for 18.9%.

For the grouping of indicators of agglomeration indicators, the measures used by externalities type in the primary studies listed at the left side and classified into subgroup (middle part) considering specific chracteristics of indicators. In order to confirm the average effect and heterogeneity of each type of agglomeration externalities, specialization and diversity were classified, but characteristics that could not be considered as the same group in meta-analysis were distinguished. As a result, it was largely divided into four groups: *'Dummy', 'Cluster', 'Specialization'*, and *'Diversity'*.

To evaluate the degree of heterogeneity in the reported estimates of the agglomeration effects on innovation, I calculate the classic Cochran's Q-statistica and the I^2 index. The Q-statistics measures the weighted sum squares of the differences between study estimates and the fixed effects average estimate. The I^2 index is equal to (Q-(n-1))/Q and quantifies the proportion of total variation in the estimates that is due to heterogeneity between studies, as opposed to sampling variability (Higgins and Thompson, 2002; Higgins et al., 2003) In baseline dataset (n=413), Q=10412.23 and I^2=99.87%, which means a very high degree of heterogeneity. This means that there is substantial variation between studies' estimation that should be accounted for. In section 3.3.3, this study employ meta-regression analysis to provide explanation for this variation.

Table 3-1 Classification of innovation measures used in primary studies for meta-analysis

GROUP NAME	VARIABLE DESCRIPTION	N	PRIMARY STUDIES
Y_COUNT	Patent count/patent stock Innovation count	164	Anokhin(2019), Ascani et al. (2020), Baptista and Sawann (1998), Beaudry and Breschi (2003), Capozza et al. (2018), Feldman and Audretsch (1999), Hornych and Schwarz (2009), Huang et al. (2012), Nieburhr et al. (2020), Panne and Beers (2006), Tavassoli and Carbonara (2014)
Y_DUMMY	Innovation dummy Product innovation dummy/process innovation dummy/service innovation dummy Radical innovation dummy	171	Aarstad et al. (2016a), Aarstad et al. (2016b), Bettiol et al. (2019), Cook et al. (2013), Beule and Beveren (2008, 2012), Shefer and Frenkel (1998), Smit et al. (2015), Zhang (2015)
Y_SALE	ln(total sales of new products) share of new product sales	78	Beugelskjik (2007), Boschma and Weterings (2005), Czarnitzki and Hottenrott (2009), Beule and Beveren (2008, 2012), Grashof (2021)

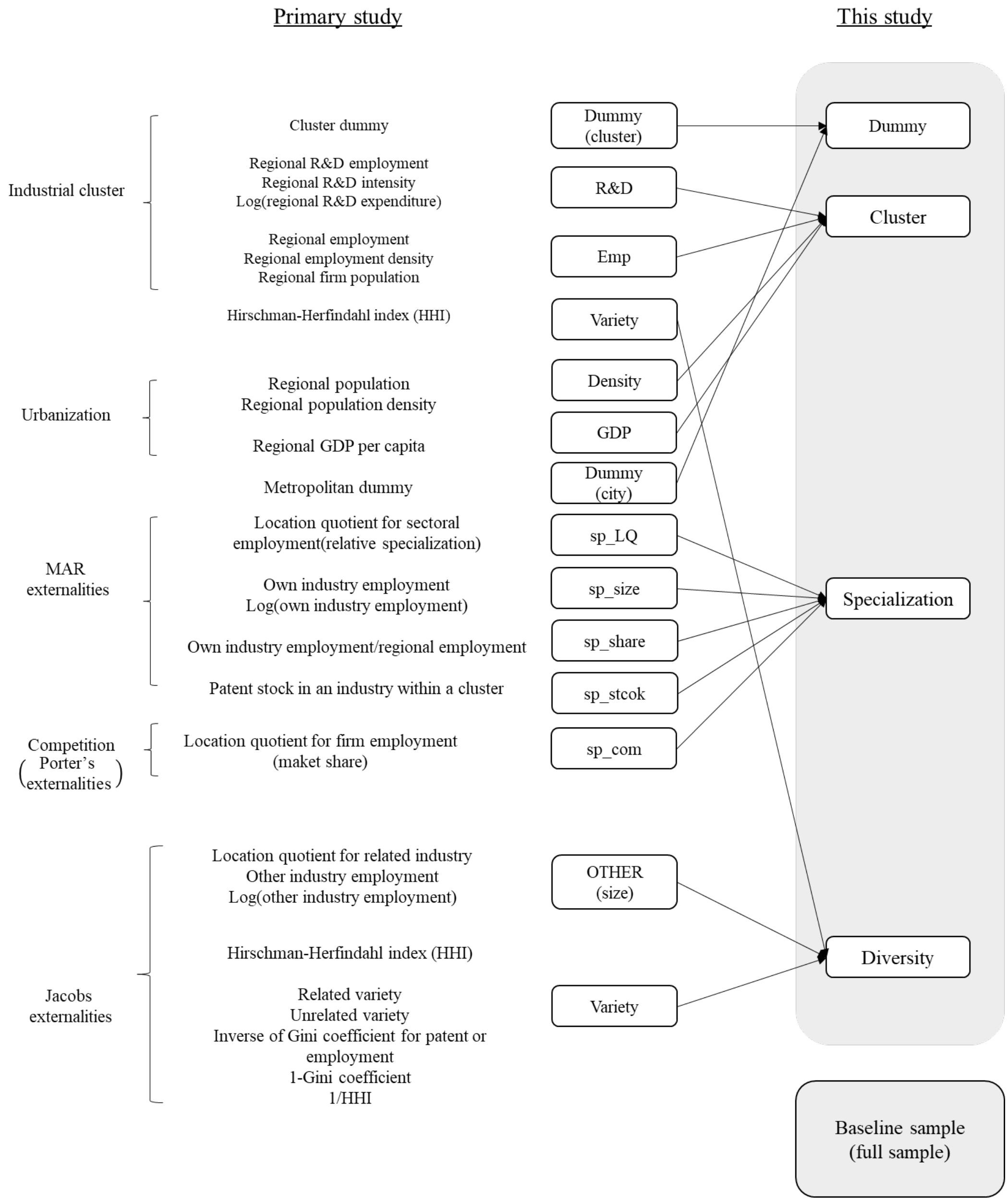


Figure 3-1 Various types and classification of measures for agglomeration externalities

Table 3-2 Description of measures for innovation and agglomeration externalities

Variables in primary studies			Variable description: equation	N		Type
innovation	Y_count		Patent/innovation/product count	164		Innovation/invention
	Y_dummy		Innovation dummy: product(112), process(35), service(24)	171		Innovation
	Y_sale		share of new products in total sales	78		Commercialization
Agglomeration	Dummy		Cluster dummy Metropolitan city dummy	36		Overall agglomeration effect
	Cluster	Density	Population density, ln(population density), ln(employment density): per km2	34	76	Scale effect (including urbanization)
		R&D	Regional R&D intensity, regional R&D employees, innovation intensity, log(regional R&D)	21		
		Emp	emp_r/emp_{nation}: (22), Total regional employment, ln(population): (7)	19		
		GDP	ln(GDP per capita)	2		
	Specializatio	sp_size	Own employment, ln(own employment), own firm population,	52	155	Marsall’s

Variables in primary studies		Variable description: equation	N		Type
n	sp_com	$location\ quotient_{firm} = \frac{(firm\ in\ region - industry)/(employment\ in\ region\ - industry)}{(firms\ within\ an\ industry\ in\ nation)/(total\ employment\ within\ an\ industry\ in\ nation)}$ Herfindahl-Hirshman index(HHI, sum of firm share)	33		externalities
	sp_LQ	$location\ quotient_{employment} = \frac{(Industry\ employment\ in\ region)/(total\ employment\ in\ region)}{(Industry\ employment\ in\ nation)/(total\ employment\ in\ nation)}$ ln(location quotient)	64		
	sp_share	Regional share of own industry	4		
	sp_stock	Patent stock in an industry within a cluster	2		
Diversity	Other	Other employment, patent stock(other industry), weighted sum of patent citation(other) Related science base-location quotient Agglomeration of services, log(other employments)	53	150	Jacobs's externalities Inter-industry spillover(size effect)

Variables in primary studies	Variable description: equation	N	Type
Variety	Inverse of Gini coefficient for patent of employment, 1-Gini coefficient $related\ variety_{i,r} = \sum_{j \in J_i} \frac{emp_{j,r}}{emp_{i,r}} \ln(\frac{emp_{i,r}}{emp_{j,r}})$, Ji=4 or 5digit industry level, i=2digit $unrelated\ variety_{i,r} = \sum_{k} \frac{emp_{k,r}}{emp_{r}} \ln(\frac{emp_{r}}{emp_{k,r}})$, HHI index(sector share), 1-HHI, 1/HHI, 1-GINI, 1/GINI	97	Jacobs's externalities Inter-industry spillover (composition effect)

3.3.3 **Publication bias and average estimate of agglomeration effects**

Before explaining the variation in studies' estimates, I compute the average estimate of the agglomeration effect with full sample, specialization and diversity and text for the presence of publication bias in this literature.

In meta-analyses the combined estimate of the effect-size is often obtained using either fixed effects or random effects estimators. They are both weighted average of the effect-sizes reported in the primary studies. The fixed effects estimator assumes that there is only one true effect-size, common to all studies, and that the observed variability in the reported estimates comes only from sampling variation. On the contrary, random effects estimator accounts for the presence of heterogeneity, as it considers that studies have different true effect sizes; consequently, the oberseved variability in the reported estimates comes not only from sampling error – within studies variation – but also from difference in studies true effect-sizes – between studies variation. Due to the heterogeneity detected in the previous section, this study uses the random effects speficiation in the estimations performed throughout the paper.

Publication bias has long been recognized as an important problem in empirical research. In is most frequent form, publication bias arises when statistically significant results are more likely to be produced and published by authors and journals than non-significant results. This lead to a distortion in empirical results, as the effect under analysis tends to be overestimated. Publication bias has been abundantly addressed in meta-analyses in many research areas, including economics (Card and Krueger, 1995; Doucouliagos, 2005; Doucouliagos et al., 2005; Stanley, 205; Stanley et al. 2008)

The funnel plot is a tool widely used to detect graphically the presence of publication bias and simultaneously to obtain an idea of the average effect. Popularized by Egger et al. (1997), the funnel plot is a scatter diagram that displays the estimates of the effect-size in the horizontal axis and their precision (usually measured by the inverse of the standard errors reported in the primary studies) on the vertical axis. As thoroughly explained by Stanley (2005), in the absence of publication bias, estimates of the effect-size will vary randomly and symmetrically around the mean, the dispersion being higher in studies with lower precision. In this case the diagram will take the shape of a symmetrical inverted funnel. But if there is publication bias favoring a certain direction, studies with higher standard errors (lower precision) tend to present estimates with a higher magnitude and biased toward that direction. In this case the diagram will be asymmetrical especially in its lower part. Thus, the (a) symmetry of the funnel plot is the key to assessing publication bias. Figure 3-1 shows the funnel plots for our dataset, one with precision=1/SE on the y-axis (Figure 3-1, left) and another in which precision appears in log scale for better visualization due to its high amplitude (Figure. 3-2, right). In addition, Figure 3-2 shows the funnel plots for the effect of specialization measures on innovation (left) and the effect of diversity measures on innovation (right).

There seems to be evidence of publication bias, as the point estimates of the agglomeration effect are asymmetrically distributed around the average (0.035). The conclusions revealed by visual inspectio of the funnel plot can be formally tested by running a simple regression of the effect-sizes on the respective standrad errors:

$$pcc_i = \beta_o + \beta_1 SE_i + u_i \tag{3.1}$$

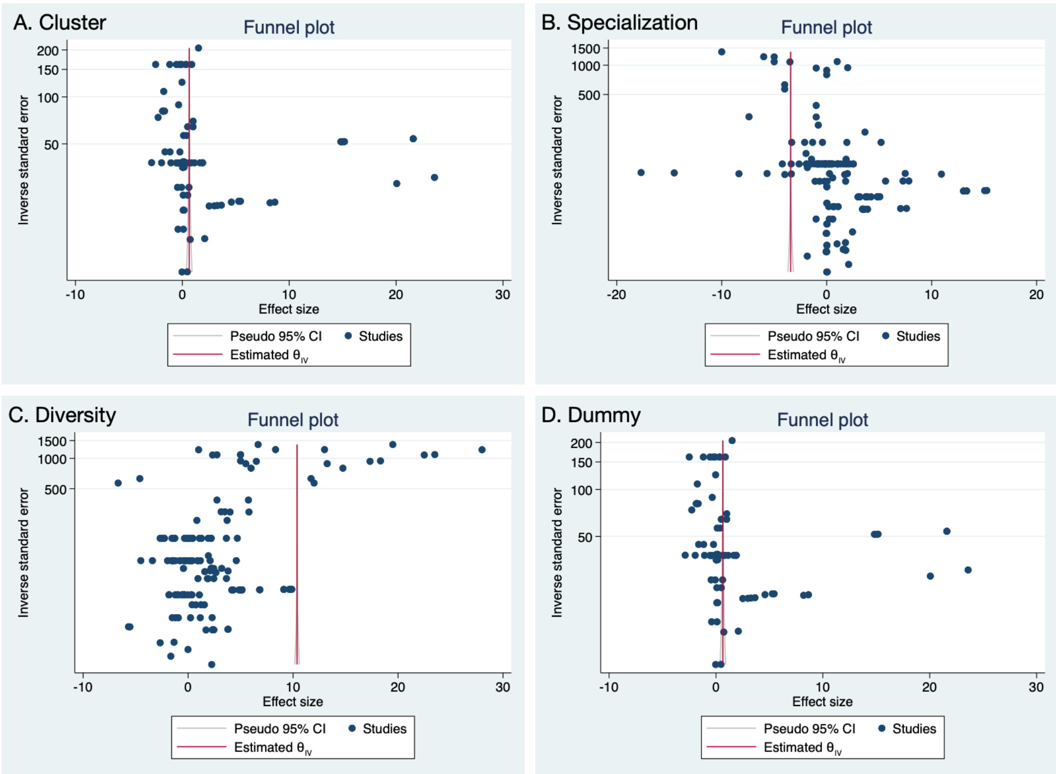


Figure 3-2 Funnel plots (by sub groups)

Note: The type of agglomeration assined by primary studies was referred to as the classification according to the index that measured agglomeration. For example, the specialization group included regression coefficients estimated using variables described as indicators for measuring specialization or MAR externalities in primary studies. Regression coefficient reflecing the cluster's holistic externalities are related to the size and agglomeration of economies activities such as production, consumption, employment, etc. within geographical units, and the Diversity group covers the regression results when the size of other industries or variety are used as independent vatiables.

In the presence of publication bias, authors of studies with small samples and higher standard errors will tend to search more intensively (from datasets, estimation techniques, and model speficiation) for higher estimates of the effect-size in order to report statistically significant results. Thus pcc will be uncorrelated with SE_i, as the reported estimates will vary randomly around the average effect, β_0, regardless of the standard error (Stanley, 2005).

Eq. (3.1) can thus be used to test for the presence of publication bias and simultaneously to estimate the average of the effect-size after controlloing for publication bias. However, its estimation by OLS has two critical problems. First, given that each reported effect has its own standard error, the disturbance u_i are heteroskedastic. This problem can be easily corrected by implementing the usual procedure of dividing both sides Eq. (3.1) by SE (Stanley, 2005), which leads to:

$$pcc_i/SE = \beta_o/SE + \beta_1 + u_i \tag{3.3}$$

$$t_i = \beta_o precision_i + \beta_1 + u_i \tag{3.4}$$

where $t_i = pcc_i/SE$ is the conventional t-statistic associated with pcc_i reported in the primary studies and precision = $1/SE_i$, Given that the coefficient are now reversed, testing for the intercept in Eq. (3.3), β_0, being equal to zero is a test for the presence of a significant average effect beyond publication bias (Precision Effect Test – PET) (Egger et al., 1997; Stanley, 2005; Ugur et al., 2016). However, some authors (e.g. Moreno et al., 2009; Stanley and Doucouliagos, 2012, Ch. 4) have suggested that a specification based on a quadratic relationship between the effect-sizes and their standard errors is more appropriate than a linear speficiation to correct for publication bias when there is significant average effect. In this case, the precision-effect estimate with standard error(PEESE) may be preferred to the PET/FAT. The equation estimated under the PEESE specification is (Stanley and Doucouliagos, 2012; Ugur et al. 2016):

$$t_i = \beta_o precision_i + \beta_1 SE_i + u_i \quad (3.5)$$

The second problem in estimating Eq. (3.1) by OLS is the presence of statistical dependence. When several observations ar drawn from the same study, they share the same datasets, specifications or estimation procedures, and therefore are likely to be correlated (Hunter and Schmidt, 1990; Nelson and Kennedy, 2009). In this case, OLS produces biased estimates. The easiest way to address this issues is to choose only one estimate from each study (Stanley, 2001; Lipsey and wilson, 2001). However, this would generally lead to a considerable reduction in the size of the meta-sample, while is not desirable when the number of studies is limited. If several observations from each study are to be sued in the meta-analysis, then multilevel models, panel data estimators, clsutered data anlaysis, or bootstrapped standard errors can be employed to address the problem of whitn-study correlation (Nelson and Kennedy, 2009; Doucouliagos and Laroche, 2009). We choose to estimate Eqs. (3.3), (3.4) using multilevel linear models, since they not only correct the standard errors for within-study correlation, but also estimate the regression coefficients allowing for the precence the heterogeneity between studies (Ugur et al. 2016). Exmples of meta-analyses in economics that have used hierachical linear models are Bateman and Jones (2003); Johnston et al. (2005) and Ugur et al. (2016)

In the multilevel models, observations are nested into groups with different characteristics. Thus, differences in individual observations can be attributed to both within-group variation and between-group variation. The model's coefficients are allowed to vary randomly between groups. In its most generic form, a multilevel random-coefficient univariate model[4] of the dependent variable Y_{ij} on explanatory variables X_{ij} can be written as:

$$Y_{ij} = (\beta_o + \gamma_{oj}) + (\beta_1 + \gamma_{1j})X_{ij} + \varepsilon_{ij} \tag{3.6}$$

where subscript i refers to observations and subscript j refers to groups; γ_{oj} and γ_{1j} are the group-specific intercept and slope, respectively, which are assumed to follow a normal distribution. This generic version is called the random coefficient model, as it allows both the intercept and the slope to vary randomly across groups. If only the intercept is allowed to vary across groups (in which case the slope is assumped to be fixed and the variance of γ_{1j} is zero), we have a random intercept model; if only the slope is allowed to vary across groups (in which case the intercept is assumed to be fixed and the variance of γ_{oj} is zero),

[4] The multilevel model was also used in the analysis of Chap6, and for a detailed explanation, see Chap6.

this study uses a random slope model.

