Empirical Studies on the Sources of Agglomeration Economies

by

Kenta Ikeuchi

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics

Graduate School of Economics
Hitotsubashi University

2015

Abstract

This thesis addresses the issues concerning the sources, or mechanisms, of agglomeration economies. It is well known that economic activities are spatially concentrated or clustered in certain areas and such agglomerated areas generate relatively higher economic growth than less agglomerated areas do. The sources of agglomeration economies, mainly the three factors proposed by Alfred Marshall, have been investigated and formalized by many subsequent scholars: input sharing, knowledge spillovers, and labor pooling. In this thesis, each of these sources of agglomeration economies is examined empirically. The structure and contents of the chapters are described below.

Chapter 1

Previous Studies on the Sources of Agglomeration Economies and Overview of the Thesis

Chapter 1 summarizes the related literature and provides an overview of the following chapters of the thesis. First, the motivations of this thesis are pointed out through literature surveys. Several pieces of empirical evidence regarding the geographic agglomeration of economic activities are discussed and then several main empirical and theoretical works related to the topic of agglomeration economies are also explained. Second, Marshall's (1920) three sources of agglomeration economies (cited by many previous studies) are introduced: input sharing, knowledge spillovers, and labor pooling. Finally, the structure of this thesis is summarized. The contents and aims of the following chapters are briefly introduced.

Chapter 2

Transportation Costs and Regional Productivity Difference in Japan: An Empirical Study of the New Economic Geography Theory

In Chapter 2, the effects of transportation costs on an agglomeration economy and the dynamics of industrial location are examined empirically. Combining a spatial demand function derived in the theoretical new economic geography (NEG) literature with a production function, a revenue production function is proposed, which captures the effects of transportation costs on a firm's revenue. The suggested revenue production function makes it possible to relate the geographic agglomeration economy with the transportation costs, which has not been done in previous empirical studies. It performs an empirical examination of the model with regional panel data of the manufacturing sector in Japan. A city level panel data constructed mainly from the Census of Manufacturers for the 1996−2006 is used for empirical analysis. The revenue function including parameters for the transportation costs of each industry is estimated. The results support the existence of positive transportation costs, and show the estimated transportation costs for the manufacturing sector are higher than those for the primary sector and lower than those for the service sector.

Chapter 3

Plant Productivity Dynamics and Private and Public R&D Spillovers: Technological, Geographic and Relational Proximity

Chapter 3 investigates the knowledge spillovers and examines the effects of R&D spillovers on total factor productivity (TFP) with a large panel of Japanese manufacturing plants matched with R&D survey data (1987–2007). This chapter simultaneously examines the role of public (university and research institutions) and private (firm) R&D spillovers, and the different effects due to technological, geographic, and relational (buyer-supplier) proximity. Estimating dynamic long difference models and allowing for a gradual convergence in TFP and geographic decay in spillover effects, the results show that technologically proximate private R&D stocks positively affect TFP growth, which decay with distance and become negligible at around 500 kilometers. In addition to knowledge spillovers from technologically proximate R&D stocks, 'relational' spillovers from buyer and supplier R&D stocks exert positive effects on TFP growth that are similar in magnitude. The elasticity of TFP is highest for public R&D (corrected for industrial relevance), in particular for plants operated by R&D-conducting firms. This chapter does not find evidence of geographic decay in the impact of public and relational spillovers. Over time, declining R&D spillovers appear to be responsible for a substantial part of the decline in the rate of TFP growth. The exit of proximate plants operated by R&D-intensive firms plays a notable role in this process and is an important phenomenon in major industrial agglomerations such as Tokyo, Osaka, and Kanagawa.

Chapter 4

Effects of Regional Human Capital on Business Entry: a Comparison of Independent Startups and New Subsidiaries in Different Industries

Chapter 4 and Chapter 5 examine the effects of labor pooling. Chapter 4 aims to investigate

the regional determinants of entry with special attention to the effects of regional human capital, using prefecture-level data from Japan. On the basis of some recent studies in the field, this chapter investigates the effects of several regional factors on business entry, distinguishing between independent startups and new subsidiaries of existing firms on the one hand and comparing different sectors on the other. Using pooled regional data at the prefecture level for the period between 1996 and 2006, it is simultaneously estimated the impact of various regional factors, including human capital, on the number of independent startups and new subsidiaries for each industry sector. Estimation results demonstrate considerable differences between independent startups and subsidiaries as well as among different industry sectors with regard to the impact of regional human capital on business entry. Considering the possible implications, the results suggest that the regional policy to activate business startups should focus more on the differences between encouraging local entrepreneurship and attracting new subsidiaries, and recognize that these differences may vary even within the service sector, depending on what type of human capital is required.

Chapter 5

R&D, Innovation, and Business Performance of Japanese Startups: A Comparison with Established Firms

Despite the importance of innovation activities in business startups, few studies have comprehensively compared these undertakings to equivalent ones in established firms. Therefore, Chapter 5 compares the determinants of R&D intensity, innovation, and firm performance in start-ups and established firms with a three-stage model, using comparable datasets in Japan. Data on start-ups is obtained from an original questionnaire survey series for Japanese start-ups that were carried out annually from 2008 to 2011 and comprises 894 firms less than 2 years of age at time of the initial survey in 2008. Comparable data of established firms were obtained from the Japanese National Innovation Survey 2009 conducted in 2009 by the National Institute of Science and Technology Policy (NISTEP), as official statistics carried out according to the Oslo Manual and the Community Innovation Survey 2010 in the EU and comprise more than 1000 firms. The empirical results suggest that 1) the local labor pooling of research-relevant workforces (professional and technical occupations) positively relates to the R&D intensity of the firms located in the neighborhood, 2) the effects of public financial support on R&D intensity are generally positive but smaller for startups, 3) the effects of research cooperation with business partners or universities on innovation are generally positive

but larger for startups, and 4) the effects of product and process innovation on labor productivity (level and growth) are positive both for startups and established firms.

Contents

Chapter 1 Previous Studies on the Sources of Agglomeration Economies and Overview of the Thesis

1. Introduction

This thesis addresses the issues concerning the sources or mechanisms of agglomeration economies. This chapter summarizes the related literature on agglomeration economies and their sources and then provides an overview of the remaining chapters. First, in Section 2, definitions of agglomeration economies are discussed and several main empirical and theoretical works related to the topic of agglomeration economies are introduced.

Second, in Section 3, Marshall's (1920) three sources of agglomeration economies (cited by many previous studies) are explained: input sharing, knowledge spillovers, and labor pooling. Since each subsequent chapter empirically examines one of the three sources of agglomeration economies, Section 3 provides readers of this thesis a guideline about where to map each chapter in the whole framework and literature on sources of agglomeration economies.

Finally, in Section 4, the structure of this thesis is described. The contents and aims of the following chapters are briefly introduced.

2. Agglomeration Economies

It is well known that economic activities are spatially concentrated or clustered in certain areas and such agglomerated areas achieve relatively higher economic growth than less agglomerated areas do. Ellison and Glaeser (1997) developed a test using a simple discrete index to measure whether observed levels of geographic concentration would be greater than would be expected to arise randomly; they empirically found that almost all industries were somewhat localized in the US manufacturing sector. Similarly, a geographic pattern of localization of industries was also found in Japan by Mori et al. (2005). Duranton and Overman (2005) developed new distance-based tests of localization to avoid problems relating to the scale and borders of geographical units, which were encountered by Ellison and Glaeser (1997), and found that almost half of four-digit industries[1] were localized, most of them at scales below 50

[1] Duranton and Overman (2005) consider 234 industries out of 239 that have more than 10 establishments and

km in the UK manufacturing sector. In this regard, Nakajima et al. (2012) found similar results for Japan.

A large body of empirical literature confirmed a positive correlation of economic agglomeration with regional productivity growth, innovation, wage rate, and entrepreneurship. Many studies in the fields of regional and urban economics found significantly positive effects of regional economic density on regional productivity (e.g., Ciccone and Hall 1996; Ciccone 2002; Tveteras and Battese 2006; Brülhart and Mathys 2008; Combes, Duranton, Gobillon and Roux 2008). Feldman and Audretsch (1996) found a positive relationship between product innovation and regional agglomeration. David B. Audretsch and his colleagues developed the *knowledge spillover theory of entrepreneurship;* their empirical analyses confirmed that agglomeration of knowledge stock accumulated by incumbent firms in a region would enhance the level of entrepreneurial activity in that area (Audretsch and Keilbach 2007; Acs et al. 2009; Acs and Audretsch 2010).

Theoretical research on the explanation of agglomeration goes back to at least 1826, when Thünen suggested comprehensive lists of centrifugal and centripetal forces of industrial agglomeration. Thünen's (1826) arguments were more formalized as the central place theory by Losch (1940) and Christaller (1933). Marshall (1920) described the mechanisms where an increase in production volume would cause an elaborate subdivision of labor through work specialization and would improve production efficiency. Then he divided the economies arising from such an increase in production scale into two classes: internal- and external-scale economies. The former involved the scale economies within a single establishment or firm. The latter (called agglomeration economies by later scholars) could often be secured by the concentration of many small businesses of a similar character in particular localities or as commonly said, the localization of industry. Quigley (1988, p. 127) identified four periods of intense study on cities:

The first of these periods, [which] occurred in the decade after World War I, included the first systematic empirical analysis of the forces affecting the location of firms and households within cities (e.g., Haig 1926) […]. The second of these periods, [which] began in the mid-1960s, formalized many of the insights about location incentives within urban areas [that] had been uncovered a half century before, mixed them with the logic of Thünen's (1826) ancient theories and applied them to the household sector (Alonso 1964; Kain 1962), [in which] the

conclude that 122 industries among those are statistically significantly localized at some distance.

widely observed pattern of decline in housing prices and a steeper decline in land prices [at a certain] distance [from] the urban center [arose] from a residential equilibrium in which higher income households live[d] farther from downtown and commute[d] longer distances, but [occupied] more housing in less dense accommodations[…]. The third concentrated period [… began from …] the late 1950s when the Regional Plan Association and a group of economists at Harvard combined [to conduct] a three-year study of the New York Metropolitan Region [… with] a hallmark […] concept of "external economies of scale" (Vernon 1962). We are in the midst of the fourth of these, […] ushered in by reconsideration in the 1980s of the nature of economic growth […]. The first two sets of developments emphasized the intra-metropolitan location patterns of households and firms. The latter two have emphasized the overall patterns of growth of cities and metropolitan regions.

Krugman (1991) pioneered the new economic geography (NEG), a novel theory to explain the formation of a large variety of economic agglomeration (or concentration) in geographical space. The NEG theory incorporates agglomeration advantages and location choice in a formal, general equilibrium framework through the interaction between scale economies and transportation costs. A growing body of literature in the field of *empirical NEG* provides evidence supporting the theory (Hansen 2005; Breinlisch 2006; Head and Mayer 2006; Pons 2007).