The multilevel structure can be applied in meta-analysis, as the observations (estimates of the effect size) are nested in roups (studies), that have different characteristics (random variation). Therefore, this study estimates Eq (3.4) and (3.5) for baseline models for the full sample and subgroups for indicators that reflect expertise and diversity using multi-level models that follow the PET/FAT and PEES specifications. Results are reported in Table 3-1.

The upper part of the table shows of the estimation of coefficient using as random the slope associated with varible precision. Estimations are obtained by maximum likelihood. In both PET/FAT and PEESE speficiations, we reject that β_1 =0, which confirms that there is evidence of publication bias in the empirical literature estimating agglomeration regardless of types of externalities. In addition, the averages of the agglomeration effect are 0.00549 (in the baseline model for FAT/PET) and 0.00613 (in the baseline model for PEESE), meaning that the overall primary studies indicate, on average, small but staitistically significant overall agglomeration effects.

The middle part of the table presents the estimates of the variances of the random slopes associated with precision, β_o, and of the residulas, as well as the respective confidence itnervals at 95%. In both speficiations the confidence interval for the variance of the slope associated with precision suggests that is is significant, which further confirms the adequacy of the random slope model adopted and the existence of heterogeneity in the reported effect-sizes. The likelihood ratio (LR) test presented in the last line clearly shows that the hierarchical models are preferred to a simple OLS model.

Before explaining the variation in studies' estimates, I compute the average estimate of the agglomeration effect with full sample, specialization and diversity and text for the presence of publication bias in this literature.

In meta-analyses the combined estimate of the effect-size is often obtained using either fixed effects or random effects estimators. They are both weighted average of the effect-sizes reported in the primary studies. The fixed effects estimator assumes that there is only

3.3.4 **Variables Setting**

The moderating variable linked with the analytical dimension of the research field are all related to the specification of the innovation production function used in the primary studies. This study include (1) a set of dummies that identify the variable used to measure innovation (dummy, sales, count); (2) A set of dummies for agglomeration measures, cluster (population density, regional R&D for firm-level anlaysis), specialization (size for own industry to which the agent belongs to, location quotient), and diversity (size of other industries, composition such as herfindahl-hershiman index, related variety, unrelated variety, theil index etc.), cluster dummy, (3) a dummy for the inclusion in primary regressions of alternative other knowledge source (i.e. collaboration, export, MNE for outside region and research institute and university for within regions) (4) absorptive capacity, for firm level analysis (firm size, R&D intensity etc.).

The moderating variables linked with the empirical dimension of the research filed are dummies for (1) the inclusion of time, industry, region, firm dummy in the estimation by the primary studies (2) the sample being composed of only panel data, (3) the esimation employing instrumental variables, 2SLS or multilevel anlaysis to control possible endogeneity or dependency (4) the sample using data (firm, region, region-industry) and (5) dummy for log form in dependent or independent variables. Finally, the number of observations and dummy for sample countries (Asia, America, Europe) included in each regression are also considered as moderating variables. Table 3-3 summarizes the description of each variable.

Table 3-3 Vatiables settings

	moderator	Type	
Variables linked with the model specification in the primary study			
innovation	Y_count	Dummy	1 if innovation measures are patent/innovation/product count variables; 0 otherwise
	Y_dummy	Dummy	1 if innovation measures are dummy variables; 0 otherwise
	Y_sale	Dummy	1 if innovation measures are sales share for innovative product; 0 otherwise
cluster	Cluster	Dummy	1 if cluster measure is related with entire size (population density, regional R&D/employment); 0 otherwise
	Specialization	Dummy	1 if cluster measure is related with size of own industry(own employment, location quotient etc.); 0 otherwise
	DIVERSITY	Dummy	1 if cluster measure is related with industrial diversity within a region(related variety, unrelated variety, HHI etc.); 0 otherwise
	Dummy	Dummy	1 if cluster measure is dummy variable; 0 otherwise
TL3(geographical unit)		Dummy	1 if cluster measure is measured in TL3/4 geographic units; 0 otherwise
TL4(geographical unit)		Duumy	1 if cluster measure is measured in TL4 geographic units; 0 otherwise
Absorptive capacity		Dummy	1 if the model specification in the primary study includes variables related with absorptive capacity; 0 otherwise
Add knowledge source (within region)		Dummy	1 if the model specification in the primary study includes another knowledge source within a region; 0 otherwise
Add knowledge source (outside region)		Dummy	1 if the model specification in the primary study includes another knowledge source outside a region; 0 otherwise
Cumulative knowledge base (internal)		Dummy	1 if the model specification in the primary study includes cumulative knowledge base (knowledge stock); 0 otherwise
Human capital		Dummy	1 if the model specification in the primary study includes human capital variables; 0 otherwise

(continued)

Variable	Type	
Variables linked with the model specification in the primary study		
lnX	Dummy	1 if the X (cluster/agglomeration measures) has the logarithm form; 0 otherwise
lnY	Dummy	1 if the Y (innovation measures) has the logarithm form; 0 otherwise.
Industry effect	Dummy	1 if the model specification in the primary study includes industry effect(industry dummy); 0 otherwise
Region effect	Dummy	1 if the model specification in the primary study includes region effect(industry dummy); 0 otherwise
Time effect	Dummy	1 if the model specification in the primary study includes time effect(industry dummy); 0 otherwise
Firm effect	Dummy	1 if the model specification in the primary study includes firm effect(industry dummy); 0 otherwise
Variables linked with the sample		
Europe	Dummy	1 if sample includes only Europe countries; 0 otherwise
America	Dummy	1 if sample includes only America countries; 0 otherwise
Manufacture	Dummy	1 if sample includes only manufacture industries; 0 otherwise
Service	Dummy	1 if sample includes only service industries; 0 otherwise
High-tech	Dummy	1 if sample includes only high-tech industries; 0 otherwise
Panel	Dummy	1 if sample is a panel database; 0 otherwise
Regional data	Dummy	1 if sample is for regional database; 0 otherwise
ML(multilevel analysis)	Dummy	1 if estimation is with multilevel method; 0 otherwise
2SLS/IV(endogeneity)	Dummy	1 if estimation is with 2SLS/IV method; 0 otherwise
observations	count	Number of observations in each estimates
Time	count	Median of analysis period in the primary study

3.4 Empirical results and discussion

In this section, the study employs multilevel meta-regression to examine the variation in reported estimates of agglomeration effects and explore how differences in the characteristics of primary studies contribute to this variation. The methodological characteristics are represented by dummy variables and categorized as analytical or empirical dimensions in Table 3-3.

To estimate the multilevel meta-regression, the study utilizes the random coefficient model. First of all, in order to determine the appropriate model specification, a two-level and three-level analysis was conducted on the research group and the innovation index group. As can be seen in Table 3-4, the variance of the error term for the type of innovation indicator was not significant, so the two-level model for each paper for 26 studies was adopted as the basic model. However, to address multicollinearity between covariates, a strategy similar to Stanley and Doucouliagos (2005) is followed, known as the general-to-specific (G-to-S) approach. Specifically, each regression eliminates covariates one by one based on their higher p-values. In this study, not only backward selection but also forward selection was performed to analyze full samples. As the results, selected variables are y-sale, cluster, diversity, absorptive capacity, logX, y_count, Middle-east, time, and industry-dummy. Among them, the variables that are determined to be robust by being selected from both forward selection and backward selection are Cluster, Diversity, Absolute capacity, and logX. (Table 3-4)

Table 3-4 Results for Meta-regression analysis for full sample by using variable selection method (Forward/Backward selection)

	FAT/PET				PEESE			
	Forward selection		Backward selection		Forward selection		Backward selection	
t_pcc	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8
precision	0.00506***	0.00516***	0.00467***	0.00501***	0.00593***	0.00598***	0.00489***	0.00512***
	(0.00143)	(0.00145)	(0.00137)	(0.00143)	(0.00141)	(0.00143)	(0.00142)	(0.00146)
y_sale	-1.532**	-1.557**			-1.518**	-1.536**		
	(0.718)	(0.729)			(0.727)	(0.737)		
Cluster	1.754***	1.752***	1.722***	1.723***	1.912***	1.903***	1.738***	1.738***
	(0.579)	(0.585)	(0.569)	(0.580)	(0.577)	(0.582)	(0.574)	(0.583)
Diveristy	2.840***	2.866***	3.054***	3.087***	2.907***	2.926***	2.919***	2.972***
	(0.664)	(0.669)	(0.668)	(0.674)	(0.666)	(0.671)	(0.664)	(0.671)
Absorptive capacity	-2.013***	-1.972***	-1.744**	-1.546*	-1.321**	-1.293*	-1.722**	-1.554*
	(0.745)	(0.764)	(0.830)	(0.880)	(0.667)	(0.684)	(0.853)	(0.894)
logX	1.539*	1.495*	1.443*	1.321	1.847**	1.801**	1.417*	1.317
	(0.808)	(0.822)	(0.769)	(0.805)	(0.796)	(0.809)	(0.786)	(0.815)
y_count			2.156**	2.314*			1.851*	1.998*
			(1.074)	(1.192)			(1.109)	(1.213)
Middleeast			-3.683**	-3.711**			-4.894***	-4.926***
			(1.586)	(1.793)			(1.620)	(1.792)
time			0.117**	0.118*			0.000631	0.000632
			(0.0568)	(0.0658)			(0.000556)	(0.000593)
Industry_dummy			-1.421*	-1.626**			-1.302*	-1.515*
			(0.746)	(0.810)			(0.776)	(0.828)
pcc_SE					18.57	17.91	10.16	8.752
					(12.14)	(12.43)	(13.69)	(14.36)
constant	1.769**	1.756**	-233.6**	-235.2*				
	(0.719)	(0.746)	(113.8)	(131.8)				
RE variance								
Var(study)	0.817***	0.878***	0.575***	0.764***	0.862***	0.915***	0.670***	0.821***
	(0.186)	(0.186)	(0.217)	(0.218)	(0.187)	(0.186)	(0.206)	(0.209)
Var(residuals)	1.331***	1.337***	1.330***	1.335***	1.333***	1.339***	1.330***	1.336***
	(0.0360)	(0.0362)	(0.0361)	(0.0362)	(0.0360)	(0.0362)	(0.0361)	(0.0363)
N	413	413	413	413	413	413	413	413
Log Likelihood								

Note: Standard error in parentheses (* p < 0.10, ** p < 0.05, p < 0.01)

Table 3-4 presents the estimation results of the specific model using the PET/FAT specification. Model 1 represents the baseline estimation obtained through maximum likelihood. In column (2), the estimation (Model 2) is performed using a three-level regression. Model 3 and 4 are estimations similar to Model 1 but exclude the PEESE model specification. The study also estimates the 'cluster' (Model 1 and 5), 'specialization' (own employment size, location quotient, population density, etc., Model 2, 6), 'diversity' (Model 3 and 7) subgroup (related variety, unrelated variety, Herfindahl and Hirschman index), and the 'dummy' (Model 4, 8) subgroup, respectively. When subgroups are analyzed separately, only the 'diversity' group shows a statistically significant positive effect (0.00981 for FAT/PET and 0.0101 for PEESE).

Analyzing the heterogeneity variables of independent and dependent variables for subgroups, as shown in Table 3-5 and Table 3-6, it is observed that the innovation/patent count variable has a higher estimate compared to other indicators in the 'cluster' and 'dummy' subgroups, and this result holds even when using the PEESE model. This suggests that using patent/innovation count as dependent variables tends to increase the agglomeration effects. In the case of indicators reflecting the scale effect of agglomeration (coefficients of 'cluster' and 'cluster dummy'), positive and significant results are observed, indicating support for urbanization rather than specialization and diversity. Based on the results presented in Tables 3-4, 3-5, and 3-6, the most significant sources of variation in agglomeration effects are the diversity measures (size of other industries, related variety, unrelated variety, etc.).

In summary, when controlling for heterogeneity by study in the two-level analysis model, only the diversity group exhibits a positive effect. Dummy variables for industrial, national, methodology, and absorption capacity-related control variables are not insignificant. However, significant variables are found among the dummy variables reflecting heterogeneity in measures for innovation and agglomeration externalities, as shown in Tables 3-5 and 3-6. Based on these results, a significant coefficient is controlled for heterogeneity by indicator, and subsequent industry/continent analysis is conducted, but no significant results are found.

Table 3-5 Estimation of FAT/PET(Eq. 3.3) and PEESE(Eq. 3.4) for full sample: dependent variable(t_pcc)

Full Sample	**FAT/PET**		**PEESE**	
t_pcc: dependent variable	Model1	Model2	Model1	Model2
precision	0.00549***	0.00660***	0.00613***	0.00703***
	(0.00147)	(0.00140)	(0.00145)	(0.00139)
SE			13.13	9.401
			(12.04)	(12.89)
constant	1.319**	1.037		
	(0.620)	(0.807)		
RE variance				
Var(study)	7.749	6.422	8.430	6.325
	[4.071;14.75]	[6.422;2.100]	[4.400;16.15]	[3.323;12.04]
Var(innovation_type)		1.031		1.634
		[0.060;17.86]		[0.148;18.10]
Var(residuals)	15.73	15.38	15.78	15.39
	[13.67;18.10]	[15.38;1.107]	[13.71;18.16]	[13.36;17.72]
N. obs. (N. Studies)	413(26)	413(26)	413(26)	413(26)
Log likelihood	-1179.62	-1176.57	-1181.15	-1176.97
Wald test	13.92***	22.26***	19.14***	25.77***
LRtest(Model2 vs. Model1)		6.08**		8.36***
LRtest(ML vs. OLS)	124.82***	130.90***	128.46***	136.82***

Note: Standard error in parentheses (* $p < 0.10$, ** $p < 0.05$, $p < 0.01$)

Table 3-6 Estimation of the multivariate meta-regression (FAT/PET) for sub groups. Dependent variable: pcc/se=t

	FAT/PET				PEESE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T_PCC	Cluster	Specialization	Diversity	Dummy	Cluster	Specialization	Diversity	Dummy
PRECISION	0.0022	-0.00279	0.00981***	0.0181	0.0145	-0.0021	0.0101***	0.0365***
	(0.0161)	(0.00181)	(0.00155)	(0.019)	(0.0143)	(0.0018)	(0.00149)	(0.0131)
SE					4.84	8.099	4.19	14.8
					-15.7	-14.98	-18.57	-15.25
CONSTANT	2.283	1.680*	0.496	1.634				
	(1.777)	(0.94)	(0.669)	(1.087)				
RE_STUDY(VAR)	1.733***	1.269***	0.692***	0.426	1.791***	1.343***	0.713***	0.510*
	(0.193)	(0.19)	(0.249)	(0.305)	(0.19)	(0.19)	(0.249)	(0.306)
RESIDUALS(VAR)	0.773***	1.029***	1.312***	-0.362***	0.771***	1.031***	1.313***	-0.359***
	(0.0909)	(0.06)	(0.0602)	(0.138)	(0.0906)	(0.0601)	(0.0603)	(0.138)
LOG LIKELIHOOD	-191.478	-399.423	-418.531	-42.6533	-191.744	-400.774	-418.793	-43.1461
WALD TEST	0.61	3.38	40.13***	1.04	2.55	2.61	46.79***	12.56***
N	76	155	150	32	76	155	150	32

Standard errors in parentheses *P<0.10 **<0.05 ***<0.01

Table 3-7 Meta-regression results for heterogeneity based on innovation measures by subgroups