Regarding the types of economic agglomeration, the distinction between localization (or specialization) and urbanization (or diversification) has been intensively discussed in the literature. There have been debates on the effects of localization and urbanization on productivity (Nakamura 1985; Henderson 2003). Feldman and Audretsch (1996) compared the impacts of specialization and diversification on product innovation introduced by local firms and concluded that the effect of diversification dominated that of specialization. Bosma et al. (2008) also found that the localization economy would work for the location of independent start-up firms and the urbanization economy for the location of new subsidiaries.[2]

3. Sources of Agglomeration Economies

What is the mechanism through which cities grow? Alfred Marshall (1920) conducted the

[2]Another explanation for the positive relationship between agglomeration and productivity is the sorting or selection mechanism through market competition (Arimoto et al. 2014).

first comprehensive analysis of the sources of external economies arising from agglomeration. As the sources of agglomeration economies, mainly the three factors were proposed by Marshall (1920) and had been investigated and formalized by many scholars: input sharing (Goldstein and Gronberg 1984; Krugman 1993; Helsley and Strange 2002), knowledge spillovers (Glaeser 1999), and labor pooling (Helsley and Strange 1990). Quigley (1988) provided further literature review of theoretical works.

Rosenthal and Strange (2001) empirically examined these three factors simultaneously by regressing the localization index of industries—developed by Ellison and Glaeser (1997) on industry characteristics that are proxies for Marshall's theories of localization economies-—based on the US data. Their results supported all these three hypotheses. Ellison et al. (2010) also obtained similar findings, using co-agglomeration indices developed by Ellison and Glaeser (1997) and Duranton and Overman (2005), based on the US and UK data.

3.1. Input Sharing

The first factor—shared inputs in production and consumption—encompasses the "economies of localized industry" described by Marshall (1920, Book IV, Chapter X, para. 8):

[T]he economic use of expensive machinery can sometimes be attained in a very high degree in a district in which there is a large aggregate production of the same kind, even though no individual capital employed in the trade be very large. For subsidiary industries devoting themselves each to one small branch of the process of production, and working it for a great many of their neighbors, are able to keep in constant use machinery of the most highly specialized character, and to make it pay its expenses, though its original cost may have been high, and its rate of depreciation very rapid.

Economies from the shared inputs arise from scale economies in the production process and transportation costs of input factors. Co-agglomeration of suppliers and customers attains efficiency of production and reduces the transportation costs for the goods traded. This means that agglomeration of demand also accelerates the scale economy for supplier industries. Krugman (1993) established the NEG theory, which is a general equilibrium model of industry location, using a monopolistic competition framework developed by Dixit and Stiglitz (1977). It assumes the economies of scale and transportation costs, as well as shows that the forward and backward linkages produce an agglomeration of the production of manufactured goods. Ellison and Glaeser (1997) found that industries appeared to co-agglomerate both with important upstream suppliers and downstream customers. Helsley and Strange (2002) specified a model where input sharing would encourage innovation by reducing the cost of realizing ideas. They showed that a dense network of input suppliers would facilitate innovation by making it less costly to bring new ideas to fruition.

Chapter 2 involves this issue. In the chapter a "tractable" model from the NEG theory is derived, in which the transportation cost and production function are incorporated. Then the proposed model is applied to a region-industry level panel data mainly from the Census of Manufacturers in Japan, and the spatial effects of transportation costs on plant-level (revenue) productivity are estimated.

3.2. Knowledge Spillovers

The second factor of the agglomeration economies involves localization of spillovers of knowledge or ideas. Technological knowledge is the most important yet invisible factor for production efficiency; it is also known that nobody can appropriate the knowledge perfectly. Marshall (1920, Book IV, Chapter X, para. 7) wrote:

[G]reat are the advantages which people following the same skilled trade get from near neighborhood to one another. The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously. Good work is rightly appreciated; inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas.

Jaffe et al. (1993) showed evidence for the localization of knowledge spillovers, using patent citations as a direct proxy for the knowledge spillovers. It is well established in the literature that the effects of research and development (R&D) spillovers on productivity are enhanced by technological and geographic proximity (Adams and Jaffe 1996; Orlando 2004; Mairesse and Mulkay 2008; Aldieri and Cincera 2009; Griffith et al. 2009; Lychagin et al. 2010; Bloom et al. 2013). Moreover, a different research stream focusing on the role of knowledge spillovers, from public research conducted at universities and research institutes, suggests the importance of such spillovers, with an explicit role of proximity (e.g., Jaffe 1989; Adams 1990; Anselin et al. 1997; Furman et al. 2005).

Despite the increasing number of large-scale, firm-level studies on R&D spillovers, existing studies have several limitations in scope and methodology (Jaffe et al., 1993; Adams and Jaffe, 1996; Aldieri and Cincera, 2009; Lychagin et al., 2010; Bloom et al., 2013; Orlando, 2004; Griffith et al., 2009; Mairesse and Mulkay, 2008). First, they typically rely on data of listed firms, aggregating over various locations and technologies in which they are active. Second, they focus on inter-firm, private spillovers, while neglecting the role of public research. Third, R&D spillovers at the firm level are in most cases modeled as a function of proximity among technology portfolios of the firms, while the role of spillovers through supplier and customer linkages receives only limited attention. Chapter 3 deals with these issues.

3.3. Labor Pooling

A third possible reason why a metropolitan area may provide greater economic efficiency arises from the reduction in transaction costs. This factor includes the possibility of better matching between worker skills and job requirements (Quigley 1998). Marshall (1920, Book IV, Chapter X, para. 9) wrote:

[A] localized industry gains a great advantage from the fact that it offers a constant market for skill. Employers are apt to resort to any place where they are likely to find a good choice of workers with the special skill which they require; while men seeking employment naturally go to places where there are many employers who need such skill as theirs and where therefore it is likely to find a good market.

Helsley and Strange (1990) developed a model with heterogeneous workers and firms and derived an agglomeration economy in the labor market from a matching process between them. Helsley and Strange (2002) argued that specialized labor agglomeration would attract entrepreneurs. Numerous studies focused on the effects of the qualitative and quantitative composition of regional labor force on regional start-up ratio. Some studies demonstrated that the ratio of white-collar to blue-collar workers (Keeble and Walker 1994; Fotopoulos and Spence 1999) and the proportions of college graduates (Guesnier 1994; Armington and Acs 2002; Acs and Armington 2004) and the workforce in professional and managerial occupations (Guesnier 1994; Hart and Gudgin 1994) would have positive effects on the start-up ratio.

Why and how does such regional human capital positively affect the start-up of new businesses? Acs and Armington (2004, 2006) indicated three reasons. First, the agglomeration of a highly educated and skilled labor force generates entrepreneurs with fresh ideas for creating new businesses (Glaeser et al. 1992). Second, it also promotes local knowledge spillovers, by which new start-ups are initiated and sustained (Reynolds et al. 1995). Third, it facilitates the search for and employment of skilled labor by founders of new firms (Rauch 1993).

Chapters 4 and 5 empirically examine this issue. In Chapter 4, the impacts of regional human capital on the start-up ratio are examined, using Japanese data at the prefecture level, focusing on the differences between independent start-ups and new subsidiaries of existing firms and across diverse industries and sectors. In Chapter 5, the impacts of accessibility to local research personnel on R&D expenditure, innovation, and productivity of the firms are examined, focusing on the differences between younger and older firms.

4. Structure of the Thesis

This thesis empirically examines each source of agglomeration economies in the following chapters. In Chapter 2, the roles of transportation costs are empirically investigated. Then the effects of knowledge spillovers on productivity are examined in Chapter 3. The impacts of labor pooling on the level of entrepreneurship are analyzed in Chapter 4, while those on R&D and innovation are examined in Chapter 5. Chapter 6 concludes the thesis.

This thesis includes materials from four papers that the author co-authored. Chapter 3 is based on Belderbos et al. (2013) and Ikeuchi et al. (2013), both co-authored with Rene Belderbos, Kyoji Fukao, YoungGak Kim, and HyeogUg Kwon. Chapters 4 and 5 are based on Ikeuchi and Okamuro (2011) and Ikeuchi and Okamuro (2013), respectively, both co-authored with Hiroyuki Okamuro.

Chapter 2 Transportation Costs and Regional Productivity Difference in Japan: An Empirical Study of the New Economic Geography Theory

1. Introduction

This chapter analyzes spatial effects of industrial geographic locations on regional productivity. According to the theoretical literature on the new economic geography (NEG), scale economies and transportation costs create an agglomeration economy (e.g., Krugman 1980; Fujita et al. 1999; Fujita and Thisse 2002). The NEG theories have implied that firms located in different places face different demand functions. Klette and Griliches (1996) suggested the inconsistency of scale estimators obtained from production function regressions when firms operated in an incomplete competitive market and prices differed among them. Levinsohn and Melitz (2002) discussed the biases in productivity estimation in the case of a product-differentiated market (industry).

Economic efficiency and optimality of industrial locations have been analyzed theoretically (e.g., Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud 2003).[3] There are also many empirical studies about locational effects on regional productivity or growth. However, few empirical studies on the topic estimate the regional productivity model that is directly derived from the NEG theoretical models.[4] For this reason, the estimation results in previous studies have not been directly linked to the theoretical models; thus, they have been unable to evaluate the efficiency and optimality of actual industrial locations.

Unfortunately, the NEG theoretical models are too complicated to estimate in a straightforward manner and their nonlinearity causes other computational issues that should be solved. This chapter challenges these issues. A "tractable" model is derived from the NEG theory. The proposed empirical framework that is directly derived from the NEG theory makes

---

[3] Baldwin et al. (2003) used simplified versions of the models of Krugman (1980) and Fujita et al. (1999).

[4] Mion (2004) and Hanson (2005) estimated the wage equation of the NEG theory. Additionally, Crozet (2004) and Pons, Paluzie, Silvestre and Tirado (2007) estimated the labor migration model, which was derived from the NEG theory. Davis and Weinstein (2008) analyzed home market effects on production location in NEG models. See Brakman, Garretsen, Gorter, Horst and Schramm (2009) and Redding (2010) for further literature review.

it possible to evaluate the efficiency and optimality of actual industrial locations, using estimation results. The proposed framework is then applied to a region-industry level panel data from the Census of Manufacturers in Japan, and the spatial effects of transportation costs on plant-level productivity are estimated.[5]

This chapter is organized as follows. Section 2 sets out a theoretical model from the literature. Section 3 discusses the typical data restriction for a researcher and the modification of the proposed model. Section 4 presents the empirical methodology and Section 5 reports the estimation results. Section 6 concludes the chapter.

2. Theoretical Model

2.1. Production Function

There are assumed to be n_I industries and n_R regions. I denotes a set of industries by $I \equiv \{1, 2, \dots, n_I\}$, a set of regions by $R \equiv \{1, 2, \dots, n_R\}$, and a set of $n_{J_{ri}}$ firms within region r and industry i by $J_{ri} \equiv \{1, 2, \dots, n_{Jri}\}$. The production function of firm $j \in J_{ri}$ is defined by:

$$q_j = \Omega_j \ell_j^{\alpha_L^i} k_j^{\alpha_K^i} \prod_{i' \in I} m_{i'j}^{\alpha_{i'}^i} \quad \forall j \in J_{ri}, \forall r \in R, \forall i \in I \ , \tag{2-1}$$

where q_j represents the quantity produced by firm j, Ω_j is the knowledge (total factor productivity [TFP]) of firm j, ℓ_j and k_j are firm j's labor input and capital stock, respectively, and $m_{i'j}$ is an aggregate of the varieties of individual intermediate inputs of firm j in industry $i' \in I$, defined by a CES (constant elasticity of substitution) function of the form:

---

[5] As another aspect of industrial location, knowledge agglomeration might create an agglomeration economy through mutual learning of firms located in close proximity, an effect also referred to as "knowledge spillovers." Thus, industrial geographic locations affect regional productivity through these two paths—transportation costs and knowledge spillovers—and these two effects determine the optimal industrial locations and geographic resource allocation. Knowledge spillover effects are investigated mainly in the literature on industrial organization (e.g., Monjon and Waelbroeck 2003; Fosfuri and Ronde 2004; Alvarez and Molero 2005; Ornaghi 2006; Henderson 2007). The effects of regional economic density on regional productivity are estimated as "agglomeration effects" in the literature on regional and urban economics (e.g., Ciccone and Hall 1996; Ciccone 2002; Tveteras and Battese 2006; Brulhart and Mathys 2008; Combes, Duranton, Gobillon and Roux 2008).

$$m_{i'j} \equiv \left[\sum_{r'\in R}\sum_{j'\in J_{r'i'}} x_{j'j}^{\frac{\sigma_{i'}-1}{\sigma_{i'}}}\right]^{\frac{\sigma_{i'}}{\sigma_{i'}-1}}, \tag{2-2}$$

where $x_{j'j}$ represents firm j's intermediate inputs of each available variety produced by firm $j' \in J_{r'i'}$ (which is located in region r' and belongs to industry i'), $\sigma_{i'} > 1$ represents the elasticity of substitution between any two intermediate varieties of industry i',[6] e.g., $x_{j'j}$ and $x_{j''j}$, where $j', j'' \in J_{i'}$; α_L^i, α_K^i, and α_h^i in (2-1) and $\sigma_{i'}$ in (2-2) are the production technology parameters to be estimated.

2.2. Demand Function

In this subsection, assumptions for consumers' preference are set and the free on board (f.o.b.) price of firm j, p_j, is related to its location. The firm's sales revenue can also be related to its location, independently of its production technology and its input level of production factors.

Consumers' demand

Following Fujita et al. (1999), consumers' utility function is assumed to be that of the Dixit-Stiglitz model of monopolistic competition (Dixit and Stiglitz 1977):

$$U_r = \prod_{i\in I} Z_{ri}^{\mu_i}, \forall r \in R, \tag{2-3}$$

where U_r represents the utility of consumers living in region r, Z_{ri} represents the consumption aggregate of commodity i, which is defined by:

$$Z_{ri} = \left(\sum_{r'\in R}\sum_{j'\in J_{r'i}} z_{rj'}^{\frac{\sigma_i-1}{\sigma_i}}\right)^{\frac{\sigma_i}{\sigma_i-1}}, \forall i \in I \quad , \tag{2-4}$$

where $z_{rj'}$ is the consumption volume of variety j' in region r, and σ_i is the elasticity of substitution between any two varieties of commodity i. Denoting total income of consumers in

---

6 In this chapter, we assume the elasticity of substitution between any two goods in the same industry to be constant because we cannot use firm-level micro data in the following empirical analysis. However, this does not imply that all firms or consumers in the same industry may purchase from all regions. Because of substantial transportation costs of the goods, they are assumed to purchase goods mostly from closely located producers.

region r by Y_r and the price of variety j' in region r by $p_{j'r}$, the consumers in region r maximize utility (2-3), subject to the budget constraint:

$$\sum_{i \in I} \sum_{r' \in R} \sum_{j' \in J_{r'i}} p_{j'r} z_{rj'} \le Y_r.$$

As the solution of such a utility maximization problem, the consumer demand function in region r' for the variety $j \in J_{ri}$ (in industry $i \in I$, which is produced in region r) is therefore derived:

$$z_{r'j} = \left(\frac{p_{jr'}}{G_{ir'}}\right)^{-\sigma_i} \left(\frac{\mu_{r'} Y_{r'}}{G_{ir'}}\right), \qquad (2\text{-}5)$$

where

$$G_{ir'} \equiv \left[\sum_{r'' \in R} \sum_{j'' \in J_{r''i}} p_{j''r''}^{-(\sigma_i - 1)}\right]^{-\frac{1}{\sigma_i - 1}} \qquad (2\text{-}6)$$

is the price index of commodity $i \in I$ in region r, corresponding to the definition of quantity indexes (2-2) and (2-4).

Intermediate demand

In the same way, in solving the profit maximization problem of firm $j' \in J_{r'i'}$, which is located in region r' and belongs to industry i', the intermediate demand of firm j' for a variety produced by firm $j \in J_{ri}$, which is located in region r and belongs to industry i, can be derived:

$$x_{jj'} = \left(\frac{p_{jr'}}{G_{ir'}}\right)^{-\sigma_i} \left(\frac{\beta_i^{i'} M_{j'}}{G_{ir'}}\right) \qquad \text{for } j \in J_{ri} \text{ and } j' \in J_{r'i'}, \qquad (2\text{-}7)$$

where $M_{j'}$ represents the total expenditure of firm j' on its intermediate inputs, and $\beta_i^{i'}$ is the share of expenditure of industry i' on intermediate inputs from industry i, which is assumed constant for the firms within the same industry.

Total demand

Taking the summation of consumer demand function (2-5) in region r' and the intermediate demand function (2-7) of all firms in region r', the total demand for the product of an individual firm $j \in J_{ri}$ located in region r and producing a variety of industry i is:

$$q_{jr'} = \left(z_{r'j} + \sum_{i' \in I} \sum_{j' \in J_{r'i'}} x_{jj'}\right) = \left(\frac{p_{jr'}}{G_{ir'}}\right)^{-\sigma_i} \left(\frac{E_{r'i}}{G_{ir'}}\right), \qquad (2\text{-}8)$$

where $E_{r'i}$ is the total expenditure in region r' on the products of industry i, which is defined by:

$$E_{r'i} \equiv \mu_i Y_{r'} + \sum_{i' \in I} \sum_{j' \in J_{r'i'}} \beta_i^{i'} M_{j'}. \tag{2-9}$$

Thus, equation (2-8) indicates that the level of demand for an individual firm j in region r' depends on the relative price of the firm (relative to the industry level price index) $p_{jr'}/G_{ir'}$ and the industry-level, total demand in a real term in region $E_{r'i}/G_{ir'}$.

2.3. Transportation Costs and Regional Pricing

According to the theoretical NEG models, a scale economy and transportation costs determine industrial geographic locations (Fujita et al. 1999). The scale economy generates a centripetal force for the industrial locations; the transportation costs generate both the centripetal and dispersion forces in case of the existence of some immobile factors. Therefore, to examine the optimal industrial locations, it is necessary to measure levels of the transportation cost of each commodity. In this subsection, a transportation cost function of firms are defined and incorporated to the demand function defined in equation (2-8).

Transportation cost to deliver the products of firm $j \in J_{ri}$, which produces a variety of industry i from region r to region r', is assumed to be:

$$TC_{jr'} = p_{jr'} q_{jr'} \left(\frac{T(d_{rr'}) - 1}{T(d_{rr'})} \right), \tag{2-10}$$

where $p_{jr'}$ represents the cost, insurance and freight (c.i.f.) price or delivered price of the product of firm j in region r', and $q_{jr'}$ represents the quantity of the product of firm j, which is transported from production region r to consuming region r'. The function $T(d_{rr'})$ represents the transportation technology firm j, and $d_{rr'} > 0$ is the (geographic) distance between region r and region r'. The transportation technology $T(d)$ of (2-10) is assumed to satisfy the following conditions:

- $T(d) > 0$ for any $d > 0$,
- $\frac{\partial T(d)}{\partial d} > 0$ for any $d > 0$, and
- $T(d) = 1$ if $d = 0$,
- If $d \to \infty$ then $T(d) \to 0$.

In this chapter, the distance weight function for transportation technology is specified as:

$$T(d_{rr'}) = e^{\tau_i d_{rr'}}. \tag{2-11}$$

Suppose firm j determines the c.i.f. price $p_{jr'}$ to maximize the profit:

$$\pi_j = \sum_{r' \in R} \left(p_{jr'} q_{jr'} - TC_{jr'} \right) - C(q_j),$$

where $C(q_j)$ is the production cost of firm j as a function of the firm's total products, $q_j = \sum_{r' \in R} q_{jr'}$. As the solution of this profit maximization problem under the monopolistic competition assumption, the c.i.f. price of firm $j \in J_{ri}$ for each region r' is given by:

$$p_{jr'} = p_j e^{\tau_i d_{rr'}}, \tag{2-12}$$

where p_j represents the mill or f.o.b. price ($p_j \equiv p_{jr}$, where r is the index of the location of firm j). This implies that the c.i.f. price $p_{jr'}$ for each region r' is proportional to the f.o.b. price of firm j, p_j.[7]

Substituting $p_{jr'}$ in the total demand function for firm j in each region (2-8) by (2-12), demand of region r' for firm j is rewritten as:

$$q_{jr'} = p_j^{-\sigma_i} E_{r'i} G_{ir'}^{\sigma_i - 1} e^{-\sigma_i \tau_i d_{rr'}} \text{ for } j \in J_{ri},$$

so the total demand for firm $j \in J_{ri}$ is:

$$q_j \equiv \sum_{r' \in R} q_{jr'} = p_j^{-\sigma_i} \left(\sum_{r' \in R} E_{r'i} G_{ir'}^{\sigma_i - 1} e^{-\sigma_i \tau_i d_{rr'}} \right). \tag{2-13}$$

Therefore, the inverse demand function is obtained as:

$$p_j = q_j^{-\frac{1}{\sigma_i}} \left(\sum_{r' \in R} E_{r'i} G_{ir'}^{\sigma_i - 1} e^{-\sigma_i \tau_i d_{rr'}} \right)^{\frac{1}{\sigma_i}}, \tag{2-14}$$

which determines the f.o.b. price of products of firm j in region r in industry i.

2.4. Revenue Function

It is assumed that the share of expenditure to the intermediate inputs from industry i' to total expenditure to intermediate inputs is determined so that the firm maximizes its profit. Then the following expression should be hold:

$$\frac{G_{i'r} m_{i'j}}{\sum_{i'' \in I} G_{i''r} m_{i''j}} = \frac{\alpha_{i'}^{i}}{\sum_{i'' \in I} \alpha_{i''}^{i}}. \tag{2-15}$$

Using equation (2-15), the aggregate index of the quantity of intermediate inputs from all industries in the production function (2-1) can be substituted by the following equation:

$$\prod_{i' \in I} m_{i'j}^{\alpha_{i'}^{i}} = \left(\frac{M_j^{\alpha_M^i}}{\prod_{i' \in I} G_{i'r}^{\alpha_{i'}^{i}}} \right) \eta_i, \tag{2-16}$$

[7] Moreover, the price-setting rule determined by equation (2-12) implies that assumption (2-10) is equivalent to the assumption of an "iceberg" form of transportation costs.

where M_j is firm j's expenditure for intermediate input goods, defined as $M_j \equiv \sum_{i' \in I} G_{i'r} m_{i'j}$, $\alpha_M^i = \sum_{i' \in I} \alpha_{i'}^i$, and η_i in equation (2-16) is the industry-specific constant term, defined as $\eta_i \equiv \left(\alpha_M^i\right)^{-\alpha_M^i} \prod_{i' \in I} \left(\alpha_{i'}^i\right)^{\alpha_{i'}^i}$.