	FAT/PET				PEESE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T_PCC	Cluster	Specialization	Diversity	Dummy	Cluster	Specialization	Diversity	Dummy
PRECISION	0.000293	-0.00183	0.0110***	0.0288**	-0.00351	-0.00146	0.0112***	0.0256***
	(0.0155)	(0.00180)	(0.00163)	(0.0128)	(0.0141)	(0.00179)	(0.00159)	(0.00794)
SE					-6.784	13.89	5.593	-11.76
					(15.98)	(14.47)	(20.91)	(9.757)
Y_COUNT	6.245**	1.732	0.509	2.973***	6.140***	2.619**	0.793	3.250***
	(2.538)	(1.515)	(1.294)	(0.840)	(2.083)	(1.184)	(1.019)	(0.739)
Y_SALE	-0.0772	-2.166**	-2.347**	0.583	-0.0954	-2.085**	-2.304**	0.573
	(1.114)	(0.939)	(1.161)	(0.464)	(1.094)	(0.927)	(1.170)	(0.462)
CONSTANT	-0.512	1.217	0.421	-0.408				
	(2.023)	(1.161)	(0.954)	(0.874)				
RE_STUDY(VAR)	1.548***	1.221***	0.705***	-0.135	1.543***	1.209***	0.701***	-0.234
	(0.197)	(0.193)	(0.253)	(0.356)	(0.196)	(0.194)	(0.255)	(0.365)
RESIDUALS(VAR)	0.774***	1.006***	1.295***	-0.406***	0.774***	1.008***	1.296***	-0.408***
	(0.0909)	(0.0601)	(0.0604)	(0.139)	(0.0907)	(0.0601)	(0.0604)	(0.139)
LOG LIKELIHOOD	-180.156	-395.84439	-416.216	-35.5696	-189.097	-395.939	-416.278	-38.0087
WALD TEST	6.25*	10.91***	45.64***	15.90***	10.24**	13.27***	53.34***	65.91***
N	76	155	150	32	76	155	150	32

Standard errors in parentheses *P<0.10 **<0.05 ***<0.01

Table 3-8 Meta-regression results for heterogeneity based on agglomeration measures by subgroups

	FAT/PET				PEESE			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
T_PCC	**CLUSTER**	**SPECIALIZATION**	**DIVERSITY**	**DUMMY**	**CLUSTER**	**SPECIALIZATION**	**DIVERSITY**	**DUMMY**
PRECISION_PCC	0.00232	-0.00296*	0.00962***	0.0181	0.00819	-0.00299*	0.00947***	0.0365***
	(0.0108)	(0.00177)	(0.00152)	(0.0190)	(0.0115)	(0.00176)	(0.00147)	(0.0131)
DENSITY	-1.366				1.854			
	(4.015)				(2.293)			
RD	-1.425				1.817			
	(4.149)				(2.510)			
GDP	-14.37***				-11.21***			
	(4.218)				(2.622)			
SP_COM		0.672				0.665		
		(0.647)				(0.644)		
SP_SIZE		3.678**				3.595***		
		(1.631)				(1.363)		
SP_SHARE		0.597				0.479		
		(4.483)				(4.275)		
SP_STOCK		8.554***				8.463***		
		(2.750)				(2.586)		
OTHER			2.874***				2.892***	
			(0.735)				(0.721)	

	FAT/PET				PEESE			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
T_PCC	**CLUSTER**	**SPECIALIZATION**	**DIVERSITY**	**DUMMY**	**CLUSTER**	**SPECIALIZATION**	**DIVERSITY**	**DUMMY**
PCC_SE					6.216	-1.352	-13.04	14.80
					(10.81)	(15.03)	(19.12)	(15.25)
CONSTANT	3.691	-0.136	-0.311	1.634				
	(3.534)	(1.322)	(0.724)	(1.087)				
STUDY(VAR)	1.913***	1.404***	0.790***	0.426	1.949***	1.400***	0.791***	0.510*
	(0.187)	(0.208)	(0.237)	(0.305)	(0.185)	(0.200)	(0.239)	(0.306)
RESIDUALS(VAR)	0.286***	0.980***	1.254***	-0.362***	0.283***	0.980***	1.255***	-0.359***
	(0.0908)	(0.0610)	(0.0605)	(0.138)	(0.0906)	(0.0606)	(0.0605)	(0.138)
LOG LIKELIHOOD	-164.531	-395.045	-411.461	-42.7203	-164.887	-395.046	-411.322	-43.2592
WALD TEST	95.87***	13.11**	57.53***	0.91	98.81	14.86**	64.41***	12.14***
N	76	155	150	32	76	155	150	32

Standard errors in parentheses *P<0.10 **<0.05 ***<0.01

3.5 Conclusions

In recent years, there has been a growing interest among researchers and policymakers in the concept of clusters as a strategy for promoting economic development. Clusters, which refer to geographically concentrated groups of interconnected firms and institutions, have been recognized for their potential positive impact on innovation, productivity, and resilience. However, the effects of clusters on innovation have been a subject of mixed findings and theoretical debates.

This study aims to investigate the effects of agglomeration, or clustering, on innovation. Innovation is a key focus because clusters are believed to facilitate knowledge spillovers that are crucial for innovative activities. Some clusters have successfully promoted innovation and contributed to local economic prosperity, such as Silicon Valley. Moreover, innovation is considered a vital driver of long-term economic growth in modern economic theories.

Improving innovation capacity has become a priority for firms, regions, and countries as they strive for better economic performance. Therefore, understanding the effects of clusters on innovation can provide valuable insights into regional economic development and offer policy implications for local authorities.

Existing knowledge on the impact of agglomeration externalities, or cluster effects, on innovation is inconsistent. Theoretical arguments suggest that clusters can both encourage innovation through knowledge spillovers and hinder it through "lock-in" effects.

Empirical studies examining the relationship between clusters and innovation have yielded conflicting results from various countries and time periods.

To overcome the limitations of individual studies, meta-analysis is proposed as a meaningful approach to combine and contrast empirical findings with diverse and contradictory outcomes. By pooling data from multiple studies, it becomes possible to reach robust and general conclusions. However, limited work has been conducted on the topic of clusters and innovation using this approach, except for notable papers by de Groot, Poot, and Smit (2010) and Fang (2015).

This study conducts a meta-analysis of relevant empirical studies on the relationship between clusters and innovation since the 1990s. It differs from previous work by de Groot, Poot, and Smit (2010) and Fang (2015) in several aspects. Firstly, while de Groot, Poot, and Smit focus on regional-level effects, this article pays equal attention to both regional-level and firm-level effects. Furthermore, this study explicitly compares results from studies conducted at the firm-level, industry-level, and regional-level, providing additional insights into whether clusters primarily impact individual firms, industries as a whole, or the entire region.

Secondly, de Groot, Poot, and Smit include studies with various dependent variables such as employment growth, productivity growth, and innovation. In contrast, this study narrows its focus to the effects of regional clusters on innovation, excluding papers that do not have an innovation-related dependent variable. This allows for a more meaningful calculation of an average effect size and facilitates clearer interpretation of the results, albeit with a smaller sample size.

Thirdly, Fang (2015) explored variables for all sources of heterogeneity, but this approach may lack theoretical relevance. Building on previous literature, this study attempts to differentiate the influence of physical dimensions in explaining the pattern of cluster agglomeration. This is achieved by classifying variables representing the size, diversity, and relatedness of agglomeration.

The study aims to answer three main questions: (1) What are the general conclusions drawn from previous studies regarding the impact of agglomeration externalities on innovation? (2) Are the estimated correlations between clusters and innovation homogeneous or heterogeneous across previous studies? (3) If there is heterogeneity, what variables act as moderators, influencing the direction and magnitude of clusters' effects on innovation?

Using a multilevel model, this article finds that, overall, clusters have statistically significant positive effects on innovation. However, significant heterogeneity is observed among individual studies, indicating the presence of moderators

In this section, the study utilizes multilevel meta-regression to explore the variation in reported estimates of agglomeration effects and understand how differences in primary study characteristics contribute to this variation. These characteristics are represented by dummy variables classified as either analytical or empirical dimensions.

According to the analysis results, the results of variable selection for the entire sample and the common conclusion from subgroup analysis are that the ‘Diversity’ group shows a relationship between positive significant innovation and agglomeration despite various heterogeneous characteristics of each study. In the overall sample analysis by variable selection, whether absorption capacity was considered or log transformation of independent variables showed significant results, but it is difficult to say that the model specification has strategic or policy implications, but it is noteworthy that the average effect of the Diversity group has a positive value. This is because it is difficult to draw general conclusions on local industrial policies as the emphasis on place-based policies or context specificity gradually intensifies in recent studies (Coe et al. 2020), but on average, it is more likely to expect positive effects when forming clusters through the size of other industries or the composition of various industries than specialization or overall size. In other words, from the results of this study, when information on places and industries is uncertain, the initial cluster policy can derive implications that the pursuit of 'Diversity' is likely to be effective.

On the other hand, the results of the sub-group did not show any significant results other than the heterogeneity between indicators of innovation and integration, indicating that the results of the 26 literature used in this study tend to be mixed or undecided. This is divided into three categories, Specialization, Diversity, and Competition, since Glaeser et al (1990), and causes reflection on existing practices that have measured and analyzed agglomeration with simple economic aggregate variables for comparison. In other words, it is argued that avoiding analyzing simple indicators by forcibly responding to abstract concepts and focusing on the direct meaning of the indicators is a necessary change in this literature.

Chapter 4. The effect of region-industry characteristics on firm innovation in different technological regime: A multilevel study

4.1 Introduction

Innovation has a clear geographic dimension that affects economic growth and technological change (Feldman & Kogler, 2010). Considering that knowledge is a crucial driving force for innovation, the relationship between intentional/unintentional knowledge flow and geographical proximity among economic actors are important in understanding the dynamics of the innovation process and thus devising regional innovation policies (Audretsch & Feldman, 2004). In this context, knowledge spillovers within the region have been actively studied based on the broadly agreed argument that knowledge transfer is localized and stimulated by geographical proximity, allowing firms operating adjacent to knowledge sources to innovate faster than those that do not (Breschi et al., 2001; Grillitsch & Nilsson, 2017).

Whether specialization (within an industry) and diversification (between industries) cause knowledge spillovers in the region or not has long been a controversial issue at the center of academic discussions in the literature on agglomeration economics

(Delgado et al., 2014; Glaeser et al., 1992; Porter, 1998; Rosenthal & Strange, 2004). A number of significant studies in this field have been analyzed but shown mixed results (Beaudry & Schiffauerova, 2009; de Groot et al., 2016). Recent overviews suggest that this could be partially fueled by the need to acknowledge the existence of heterogeneous firms in the same region, which means using data with a hierarchical structure and the previous findings that the patterns of regional knowledge spillovers are sector-specific and dependent on industrial attributes (Liang & Goetz, 2018; van Oort et al., 2012). However, there has been little empirical evidence to confirm or refute a precise relationship between industrial characteristics and actual benefit from regional knowledge spillovers, particularly in firm innovations considering the hierarchical structure of data. In addition, although significant studies have primarily focused on the benefits of regional knowledge spillovers, recent studies have reported that regional knowledge spillovers cause limited or even negative effects in some cases (Grillitsch & Nilsson, 2017). Nevertheless, few studies have suggested an integrated explanation of such bidirectional nature of regional knowledge spillover.

This paper aims to fill the aforementioned gaps in the literature on regional knowledge spillovers. Based on firm-level data of 3,664 Korean manufacturing companies in 24 industries and 17 regions, we explore the effect of regional and technological characteristics on firm innovation using the concept of technological regimes consisting of at least three dimensions (Malerba & Orsenigo, 1993; Winter, 1984): appropriability, technological opportunity, and technological cumulativeness. To this end, three-level multilevel modeling is applied, and the interaction effects between regional knowledge spillovers and technological regimes are tested. Agglomeration economies can also cause economies of scale and pecuniary externalities due to regional size or population density (Krugman, 1991; Martin & Sunley, 1998). In addition, the firm's knowledge acquisition channel is not only knowledge spillovers within the same region but also export or collaboration (J. Aarstad et al., 2016; Rodríguez-Gulías et al., 2021). Therefore, this paper considers regions' pecuniary externalities, economies of scale, and firms' other knowledge acquisition channels by using regional population density and collaboration dummy, export dummy at the firm level as independent variables. To model specialization and diversification, we apply the location quotient method for specialization and the entropy measurement of Shannon (1948) for diversification at the industry-region level. In doing so, regional knowledge spillovers that affect firm innovations in a specific industry within a specific region are intended to be more elaborated.

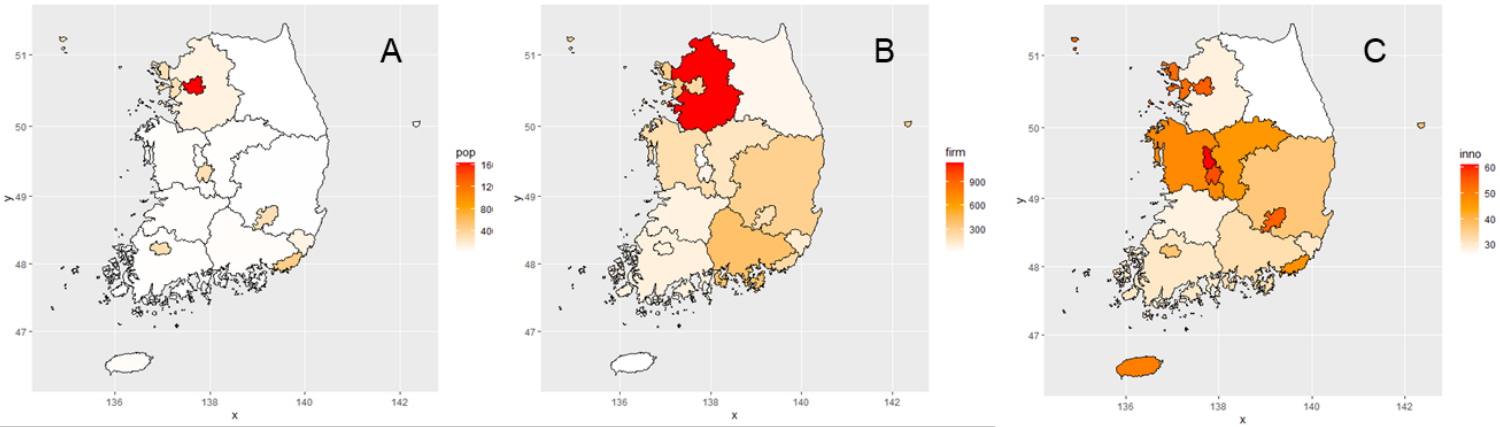
A technological regime is a framework that explains the technological conditions under which a firm implements innovation, defining the knowledge and learning environment for opportunities and constraints that seek to undertake innovative activities (Castellacci & Zheng, 2010; Malerba & Orsenigo, 1993; Revilla & Fernández, 2012). There have been studies examining the relationship between firm performance and various knowledge acquisition channels such as cooperation between firms, technology licensing, and networks under different characteristics of the technological regime in the firm-level strategic management literature (Herstad et al., 2014; Lee et al., 2017; Seo et al., 2017).

Because knowledge within a region is one of the knowledge acquisition channels, this paper argues that three dimensions of the technological regime may be significant contingency factors for regional knowledge spillovers by explaining their bidirectional characteristics and consequently can elucidate mixed results related to the debate on specialization and diversification.

This paper contributes to the literature in several ways. First, multilevel analysis allows the separation of the impacts of the regional, industrial (within a region) and firm characteristics on *firm* innovations. In this respect, this paper responds to the claim of integrating the discussion on regional and sectoral contexts at the firm level, gaining insights into how sectoral context (measured by technological regimes) shapes the effects of regional knowledge spillovers on firm innovations (Beaudry & Schiffauerova, 2009; Carreira & Lopes, 2018; Liang & Goetz, 2018). Second, our analysis shows that the characteristics of the technological regime can act as a boundary condition for the bidirectional characteristics of regional knowledge spillovers and maybe as another dimension describing the mixed results on specialization and diversification. Third, most literature used measures aggregated at the regional level to model specialization and diversification. However, we tried to accurately describe regional knowledge spillovers that affect firms by measuring specialization and diversification concerning industries within a region. In addition, we consider the pecuniary externalities and other firm-level knowledge acquisition channels to reduce the perils of confounding the effects of regional knowledge spillovers. Fourth, most studies have dealt with regional knowledge spillover on specialization and diversification in the context of the United States or European countries. This paper expanded the geographical category of the discussion using Korean data. Seoul, the capital city of South Korea, accounts for 0.6% of the total land, but about 20% of the total population lives in this city, about 13 times higher than the average population density in other regions. The red area shown in Figure is Seoul. Because of this specificity, the capital area (Seoul, Gyeonggi-do) dummy variable was included in the analysis

This paper is organized as follows. The following section addresses theoretical perspectives. The data and the empirical methods are subsequently presented. We then report our results and findings, and the final section concludes and provides the limitations of this study.

Figure 4-1 Population density (A), number of manufacturing companies (B), and proportion of innovative companies (C) at the si-do level in Korea (Sources: Author's work using data from 2016 Korean Innovation Survey (KIS): Manufacturing industry



4.2 Literature review

4.2.1 Geographical proximity and knowledge spillover

Knowledge dynamics within the region and the interrelation between these dynamics and firm performance have been discussed extensively for the last decades (J. Aarstad et al., 2016; Feldman & Kogler, 2010). Although different streams of literature have provided different arguments about what mechanisms are (i.e., networks, knowledge spillovers, labor mobility, local buzz), regional knowledge spillovers have been actively studied based on the broadly agreed argument. It is that knowledge transfer is localized and stimulated by geographical proximity, allowing firms operating adjacent to knowledge sources to innovate faster than those that do not (Audretsch & Feldman, 1996; Roper et al., 2017).

One explanation for the positive effects caused by geographically adjacent sources of knowledge relates to the spatially bounded transmission of tacit knowledge due to the contextualized and locally embedded (or sticky) nature of such knowledge (Audretsch & Dohse, 2007; Fritsch & Franke, 2004; Storper & Venables, 2004). This literature emphasizes that geographical proximity can provide face-to-face contacts, shared socio-cultural and institutional contexts, mutual trust (Bathelt et al., 2004; Boschma, 2005; Nilsson & Mattes, 2015), which not only stimulate the circulation and transmission of codified knowledge but also encourage spillovers of tacit knowledge (Gertler, 1995). The other explanation emphasizes the role of labor market dynamics, including the regional knowledge spillovers by the mobility of skilled labor and inventors (Angeli et al., 2014;

Breschi & Lenzi, 2013; Eriksson & Lindgren, 2008).