Combining the inverse demand function (2-14) and equation (2-16) with the production function (2-1) yields a revenue function:

$$V_j = \phi_{ri}^{\frac{1}{\sigma_i}} \left(\Omega_j \ell_j^{\alpha_L^i} k_j^{\alpha_K^i} M_j^{\alpha_M^i} \prod_{i' \in I} G_{i'r}^{-\alpha_{i'}^i} \eta_i \right)^{\frac{\sigma_i - 1}{\sigma_i}}, \tag{2-17}$$

where $V_j \equiv p_j q_j$ represents the nominal revenue of firm j, and ϕ_{ri} denotes the agglomeration of the demand for industry i in region r, which is appeared in demand function (2-13), defined as:

$$\phi_{ri} \equiv \sum_{r' \in R} E_{r'i} G_{ir'}^{\sigma_i - 1} e^{-\sigma_i \tau_i d_{rr'}}, \tag{2-18}$$

Ω_j is the knowledge (or TFP) of firm j, and ℓ_j and k_j are the amounts of labor inputs and capital stock, respectively.

Substituting equation (2-12) for $p_{jr'}$ in equation (2-6), the price index of industry i in region r, G_{ir} can be rewritten as:

$$G_{ir} = \left[\sum_{r' \in R} \sum_{j \in J_{r'i}} \left(p_j e^{\tau_i d_{rr'}} \right)^{-(\sigma_i - 1)} \right]^{-\frac{1}{\sigma_i - 1}}. \tag{2-19}$$

Although G_{ir} is defined by equation (2-19), this data is rarely obtained. Therefore, multiplying p_j to the both sides of the demand function (2-13) and rearranging the equation, the f.o.b. price of firm j, p_j can be rewritten as:

$$p_j = \left(\frac{\phi_{ri}}{V_j} \right)^{\frac{1}{\sigma_i - 1}} \text{ for } j \in J_{ri}.^{8} \tag{2-20}$$

By inserting equation (2-20) into (2-19), the price index of commodity i in region r can be rewritten as:

$$G_{ir} = \left[\sum_{r' \in R} V_{r'i} \phi_{r'i}^{-1} e^{-(\sigma_i - 1)\tau_i d_{rr'}} \right]^{-\frac{1}{\sigma_i - 1}} \tag{2-21}$$

where $V_{r'i} = \sum_{j \in J_{r'i}} V_j$ is the total revenue of all firms located in region r' and producing each variety of industry i. Equation (2-21) suggests that the regional price index, G_{ir}, for a

[8] Using equation (2-18), the demand function (2-13) can be rewritten as $q_j = p_j^{-\sigma_i} \phi_{ri}$. Multiplying both sides of this equation by p_j, $p_j q_j = p_j^{-(\sigma_i - 1)} \phi_{ri}$. Replacing $p_j q_j$ by V_j and rearranging the equation to solve for p_j, equation (2-20) can be obtained.

commodity i is higher in a region r, where there is a larger supply of the commodity V_{ri} and less demand for the commodity ϕ_{ri} in the region itself, r, or its neighboring regions, r' with small $d_{rr'}$. This indicates that the lack of geographic competition in a region leads to a higher price in the region.

Given equation (2-18) and (2-21), expenditures and revenues for each region and industry, E_{ri} and V_{ri}, respectively, and the total factor productivity of each firm, Ω_j, each of firms determines the level of labor and capital inputs, ℓ_j and k_j, respectively, and the expenditure for intermediate inputs, M_j, so that it maximize the profit defined by revenue function (2-17) minus total cost, which is determined by quantities and prices of factor inputs. In this study, equilibrium in the labor market and capital market are both given.[9]

2.5. Special Cases

To determine the characteristics of the derived revenue function (2-17), it should be a good examination to review several extreme, special cases as follows. In several cases, it is not able to identify τ_i in revenue function (2-17).

Case 1

First, assuming there is no transportation cost, $\tau_i = 0$, for all $\in I$, then the demand and supply condition (price index) is constant across regions:

$$\phi_{ri} = \sum_{r' \in R} E_{r'i} \bar{G}_i^{\sigma_i - 1}, \forall r \in R$$

and

$$\bar{G}_i = G_{ir} = \left[\sum_{r' \in R} \sum_{j \in J_{ri}} (p_j)^{-(\sigma_i - 1)} \right]^{-\frac{1}{\sigma_i - 1}}, \forall r \in R\,.$$

Thus, revenue function becomes:

$$V_j = \tilde{\Omega}_j \ell_j^{\tilde{\alpha}_L^i} k_j^{\tilde{\alpha}_K^i} M_j^{\tilde{\alpha}_M^i} \tilde{A}_i\ ,$$

where $\tilde{A}_i = \bar{\phi}_i^{\frac{1}{\sigma_i}} \left(A_i \prod_{i' \in I} \bar{G}_{i'}^{-\alpha_{i'}^i} \right)^{\frac{\sigma_i - 1}{\sigma_i}}$, $\tilde{\Omega}_j = \Omega_j \frac{\sigma_i - 1}{\sigma_i}$, $\tilde{\alpha}_L^i = \alpha_L^i \frac{\sigma_i - 1}{\sigma_i}$, $\tilde{\alpha}_K^i = \alpha_K^i \frac{\sigma_i - 1}{\sigma_i}$, and

[9] General equilibrium analysis of the model is conducted in Fujita et al. (1999).

$\tilde{\alpha}_M^i = \alpha_M^i \frac{\sigma_i - 1}{\sigma_i}$. In this case, the revenue function is almost the same as the production function (2-1). Thus, it is difficult to identify the parameter for elasticity of substitution by using the production function estimation under the condition of extremely low transportation costs.

Case 2

In the next case, the products are perfectly homogeneous and the price elasticity of demand is infinite, $\sigma_i = \infty$ for all $i \in I$, then $G_{ir} = 1$ and $\phi_{ri} = E_{ri}$ for all $r \in R$. Thus,

$$V_j = \Omega_j \ell_j^{\alpha_L^i} k_j^{\alpha_K^i} x_j^{\alpha_M^i} A_i.$$

Moreover, in Case 2, similar to Case 1, revenue function is exactly the same as production function (2-1). Thus, the transportation cost parameter cannot be identified from the production function estimation if the products are perfectly homogeneous and the price elasticity of demand $\sigma_i = \infty$. Because the revenue function in Case 2 has the same form of that function in Case 1, unless other information is not available, it is also not possible to distinguish Case 2 from Case 1. In other words, in both these cases, firm revenues do not depend on demand or supply agglomeration. Thus, in the empirical analysis, these cases are treated as a null model.

Case 3

The opposite assumption to Case 1 is infinite transportation cost, $\tau_i = \infty$ for all $i \in I$. Then $\phi_{ri} = E_{ri} G_{ir}^{\sigma_i - 1}$ and $G_{ir}^{-(\sigma_i - 1)} = \sum_{j \in J_{ri}} p_j^{-(\sigma_i - 1)}$ for all $r \in R$. Thus:

$$V_j = \left(\frac{E_{ri} G_{ir}^{\sigma_i - 1}}{\prod_{i' \in I} G_{i'r}^{\alpha_{i'}^i (\sigma_i - 1)}} \right)^{\frac{1}{\sigma_i}} \left(\Omega_j \ell_j^{\alpha_L^i} k_j^{\alpha_K^i} x_j^{\alpha_M^i} A_i \right)^{\frac{\sigma_i - 1}{\sigma_i}}.$$

In this case, although regional expenditure and price index, E_{ri} and G_{ir}, only affect the revenues of firms within the region itself, σ's and α's are recovered by an estimation of the equation.

3. Data

3.1. Data Sources

The data used in empirical analyses is obtained from some statistics in Japan for the 1996–2006 period. First, information on the production inputs and outputs of the firms in the manufacturing industries at the regional level was collected from the *Census of Manufacturers*, which is conducted by the Ministry of Economy, Trade and Industry in Japan. This data contains information on all plants located in Japan, with at least four employees. Data at the two-digit Japanese Standard Industrial Classification (JSIC) level by city (*shi, ku, cho*, and *son* levels in Japanese) are available. The two-digit JSIC includes 22 industries. The number of cities in Japan almost reached 2,000 in the latest year of the observation period.[10]

Second, for the estimates of the input coefficients of each industry, the *Input-Output Tables for Japan* prepared by the Ministry of Internal Affairs and Communications are used. Because these tables are produced every five years, the input coefficients in the intermediate years are interpolated.

Third, for the regional distribution of the workers in non-manufacturing industries, the *Establishment and Enterprise Census* is used, which is conducted by the Ministry of Internal Affairs and Communications in Japan. This census is updated every three or five years and contains information on all establishments (excluding self-employment in the primary sector; agriculture, forestry, and fishery). The shares of the number of workers employed in each industry in the region in the intermediate years are also interpolated.

Fourth, for the regional distribution of the population, the *Population Census* is used, which is conducted by the Ministry of Internal Affairs and Communications in Japan. This census is updated every five years and contains the information on population of each region. The data in the intermediate years are also interpolated.

3.2. Data Description

A panel of Japanese regional panel data from the *Census of Manufacturers*, *Establishment*

[10]From the mid-1990s, the Japanese administrative division of the regions had been restructured and hundreds of regions were merged with each other during the 1996–2006 period. To ensure consistency, I used the latest (and the largest-meshed) classification for the whole period.

and Enterprise Census, *Population Census* and *Input-Output Tables* is constructed. This dataset was composed of 1,928 Japanese regions, 22 manufacturing industries (two digits), and 15 non-manufacturing sectors (one digit), spanning the 1996–2006 period. Summary statistics are provided in tables in the appendix.

3.3. Further Assumptions for Data Restrictions

This chapter has several data restrictions (which may be common among other researchers):

1. Individual, firm-level (micro-level) data could not be obtained, only an aggregate (region-industry level) data.
2. Regional input and output data could be obtained only for manufacturing industries, not for nonmanufacturing industries.