On the other hand, several studies have pointed out that technological knowledge spillovers within clusters tend to be limited or do not seem to generate significant advantages (Grillitsch & Nilsson, 2017; Huber, 2011). A key argument in these studies which emphasizes the existence of negative effects caused by knowledge spillovers within regions or clusters is "that firms are not only receivers but also sources of knowledge spillovers" (Grillitsch & Nilsson, 2017, p. 1222). Direct interaction, such as face-to-face contact, can stimulate the transmission of complex knowledge, including tacit knowledge, and generate negative effects by intentionally or unintentionally leaking knowledge (Sammarra & Biggiero, 2008). The other explanation is related to labor poaching. In addition, the negative effects of regional knowledge spillovers can also be inferred from the framework of local buzz and global pipeline addressed by Bathelt et al. (2004). It has been argued that while extra-regional pipelines, called global pipelines, allow firms to gain numerous fruitful opportunities for access to novel and non-redundant information that can encourage innovation, local buzz (or regional knowledge) can induce lock-in and inertia because of similarity among regional knowledge base[5] (Bathelt et al., 2004; Benneworth & Hospers, 2007; Breschi & Lenzi, 2013). Fitjar and Rodríguez-Pose (2015) find that global pipelines contribute more to value creation than local buzz by examining the

[5] Actors participating in the local buzz "continuously contribute to and benefit from the diffusion of information, gossip, and news by just being there" (Bathelt, 2004, p.38) and benefit from the application of the same interpretative schemes and mutual common experience of problem-solving based on mutual trust and shared cultural traditions, norms, institutions, and habits. While these characteristics of the local buzz can facilitate knowledge spillovers in regions, it is likely to cause lock-in, which means obscuring the view on new technologies or new market possibilities (Boschma, 2005).

interaction effects on innovation from international collaboration and regional R&D investments.

In summary, knowledge within the region is likely conducive to firm performance, but negative externality can be induced. The positive and negative effects of regional knowledge spillovers may differ depending on *certain conditions* that affect a firm's innovation activities and knowledge acquisition process. *The net effect of regional knowledge spillovers on a firm will depend on whether the firm is primarily a source or a receiver.* It can be related to *firms' incentives* to acquire external knowledge. In addition, when state-of-the-art knowledge is required in a specific industry, the effect of regional knowledge spillover can be limited because the knowledge shared within the region is likely to be somewhat *tacit* and *redundant*.

4.2.2 Regional knowledge spillovers and technological characteristics

As mentioned above, regional knowledge spillovers are not always beneficial for economic performance. If then, *under what condition does this effect become stronger or weaker?* What regional characteristics cause positive knowledge spillovers has also been considerably studied. (Beaudry & Schiffauerova, 2009; Kemeny & Storper, 2015). In particular, whether regions benefit more from pursuing industrial specialization or diversification has been at the center of scholarly discussion for a long time in agglomeration economics. In a seminal paper on the growth of cities, Glaeser et al. (1992) introduced a line of inquiry known as MAR versus Jacobs. The MAR theory emphasizes intra-industry knowledge spillovers and the spatial concentration of same-industry firms. In contrast to MAR, Jacobs et al. (1969) argued that knowledge spillovers between rather than within industry can more encourage recombining different ideas, incubating innovations, and providing for technology breakthroughs. Using panel data on the growth of large industries in US cities, Glaeser et al. (1992) find support for the argument related to the ideas of Jacobs (1969). This seminal work was followed by extensive literature which has examined whether regional specialization or diversification is beneficial, and this literature has shown mixed empirical evidence (Beaudry & Schiffauerova, 2009; de Groot et al., 2016).

Recent overviews show that these mixed results are fueled by measurement issues of specialization and diversification, the aggregation level defined as the 'same' industry, heterogeneity in terms of the scale of time and space, research orientation and the characteristics of firms (Ooms et al., 2015; van Oort et al., 2012). While Liang and Goetz (2018) calculated the related variety as the diversification measure at a three-digit level, J. Aarstad et al. (2016) identified the same variable at a two-digit level. Frenken et al. (2007) introduced the concept of related and unrelated variety, pointing out the oversimplified dichotomy related to specialization and diversification. These concepts provide more disentangled views of diversified industry structures than previous conceptions, pointing out the importance of the industries' "interrelationship (how similar/different industries)."

Explaining the diversification of industries whose cognitive distance is neither too large nor too small through the concept of related variety has somewhat resolved the problem of mixed results.

Another cause of mixed results is related to the argument that regional knowledge spillovers are *sector-specific* and *dependent on industrial attributes* (Beaudry & Schiffauerova, 2009; Liang & Goetz, 2018). Technology has path-dependent characteristics that progress along the technology trajectory (Nelson & Winter, 1982), and industry-specific technological regimes have a strong impact on potential learning effects (Carreira & Lopes, 2018; Marsili, 2002). Firm learning varies from industry to industry. This implies that industrial and technological characteristics are important factors when understanding the regional knowledge spillovers affected by firm learning. Liang and Goetz (2018) analyzed the moderating effect of technology intensity on the effect of specialization and related variety on industrial employment growth using employment data from the 3-digit NAICS industry in the United States. The technology intensity is the value obtained by dividing the industry's R&D investment by the total sales, and it was shown that the influence of related variety increased at high technology intensity, and specialization effects increased at low technology intensity. Carreira and Lopes (2018) used the fixed-effect model and firm-level panel data from the Portuguese manufacturing industry to confirm the non-linear relationship between the regional knowledge spillovers and firm productivity. The characteristics of the industry were analyzed by dividing it into high, medium, and low-tech industries with R&D intensity. As a result, it was argued that policies for low technology sectors aimed at new investors in similar industries need to be devised. Although it analyzes firms' TFP, it confirms similar results to Liang and Goetz (2018).

Carreira and Lopes (2018) are somewhat similar to the approach attempted in this paper, but total factor productivity, not direct measures for firm innovation, was used as a dependent variable. Moreover, the fixed-effect model cannot estimate the variance separately for each level using data with a hierarchical structure. Furthermore, according to Beaudry and Schiffauerova (2009), reviewing several related studies, Marshallian externalities were more substantial in low-tech sectors, and the higher the technology intensity, the more research tends to report the positive impact of Jacobs externalities. However, both types appeared in all industrial groups (high-tech, medium-tech, low-tech). This means that both types can induce valid knowledge spillovers, but their effects can vary depending on the *more detailed characteristics of the industry*. Technology intensity alone is insufficient to capture the characteristics of industry and technology for regional knowledge spillovers. In addition, existing studies focused only on whether regional knowledge spillover affects the performance or not and did not consider mechanisms of net effects due to relative size changes in positive or negative bidirectional effects depending on various technological environments.

A technological regime defines a technological environment, which characterizes the key features of technological conditions where firms implement innovation (Castellacci & Zheng, 2010; Malerba & Orsenigo, 1993; Revilla & Fernández, 2012). In other words, the degree to which a firm can influence the process of acquiring knowledge in the region, learning, and creating innovation may differ by industry classified according to the characteristics of the technological regime. There are three dimensions: appropriability, technological opportunity, and cumulativeness, which affect firms' incentives for innovations, requisite knowledge characteristics, and learning patterns. In line with this, studies examining the effects of the technological regime on cooperation, licensing, and network formation have been studied a lot in firm-level strategic management literature. (Herstad et al., 2014; Lee et al., 2017; Seo et al., 2017) Few studies have analyzed the relationship between regional knowledge spillovers and firm innovation, considering the technological environment's characteristics with multi-dimensions to the best of the authors' knowledge. It is not easy to know what mechanisms in regional knowledge spillover affect firm innovation. However, when analyzing the interaction between technological regimes and regional knowledge spillover, we can *infer* how regional knowledge spillover strengthened or weakened under certain technological conditions. These inferences are expected to help expand our understanding of the black box regarding regional knowledge spillover.

In addition, knowledge acquisition and learning capabilities vary from company to company (Rodríguez-Gulías et al., 2021; Roper et al., 2008). Therefore, to expand the understanding of the regional knowledge spillovers, it is necessary to consider the heterogeneity of companies, such as asorptive capacity and other knowledge acquisition channels. Agglomeration economies can also cause economies of scale and pecuniary externalities due to regional size or population density, as well as regional knowledge spillovers (Krugman, 1991; Martin & Sunley, 1998). This paper considers the above factors to reduce the peril of confounding the effects of regional knowledge spillovers.

4.3 Methodology

4.3.1 Data

This study utilizes the ‘2016 Korean Innovation Survey (KIS): Manufacturing industry’ from the Science and Technology Policy institute(STEPI) of South Korea, which involves the firm-level data on the innovative activities for the period 2013-2015. The KIS dataset entails financial information based on the Community Innovation Survey (CIS), and the methodology and questionnaires serve the OECD Oslo manual. The Community Innovation Survey has been used extensively in innovation, economics, and strategic studies (Cassiman & Veugelers, 2006; Lee et al., 2017; Roper et al., 2008; Seo et al., 2017). The KIS consists of data on 4,000 firms in 24 industries. The sample is composed of manufacturing firms with more than ten employees. The KIS data identify the economic-geographical region where each surveyed firm is located. As for the additional regional data on population and labor size in each region, we used the database of Statistics Korea in 2013. (Kostat, 2013)[6]

Due to the missing values of the variables used in this study, 336 firms were removed from our sample. As a result, we narrowed our final sample for the empirical

[6] For detail of data sources see: http://kostat.go.kr/portal/eng/pressReleases/1/index.board

analysis to 3,664 firms that included all related information for the variables. We assume a partial lag between the independent and dependent variables by relating innovative performance in 2015 to the explanatory variables measured from 2013 to 2015. Although some cross-sectional nature limitations exist in the data explaining the results, this assumption follows prior CIS studies. (J. Aarstad et al., 2016; Branstetter, 2001; Klingebiel & Rammer, 2014; Laursen & Salter, 2014).

4.3.2 Definition and measurement of the variables

We operationalize innovation as a dummy variable to measure the dependent variable (INNO). The KIS respondents were requested to indicate whether the firm had product, process, or organizational innovations new to the market between 2013 and 2015. It takes the value of 1 if the firm is performing any product or process innovation or 0 otherwise. More than 38% of firms observed during the period 2013-2015 had at least a product innovation, as shown in descriptive statistics (Table 1). The number of innovative companies is 1,415, and the number of non-innovative companies is 2,249.

Table 4-1 Descriptive statistics

	Full samples (N=3664)				Innovation=1 (N=1415)				Innovation=0 (N=2249)			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
(1) INNO	0.386	0.487	0	1	1.000	0	1	1	0	0	0	0
(2) SP	0	1	-1.136	9.199	-0.114	0.745	-1.136	9.200	0.065	1.132	-1.131	9.200
(3) DIV	0	1	-3.435	1.652	0.135	0.975	-3.435	1.652	-0.089	1.006	-3.435	1.652
(4) CAPITAL	0.376	0.485	0	1	0.322	0.468	0	1	0.410	0.492	0	1
(5) POP	2.295	3.846	0.089	16.14	2.840	4.419	0.089	16.14	1.952	3.394	0.089	16.14
(6) APP	0.466	0.499	0	1	0.587	0.493	0	1	0.390	0.488	0	1
(7) OPP	0.508	0.500	0	1	0.637	0.481	0	1	0.427	0.495	0	1
(8) CUM	0.325	0.469	0	1	0.291	0.454	0	1	0.347	0.476	0	1
(9) SIZE	3.678	1.065	2.303	9.557	4.183	1.101	2.303	8.006	3.361	0.908	2.303	9.887
(10) EXPORT	0.282	0.450	0	1	0.454	0.498	0	1	0.174	0.379	0	1
(11) COLL	0.096	0.295	0	1	0.215	0.411	0	1	0.021	0.145	0	1
(12) RD	3.672	9.355	0	250	7.850	13.28	0.035	250	1.043	3.712	0	57.28

Five main explanatory variables (level 2) were included as regional knowledge spillovers and technological characteristics at the region-industry level. The first two variables in level 2 (specialization and diversification)[7] were the standardized value of the calculated index following equations in Table 2 (Bishop & Gripaios, 2010; Frenken et al., 2007; Liang & Goetz, 2018). Diversification was measured by 'related variety' at industries within a region, paying attention to inter-industry connections[8] (Bishop & Gripaios, 2010). The specialization and the related variety were calculated as a weighted sum of these indexes in most literature. It means that specialization and related variety are the same for all firms in a region, regardless of industry. However, the variance of the industry-region level is significant, and it is higher than the regional level one in our data. That is to say, that information on the variance between industries in a region is lost when aggregated regional-level data are used. Hence, we used the specialization and related variety at the industry-region level.

The other three variables for characteristics of technological regimes: appropriability, technological opportunity, and cumulativeness, are three dummies that are

[7] In the analysis of firm-level performance on regional characteristics of management literature, regional characteristics are measured by regional R&D intensity, the number of the patent application, and the number of skilled labor within a region. Because this paper focused on the relationship between regional knowledge spillover and industrial attributes, we adopt variables in terms of industrial structure for regional knowledge spillovers (specialization and diversification). Population density is used for pecuniary externalities, economies of scale, or labor pooling.

[8] This is based on the arguments of previous studies that related variety is more suitable for explaining potential benefits from Jacobs spillover compared to other diversification indicators, and many empirical studies have reported that related diversity has a significant effect on regional growth, innovation, and productivity. (J. Aarstad et al., 2016; Boschma & Iammarino, 2009; Hartog et al., 2012). There is increasing support for the notion that related variety is associated with innovation and growth. In addition, a small number of empirical studies have shown that unrelated variety is positively associated with high-impact breakthrough innovation

divided into high and low groups by the mean value. *Appropriability(APP)* reflects the possibilities of protecting innovations from imitation and profiting from innovative activities (Malerba & Orsenigo, 1993). The KIS survey asked firms to evaluate the effectiveness of four appropriability mechanisms on a five-point Likert scale – patents, secrecy, design complexity, lead-time advantage – in protecting their product and process innovations. We calculate the mean values of the maximum score received by any one of the five appropriability mechanisms for product and process innovation of firms within the two-digit industry level. *Technological opportunity(OPP)* refers to the easiness of innovating for any given amount of R&D investment in search. A high level of technological opportunity reflects a powerful incentive to conduct innovative activities. Thus, this variable indicates the effort and resources invested in research and development (R&D) activities. Previous studies have regarded average industry R&D intensity or a dummy variable as a good proxy for an industry-specific technological opportunity (Castellacci & Zheng, 2010; Kim & Lee, 2016). *Cumulativeness(CUM)* is related to "today's knowledge and innovative activities form the base and the building blocks of tomorrow's innovations: an innovation generates a stream of subsequent innovations" (Breschi et al., 2001; Peneder, 2010). Given the abstract nature of the concept, the KIS does not provide any measure of cumulativeness directly. However, we can relate technological cumulativeness to the concept of innovation radicalness (Peneder, 2010; Revilla & Fernández, 2012). Radical innovations generate technological discontinuities and destroy pre-existing knowledge bases (Henderson & Clark, 1990). There is thus an inverse relationship between innovation radicalness and technological cumulativeness. According to Revilla and Fernández (2012), radicalness is measured as the mean value of the number of firms that have a totally new product or process innovation in the market at the two-digit industry level using KIS data. We take the reciprocal of this radicalness for cumulativeness. Table 3 shows the sectoral classification for 2-digit industries according to three dimensions of the technological regime.

To account for the agglomeration externalities in terms of scale, we use the population density and the dummy variable for the capital area[9]. In Korea, about 40% of the country's population lives in Seoul and Gyeonggi-do. In other words, it means that 10.7% percent of the country's land area accounts for about half its population. Thus, we introduce the capital area dummy. In addition, we control other possible knowledge transmission channels such as export and collaboration with whom to isolate the effects of regional knowledge spillover via face-to-face contacts and social networks bounded space. To control for firm characteristics relevant to innovative performance, we also use the internal R&D expenditure per employee (RD) and the log of firm size (SIZE) to capture the firm's absorptive capacity.

[9] The number of innovative companies is the largest in Gyeonggi-do, and Sejong City has the highest proportion of innovative companies (Figure A1 and A2).

Table 4-2 Description of variables

Variables	Description	Source
Depedent variables		
INNO	1 for companies that had any product innovation and 0 otherwise	KIS
Level 1 variables		
SIZE	Natural logarithm of the number of employee	KIS
EXPORT(dummy)	1 for companies that export and 0 otherwise	KIS
COLL (dummy)	1 for companies that had any collaboration and 0 otherwise	KIS
RD	R&D expenditures/employee	KIS
Level 2 variables		
SP	$Specialization(SP_{jk}) = \frac{emp_{jk} / emp_{k}}{emp_{j,nation} / emp_{nation}}$ where emp_{jk} is the number of employee for industries j within region k, emp_{k} is the number of employee within region k, $emp_{j,nation}$ is the number of employee for industries j within the whole nation, and emp_{nation} is the number of employee within the whole nation.	Kostat
DIV	$Diversification(DIV_{jk}) = \sum_{l \in L_i} \frac{emp_{lk}}{emp_{jk}} \ln(\frac{empi_{jk}}{emp_{lk}})$ where L_i is the set of all sub-industry (4-digit) in an industry j (2-digit),	Kostat

	emp_{lk} is the number of employee for 4-digit industry l in a region k, and emp_{jk} is the number of employee for 2-digit industry j in a region k.	
APP	1 for industries within regions with higher appropriability than the mean value at industry level and 0 otherwise	KIS
OPP	1 for industries within regions with higher opporunity than the mean value at industry level and 0 otherwise	KIS
CUM	1 for industries within regions with higher cumulativeness than the mean value at industry level and 0 otherwise	KIS
Level 3 variables		
CAPITAL	1 for firms that located in capital area(Seoul and Gyeonggi-do and 0 otherwise	
POP	Population size (per thousands)/geographical size (square kilometers) at regional level	

4.3.3 Model specification and estimation strategies

The structure of our specification is hierarchical since firms are nested in region-industry and regions. Multilevel models acknowledge the existence of such data hierarchies by allowing for residual components of slopes and intercepts at each level in the hierarchy. The equation for a multilevel model contains the fixed coefficients; it is the fixed (or deterministic) part of the model, and the random error terms for each level; it is the random (or stochastic) part of the model (Maas & Hox, 2005). An appropriate approach to analyze relations identified at different levels is multilevel modeling due to several theoretical reasons (Hox et al., 2017; Srholec, 2010; Tojeiro-Rivero & Moreno, 2019). First, the use of single-level models is based on the assumption of independence of each observation. If data have a hierarchical structure, firms within the same region are more likely to be similar among them than those in different regions, and independent assumption tends to be violated. Multilevel modeling enables researchers to analyze the extent to which specific differences between regions explain the firm-level outcomes by relaxing the independence assumption. Second, the multilevel approach allows us to identify firm-level effects and regional effects from the total effect by using model variance through random intercept, accounting for the unobserved heterogeneity. Third, while many empirical studies on knowledge agglomeration use aggregated data with regions or region-industry, those results about regional-level relationships are not necessarily reproduced at the firm level because information on the variance between firms is eliminated(van Oort et al., 2012). This micro-macro problem, referred to as the 'ecological fallacy', can be better addressed using multilevel modeling.

The number of regions in this analysis is 17 groups, which is not too high. In this case, it has been argued that estimates for the regional variance component can be biased(Maas & Hox, 2005). Following Stegmueller (2013) and Tojeiro-Rivero and Moreno (2019), when the number of the highest level group is between 15 and 20, it is most appropriate to use the random intercept model. Furthermore, since the data involves firm-level observations (level 1) for region-industry combinations (level 2) nested in the region (level3), we used a three-level multilevel model. In addition, given that the dependent variable for innovation ($INNO_{ijk}$) is dummy, we assume a logistic model. More specifically, we suppose fixed slopes and estimate *a three-level logistic random intercept model*, which can be as the reduced form specification as follows:

Model 1 (with level 1 variables)

$$INNO_{ijk} = \beta_0 + \beta_1 SIZE_{ijk} + \beta_2 EXPORT_{ijk} + \beta_3 RD_{ijk} + \beta_4 COLL_{ijk} + u_k + v_{jk} + e_{ijk}$$

Model 2 (with level 1, 2 and 3 variables)

$$\begin{aligned} INNO_{ijk} &= \alpha_0 + \sum_{n=1}^{4} \alpha_n X_{ijkn} \\ &+\beta_1 CAPITAL_k + \beta_2 POP_k \\ &+\gamma_1 SP_{jk} + \gamma_2 DIV_{jk} + \gamma_1 APP_{jk} + \gamma_2 OPP_{jk} + \gamma_3 CUM_{jk} \\ &+u_k + v_{jk} + e_{ijk} \end{aligned}$$

Model 3 (with interaction terms in Model 2)

$$
\begin{aligned}
INNO_{ijk} = \alpha_0 &+ \sum_{n=1}^{4} \alpha_n X_{ijkn} + \sum_{m=1}^{2} \beta_m Y_{km} + \sum_{l=1}^{5} \gamma_l Z_{jkl} \\
&+ \lambda_1 APP_{jk} SP_{jk} + \lambda_2 APP_{jk} DIV_{jk} \\
&+ \lambda_3 OPP_{jk} SP_{jk} + \lambda_4 OPP_{jk} DIV_{jk} \\
&+ \lambda_5 CUM_{jk} SP_{jk} + \lambda_6 CUM_{jk} DIV_{jk} + u_k + v_{jk} + e_{ijk}
\end{aligned}
$$

where $INNO_{ijk}$ refers to our dependent variable observed product innovation for firm i in the industries within regions j in the region k, α_0 is the overall mean of $INNO_{ijk}$ (across all groups), u_k is the region random effect (level 3 random effect), v_{jk} is the industries within a region random effect (level 2 random effect), and e_{ijk} is the firm residual (level 1 residual) with mean zero and variance σ^2_e (assuming a logistic distribution) (Rodríguez-Gulías et al., 2021; Srholec, 2010). X_{ijkn} refers to firm-level variables (level 1: $SIZE_{ijk}$, $EXPORT_{ijk}$, RD_{ijk}, $COLL_{ijk}$), Y_{km} are variables for regional characteristics (level 3: $CAPITAL_k$, POP_k), and the Z_{jkl} will be measures for specialization/diversification (level 2: SP_{jk}, DIV_{jk}) and three dimensions of technological regimes(level 2: APP_{jk}, OPP_{jk}, CUM_{jk}) that are our key industries within regions-level variables.

Model 1 is the base model with a random intercept and firm-level variables. Model 2 adds the level 2 and 3 variables, and Model 3 is simultaneously defined with level 1, 2, and 3 variables and six interaction terms. The correlation matrix (Table 4) shows that the correlation coefficient between level 2 variables is not ignorable, so the correlation between interaction terms may cause multicollinearity problems. Although the level 2 variables were standardized, the variation inflation factor (VIF) of model 3 was quite high, so Model 4.1 to 4.3 include two interaction terms with appropriability in Model 4.1, opportunity in

Model 4.2, and cumulativeness in Model 4.3 for robustness check of Model 3 estimation.

Table 4-3 Correlation matrix of the variables

	(1)	(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) INNO	1.000											
(2) SP	-0.087	1.000										
(3) DIV	0.109	-0.376	1.000									
(4) CAPITAL	-0.089	-0.142	0.166	1.000								
(5) POP	0.112	-0.101	-0.042	0.296	1.000							
(6) APP	0.232	-0.178	0.422	0.129	0.087	1.000						
(7) OPP	0.235	-0.204	0.321	0.086	0.021	0.682	1.000					
(8) CUM	-0.106	0.369	-0.285	-0.083	0.023	-0.462	-0.480	1.000				
(9) SIZE	0.376	0.100	-0.119	-0.100	-0.020	0.038	0.046	0.021	1.000			
(10) EXPORT	0.302	-0.081	0.076	-0.024	0.041	0.193	0.164	-0.130	0.358	1.000		
(11) COLL	0.320	-0.037	0.018	-0.028	0.060	0.113	0.132	-0.072	0.154	0.156	1.000	
(12) RD	0.354	-0.055	0.069	0.050	0.046	0.137	0.201	-0.094	0.027	0.151	0.231	1.000

4.4 Empirical Results and Discussion

Table 5 contains six different estimations to analyze how characteristics of technological regimes and regions affect firms' innovative performance using multilevel mixed-effects logistic regression. The results of multilevel analyses, as mentioned before, provide both fixed effects and random effects. Fixed effects represent regression estimates, whereas random-effects represent estimated variance components. Residuals represent the estimated standard deviation of the overall error term.

4.4.1 The effect of regional and technological characteristics on firm innovation

In our first specification (Model 1), we only include firm characteristics to elaborate the variability of our dependent variable at the level of region-industry and region. As observed by the results of variance components, it is worth drawing several conclusions. Firstly, both the variance of the region-industry and the variance of the region are highly significant, which means that it is necessary to use the multilevel methodology. Our estimation method considers the interdependencies among the observations for a given industry within a region and a given region by employing the multilevel model. Another interesting result is that while both are significant, the variance of different industries nested within regions is higher than the regional level one. This indicates not only that regional characteristics are relevant for the innovativeness of firms but also that industrial characteristics are important variables to be necessarily considered.

This first specification illustrates that all the variables at the firm level show the expected sign. R&D intensity has a strong positive effect on the firm's innovative performance, validating the idea that more absorptive capacities allow the development of ideas that can be transformed into innovations. (Cohen & Levinthal, 1990) We also find that large firms (in terms of the number of employees) are significantly more innovative than small firms, probably because of economies of scale at the firm level. We observe that Export and Collaboration, which can be other channels for knowledge acquisition, positively and significantly affect firm innovation. The sign and magnitude of the firm-level control variables' parameters are almost the same in all our specifications in Table 5.

Model 2 considers the region-industry variables such as Specialization(SP), Diversity(DIV), including other control variables at the level of the region (CAPITAL and POP), and three characteristics of the technological regime such as Appropriability(APP), Opportunity(OPP), Cumulativeness(CUM). The Wald χ^2 is significant in Model 2 and all reported models in Table 5, confirming a robust model fit. Although the random effect at the regional level is significant in Model 1, it is zero and insignificant in all reported models in Table 5 except for Model 1. The previous literature has discussed zero random effects (J. Aarstad et al., 2016; Andrews, 1999; Self & Liang, 1987), and it implies that the regional and technological variables in our model are accountable for a significant part of the regional variability. Model 2 provides significant support for a positive relationship between diversification and firm innovations, while specialization does not generate significant effects.

Table 4-4 Multilevel logstic regression analysis, with innovation as the dependent variable

	Model 1	Model 2	Model 3	Model 4.1	Model 4.2	Model 4.3
SIZE	0.840***	0.841***	0.839***	0.839***	0.839***	0.843***
	(0.0514)	(0.0510)	(0.0508)	(0.0510)	(0.0509)	(0.0509)
EXPORT	0.563***	0.505***	0.508***	0.524***	0.500***	0.489***
	(0.113)	(0.112)	(0.112)	(0.112)	(0.112)	(0.112)
RD	0.251***	0.245***	0.244***	0.244***	0.246***	0.244***
	(0.0145)	(0.0144)	(0.0143)	(0.0143)	(0.0144)	(0.0143)
COLL	1.746***	1.714***	1.720***	1.721***	1.709***	1.712***
	(0.199)	(0.197)	(0.197)	(0.197)	(0.197)	(0.197)
CAPITAL		-0.858***	-0.900***	-0.901***	-0.868***	-0.883***
		(0.223)	(0.205)	(0.214)	(0.218)	(0.218)
POP		0.116***	0.126***	0.120***	0.118***	0.123***
		(0.0225)	(0.0213)	(0.0218)	(0.0223)	(0.0224)
SP		-0.0107	0.0743	-0.155*	0.0423	0.177
		(0.0867)	(0.177)	(0.0938)	(0.0999)	(0.147)
DIV		0.192**	-0.152	-0.0599	0.377***	0.0493
		(0.0877)	(0.181)	(0.124)	(0.128)	(0.113)
CUM		0.0832	-0.201	-0.0549	-0.0323	0.130
		(0.176)	(0.204)	(0.180)	(0.183)	(0.182)
APP		0.782***	0.799***	0.952***	0.755***	0.713***

	Model 1	Model 2	Model 3	Model 4.1	Model 4.2	Model 4.3
		(0.168)	(0.177)	(0.177)	(0.166)	(0.170)
OPP		0.380**	0.312*	0.333**	0.290*	0.486***
		(0.165)	(0.180)	(0.160)	(0.171)	(0.173)
APP*SP			0.441*	0.583**		
			(0.242)	(0.227)		
APP*DIV			0.698***	0.439**		
			(0.190)	(0.177)		
OPP*SP			-0.171		-0.0401	
			(0.212)		(0.204)	
OPP*DIV			-0.370**		-0.348**	
			(0.181)		(0.173)	
CUM*SP			-0.170			-0.236
			(0.191)			(0.182)
CUM*DIV			0.372**			0.321*
			(0.186)			(0.181)
Constant	-4.655***	-5.248***	-5.171***	-5.274***	-5.145***	-5.243***
	(0.242)	(0.261)	(0.261)	(0.258)	(0.261)	(0.261)
Industries within regions	0.978***	0.766***	0.670***	0.717***	0.742***	0.741***
	(0.102)	(0.0952)	(0.0920)	(0.0923)	(0.0960)	(0.0939)
Regional effect	0.394***	0.000000372	2.53e-11	5.45e-10	-1.84e-09	-9.45e-10

	Model 1	Model 2	Model 3	Model 4.1	Model 4.2	Model 4.3
	(0.120)	(0.140)	(0.124)	(0.126)	(0.134)	(0.130)
Wald χ^2 (fixed effects, regressors)	723.21***	750.48***	766.37***	755.15***	756.39***	755.47***
No. observations	3664	3664	3664	3664	3664	3664
No. Superclusters (Regions)	17	17	17	17	17	17
No. Clusters (industries within regions)	408	408	408	408	408	408
Log likelihood	-1538.09	-1509.69	-1497.31	-1503.78	-1507.75	-1506.77

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This shows that diversification, which measures the variety of related industries, has a significant correlation with firm innovations. Population density is positively associated with firm innovation, and the dummy variable for the capital area has a highly significant negative effect. It has been argued that agglomeration economies can induce pecuniary externalities (Martin & Sunley, 1998) as a function of regional size or population density (Krugman, 1991). In line with this, the positive effect of population density on innovation can be interpreted as supporting the existence of those pecuniary externalities[10]. However, a significant negative effect of a location in the capital area means that when the population density reaches a certain level, congestion costs caused by traffic jams, pollution, and high housing prices (Hanlon & Miscio, 2017) can outweigh the benefit of pecuniary externalities from agglomeration in Korea.

[10] The inverted U-shaped relationship between population density and innovation was assessed by including the squared terms of population density in the regression, but the result was not significant.

4.4.2 The role of technological regimes in shaping the effect of regional knowledge spillovers on firm innovation the interaction effects

Model 3 refers to the regression including six combinations of two variables for regional knowledge externalities (specialization and diversity) and three variables of the technological regime (appropriability, opportunity, cumulativeness). According to the results of interaction terms, the positive effect of diversity on innovation increases under high appropriability conditions. As mentioned in the literature review, technological knowledge spillovers within the cluster are likely to be negative (Antonietti & Cainelli, 2011) because of knowledge leakage and labor poaching (Combes & Duranton, 2006). For example, Yoffie (1993) argues that firms in the semiconductor industry tend to avoid being located in the same region as rival firms for fear of leakage of internal knowledge. Sammarra and Biggiero (2008) also find that collocation and direct interaction stimulate the transfer of complex knowledge and increase the risk of negative knowledge externalities in the form of knowledge leakage. The appropriability refers to how innovations are protected from imitation and firms can extract profits from innovative outcomes. Therefore, it can be argued that knowledge leakages are suppressed under high appropriability conditions. Studies of other knowledge acquisition channels such as collaboration and licensing have emphasized that knowledge transfers are highly vulnerable to opportunistic behavior that seeks to use knowledge without having enough to pay for such knowledge in a market transaction. (Belderbos et al., 2004; Cassiman & Veugelers, 2006; Lee et al., 2017) In this regard, several empirical studies suggest that a strong appropriability can effectively suppress opportunistic behavior and knowledge leakage because high appropriability conditions reduce uncertainty, imitation risk, and transaction cost (Kim & Lee, 2016; Lee et al., 2017) and can increase the effect of strategy for knowledge acquisition. (Belderbos et al., 2004; Lee et al., 2017; Seo et al., 2017). However high appropriablity may reduce the positive effect of a firm from the knowledge receiver's position, just like both sides of the coin. Nevertheless, it is inferred that high appropriability can be formed an environment in the direction of increasing the firm's banefit from innovation through net regional knowledge spillover effects.

Interestingly, the interaction term between specialization and appropriability is significantly positive, while the effect of specialization on innovation is insignificant. It can be inferred that it is not that the effects of regional knowledge spillovers induced by specialization itself do not exist but that the positive effects caused by knowledge spillovers and the negative effects caused by knowledge leakage are offset so that the overall effect does not appear. If the interaction effect between regional knowledge spillovers and technological environments was not considered, it would have been concluded that knowledge spillovers exist within the region due to diversification rather than specialization through the estimation results of Model 2 and would have supported Jacobs (1969)'s argument on the related debate. Thus, this is interpreted as a result that can emphasize that it is more necessary to understand the different patterns of regional knowledge spillovers under different technological environments.

On the other hand, the interaction term of diversification and technological opportunity was negative at the 5% significance level, confirming that the knowledge externalities by diversification decrease under a high technological opportunity environment. Technological opportunity refers to the ease of innovating for any given amount of R&D investment (Breschi et al., 2001). High opportunities represent a strong incentive to engage in innovative activities and denote an economic environment in which firms bear greater fruition after any given of money invested in search of relatively abundant possibilities (Malerba & Orsenigo, 1993). In other words, under high technological opportunity conditions, incentives for acquiring valuable knowledge are strong and obtainable knowledge is abundant, so those conditions can be conducive to innovation. It can also be confirmed from the results of this study that technological opportunities have positive effects on technological innovation. Many studies argued that the performance of external knowledge sources such as cooperation and licensing would increase in an environment with a high level of technological opportunity and found empirical evidence for that. (Lee et al., 2017; Seo et al., 2017)

However, there is a need to condition that the knowledge acquired through channels such as cooperation, licensing, and regional sources is valid. The high geographical proximity provides a similar cultural, social, and institutional context (Boschma, 2005), so it is highly likely to have difficulty acquiring novelty knowledge, as mentioned in section 2. Regional knowledge is more likely to cause lock-in or inertia than knowledge acquired outside the region, and knowledge acquired through the global pipeline contributes more to value creation (Breschi & Lenzi, 2013; Fitjar & Rodríguez-Pose, 2015). Herstad et al. (2014) showed that the global innovation linkage measured by cooperation with overseas partners positively correlates with technology opportunities, and Bathelt et al. (2004) found that strong regional knowledge spillover and lack of extra-regional pipeline can increase the risk of lock-in and loss of creativity. These studies ultimately imply that regional knowledge may be vulnerable to radical search and exploration. In other words, under high technological opportunity conditions, regional knowledge can provide high accessibility but not availability, such as poverty in the midst of plenty, so positive regional knowledge externalities on innovation can decrease.