3.4. Homogeneity of Firms

The following assumptions are needed in order to estimate the models discussed in the previous section from an aggregated, region-industry level dataset (instead of firm-level micro data):

- Production technologies of the firms are the same in each industry:
  $$\alpha_L^j = \alpha_L^i, \alpha_K^j = \alpha_K^i, \alpha_h^j = \alpha_h^i, \forall h \in I, \forall j \in J_{ri}, r \in R. \tag{2-22}$$
- The production input quantity for each factor is the same across the firms located in the same region and belonging to the same industry:
  $$\ell_j = \ell_{j'}, k_j = k_{j'}, m_{ij} = m_{i'j'}, \forall i \in I, \forall j, j' \in J_{ri}. \tag{2-23}$$
- Unobserved efficiency is the same across the firms located in the same region and belonging to the same industry:
  $$\Omega_j = \Omega_{ri}, \forall j \in J_{ri}. \tag{2-24}$$

Then the region-industry level aggregation of the model (2-1) is:

$$q_{ri} = \frac{Q_{ri}}{N_{ri}} = \Omega_{ri} \ell_{ri}^{\alpha_L^i} k_{ri}^{\alpha_K^i} \prod\nolimits_{i' \in I} m_{i'ri}^{\alpha_{i'}^i}, \forall j \in J_{ri}, \tag{2-25}$$

where q_{ri}, ℓ_{ri}, k_{ri}, and m_{ri} represent the quantities of the total output and production inputs of the firms in industry i in region r. From equations (2-17) and (2-25), the region-industry, aggregated revenue function is obtained:

$$V_{ri} = A_i^{\frac{\sigma_i-1}{\sigma_i}} \phi_{ri}^{\frac{1}{\sigma_i}} \prod_{i' \in I} G_{i'r}^{-\alpha_{i'}^i \frac{\sigma_i-1}{\sigma_i}} \left(\Omega_{ri} \ell_{ri}^{\alpha_L^i} k_{ri}^{\alpha_K^i} M_{ri}^{\alpha_M^i}\right)^{\frac{\sigma_i-1}{\sigma_i}}. \tag{2-26}$$

3.5. Non-manufacturing Industries

Typically, detailed data of production inputs and outputs for the non-manufacturing sector cannot be obtained at the regional level. Even if estimating the production function of non-manufacturing firms is ignored, in order to estimate the regional revenue function for the manufacturing sector, regional demand potential ϕ and regional factor price index G for the manufacturing industries, as well as the regional revenue and expenditure for intermediate inputs of the non-manufacturing sector are needed.

Assume that national level revenue V_i and total intermediate expenditure M_i of nonmanufacturing industry i are observable and that their regional level amounts are proportionate to the number of workers employed in the industry in the region. Specifically, the revenue of the firms in industry i in region r is approximated by:

$$V_{ri} = V_i \frac{L_{ri}}{\sum_{r' \in R} L_{r'i}}, \forall i \in I_S \tag{2-27}$$

and the expenditure for intermediate inputs of the firms in industry i in region r by

$$M_{ri} = M_i \frac{L_{ri}}{\sum_{r' \in R} L_{r'i}}, \forall i \in I_S, \tag{2-28}$$

where I_S is the set of non-manufacturing industries. In most cases, researchers can observe the number of workers at the detailed regional level, even for non-manufacturing industries.

4. Estimation Method

4.1. Revenue Function Estimation

Taking natural logarithms of both sides of equation (2-26) and adding a time dimension t, the following equation is obtained for all regions $r \in R$ and all manufacturing industries $i \in I$:

$$\ln V_{ri} = \tilde{\alpha}_0^i + \tilde{\alpha}_L^i \ln \ell_{ri}^{(t)} + \tilde{\alpha}_K^i \ln k_{ri}^{(t)} + \tilde{\alpha}_M^i \ln M_{ri}^{(t)} + \frac{\ln \phi_{ri}^{(t)}}{\sigma_i} - \tilde{\alpha}_M^i \sum_{i' \in I} \hat{\beta}_{i'}^i \ln G_{i'r}^{(t)} + \alpha_{ri} + \alpha_i^{(t)} + \omega_{ri}^{(t)} \tag{2-29}$$

for $t = 1, 2, \ldots, T$, where $\tilde{\alpha}_0^i = \frac{\sigma_i-1}{\sigma_i} \ln A_i$, $\tilde{\alpha}_L^i = \frac{\sigma_i-1}{\sigma_i} \alpha_L^i$, $\tilde{\alpha}_K^i = \frac{\sigma_i-1}{\sigma_i} \alpha_K^i$, $\tilde{\alpha}_M^i = \frac{\sigma_i-1}{\sigma_i} \alpha_M^i$, $\omega_{ri}^{(t)} =$

$\frac{\sigma_i-1}{\sigma_i}\ln\Omega_i^{(t)}$. ϕ_{ri} is the agglomeration of demand for the firms in region r in industry i, defined by equation (2-18):

$$\phi_{ri}^{(t)} \equiv \sum_{r'\in R} e^{-\sigma_i\tau_i d_{rr'}} E_{r'i}^{(t)} \left[G_{r'i}^{(t)}\right]^{\sigma_i-1}, \tag{2-30}$$

where $E_{ri}^{(t)}$ is the expenditure for commodity i in region r, defined by equation (2-9):

$$E_{ri}^{(t)} = \mu_i Y_r + \sum_{i'\in I} \beta_i^{i'} M_{ri'}^{(t)}. \tag{2-31}$$

$G_{i'r}^{(t)}$ is the price index of commodity i' in region r, defined by equation (2-21):

$$G_{i'r}^{(t)} = \left[\sum_{r'\in R} e^{-(\sigma_{i'}-1)\tau_{i'} d_{rr'}} \frac{V_{r'i'}^{(t)}}{\phi_{r'i'}^{(t)}}\right]^{-\frac{1}{\sigma_{i'}-1}}. \tag{2-32}$$

Since $E_{r'i}^{(t)}$ correlates with $V_{r'i}^{(t)}$, $\phi_{ri}^{(t)}$ can be a function of $V_{r'i}^{(t)}$ for all r'. G_{ir} is also determined by $V_{r'i'}^{(t)}$ for all r' and i'. This means that the left hand side of equation (2-29), the revenue of industry i in region r, $V_{ri}^{(t)}$, is related with that in any other regions, $V_{r'i}^{(t)}$. Therefore, the proposed model is a special case of the spatial auto regressive (SAR) models.[11]

α_{ri} is a region-industry specific effect which is estimated by including region-industry dummies. It may capture the effects of distance from network hubs such as ports, airports, highway interchanges, and train stations. $\alpha_i^{(t)}$ is an industry-year specific effect which is estimated by including industry-year dummies. It may include the influence from the international market through export and import.

4.2. Econometric Issues

Consistency of the estimator for the parameters relies on assumptions about the conditional mean of the unobserved efficiency term, $\omega_{ri}^{(t)}$. According to Wooldridge (2002), in the nonlinear regression, if:

$$\mathrm{E}\left(\omega_{ri}^{(t)} \middle| \frac{\partial \ln V_{ri}^{(t)}}{\partial \boldsymbol{\theta}}, \ln L_{ri}^{(t)}, \ln K_{ri}^{(t)}, \ln M_{ri}^{(t)}\right) = 0, \tag{2-33}$$

[11] Indeed, the right hand side of equation (2-29) also includes the variable in the left hand side, $V_{ri}^{(t)}$. Thus, this is a simultaneous equation of the revenue in all regions and industries. However, we could not control for this simultaneity because of the computational burden. Therefore, estimation results should be biased to some extent. This is one of the important technical challenges remained for the future research.

where $\boldsymbol{\theta} = (\tau_1\ \tau_2\ \dots\ \tau_I\ \sigma_1\ \sigma_2\ \dots\ \sigma_I)'$, then the nonlinear least squares (NLS) estimator is consistent, where the NLS estimator minimizes the objective function:

$$O_{NLS}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_r \sum_i \sum_t \left(\ln V_{ri}^{(t)} - h_{ri}^{(t)}(\boldsymbol{\theta}) - \mathbf{X}_{ri}^{(t)} \boldsymbol{\alpha}_i\right)^2, \tag{2-34}$$

where $\boldsymbol{\alpha}_i = \left(\tilde{\alpha}_L^i\ \tilde{\alpha}_K^i\ \tilde{\alpha}_M^i\right)'$, $h_{ri}^{(t)}(\boldsymbol{\theta}) = \frac{\ln \phi_{ri}^{(t)}(\tau_i,\sigma_i)}{\sigma_i} - \tilde{\alpha}_M^i \sum_{i' \in I} \hat{\beta}_{i'}^i \ln G_{i'r}^{(t)}(\tau_i, \sigma_i)$, and $\mathbf{X}_{ri}^{(t)} = \left(\ln L_{ri}^{(t)}\ \ln K_{ri}^{(t)}\ \ln M_{ri}^{(t)}\right)$.

As suggested in the large body of the econometric literature on production function estimation, the condition in which equation (2-33) holds true cannot be met in practice. To obtain consistent estimators when the condition of equation (2-33) is violated, the term of unobserved efficiency of the firms in region *r* in industry *i* in year *t* is decomposed into three components, as follows:

$$\omega_{ri}^{(t)} = \mu_i^{(t)} + \bar{\omega}_{ri} + u_{ri}^{(t)}, \tag{2-35}$$

where $\mu_i^{(t)}$ is the industry-year-specific efficiency shock, $\bar{\omega}_{ri}$ is the persistent efficiency difference across regions in each industry, and $u_{ri}^{(t)}$ is a time-dependent, region-specific efficiency shock for each industry.

By including industry-year dummies as explanatory variables, the first component is controlled out. Next, if:

$$\mathrm{E}\left(\bar{\omega}_{ri} \middle| \frac{\partial \ln V_{ri}^{(t)}}{\partial \boldsymbol{\theta}}, \mathbf{X}_{ri}^{(t)}\right) \neq 0, \tag{2-36}$$

then the NLS estimator is no longer consistent. To conduct consistent estimation, the fixed effects (FE) or the first difference (FD) estimator should at least be used in this case. The FE estimator minimizes the objective function:

$$O_{FE}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_r \sum_i \sum_t \left(\ln \ddot{V}_{ri}^{(t)} - \ddot{h}_{ri}^{(t)}(\boldsymbol{\theta}) - \ddot{\mathbf{X}}_{ri}^{(t)} \boldsymbol{\alpha}_i\right)^2, \tag{2-37}$$

where the variable with " $\ddot{}$ " is demeaned, e.g., $\ddot{y}_{ri}^{(t)} = \mathbf{y}_{ri}^{(t)} - \overline{y_{ri}}$.