Another explanation is related to the fact that the high technological opportunity environment is highly correlated with the early stage of industrial development (Malerba & Orsenigo, 1993). Since the initial stage of industrial development is before the creation of a dominant design, R&D for products of relatively primitive design is conducted, and the creation of various exploratory technologies is attempted without specialized manufacturing processes (Klepper, 1997; Williamson, 2007). In this regard, Cozzi (2001) argued that in the early stages of industrial development, various technological ‘hint’ level knowledge that is difficult to prove novelty or progressiveness of technology is first created, and then innovation emerges through these hints. Moreover, since the 'hint' in the exploratory search stage is a form in which it is difficult to obtain protection, the R&D sector has a large incentive for conducting imitation by spying and labor poaching. Thus the impact of regional knowledge externalities can decrease under such conditions due to knowledge leakage. Consequently, the overall effects of regional knowledge spillovers can decrease. In this context, it can be explained that there is a negative interaction between technological opportunity and regional knowledge spillovers induced by diversification.

Finally, it is shown that cumulativeness, the last characteristic of the technological regime, interacts positively with diversification at the 5% significance level. Technological cumulativeness refers to the fact that today’s innovations and innovative activities build the foundation and the building blocks of tomorrow’s innovations (Lee et al., 2017; Malerba & Orsenigo, 1993). This dimension represents what extent the existing knowledge base is utilized when creating new knowledge. According to Dosi and Nelson (2010), technological progress is based on scientific discipline and operating experience. The characteristics of knowledge related to operating experience are tacit and cumulative because it reflects the path-dependency of the development process, the incomplete understanding of technology, and the heterogeneity of agents using technology. The radical technological innovation that destroys the existing knowledge base is highly likely to have low relevance to tacit knowledge built through operating experience and maybe hardly affected by regional knowledge externalities related to the tacitness and embeddedness of knowledge. In this regard, it has been argued that cumulativeness is closely associated with complexity and system embeddedness (Bassanini & Ernst, 2002; Herstad et al., 2014). In line with this, under low cumulativeness conditions where radical innovation takes place, the contribution of tacit knowledge in a region to innovation may below. On the other hand, it can be inferred that when a firm acquires the same amount of tacit knowledge, the degree to which it is used to create innovations will be a higher cumulativeness environment. For these reasons, the cumulativeness of technology and regional knowledge externalities by diversification can have a positive interaction relationship.

In summary, in Model 2, diversification measured as related variety increases the tendency of firm innovation. We find that the positive influence on technological innovation of firms by regional knowledge externalities increases under an environment with higher appropriability, lower technological opportunity, and higher cumulativeness. Table 6 summarises the interaction effects between regional knowledge spillovers and technological regimes. Additionally, we find the positive effect of pecuniary externalities asserted by Krugman (1991) by confirming the tendency for firm innovation to increase as the population density increases up to a certain level. It is shown that the effects of regional knowledge spillovers can decrease when the level of population density is excessively high, such as in Seoul.

Models 4.1 to 4.3 were analyzed by including interaction terms with appropriability in Model 4.1, opportunity in Model 4.2, and cumulativeness in Model 4.3. The reason for this is that interaction terms can correlate strongly with each other and this may cause multicollinearity problems. Calculating the variance inflation factor (VIF) for the interaction terms in Model 3 returns values above the range of critical values between 4 and 10, as suggested in the literature. (c.f. O'brien (2007)) Therefore, in this paper, the interaction terms were divided and analyzed to avoid the multicollinearity issue as much as possible, and the same results were obtained as shown in Model 3.

Additional analysis was conducted together in order for the robustness of the analysis. Seoul, the capital of South Korea, accounts for 0.6% of the total land, but 20% of the total population lives, which is about 13 times higher than the average population density in other regions. In addition, this paper analyzed firms in the manufacturing industry, and if the location is marked as Seoul, it is likely the location of the headquarters, not the factory. Dummy variables for the capital area were inserted into the model to control and analyze these parts, but the analysis was performed by removing samples from Seoul, showing significantly different characteristics compared to other regions. (The capital area is a dummy variable, including Seoul and its surrounding areas, but if all samples corresponding to this area are removed, the number of samples decreases considerably, so it was analyzed except only for firms in Seoul.) while there was a slight change in the size of all coefficients, all results showed the same trend and significance. In addition, if a firm is not an independent company but an affiliate, the connection outside the region or abroad can be formed internally in a multidivisional organization. Accordingly, the analysis was performed by adding a dummy of affiliates, but it was not significant, and the tendency and significance of the results of major variables were the same.

Table 4-5 Summaries of the results for three-level logistic random intercept model: cross-interaction effects

			TECHNOLOGICAL REGIME		
			APP	OPP	CUM
DIRECT EFFECT			(+)	(+)	
REGIONAL KNOWLEDGE SPIILOVERS	SP		(+)		
	DIV	(+)	(+)	(-)	(+)

Notes: (+/-) denote a positive/negative effect on firm innovation. Blank means non-significant.

4.5 Conclusions

This paper analyzed a three-level logistic model to examine how knowledge spillovers within the region can differ under the multi-faceted technological environment defined as a technological regime for Korean manufacturing firms using various control variables. This paper showed that diversification measured by related variety significantly affected firm innovation, as mentioned in the existing literature (J. Aarstad et al., 2016; Ejdemo & Örtqvist, 2020). The diversification effect was significantly different depending on the technological environment of the industry to which the firm belongs. Although specialization was not significant in Model 2, the effect increased under high appropriability.

The types of spatial externalities studied in agglomeration economics and economics of geography can be divided into knowledge spillovers and pecuniary externalities (Martin & Sunley, 1998). Pecuniary externalities indirectly confirmed the effect through population density variables. At the level of population density in the capital area of Korea (40% of the population is concentrated in the capital area, including Seoul and Gyeonggi), the congestion effect can reduce the innovation tendency of firms. Model 1 confirms that characteristics (SIZE, RD) and knowledge acquisition channels (EXPORT, COLL) at the firm level were all significant, and that the variance at the industries within a region was much more subtantive than that at the regional level. Given these results, policy-makers should be conscious of the seriousness of the concentration problem in the capital area and the need to keep the regional industrial policy, not the "ideal model" of policy regarding regional innovation.

According to the results, if a firm has a high appropriability, low technological opportunity, high cumulativeness environment, the effects of regional knowledge spillovers on firm innovation can increase, while it can be not significant or decrease in case of firms included in low appropriability, high technological opportunity, and low cumulativeness. We confirm once again that industrial and technological features of firms in a region (industry and technology heterogeneity within a region), which have not been actively discussed as much as industrial structures within a region (specialization and diversity), should not be overlooked. We also prove that technological regimes provide significant dimensions for environments explaining regional knowledge spillovers. Therefore, it was confirmed that the technological regime is a more systematic and detailed contingency factor that can explain regional knowledge spillovers beyond technology intensity, fragmentary technological characteristics used to explain the heterogeneity of regional knowledge spillovers effects in previous studies.

Breschi et al. (2001) proposed that the specific pattern of innovative activities in an industry (Schumpeterian Mark I and Mark II) can be explained as the outcome of different technological (learning) regimes and showed that fewer technological opportunities, better appropriability conditions, more cumulative knowledge work in the direction of *creative accumulation* patterns (or Schumpeter Mark II). Given Breschi et al. (2001)'s argument and the results of this paper, it seems that the effect of regional knowledge spillover on firm innovation is more significant in industries with the Schumpeter Mark 2 pattern of innovative activities. A Schumpeter mark II pattern is characterized by a high concentration of innovative activities, a stable hierarchy of innovators, and high entry barriers, so could be labeled 'deepening'. Malerba and

Orsenigo (1993) said that the deepening pattern of innovation is associated with the dominance of a few firms which are continuously innovative through the accumulation over time of technological and innovative capabilities. It implies that regional policies that focus only on facilitating the flow of knowledge within the region are not sufficient to initiate radical innovation, facilitate the entry of new innovators, or increase industry diversification. Therefore, policymakers should not overlook the importance of global pipelines (or networking with the outside region) when establishing regional industrial policies.

This study provides some implications for policymakers, but there are also limitations. The data used in this paper is cross-sectional data from 2013 to 2015. It has limitations in that it cannot consider the dynamic aspect of regional knowledge spillovers on firm innovation (or deepening pattern within a region), although it exists in related studies using cross-sectional data (e.g., J. Aarstad et al. (2016)). In addition, firm productivity and innovation can have long-term effects on the industrial structure and resources in the region (Martin & Sunley, 1998), and the endogenous relationship between the growth of industries within a region and firm performance has not been considered. Of course, regional characteristics change slowly over time (Wixe, 2015), and J. Aarstad et al. (2016) also mentioned that it might be reasonable to assume that the relatively fast-changing characteristics of firms do not have a significant impact on regional characteristics. However, in future studies, it is necessary to analyze using longitudinal data.

The generalization of results should also be carefully considered. Although almost all manufacturing industries have been used for analysis, it is the result of analysis through data on only one country. Since the level of economic development and cultural characteristics differ from country to country, companies in each country have different characteristics for innovation activities, and accordingly, the characteristics of knowledge spillovers within the region may also differ. In addition, the dependent variable of this analysis was a dummy variable for innovation tendency. Although it has been widely used in previous studies (Jarle Aarstad et al., 2016; Crescenzi & Gagliardi, 2018), it may not fully reflect the quantitative and qualitative aspects of innovation. For example, Castaldi et al. (2015) and Miguelez and Moreno (2018) reported that unrelated variety affects breakthrough innovation, but in this study, it was not possible to confirm the significant effect of unrelated variety on firm innovation[11]. Therefore, the analysis of the longitudinal data considering the quantitative and qualitative characteristics of innovation could also be a future study.

[11] Unrelated variety was not included in the main analysis of this paper because it is mainly referred to as a factor for a portfolio effect at the regional level in previous literature, and several empirical studies have shown that unrelated variety is positively associated with high-impact breakthrough innovation (Castaldi et al., 2015; Miqguelez and Moreno, 2018). However, since the dependent variable is a dummy in this study, it does not consider the newness of innovation. Nevertheless, unrelated variety was also analyzed in the additional analysis, but it was not significant as expected.

Chapter 5. **Overall Conclusions**

In recent years, clusters have emerged as a prominent concept in discussions on economic development strategies. These clusters are geographically concentrated groups of interconnected firms and institutions that are believed to have the potential to drive innovation, increase productivity, and enhance regional resilience. However, the relationship between clusters and innovation has been a subject of debate among researchers and policymakers, with mixed findings and theoretical arguments.

The objective of this study is to investigate the effects of agglomeration, or clustering, on innovation. Innovation is a crucial focus because clusters are thought to facilitate knowledge spillovers that are essential for innovative activities. Examples like Silicon Valley have demonstrated the success of clusters in promoting innovation and contributing to local economic prosperity. Moreover, innovation is widely recognized as a key driver of long-term economic growth in modern economic theories.

Given the significance of innovation for economic performance, enhancing innovation capacity has become a priority for firms, regions, and countries. Understanding the effects of clusters on innovation can provide valuable insights into regional economic development and offer policy implications for local authorities.

However, the existing knowledge on the impact of agglomeration externalities, or cluster effects, on innovation is inconsistent. Theoretical arguments suggest that clusters can both encourage innovation through knowledge spillovers and hinder it through "lock-in" effects. Empirical studies examining the relationship between clusters and innovation have produced conflicting results from different countries and time periods.

To overcome the limitations of individual studies and provide more robust conclusions, this study adopts a meta-analysis approach in Chapter 3. By pooling data from multiple studies, it becomes possible to reach more generalizable conclusions. However, there has been limited research applying this approach to the study of clusters and innovation, except for notable works by de Groot, Poot, and Smit (2010) and Fang (2015).

The study in chapter 3 conducts a meta-analysis of relevant empirical studies on the relationship between clusters and innovation since the 1990s. It differs from previous works by de Groot, Poot, and Smit (2010) and Fang (2015) in several aspects. Firstly, while the former focuses on regional-level effects, this study pays equal attention to both regional-level and firm-level effects. This provides a comprehensive understanding of whether clusters primarily impact individual firms, industries as a whole, or the entire region.

Secondly, previous works included studies with various dependent variables such as employment growth, productivity growth, and innovation. In contrast, this study narrows its focus to the effects of regional clusters on innovation, excluding papers without an innovation-related dependent variable. This allows for a more meaningful calculation of the average effect size and facilitates clearer interpretation of the results, although it leads to a smaller sample size.

Thirdly, building on previous literature, this study attempts to differentiate the influence of physical dimensions in explaining the pattern of cluster agglomeration. This is achieved by classifying variables representing the size, diversity, and relatedness of agglomeration.

The study aims to address three main questions: (1) What are the general conclusions drawn from previous studies regarding the impact of agglomeration externalities on innovation? (2) Are the estimated correlations between clusters and innovation homogeneous or heterogeneous across previous studies? (3) If there is heterogeneity, what variables act as moderators, influencing the direction and magnitude of clusters' effects on innovation?

Using a multilevel model, this study finds that, overall, clusters have a statistically significant positive impact on innovation. However, significant heterogeneity is observed among individual studies, indicating the presence of moderators. The study employs multilevel meta-regression to explore the variation in reported estimates of agglomeration effects and understand how differences in primary study characteristics contribute to this variation.

In conclusions, the results in Chapter 3 reveal that the "Diversity" group consistently shows a positive relationship between clusters and innovation, despite the heterogeneous characteristics of primary studies.

In Chapter 4, the study examines regional knowledge spillovers and their impact on firm innovation in the context of the multi-faceted technological environment of Korean manufacturing firms. The study finds that diversification, measured by related variety, significantly affects firm innovation, with varying effects depending on the technological environment of the industry. Additionally, the study explores the types of spatial externalities in agglomeration economics and economics of geography, distinguishing between knowledge spillovers and pecuniary externalities.

The results indicate that population density, particularly in the capital area of Korea, can have a congestion effect that reduces firms' tendency to innovate. The study also confirms the significance of firm-level characteristics (such as size and research and development) and knowledge acquisition channels (such as exports and collaborations) in influencing innovation. Furthermore, the variance at the industry level within a region is more substantial than at the regional level.

The findings emphasize the importance of considering industrial and technological features, in addition to industrial structures, within a region when studying knowledge spillovers. The study introduces the concept of technological regimes as a systematic and detailed factor that explains regional knowledge spillovers beyond traditional measures. It suggests that industries with a Schumpeter Mark 2 pattern of innovative activities exhibit a stronger effect of regional knowledge spillovers on firm innovation.

The study highlights the need for regional policies that go beyond facilitating knowledge flow within the region. It suggests that policymakers should consider global pipelines and networking with external regions to promote radical innovation, facilitate the entry of new innovators, and increase industry diversification.

While the study provides implications for policymakers, it acknowledges some limitations. The use of cross-sectional data restricts the consideration of dynamic aspects and the endogenous relationship between regional characteristics and firm performance. Longitudinal data analysis is recommended for future research. The generalization of the results should also be approached cautiously, considering the specific characteristics of each country and the limitations of using a dummy variable to capture innovation tendency. Future studies could explore longitudinal data and consider the quantitative and qualitative aspects of innovation.

Bibliography

Aarstad, J., Kvitastein, O. A., & Jakobsen, S.-E. (2016). Local buzz, global pipelines, or simply too much buzz? A critical study. *Geoforum*, *75*, 129-133. https://doi.org/https://doi.org/10.1016/j.geoforum.2016.07.009

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. <Go to ISI>://WOS:000371943800008

Andrews, D. W. K. (1999). Estimation When a Parameter is on a Boundary [https://doi.org/10.1111/1468-0262.00082]. *Econometrica*, *67*(6), 1341-1383. https://doi.org/https://doi.org/10.1111/1468-0262.00082

Angeli, F., Grandi, A., & Grimaldi, R. (2014). Directions and Paths of Knowledge Flows through Labour Mobility: A Social Capital Perspective. *Regional Studies*, *48*(11), 1896-1917. https://doi.org/10.1080/00343404.2012.756579

Antonietti, R., & Cainelli, G. (2011). The role of spatial agglomeration in a structural model of innovation, productivity and export: a firm-level analysis. *The Annals of Regional Science*, *46*(3), 577-600. https://doi.org/10.1007/s00168-009-0359-7

Audretsch, D. B., & Dohse, D. (2007). Location: A Neglected Determinant of Firm Growth. *Review of World Economics*, *143*(1), 79-107. https://doi.org/10.1007/s10290-007-0099-7

Audretsch, D. B., & Feldman, M. P. (1996). R&D Spillovers and the Geography of Innovation and Production. *The American Economic Review*, *86*(3), 630-640. http://www.jstor.org/stable/2118216

Audretsch, D. B., & Feldman, M. P. (2004). Chapter 61 - Knowledge spillovers and the geography of innovation. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2713-2739). Elsevier. https://doi.org/https://doi.org/10.1016/S1574-0080(04)80018-X

Bassanini, A., & Ernst, E. (2002). Labour market regulation, industrial relations and technological regimes: a tale of comparative advantage. *Industrial and Corporate Change*, *11*(3), 391-426.

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, *28*(1), 31-56. https://doi.org/10.1191/0309132504ph469oa

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/https://doi.org/10.1016/j.respol.2008.11.010

Belderbos, R., Carree, M., & Lokshin, B. (2004). Cooperative R&D and firm performance. *Research Policy*, *33*(10), 1477-1492. https://doi.org/https://doi.org/10.1016/j.respol.2004.07.003

Benneworth, P., & Hospers, G.-J. (2007). The New Economic Geography of Old Industrial Regions: Universities as Global — Local Pipelines. *Environment and Planning C: Government and Policy*, *25*(6), 779-802. https://doi.org/10.1068/c0620

Bishop, P., & Gripaios, P. (2010). Spatial Externalities, Relatedness and Sector Employment Growth in Great Britain. *Regional Studies*, *44*(4), 443-454. https://doi.org/10.1080/00343400802508810

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, *39*(1), 61-74. https://doi.org/10.1080/0034340052000320887

Branstetter, L. G. (2001). Are knowledge spillovers international or intranational in scope?: Microeconometric evidence from the US and Japan. *Journal of International Economics*, *53*(1), 53-79.

Breschi, S., & Lenzi, C. (2013). Local buzz versus global pipelines and the inventive productivity of US cities. In *The geography of networks and R&D collaborations* (pp. 299-315). Springer.

Breschi, S., Malerba, F., & Orsenigo, L. (2001). Technological Regimes and Schumpeterian Patterns of Innovation. *The Economic Journal*, *110*(463), 388-410. https://doi.org/10.1111/1468-0297.00530

Carreira, C., & Lopes, L. (2018). Regional knowledge spillovers: a firm-based analysis of non-linear effects. *Regional Studies*, *52*(7), 948-958. https://doi.org/10.1080/00343404.2017.1360484

Cassiman, B., & Veugelers, R. (2006). In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition. *Management Science*, *52*(1), 68-82. https://doi.org/10.1287/mnsc.1050.0470

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

Castellacci, F., & Zheng, J. (2010). Technological regimes, Schumpeterian patterns of innovation and firm-level productivity growth. *Industrial and Corporate Change*, *19*(6), 1829-1865.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 128-152.

Combes, P.-P., & Duranton, G. (2006). Labour pooling, labour poaching, and spatial clustering. *Regional Science and Urban Economics*, *36*(1), 1-28. https://doi.org/https://doi.org/10.1016/j.regsciurbeco.2005.06.003

Cozzi, G. (2001). Inventing or Spying? Implications for Growth. *Journal of Economic Growth*, *6*(1), 55-77. https://doi.org/10.1023/A:1009898011453

Crescenzi, R., & Gagliardi, L. (2018). The innovative performance of firms in heterogeneous environments: The interplay between external knowledge and internal absorptive capacities. *Research Policy*, *47*(4), 782-795. <Go to ISI>://WOS:000430034500007

de Groot, H. L. F., Poot, J., & Smit, M. J. (2016). Which agglomeration externalities matter most and why? *Journal of Economic Surveys*, *30*(4), 756-782. https://doi.org/https://doi.org/10.1111/joes.12112

Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy*, *43*(10), 1785-1799. https://doi.org/https://doi.org/10.1016/j.respol.2014.05.007

Dosi, G., & Nelson, R. R. (2010). Technical change and industrial dynamics as evolutionary processes. *Handbook of the Economics of Innovation*, *1*, 51-127.

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/https://doi.org/10.1016/j.respol.2020.104073

Eriksson, R., & Lindgren, U. (2008). Localized mobility clusters: impacts of labour market externalities on firm performance. *Journal of Economic Geography*, *9*(1), 33-53. https://doi.org/10.1093/jeg/lbn025

Feldman, M. P., & Kogler, D. F. (2010). Stylized facts in the geography of innovation. *Handbook of the Economics of Innovation*, *1*, 381-410.

Fitjar, R. D., & Rodríguez-Pose, A. (2015). Networking, context and firm-level innovation: Cooperation through the regional filter in Norway. *Geoforum*, *63*, 25-35. https://doi.org/https://doi.org/10.1016/j.geoforum.2015.05.010

Frenken, K., Van Oort, F., & 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., & Franke, G. (2004). Innovation, regional knowledge spillovers and R&D cooperation. *Research Policy*, *33*(2), 245-255. <Go to ISI>://WOS:000220259600005

Gertler, M. S. (1995). "Being There": Proximity, Organization, and Culture in the Development and Adoption of Advanced Manufacturing Technologies. *Economic Geography*, *71*(1), 1-26. https://doi.org/10.2307/144433

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

Grillitsch, M., & Nilsson, M. (2017). Firm performance in the periphery: on the relation between firm-internal knowledge and local knowledge spillovers. *Regional Studies*, *51*(8), 1219-1231. <Go to ISI>://WOS:000405183800008

Hanlon, W. W., & Miscio, A. (2017). Agglomeration: A long-run panel data approach. *Journal of Urban Economics*, *99*, 1-14. https://doi.org/https://doi.org/10.1016/j.jue.2017.01.001

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9-30.

Herstad, S. J., Aslesen, H. W., & Ebersberger, B. (2014). On industrial knowledge bases, commercial opportunities and global innovation network linkages. *Research Policy*, *43*(3), 495-504. <Go to ISI>://WOS:000331502100003

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Huber, F. (2011). Do clusters really matter for innovation practices in Information Technology? Questioning the significance of technological knowledge spillovers. *Journal of Economic Geography*, *12*(1), 107-126. https://doi.org/10.1093/jeg/lbq058

Jacobs, J., Jane, J., & House, R. (1969). *The Economy of Cities*. Random House. https://books.google.co.kr/books?id=GgyxAAAAIAAJ

Kemeny, T., & Storper, M. (2015). Is specialization good for regional economic development? *Regional Studies*, *49*(6), 1003-1018. https://doi.org/10.1080/00343404.2014.899691

Kim, J., & Lee, C. Y. (2016). Technological regimes and firm survival. *Research Policy*, *45*(1), 232-243. <Go to ISI>://WOS:000367484100018

Klepper, S. (1997). Industry Life Cycles. *Industrial and Corporate Change*, *6*(1), 145-182. https://doi.org/10.1093/icc/6.1.145

Klingebiel, R., & Rammer, C. (2014). Resource Allocation Strategy for Innovation Portfolio

Management. *Strategic Management Journal*, *35*(2), 246-268. https://doi.org/10.1002/smj.2107

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, *99*(3), 483-499. https://doi.org/10.1086/261763

Laursen, K., & Salter, A. J. (2014). The paradox of openness: Appropriability, external search and collaboration. *Research Policy*, *43*(5), 867-878. https://doi.org/10.1016/j.respol.2013.10.004

Lee, J. S., Park, J. H., & Bae, Z. T. (2017). The effects of licensing-in on innovative performance in different technological regimes. *Research Policy*, *46*(2), 485-496. https://doi.org/10.1016/j.respol.2016.12.002

Liang, J. C., & Goetz, S. J. (2018). Technology intensity and agglomeration economies. *Research Policy*, *47*(10), 1990-1995. <Go to ISI>://WOS:000450134200009

Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology*, *1*(3), 86-92. https://doi.org/10.1027/1614-2241.1.3.86

Malerba, F., & Orsenigo, L. (1993). Technological Regimes and Firm Bebavior. *Industrial and Corporate Change*, *2*(1), 45-71. https://doi.org/10.1093/icc/2.1.45

Marsili, O. (2002). Technological Regimes and Sources of Entrepreneurship. *Small Business Economics*, *19*(3), 217-231. https://doi.org/10.1023/A:1019670009693

Martin, R., & Sunley, P. (1998). Slow Convergence? The New Endogenous Growth Theory and Regional Development*. *Economic Geography*, *74*(3), 201-227. https://doi.org/https://doi.org/10.1111/j.1944-8287.1998.tb00113.x

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

Nelson, R., & Winter, S. (1982). An Evolutionary Theory of Economic Change, . In: Harvard University Press, Cambridge. .

Nilsson, M., & Mattes, J. (2015). The spatiality of trust: Factors influencing the creation of trust and the role of face-to-face contacts. *European Management Journal*, *33*(4), 230-244. https://doi.org/https://doi.org/10.1016/j.emj.2015.01.002

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673-690.

Ooms, W., Werker, C., Caniëls, M. C. J., & Bosch, H. v. d. (2015). Research orientation and agglomeration: Can every region become a Silicon Valley? *Technovation*, *45-46*, 78-92. https://doi.org/https://doi.org/10.1016/j.technovation.2015.08.001

Peneder, M. (2010). Technological regimes and the variety of innovation behaviour: Creating integrated taxonomies of firms and sectors. *Research Policy*, *39*(3), 323-334. https://doi.org/https://doi.org/10.1016/j.respol.2010.01.010

Porter, M. E. (1998). Clusters and competition: New agendas for companies, governments, and institutions. *On competition*.

Revilla, A. J., & Fernández, Z. (2012). The relation between firm size and R&D productivity in different technological regimes. *Technovation*, *32*(11), 609-623. https://doi.org/https://doi.org/10.1016/j.technovation.2012.06.004

Rodríguez-Gulías, M. J., Rodeiro-Pazos, D., Fernández-López, S., & Nogueira-Moreiras, M. Á. (2021). The effect of regional resources on innovation: a firm-centered approach. *The Journal of Technology Transfer*, *46*(3), 760-791. https://doi.org/10.1007/s10961-020-09811-8

Roper, S., Du, J., & Love, J. H. (2008). Modelling the innovation value chain. *Research Policy*, *37*(6-7), 961-977. <Go to ISI>://WOS:000257560500001

Roper, S., Love, J. H., & Bonner, K. (2017). Firms' knowledge search and local knowledge externalities in innovation performance. *Research Policy*, *46*(1), 43-56. <Go to ISI>://WOS:000390743400004

Rosenthal, S. S., & Strange, W. C. (2004). Chapter 49 - Evidence on the Nature and Sources of Agglomeration Economies. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2119-2171). Elsevier. https://doi.org/https://doi.org/10.1016/S1574-0080(04)80006-3

Sammarra, A., & Biggiero, L. (2008). Heterogeneity and Specificity of Inter-Firm Knowledge Flows in Innovation Networks. *Journal of Management Studies*, *45*(4), 800-829. https://doi.org/https://doi.org/10.1111/j.1467-6486.2008.00770.x

Self, S. G., & Liang, K.-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions. *Journal of the American Statistical Association*, *82*(398), 605-610. https://doi.org/10.1080/01621459.1987.10478472

Seo, H., Chung, Y., & Yoon, H. (2017). R&D cooperation and unintended innovation performance: Role of appropriability regimes and sectoral characteristics. *Technovation*, *66-67*, 28-42. <Go to ISI>://WOS:000416189000004

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379-423.

Srholec, M. (2010). A Multilevel Approach to Geography of Innovation. *Regional Studies*, *44*(9), 1207-1220. <Go to ISI>://WOS:000283317800007

Stegmueller, D. (2013). How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches. *American Journal of Political Science*, *57*(3), 748-761. https://doi.org/https://doi.org/10.1111/ajps.12001

Storper, M., & Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*(4), 351-370. https://doi.org/10.1093/jnlecg/lbh027

Tojeiro-Rivero, D., & Moreno, R. (2019). Technological cooperation, R&D outsourcing, and innovation performance at the firm level: The role of the regional context. *Research Policy*, *48*(7), 1798-1808. <Go to ISI>://WOS:000470952000013

van Oort, F. G., Burger, M. J., Knoben, J., & Raspe, O. (2012). Multilevel approaches and the firm-agglomeration ambiguity in economic growth studies. *Journal of Economic Surveys*, *26*(3), 468-491. https://doi.org/10.1111/j.1467-6419.2012.00723.x

Williamson, O. E. (2007). The economic institutions of capitalism. Firms, markets, relational contracting. In *Das Summa Summarum des Management* (pp. 61-75). Springer.

Winter, S. G. (1984). Schumpeterian competition in alternative technological regimes. *Journal of Economic Behavior & Organization*, *5*(3), 287-320. https://doi.org/https://doi.org/10.1016/0167-2681(84)90004-0

Wixe, S. (2015). The Impact of Spatial Externalities: Skills, Education and Plant Productivity. *Regional Studies*, *49*(12), 2053-2069. https://doi.org/10.1080/00343404.2014.891729

Yoffie, D. B. (1993). Foreign direct investment in semiconductors. In *Foreign direct investment* (pp. 197-230). University of Chicago Press.

Appendix: Tables and figures

Table A1. Summary of studies' main characteristics

Code	Study	Data level	Data stucture	Data period	No. of estimates	Mean of PCC	Mean of SE of estimates	Sample of countries	Estimation method
1	Aarstad et al. (2016a)	Firm	Cross	2008-2010	9	0.0091	0.0126	Norway	Multilevel-logistic
2	Aarstad et al. (2016b)	Firm	Cross	2008-2010	4	0.0024	0.0123	Norway	Multilevel-logistic
3	Anokhin et al. (2019)	Region	Panel	2002-2006	5	0.2482	0.0468	US	Panel(AR1)
4	Antonietti and Cainelli (2011)	Firm	Cross	2001-2003	12	-0.0002	0.0380	Italy	Probit Structural eq.
5	Ascani et al. (2020)	Region	Panel	2007-2012	12	0.1239	0.0300	Italy	Fixed effect

Code	Study	Data level	Data stucture	Data period	No. of estimates	Mean of PCC	Mean of SE of estimates	Sample of countries	Estimation method
6	Baptista and Swann (1998)	Firm	Panel	1982-1979	28	0.0244	0.0225	UK	OLS Poisson Negative binomial
7	Breaudry and Breschi (2003)	Firm	Cross	1990-1998	36	0.0009	0.0062	UK and Italy	Negative binomial
8	Bettiol et al. (2019)	Firm	Cross	2016	14	0.0145	0.0698	Italy	Logit
9	Beugelskijk (2007)	Firm	Panel	1998	14	0.0015	0.0265	Netherland	OLS
10	Boschma and Weterings (2005)	Firm	Cross	2000-2002	6	0.0868	0.0787	Netherland	Tobit
11	Capozza et al. (2018)	Region-industry	Cross	2012-2016	20	0.0440	0.0155	Italy	Poisson Negative-binomial Zero-inflated negative

Code	Study	Data level	Data stucture	Data period	No. of estimates	Mean of PCC	Mean of SE of estimates	Sample of countries	Estimation method
									binomial
12	Cook et al. (2013)	Firm	Cross	2004-2006	4	-0.0075	0.0092	UK	2SLS
13	Czarnitzki and Hottenrott (2009)	Firm	Cross	2002-2004	12	-0.0023	0.0283	Belgium	Tobit IV
14	Beule and Beveren (2012)	Firm	Panel	2002-2004	28	0.0117	0.0607	Belgium	Logit Tobit
15	Beule and Beveren (2008)	Firm	Cross	2002-2004	8	0.0421	0.0177	Belgium	Tobit Logit
16	Feldman and Audretsch (1999)	Region-industry	Cross	1982	12	-0.0311	0.0129	US	Poisson

Code	Study	Data level	Data stucture	Data period	No. of estimates	Mean of PCC	Mean of SE of estimates	Sample of countries	Estimation method
17	Grashof (2021)	Firm	Cross	1997-2013	2	0.0184	0.0093	Germany	OLS with robust SE
18	Greunz (2004)	Region-industry	Cross	1997-1998	28	0.2062	0.0197	Mixed (Europe)	Gernalized MLE
19	Hornych and Schwarz (2009)	Region-industry	Cross	2000-2005	6	0.4257	0.0444	Germany	OLS
20	Huang et al. (2012)	Firm	Cross	2003-2008	3	0.1558	0.0794	Taiwan	OLS
21	Niebuhr et al. (2020)	firm	Panel	1999-2010	8	0.0009	0.0096	Germany	Multilevel
22	Panne and Beers (2006)	Firm and region	Cross	2000-2002	8	0.0066	0.0401	Netherland	Negative binomial

Code	Study	Data level	Data stucture	Data period	No. of estimates	Mean of PCC	Mean of SE of estimates	Sample of countries	Estimation method
23	Shefer and Frenkel (1998)	Firm	Cross	1994	4	0.0402	0.1009	Israel	Logit
24	Smit et al. (2015)	firm	Cross	2002-2004	72	0.0006	0.0103	Netherland	Probit
25	Tavassoli and Carbonara (2014)	Firm	Panel	2002-2007	10	0.1437	0.0498	Sweden	Negative binomial
26	Zhang (2015)	Firm	Panel	1998-2007	48	0.0054	0.0017	China	Fixed effect

Notes: Code is the identifier in the database. SE stands for standard-error. In case of studies with cross-section data type even though data period is not one year, the average value of the period was used, or the number of innovations or patents over the entire period was used as a dependent variable. For example, in Grashof (2021), dependent variable is calculated by the average share of the firm's product innovations in 1993-2013. The mean and SE of pcc were rounded to the fifth decimal place