The FD estimator minimizes the least squares of the error term, using the FD of the original dataset:

$$O_{FD}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_r \sum_i \sum_t \left(\Delta \ln V_{ri}^{(t)} - \Delta h_{ri}^{(t)}(\boldsymbol{\theta}) - \Delta \mathbf{X}_{ri}^{(t)} \boldsymbol{\alpha}_i\right)^2. \tag{2-38}$$

If:

$$\mathrm{E}\left(u_{ri}^{(t)} \middle| \frac{\partial \ln V_{ri}^{(t)}}{\partial \boldsymbol{\theta}}, \mathbf{X}_{ri}^{(t)}\right) \neq 0, \tag{2-39}$$

then the FE and FD estimators are no longer consistent, as well as the NLS estimator. In addition to equation (2-36), in order to conduct consistent estimation under the condition of (2-39), an instrumental variable estimator via the GMM (general method of moments) estimation technique should be used. The GMM estimator minimizes the objective function:

$$O_{GMM}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = (\textstyle\sum_r \sum_i \mathbf{e}'_{ri}\mathbf{W}_{ri})\mathbf{A}(\sum_r \sum_i \mathbf{e}'_{ri}\mathbf{W}_{ri})', \tag{2-40}$$

where $\mathbf{e}_{ri}$ is a $(T \times 1)$ vector of residuals for region r in industry i, $\mathbf{W}_{ri}$ is a $(T \times P)$ matrix of instrumental variables for region r in industry i, and $\mathbf{A}$ is a $(P \times P)$ positive definite matrix called weighting matrix. For the specification of residual vector $\mathbf{e}$, an effective choice of instruments $\mathbf{W}$, and an efficient estimation of weighting matrix $\mathbf{A}$, the difference GMM (DGMM) developed by Arellano and Bond (1991) and the system GMM (SGMM) developed by Blundell and Bond (1998) are used. In summary, the DIF-GMM method uses the FD residuals $\Delta u_{ri}^{(t)}$ for the element of $\mathbf{e}_{ri}$ and lagged instrumental variables in level $\mathbf{X}_{ri}^{(t-s)}$, $\mathbf{X}_{ri}^{(t-s-1)}$, ..., $\mathbf{X}_{ri}^{(1)}$ for $\mathbf{W}_{ri}^{(t)}$. Additionally, the SYS-GMM method adds the residuals in level $u_{ri}^{(t)}$ to $\mathbf{e}_{ri}$ and the lagged difference of instrumental variables $\Delta\mathbf{X}_{ri}^{(t-s)}$ to $\mathbf{W}_{ri}^{(t)}$. Estimation Procedure

The parameters of the model can be calculated by the following estimation procedure. Since both equations (2-30) and (2-32) are nonlinear functions and are dependent on the unknown variable the agglomeration of demand, ϕ, and the price index, G, each other, the Gauss-Newton method is employed and a fixed-point iteration is nested.

1. Initialize the parameters for transportation cost and elasticity of substitution, $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\sigma}}$, respectively.
2. Loop the following steps until objective function $O(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\alpha}})$ is minimized.
   A) Initialize $\hat{\phi}_{ri}^{(t)}$ for each r, i, and t. Then loop the following steps until $\hat{\phi}_{ri}^{(t)}$ and $\hat{G}_{ri}^{(t)}$ do not change for all r, i, and t.
      (a) Given $\hat{\phi}_{ri}^{(t)}$ for all r and i, calculate $\hat{G}_{ri}^{(t)}$ for each r, i, and t, using equation (2-33).
      (b) Given $\hat{G}_{ri}^{(t)}$ for all r and i, update $\hat{\phi}_{ri}^{(t)}$ for each r, i, and t, using equation (2-31).
   B) Using $\hat{\boldsymbol{\phi}}$, $\hat{\mathbf{G}}$, and $\hat{\boldsymbol{\sigma}}$, estimate the parameters of the linear part of revenue function (2-30), i.e., $\tilde{\alpha}_L^i\ \tilde{\alpha}_K^i\ \tilde{\alpha}_M^i$, $\forall i \in I_M$, to minimize the augmented objective function $O(\hat{\boldsymbol{\alpha}}|\hat{\boldsymbol{\phi}}, \hat{\mathbf{G}}, \hat{\boldsymbol{\sigma}})$.
   C) Calculate the value of objective function $\hat{O} = O(\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\alpha}})$.
   D) Calculate the derivatives of the objective function:

$$\hat{\mathbf{J}} = \left(\frac{\partial O(\hat{\boldsymbol{\tau}},\hat{\boldsymbol{\sigma}},\hat{\boldsymbol{\alpha}})}{\partial \boldsymbol{\tau}'} \; \frac{\partial O(\hat{\boldsymbol{\tau}},\hat{\boldsymbol{\sigma}},\hat{\boldsymbol{\alpha}})}{\partial \boldsymbol{\sigma}'} \; \frac{\partial O(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})}{\partial \boldsymbol{\alpha}'}\right). \tag{2-41}$$

E) Update parameters $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\sigma}}$, using objective value $\hat{O}$ and the Jacobian $\hat{\mathbf{J}}$.

5. Results

5.1. Results of Revenue Function Estimation

Table 2.1 reports the estimation results of the revenue function, specifically the estimated values of the parameters in the nonlinear part of the model, τ and σ. Although the elasticity of outputs with respect to the factors of labor, capital, and intermediate inputs (α_L^i, α_K^i and α_M^i) are jointly estimated for each two-digit industry, their estimation results do not appear in the table to avoid excessive complexity. The first column presents the results of the NLS estimation, the second column shows the results of the FE model, and the third column provides the system GMM (SGMM) results. The transportation cost parameters are significantly positive for the manufacturing sector in all methods. Although the transportation cost parameters for the primary and service sectors[12] are negative in the NLS and FE models, respectively, in the SGMM estimation results, the parameter for the service sector becomes significantly positive[13] and its magnitude is higher than that for the manufacturing sector. Figure 2.1 illustrates the estimated transportation cost function for each sector.

The elasticity of substitution σ is also estimated for each sector. In the SGMM results, the value ranges from 16 for manufacturing to 58 for the service sector, indicating a higher degree of differentiation of the products in the manufacturing sector than in the primary and service sectors. However, the results of the GMM distance statistic test indicate that a null hypothesis, H0: $\tau = 0$ or $\sigma = \infty$, cannot be rejected for all industries, while it is rejected in the NLS and FE models.

[12] The primary sector includes agriculture, forestry, fisheries, and mining industry. Service sector includes all remaining industries not included in the primary and the manufacturing sector.

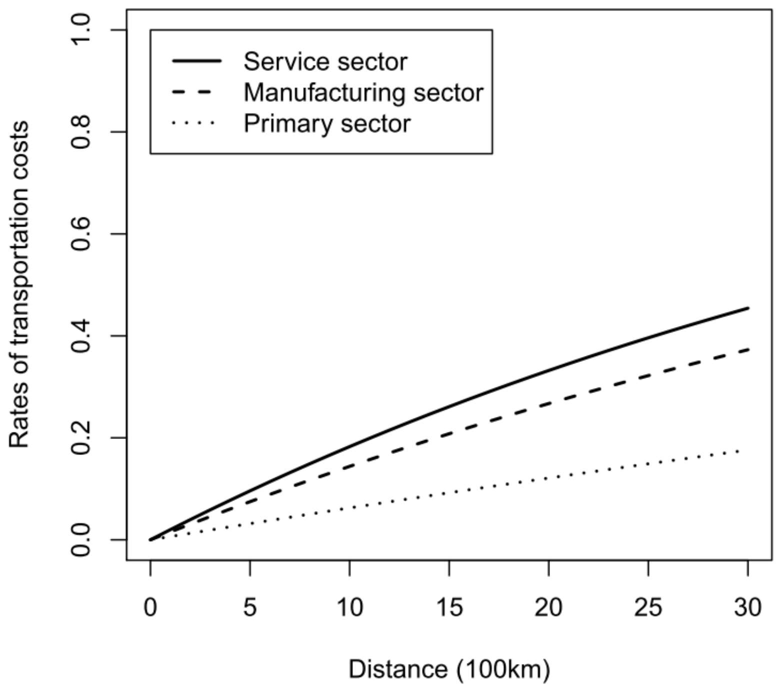
[13] We assume that the transportation costs on services imply the commuting and communication cost of sales persons and customers.

Table 2.1. Estimation results of the revenue function

		NLS	FE	SGMM
Primary sector	τ	-0.0007	-0.0247	0.0065^{***}
		[0.0101]	[0.0306]	[0.0022]
	σ	33.3211^{***}	29.9847^{***}	31.0388
		[0.0001]	[0.0777]	[25.4835]
Manufacturing	τ	0.0116^{***}	0.0166^{***}	0.0155^{***}
		[0.0009]	[0.006]	[0.0003]
	σ	34.0455^{***}	30.0461^{***}	16.1425^{***}
		[2.2607]	[0.2249]	[0.0793]
Service	τ	0.0406^{***}	-0.0354^{*}	0.0202^{***}
		[0.0038]	[0.0204]	[0.0019]
	σ	11.5428^{***}	30.0095^{***}	58.0532^{***}
		[2.4326]	[9.0884]	[16.1928]
H0: $\tau = 0$ or $\sigma = \infty$		$F(6, 116197) = 8920.6^{***}$	$F(6, 102782) = 2.6^{**}$	$\chi^2(6) = 1.2$

Notes: Robust standard errors for NLS and FE and ordinary standard errors for SGMM are enclosed in brackets. ***, **, and * are significant at $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively. In the SGMM estimation, 6 years and more lagged lnL, lnK, and lnM are used as "GMM-type" instruments and derivatives with respect to τ, and σ are treated as standard instruments without any lags. Sargan test statistic for SGMM is 2006.691 ($p = 0.996$) and the p-value of the Arellano-Bond test for AR(4) is 0.0292.

Figure 2.1 Estimated transportation function



5.2. Decomposition of Total Factor Productivity

The logarithm of the total factor productivity (TFP) for each region r and industry i is defined as:

$$\ln TFP_{ri} = \ln V_{ri} - \frac{\hat{\sigma}_i - 1}{\hat{\sigma}_i}\left(\hat{\alpha}_L^i \ln \ell_{ri} - \hat{\alpha}_K^i \ln k_{ri} - \hat{\alpha}_K^i \ln m_{ri}\right). \quad (2\text{-}42)$$

From equation (2-29), these TFPs can be decomposed into the demand potential and supply potential:

$$\ln TFP_{ri} = \frac{1}{\hat{\sigma}_i} \ln \hat{\phi}_{ri} + \frac{\hat{\sigma}_i - 1}{\hat{\sigma}_i} \hat{\alpha}_M^i \ln \hat{\psi}_{ri} + \omega_{ri}, \quad (2\text{-}43)$$

where $\hat{\phi}_{ri} = \sum_{r' \in R} e^{-\hat{\sigma}_i \hat{\tau}_i d_{rr'}} \frac{E_{ir'}}{\hat{G}_{ir'}}$ and $\hat{G}_{ir} = \sum_{r' \in R} e^{-(\hat{\sigma}_{i'} - 1)\hat{\tau}_{i'} d_{rr'}} \frac{V_{r'i'}}{\hat{\phi}_{r'i'}}$ for $\forall r \in R, \forall i \in I$

and $\hat{\psi}_{ri} = \prod_{i' \in I} \left(\hat{G}_{i'r}\right)^{-\hat{\beta}_{i'}^i}$ for $\forall r \in R, \forall i \in I$.

The first term in equation (2-43) indicates a demand potential for industry i in region r, and the second term in equation (2-43) reveals a supply potential for intermediate inputs of industry i in region r. Table 2.2 shows the difference in TFP between the region with a demand and supply potential above the 90 (or 50) percentile and the region with those below the 10 (or 50) percentile. The results indicate significant differences between those regions. The TFP in a region with the top 10% demand potential is 7.2% higher than that in a region with the bottom 10% demand potential on average. The TFP in a region with the top 10% supply potential is 7.0% higher than that in a region with the bottom 10% supply potential on average.

Table 2.2 Demand and supply agglomeration and TFP differences

		TFP difference in log
Demand potential	Top 10% - Bottom 10%	0.072***
		[0.004]
	Top 50% - Bottom 50%	0.057***
		[0.001]
Supply potential	Top 10% - Bottom 10%	0.070***
		[0.004]
	Top 50% - Bottom 50%	0.051***
		[0.001]

Notes: Standard errors are enclosed in brackets. *** is significant at $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively. The demand potential is $\ln \phi_{ri}$ and the supply potential is $\ln \hat{\psi}_{ri}$. The SGMM results are used for the calculations.