Table A2. Summary of studies' indicators of agglomeration externalities

No.	Innovation measures	cluster	specialization	competition	Jacobs	urbanization
1	product innovation dummy				0.205***~0.294*** (RV2digit)	-0.233~-0.037 (POPD)
2	product innovation dummy				0.230***~0.829*** (RV2digit)	-0.557~-0.177 (POPD)
3	log(patent count per capita)	0.52~0.55 (HHI)				1.42~1.75 (log(POP))
4	innovation dummy product innovation dummy process innovation dummy		0.0172 (innovation_LQ2digit) -0.0692 (product_LQ2digit) 0.0404 (process_LQ2digit)		-0.3056 (innovation_RV2digit) -0.3854 (product_RV2digit) -0.4934 (process_RV2digit) 0.1898 (innovation_URV2digit) -0.3854 (product_URV2digit) 0.3810*	-0.0018 (innovation_POPD) 0.0160 (product innovation_POPD) -0.0118 (process innovation_POPD)

No.	Innovation measures	cluster	specialization	competition	Jacobs	urbanization
					(process_URV2digit)	
5	patent count		19.302***~26.984*** (own_emp)		6.713***~8.787*** (RV2digit) -10.024***~-5.407*** (URV2digit)	
6	innovation count		0.00168***~0.114*** (own emp)		-0.000382~-0.00000708 (No. of other_emp)	-0.0000551***~-0.0000111*** (POP)
7	patent count		-0.107**~-0.063(UK) (log(No. of own emp)) -0.130**~-0.014(Italy) (log(No. of own emp))		-0.160**~0.066 (UK) (log(No. of other emp)) -0.164**~0.086***(Italy)	-0.689* (POP)

No.	Innovation measures	cluster	specialization	competition	Jacobs	urbanization
8	product innovation dummy	-0.112~0.161 (cluster dummy)				
9	share of new product sales share of radical product sales	-3.3*~0.19 (regional R&D emp) -0.66~11.46* (regional R&D intensity)				
10	share of new product sales		-0.005~0.264* (log(LQ2digit))		0.085 (HHI)	
11	start-ups counts		0.1547***~0.3803*** (LQ2digit)	-0.0312~0.001 (Firm_LQ2digit)	0.4959 (RV2digit) 1.9529 (URV2digit) 0.1762*~0.2513** (DIVERSITY)	0.7564***~0.8105*** (metropolitan dummy)
12	product innovation dummy	-0.00000000628* (total regional emp)	-0.0395 (LQ3digit) 0.2657* (log(own_emp))		-19.9041* (HHI)	

No.	Innovation measures	cluster	specialization	competition	Jacobs	urbanization
13	ln(total sales of new products)	0.782~5.048 (regional R&D intensity)	0.09~0.57 (own_emp)		-80.182~-20.889 (HHI)	
14	product innovation dummy share of new product sales	-0.404~2.096** (dummy_emp_r/emp) 0.118**~0.736** (sale_empr/emp)	-0.056~0.044** (dummy_LQ2digit) 0.007**~0.012** (sale_LQ2digit)			
15	product innovation dummy share of new product sales	0.068*~0.095*** (dummy_cluster dummy) 0.02**~0.028*** (sale_cluster dummy)				
16	innovation counts		-0.527~-0.142 (LQ4digit)	-0.175***~0.576*** (Firm_LQ2digit)	0.069**~0.104** (related_LQ4digit)	
17	share of new product sales	0.639**~0.642** (cluster dummy)				

No.	Innovation measures	cluster	specialization	competition	Jacobs	urbanization
18	patent count		0.4***~0.42*** (ln(LQ4digit))		0.63***~0.89*** (ln(KIS)) 0.77***~0.92*** (ln(patent_inverse_gini)) 0.87***~0.9*** (ln(emp_inverse_gini) 2.62**~2.68** (ln(TD))	0.71***~0.85*** (ln(POP))
19	patent count		0.182*** (own_emp) 0.169** (concentrated industry dummy)			1.281*** (ln(GDPpercapita)) 1.023***~1.156*** (ln(POPD))
20	patent count stock	-0.469** (science park dummy) -0.621 (industry park dummy) 0.849*** (spontaneous cluster dummy)				

No.	Innovation measures	cluster	specialization	competition	Jacobs	urbanization
21	innovation dummy	-0.056~0.335 (ln(empdensity))	-0.0007*~0.0051*** (own_emp/empr)			
22	innovation counts share of new product sales radical innovation dummy	1.85*** (count_total firm)	1.15*** (count_LQ2digit) 0.09** (share_LQ2digit) 1.14 (radical_LQ2digit)	-0.34* (count_Firm_LQ2digit)	1.1 (count_1-gini) -0.08 (share_1-gini) -0.68* (radical_1-gini)	
23	innovation dummy		8.4e-8* (own_emp)		5.58e-4~1.72e-3** (No. of emp in service industries)	

Table A3. Statistical description by subgroup

			N	MEAN	SD	MIN	MAX
total			413	0.039	0.097	-0.224	0.774
innovation	Y_count		164	0.087	0.131	-0.224	0.774
	Y_dummy		171	0.005	0.44	-0.179	0.248
	Y_sale		78	0.013	0.04	-0.089	0.145
agglomeration externalities	cluster dummy		32	0.048	0.059	-0.75	0.215
	concentration(HHI)*		67	0.007	0.048	-0.047	0.265
	SIZE	SIZE	80	0.077	0.159	-0.076	0.774
		SIZE_R&D	21	0.024	0.064	-0.076	0.181
		SIZE_emp	19	0.017	0.043	-0.028	0.17
		SIZE_GDP	2	0.4	0.013	0.391	0.409
		SIZE_urban	34	0.093	0.22	-0.036	0.774
		SIZE_dummy(city)	4	0.089	0.013	0.078	0.107
		SIZE_share*	22	0.014	0.04	-0.028	0.17
	OWN	OWN	155	0.031	0.079	-0.224	0.295
		Own_LQ	64	0.032	0.098	-0.224	0.295
		Own_size	52	0.053	0.072	-0.169	0.232
		Own_share	4	-0.003	0.019	-0.021	0.018
		Own_stock	2	0.022	0.014	0.012	0.032
		Own_com	33	-0.0001	0.026	-0.053	0.097
	OTHER		53	0.027	0.073	-0.151	0.248
	Diversity		97	0.027	0.063	-0.179	0.265
Industry	manufacture+service		135	0.034	0.07	-0.179	0.265
	service		28	0.09	0.035	-0.027	0.17
	manufacture		250	0.045	0.112	-0.224	0.774
	high technology		66	0.066	0.152	-0.047	0.774
country	Belgium		48	0.133	0.0326	-0.049	0.17
	Germany		16	0.162	265	-0.021	0.774
	Italy		86	0.031	0.055	-0.075	0.224

	Netherland	100	0.006	0.043	-0.179	0.145
	Norway	13	0.007	0.027	-0.031	0.039
	Sweden	10	0.144	0.035	0.086	0.189
	UK	40	0.015	0.047	-0.041	0.113
	China	48	0.005	0.01	-0.025	0.023
	Taiwan	3	0.156	0.066	0.085	0.215
	Israel	4	0.04	0.231	-0.169	0.248
	US	17	0.051	0.155	-0.224	0.265
	Mixed	28	0.206	0.094	0.084	0.401
geographical unit	Level2	58	0.023	0.07	-0.169	0.248
	Level3	177	0.064	0.128	-0.224	0.774
	Level4	178	0.02	0.053	-0.179	0.215
Data type	Panel	153	0.036	0.068	-0.076	0.265
	Cross	260	0.041	0.11	-0.224	0.774

Table A4. Industry classification by characteristics of technology regime

Technological regimes	HIgh	Low
Appropriability	Manufacture of Food Products (10) Manufacture of Beverages (11) Manufacture of wearing apparel, Clothing Accessories and Fur Articles (14) Manufacture of Wood Products of Wood and Cork; Except Furniture (16) Manufacture of Phamaceuticals, Medicinal Chemicals and Botanical Products (21) Manufacture of Rubber and Plastic Products Manufacture of Electronic Components, Computer, Radio, Television and Communication Equipment and Apparatuses (26) Manufacture of Medical, Precision and Optical Instruments, Watches and Clocks (27) Manufacture of Other Machinery and Equipment (29)	Manufacture of Textiles, Except Apparel (13) Tanning and Dressing of Leather, Manufacture of Luggage and Footwear (15) Manufacture of Pulp, Paper and Paper products (17) Printing and Reproduction of Recorded Media (18) Manufacture of Coke, hard-coal and lignite fuel briquettes and Refined Petroleum Products (19) Manufacture of chemicals and chemical products except pharmaceuticals, medicinal chemicals (20) Manufacture of Other Non-metallic Mineral Products (23) Manufacture of Basic Metal Products (24) Manufacture of Fabricated Metal Products, Except Machinery and Furniture (25) Manufacture of electrical equipment (28) Manufacture of Motor Vehicles, Trailers and Semitrailers (30) Manufacture of Other Transport Equipment (31)

		Manufacture of Furniture (32) Other manufacturing (33)
Technological opportunities	Tanning and Dressing of Leather, Manufacture of Luggate and Footwear (15) Printing and Reproduction of Recorded Media (18) Manufacture of Coke, Hard-coal and Lignite fuel briquettes and Refined Petroleum Products (19) Manufacture of Chemicals and Chemical products except pharmaceturicals, medicinal chemicals (20) Manufacture of Pharmaceuticals, Medicinal Chemicals and Botanical products (21) Manufacture of Electronic Components, Computer, Radio, Television and Communication Equipment and Apparatuses (26) Manufacture of Medical, precision and Optical Instruments, Watches and Clocks (27) Manufacture of electrical equipment (28) Manufacture of Other machinery and Eqipment (29) Manufacture of Motor, Vehicles, Trailers and Semitrailers (30)	Manufacture of Food Products (10) Manufacture of Beverages (11) Manufacture of Textiles, Except Apparel (13) Manufacture of wearing apparel, Clothing Accessories and Fur Articles (14) Manufacture of Wood Products of Wood and Cork ; Except Furniture (16) Manufacture of Pulp, Paper and Paper products (17) Manufacture of Coke, hard-coal and lignite fuel briquettes and Refined Petroleum Products (19) Manufacture of Rubber and Plastic Products (22) Manufacture of Other Non-metallic Mineral Products (23) Manufacture of Basic Metal Products (24) Manufacture of Fabricated Metal Products, Except Machinery and Furniture (25) Manufacture of Other Transport Equipment (31) Manufacture of Furniture (32)

		Other manufacturing (33)
Cumulativeness	Manufacture of Food products (10) Manufacture of Textile, Except Apparel (13) Manufacture of wearing apparel, Clothing Accessories and Fur Articles (14) Manufacture of Pulp, Paper and Paper Products (17) Printing and Reproduction of Recorded Media (18) Manufacture of Basic Metal Products (24) Manufacture of Motor Vehicles, Trailers and Semitrailers (30) Manufacture of Other Transport Equipment (31)	Manufacture of Beverages (11) Tanning and Dressing of Leather , Manufacture of Luggage and Footwear (15) Manufacture of Wood Products of Wood and Cork ; Except Furniture (16) Manufacture of Coke, hard-coal and lignite fuel briquettes and Refined Petroleum Products (19) Manufacture of chemicals and chemical products except pharmaceuticals, medicinal chemicals (20) Manufacture of Pharmaceuticals, Medicinal Chemicals and Botanical Products (21) Manufacture of Rubber and Plastic Products (22) Manufacture of Other Non-metallic Mineral Products (23) Manufacture of Fabricated Metal Products, Except Machinery and Furniture (25) Manufacture of Electronic Components, Computer, Radio, Television and Communication Equipment and Apparatuses (26) Manufacture of Medical, Precision and Optical Instruments, Watches

		and Clocks (27) Manufacture of electrical equipment (28) Manufacture of Other Machinery and Equipment (29) Manufacture of Furniture (32) Other manufacturing (33)

Abstract (Korean)

혁신은 경제 성장과 산업 발전을 주도하는 중요한 동력으로 간주된다. 포터가 제안한 클러스터는 지역 내 혁신을 촉진하고 국가 경쟁력을 강화하기 위한 전략으로 많은 학자들의 관심을 끌었으나, 클러스터의 외부성에 대한 체계적이고 포괄적인 분석의 부족으로 인해 클러스터 전략의 효과성에 대해 의문을 표하는 목소리 또한 높다. 더불어 집적의 외부성과 혁신이 가지는 다양한 특성과 그 관계에 대한 다면적인 특성 이해에 대한 필요성이 점차 증가하고 있다. 따라서 본 논문에서는 혁신과 집적의 외부성이 가지는 다면적인 특성과 이로 인한 혼합된 연구 결과를 이해하기 위해 한국 기업 수준 데이터를 활용한 다수준 분석을 통해 혁신과 집적의 외부성과의 관계를 기존 문헌의 방법론을 적용하여 정량적으로 분석하고(제3장), 관련 문헌 중 선택된 26건의 연구를 기반으로 메타 회귀 분석을 수행하였다 (제4장).

제3장에서는 한국 기업의 데이터를 활용하여 지역-산업 단위와 지역 단위의 집적과 관련된 지표와 혁신 간의 상관 관계를 탐구하기 위해 기업 수준의 다수준 분석 모델을 사용하였다. 기존 문헌에서 흔히 채택되는 전문화와 다양성 개념에 근거한 이론적 프레임워크 내에서, 본 연구는 주로 전문화와 다양성과 관련된 지표를 활용하여 집적의 외부성을 측정하였다. 더욱이 분석은 집적의 규모 효과를 살펴보기 위해 한국의 수도권에 해당하는 지역의 인구 밀도와 독특한 특성의 더미 변수를 도입했다. 그 결과 다양성과 혁신 간의 관계를 확인하였고, 인구밀도로 측정되는 집적의 규모에 대한 양의

효과를 확인하였으나, 수도권 더미에 대한 음의 계수를 통해 수도권이 집적의 양의 효과를 상쇄할 정도로 높은 부정적 효과를 가짐을 확인하였다. 특히, 본 연구는 기술체제에 대한 세 개의 조절 변수를 도입하였는데, 이는 기술적 특성을 포착하여 집적과 혁신 간의 관계에 영향을 미치는 맥락적 요인(contextual factors)을 파악하고 이질성을 연구하기 위한 목적이다. 본 분석을 통해, 관련 문헌에서 널리 활용되고, 잘 알려진 이론적 프레임워크의 기본 가정들을 검토하고, 집적과 혁신 간의 관계를 분석하는 데 사용되는 기존 방법론의 한계를 제시한다.

제4장은 광범위한 문헌 조사를 토대로 혁신과 집적의 외부성에 초점을 맞춘 26건의 연구를 선택하여 메타 회귀 분석을 수행했다. 제2장의 문헌 조사와 제3장의 경험적 분석 결과에서 얻은 통찰력을 기반으로, 이 장은 이질성을 유발하는 이론적 및 경험적 요인을 규명하고 분류하였다. 특히, 본 연구에서는 관련 문헌에서 흔히 사용되는 종속 변수인 혁신 지표와 집적의 외부성을 포착하는 대리 지표 간의 이질성을 강조한다. 이 분류를 기반으로, 본 연구는 각 이질성의 요인과 관련된 조절 변수를 코드화하여 관련 문헌의 연구 결과에서 혁신과 집적 간의 관계에 어떤 방향과 정도로 영향을 미치는지를 정량적으로 확인하고자 하였다. 또한 이 장의 목적은 메타 회귀 분석을 통해 혁신과 집적 간의 관계에 대한 다양한 결과로부터 일관된 패턴이나 공유된 통찰력이 추출될 수 있는지를 탐구한다. 그 결과 다양성이 환경적 요인이 달라져도 일관된 효과를 나타냄을 확인하였고, 문헌결과의 이질성의 대부분은 분석 모형에서 사용되는 지표들 간의 이질성에 의한

것이라는 결과를 제시하였다.

제5장에서는 본 논문의 주요 결과를 요약하며, 제2장의 문헌 조사와 제3장의 분석에서 제기된 혁신과 집적 간의 관계에 대한 이질성과, 제4장에서 메타 회귀 분석을 통해 확인된 이질성을 분석 결과와 통합하여 정리한다. 그 결과, 유사한 연구 질문을 다루는 연구들이 사용하는 지표 간의 이질성이 해당 연구 분야에서 관측되는 혼재된 결과에 중요하게 기여했음을 강조한다. 더욱이, 이 장은 제3장의 분석 결과를 철저하게 재검토하며, 혁신과 집적의 외부성에 대한 현재의 이론적 및 방법론적 프레임워크의 한계와 결함을 밝히고 개선 가능한 방향을 제안하고자 하였다. 게다가, 본 장은 이러한 제한을 극복하기 위해 최근 문헌 연구에서 수행되는 관련성(relatedness) 및 복잡계 관점(complex system perspective)을 소개하며, 이 분야가 이러한 한계에 어떻게 대처하며 발전하고 있는지 제시한다. 더욱이, 이 장은 본 학위 논문이 관찰된 이질성에 관한 가치 있는 시각을 제시하고, 연구 방법론을 더욱 정제하는 방향을 제안하는 측면에서 기여점이 있음을 강조한다. 추가적으로 이러한 통찰력과 기여를 통해 본 학위 논문의 연구는 경제적 변화와 지식 생산 과정의 변화하는 동학을 고려한 집적과 혁신 사이의 복잡한 상호작용에 대한 세심한 관점을 제공하며, 관련 연구에 체계적으로 포괄적인 결론을 도출하기 위한 연구방향을 제안한다.

주요어 : 집적, 외부성, 파급효과, 메타회귀분석, 다수준분석, 기술레짐
학 번 : 2014-30281