From equation (2-43), the variance of TFP in the natural logarithm across regions for each industry can be decomposed into six terms as follows:

$$\sum_{i\in I}\frac{\sum_{r\in R_i}\left(\ln TFP_{ri}-\overline{\ln TFP_i}\right)^2}{N_{R_i}}=\sum_{i\in I}\frac{\sum_{r\in R_i}\left(\frac{1}{\hat{\sigma}_i}\left(\frac{\ln\hat{\phi}_{ri}}{\hat{\sigma}_i}-\frac{\overline{\ln\hat{\phi}_i}}{\hat{\sigma}_i}\right)\right)^2}{N_{R_i}}+\sum_{i\in I}\frac{\sum_{r\in R_i}\left(\frac{\hat{\sigma}_i-1}{\hat{\sigma}_i}\hat{\alpha}_M^i\left(\ln\hat{\psi}_{ri}-\overline{\ln\hat{\psi}_i}\right)\right)^2}{N_{R_i}}+$$

$$\sum_{i\in I}\frac{\sum_{r\in R_i}(\omega_{ri}-\overline{\omega_i})^2}{N_{R_i}}+2\sum_{i\in I}\frac{\sum_{r\in R_i}\frac{1}{\hat{\sigma}_i}\left(\frac{\ln\hat{\phi}_{ri}}{\hat{\sigma}_i}-\frac{\overline{\ln\hat{\phi}_i}}{\hat{\sigma}_i}\right)\frac{\hat{\sigma}_i-1}{\hat{\sigma}_i}\hat{\alpha}_M^i\left(\ln\hat{\psi}_{ri}-\overline{\ln\hat{\psi}_i}\right)}{N_{R_i}}+$$

$$2\sum_{i\in I}\frac{\sum_{r\in R_i}\frac{1}{\hat{\sigma}_i}\left(\frac{\ln\hat{\phi}_{ri}}{\hat{\sigma}_i}-\frac{\overline{\ln\hat{\phi}_i}}{\hat{\sigma}_i}\right)(\omega_{ri}-\overline{\omega_i})}{N_{R_i}}+2\sum_{i\in I}\frac{\sum_{r\in R_i}\frac{\hat{\sigma}_i-1}{\hat{\sigma}_i}\hat{\alpha}_M^i\left(\ln\hat{\psi}_{ri}-\overline{\ln\hat{\psi}_i}\right)(\omega_{ri}-\overline{\omega_i})}{N_{R_i}}, \qquad (2\text{-}44)$$

where N_{R_i} is the number of regions in which one or more establishments in industry i exist. The first three terms in equation (2-44) denote the contribution of the variance of demand potential, supply potential and other factors, respectively, in the variance of TFP while the last three terms denotes the contribution of the covariance among these three terms. Table 2.3 shows the results of variance decomposition using equation (2-44). Demand and supply potential explain almost 1% while the other factors do almost 99% of the variance of TFP across regions. These results show that the other factors unrelated to the transportation costs on the products, such as technological knowledge or agglomeration of knowledge, might play important roles in the regional dispersion of TFP.

Table 2.3 Relative importance of demand and supply potential

	lnTFP	Demand potential	Supply potential	Other factors	Demand potential * Supply potential	Demand potential * Other factors	Supply potential * Other factors
Variance	0.04968	0.00036	0.00013	0.04915	0.00041	-0.00024	-0.00013
(%)	100.0	0.7	0.3	98.9	0.8	-0.5	-0.3

Notes: The demand potential is $\ln\phi_{ri}$ and the supply potential is $\ln\hat{\psi}_{ri}$. The SGMM results are used for the calculations. Calculations are based on deviations from industry-year mean for each variable.

6. Conclusion

In this chapter, the effects of transportation costs on agglomeration economy were examined empirically. Combining a spatial demand function derived from the theoretical NEG literature (e.g., Krugman 1980; Fujita et al. 1999) with a production function, this chapter proposed a revenue function, which would capture the effects of transportation costs on a firm's revenue. Since transportation costs generate these spatial effects on a firm's own products and its intermediate goods, the proposed revenue function makes it possible to relate the geographic agglomeration economy to the transportation costs, something not done in previous empirical studies.

This chapter empirically examined the model with the regional panel data of the manufacturing sector in Japan. The estimation results of the revenue function show significant, robust, and positive transportation costs for the manufacturing products. Additionally, consistent GMM estimation results provide evidence of positive transportation costs, not only for the outputs of the manufacturing sector but also for those of the primary and service sectors. These findings indicate that the efficiency of the manufacturing firms depends on their access to the markets and the intermediate inputs supply. Because the proposed model can simulate the value added given a set of geographic distribution of the firms, the empirical results can be used to evaluate the optimality of the actual industrial location and the effects of some policy intervention on the optimality.

However, several issues remain for future research. First, to investigate effective regional and location policies, empirical analyses of the dynamics of the location of firms and labor supply are additionally needed. From the results presented in this chapter, only comparative statics can be performed. From the dynamic perspective, other important aspects, such as relocation cost, entry cost, or time lags, should be taken into account. They can be analyzed only in dynamic models of the location choice. Second, this chapter ignores the export and import activities and roles of trade hubs, for example, harbors, airports, or train stations. Since the distance from such trade hubs should affect the transportation costs, as well as the efficiency of firms, it should be necessary to control for such effects. Third, this chapter excludes the spillover effects of knowledge or of R&D investments on productivity. Knowledge spillover effects may be correlated with the demand and supply access, which are examined in this chapter. Thus, it should be necessary to control for the effects of knowledge agglomeration.

Appendix

Appendix 2.1 Industries (2 digit Japan Standard Industry Classification)

Manufacturing sector	F09	Food
	F10	Beverages, Tobacco and Feed
	F11	Textile Mill Products
	F12	Apparel and Other Finished Products Made From Fabrics and Similar Materials
	F13	Lumber and Wood Products, Except Furniture
	F14	Furniture and Fixtures
	F15	Pulp, Paper and Paper Products
	F16	Printing and Allied Industries
	F17	Chemical and Allied Products
	F18	Petroleum and Coal Products
	F19	Plastic Products, Except Otherwise Classified
	F20	Rubber Products
	F21	Leather Tanning, Leather Products and Fur Skins
	F22	Ceramic, Stone and Clay Products
	F23	Iron and Steel
	F24	Non-Ferrous Metals and Products
	F25	Fabricated Metal Products
	F26	General Machinery
	F27	Electrical Machinery, Equipment and Supplies
	F30	Transportation Equipment
	F31	Precision Instruments and Machinery
	F32	Miscellaneous Manufacturing Industries
Primary sector	A	Agriculture
	B	Forestry
	C	Fisheries
	D	Mining
Service sector	E	Construction
	G	Electricity, Gas, Heat Supply and Water
	H	Information and Communications
	I	Transport
	J	Wholesale and Retail Trade
	K	Finance and Insurance
	L	Real Estate
	M	Eating and Drinking Places, Accommodations
	N	Medical, Health Care and Welfare
	O	Education, Learning Support
	Q	Services, N.E.C.

Appendix 2.2 Total number of establishments (1 thousand)

Industry	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
F09	27.6	26.7	28.4	26.8	26.5	25.3	24.1	24.6	24.1	27.5	27.5
F10	2.6	2.5	2.7	2.5	2.6	2.4	2.3	2.4	2.4	3.1	3.2
F11	10.1	9.5	9.1	8.2	7.6	6.9	6.1	6.1	6.0	6.7	6.4
F12	20.1	18.8	19.4	17.1	15.8	13.7	11.9	12.0	10.8	12.2	11.3
F13	8.4	7.9	7.8	7.2	6.9	6.2	5.6	5.7	5.5	6.8	6.5
F14	11.0	10.6	10.4	9.6	9.2	8.6	7.5	7.7	7.0	8.0	7.3
F15	8.2	7.8	8.3	7.7	7.6	7.1	6.6	6.7	6.5	7.0	6.8
F16	23.3	22.7	24.8	22.8	22.4	20.8	17.3	17.6	16.3	16.7	15.6
F17	3.7	3.7	3.9	3.7	3.7	3.6	3.5	3.5	3.6	3.8	3.9
F18	0.4	0.4	0.4	0.4	0.3	0.3	0.3	0.3	0.3	0.3	0.3
F19	13.5	13.2	14.2	13.3	13.4	12.7	12.1	12.7	12.5	14.2	13.9
F20	3.3	3.2	3.3	3.1	3.0	2.8	2.5	2.6	2.4	2.7	2.6
F21	3.4	3.2	3.1	2.9	2.6	2.5	2.1	2.1	1.8	1.9	1.7
F22	10.9	10.5	10.8	10.2	9.9	9.4	8.8	8.8	8.9	10.5	10.6
F23	4.2	4.1	4.2	3.8	3.8	3.6	3.4	3.5	3.3	3.7	3.7
F24	2.4	2.4	2.5	2.3	2.3	2.1	2.0	2.1	1.9	2.2	2.3
F25	36.1	35.1	36.6	33.3	33.5	31.3	29.0	30.3	28.7	32.1	30.8
F26	32.3	31.9	33.7	30.7	31.5	29.1	27.1	28.5	27.6	31.3	30.8
F27	21.0	20.4	21.1	19.5	19.4	17.4	15.7	15.9	15.4	17.3	17.0
F30	9.8	9.5	10.0	9.4	9.3	8.9	8.6	9.1	8.9	10.4	10.4
F31	4.2	4.1	4.4	4.0	4.0	3.7	3.4	3.5	3.3	3.6	3.4
F32	7.8	8.2	9.3	8.2	8.4	7.3	7.1	7.6	8.0	8.9	6.8

Appendix 2.3 Number of regions with one or more establishments

Industry	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
F09	752	750	753	750	752	754	756	763	798	847	876
F10	356	344	366	360	365	346	321	334	336	394	410
F11	321	312	315	301	301	292	276	278	272	294	307
F12	682	668	677	664	662	643	623	628	635	708	691
F13	583	569	570	555	551	521	497	508	498	562	567
F14	612	601	600	586	580	574	540	548	529	595	580
F15	500	496	505	492	488	484	472	470	481	521	520
F16	681	679	693	685	688	682	669	673	676	721	717
F17	378	377	390	384	384	381	376	376	386	407	422
F18	92	90	98	89	84	84	72	76	69	77	80
F19	585	587	598	589	590	590	580	588	609	664	677
F20	251	238	250	248	242	245	232	236	223	257	264
F21	172	162	169	158	144	137	119	117	96	117	112
F22	672	663	679	667	660	657	647	647	672	735	757
F23	342	334	349	332	330	315	302	315	304	354	366
F24	282	276	284	268	263	264	255	254	249	286	295
F25	717	715	716	713	714	711	706	719	737	799	815
F26	672	678	679	676	679	673	665	679	701	764	781
F27	665	661	665	662	664	655	600	601	625	682	693
F30	498	491	494	490	491	493	474	488	508	561	586
F31	336	334	348	332	325	328	310	306	314	345	350
F32	500	516	548	533	529	520	532	542	565	621	564

Appendix 2.4 Basic statistics for the manufacturing industries (city-level)

Industry		N	ln*V*				ln*L*				ln*K*				ln*M*			
			Mean	S.D.	Min.	Max.	Mean	S.D.	Min.	Max.	Mean	S.D.	Min.	Max.	Mean	S.D.	Min.	Max.
F09	Food	8,551	-0.799	1.006	-4.341	2.915	3.303	0.667	1.651	6.062	-2.328	1.385	-7.772	1.548	-1.424	1.086	-5.590	1.988
F10	Beverages, Tobacco and Feed	3,932	-0.451	1.647	-4.669	4.222	2.956	0.738	1.386	5.490	-2.621	1.806	-5.542	3.179	-1.416	1.720	-6.220	2.587
F11	Textile Mill Products	3,269	-1.634	1.119	-4.881	1.711	2.764	0.714	0.956	5.507	-3.642	1.858	-8.740	2.081	-2.460	1.310	-7.114	1.370
F12	Apparel	7,281	-2.249	0.785	-4.651	1.429	2.707	0.561	1.335	5.729	-4.557	1.623	-9.028	0.472	-3.183	1.072	-7.677	0.540
F13	Lumber and Wood Products	5,981	-1.741	0.896	-4.395	3.067	2.472	0.536	1.386	5.710	-4.053	1.560	-7.100	1.191	-2.373	1.029	-6.444	2.721
F14	Furniture and Fixtures	6,345	-2.041	0.965	-4.359	1.989	2.354	0.628	0.875	6.644	-4.151	1.246	-7.217	1.550	-2.809	1.071	-6.635	1.806
F15	Pulp, Paper and Paper Products	5,429	-0.807	1.188	-4.227	3.268	3.089	0.691	1.466	6.056	-2.416	1.736	-5.467	3.302	-1.445	1.297	-6.167	2.619
F16	Printing and Allied Industries	7,564	-1.518	0.967	-4.492	2.546	2.777	0.564	1.386	5.492	-3.371	1.627	-6.896	1.865	-2.446	1.096	-6.571	1.881
F17	Chemical and Allied Products	4,261	0.616	1.274	-5.046	4.653	3.870	0.821	1.386	7.376	-0.851	1.741	-5.054	3.261	-0.178	1.316	-8.230	3.721
F18	Petroleum and Coal Products	911	0.149	1.897	-3.036	6.295	2.822	0.993	1.540	6.047	-2.200	1.871	-4.241	4.661	-0.366	1.959	-4.413	6.099
F19	Plastic Products	6,657	-0.937	0.988	-4.315	2.789	3.081	0.632	0.916	5.674	-3.114	2.569	-10.260	2.026	-1.630	1.092	-5.901	2.323
F20	Rubber Products	2,686	-1.059	1.334	-3.979	3.380	3.182	0.949	1.466	6.218	-3.963	2.284	-6.741	2.023	-1.878	1.407	-6.041	2.408
F21	Leather Products	1,503	-2.060	0.963	-4.867	1.610	2.445	0.527	1.386	5.165	-4.810	1.747	-8.517	-0.212	-2.835	1.274	-7.706	0.651
F22	Ceramic, Stone and Clay Products	7,456	-0.892	0.833	-4.481	2.646	2.933	0.593	1.281	6.731	-2.568	1.851	-8.405	1.884	-1.665	0.879	-5.828	1.670
F23	Iron and Steel	3,643	0.002	1.313	-3.752	4.725	3.387	0.888	1.466	6.825	-1.401	1.694	-4.294	4.354	-0.572	1.433	-5.648	4.135
F24	Non-Ferrous Metals and Products	2,976	-0.358	1.549	-4.862	4.249	3.322	0.974	1.386	7.113	-2.338	1.878	-5.391	2.705	-0.987	1.754	-6.510	4.138
F25	Fabricated Metal Products	8,062	-1.294	0.827	-4.177	2.899	2.779	0.520	1.335	5.685	-2.700	1.086	-6.757	0.940	-2.044	0.951	-5.982	1.683
F26	General Machinery	7,647	-0.673	1.021	-3.835	3.682	3.195	0.669	1.253	6.343	-2.224	1.322	-5.876	2.093	-1.403	1.182	-6.027	3.657
F27	Electrical Machinery	7,173	-0.063	1.352	-4.423	4.326	3.830	0.863	0.619	7.435	-1.986	1.990	-7.145	3.359	-0.707	1.495	-7.277	4.034
F30	Transportation Equipment	5,574	-0.281	1.530	-5.180	5.348	3.540	0.996	1.386	7.751	-2.580	2.661	-7.799	3.172	-0.949	1.745	-8.092	5.015
F31	Precision Instruments and Machinery	3,628	-0.967	1.198	-4.373	4.063	3.172	0.824	1.466	6.890	-3.482	1.959	-8.354	1.790	-1.829	1.366	-6.645	3.571
F32	Miscellaneous Manufacturing Industries	5,970	-1.813	1.177	-5.282	4.332	2.531	0.681	-0.440	6.450	-4.254	1.636	-8.262	1.607	-2.678	1.339	-6.567	4.309

Appendix 2.5 Basic statistics for the all industries (prefecture-level, 1 billion yen)

Industry		n	*E*				*V*			
			Mean	S.D.	Min.	Max.	Mean	S.D.	Min.	Max.
F09	Food Products	517	871.6	915.5	185.9	4997.9	4.7	4.1	0.9	24.9
F10	Beverages, Tobacco and Feed	517	388.7	409.7	83.4	2286.2	13.0	9.0	2.5	40.4
F11	Textile Mill Products	517	65.1	60.6	9.0	425.2	1.0	0.6	0.1	3.7
F12	Apparel	517	256.2	293.2	42.6	2047.9	1.1	0.9	0.2	7.2
F13	Lumber and Wood Products	517	115.6	103.2	22.2	646.1	1.5	1.5	0.3	14.0
F14	Furniture and Fixtures	517	81.0	90.9	17.0	631.7	0.8	0.6	0.2	3.0
F15	Pulp, Paper and Paper Products	517	215.0	262.6	40.9	1665.5	4.8	3.2	0.8	14.5
F16	Printing and Allied Industries	517	235.3	347.4	22.9	2467.0	1.6	1.2	0.3	6.5
F17	Chemical and Allied Products	517	689.4	736.6	106.8	4394.5	15.2	11.2	0.3	42.3
F18	Petroleum and Coal Products	517	425.4	497.2	64.3	3390.0	30.9	35.2	2.6	197.3
F19	Plastic Products	517	259.1	275.9	37.8	1422.6	4.0	2.1	1.0	14.2
F20	Rubber Products	517	77.1	89.6	10.9	420.5	3.0	2.2	0.0	10.3
F21	Leather Products	517	43.7	51.8	8.2	340.3	0.5	0.3	0.0	2.0
F22	Ceramic, Stone and Clay Products	517	227.2	210.2	42.5	1273.2	3.5	2.6	0.9	15.3
F23	Iron and Steel	517	462.0	563.6	36.0	3483.6	12.1	10.3	1.3	53.4
F24	Non-Ferrous Metals and Products	517	189.6	205.6	14.4	973.1	6.9	6.4	0.1	41.3
F25	Fabricated Metal Products	517	332.3	314.5	66.9	1783.4	1.8	1.0	0.5	4.8
F26	General Machinery	517	226.7	260.4	26.6	1278.5	4.3	2.3	0.4	12.3
F27	Electrical Machinery	517	760.0	794.2	96.1	4244.5	12.3	5.8	1.1	30.8
F30	Transportation Equipment	517	707.5	1033.5	58.1	7389.5	12.8	10.2	0.6	59.0
F31	Precision Instruments and Machinery	517	65.0	70.9	11.2	397.8	2.5	1.9	0.1	8.5
F32	Miscellaneous Manufacturing Industries	517	169.0	190.2	28.0	1100.7	1.0	0.6	0.1	2.8
A	Agriculture	517	344.6	296.4	79.4	1472.3	7.7	5.5	1.9	42.4
B	Forestry	517	29.1	23.1	3.4	199.7	3.4	4.0	0.2	28.7
C	Fisheries	517	65.9	60.0	15.0	339.1	1.9	1.8	0.1	13.0
D	Mining	517	289.3	369.2	31.3	2890.7	2.0	1.7	0.3	15.5
E	Construction	517	176.7	242.4	26.4	1629.5	1.2	1.2	0.4	10.2
G	Electricity, Gas, Heat Supply and Water	517	474.1	534.7	88.7	3168.4	32.1	30.9	6.3	152.6
H	Information and Communications	517	544.8	855.4	58.0	7731.3	1.4	1.8	0.1	14.3
I	Transport	517	235.6	260.5	50.2	1484.9	3.0	2.1	0.9	12.5
J	Wholesale and Retail Trade	517	23.8	29.1	2.5	203.2	0.4	0.3	0.1	1.9
K	Finance and Insurance	517	777.9	977.8	146.5	6561.8	2.6	1.8	0.8	9.9
L	Real Estate	517	1290.3	1463.9	270.9	8673.8	3.2	2.8	0.5	13.8
M	Restaurants	517	406.6	449.3	81.6	2423.4	0.2	0.2	0.1	1.0
N	Medical, Health Care and Welfare	517	231.6	255.4	43.1	1576.7	1.6	1.0	0.4	7.3
O	Education, Learning Support	517	145.3	164.4	30.3	1009.4	0.5	0.4	0.1	2.1
Q	Services, N.E.C.	517	2021.9	2458.4	371.3	15921.0	0.6	0.5	0.2	3.2

Chapter 3 Plant Productivity Dynamics and Private and Public R&D Spillovers: Technological, Geographic and Relational Proximity[14]

1. Introduction

It is well established in the literature that the productivity effects of R&D spillovers are enhanced by technological proximity and geographic proximity (Jaffe et al., 1993; Adams and Jaffe, 1996; Aldieri and Cincera, 2009; Lychagin et al., 2010; Bloom et al., 2013; Orlando, 2004; Griffith et al., 2009; Mairesse and Mulkay, 2008). Despite the increasing number of large-scale firm-level studies on R&D spillovers,[15] existing studies have a number of limitations in scope and methodology. First, they typically relied on data on publicly listed firms, aggregating over the various locations and technologies in which firms are active.[16] Second, the focus has been on inter-firm private spillovers while abstracting from the role of public research. A different research stream focusing on the role of knowledge spillovers from public research conducted at universities and research institutes has however suggested the importance of such spillovers, with an explicit role of proximity (e.g. Jaffe, 1989; Adams, 1990; Anselin et al., 1997; Furman et al., 2005). Third, R&D spillovers at the firm level have in most cases been modelled as knowledge spillovers as a function of proximity between technology portfolios of the firm, while the role of spillovers through supplier and customer linkages has only received limited attention.[17] A separate literature on the role of spillovers in the context of foreign direct

[14] This chapter is based on Ikeuchi et al. (2013) and Belderbos et al. (2013), both co-authored with Rene Belderbos, Kyoji Fukao, YoungGak Kim, and HyeogUg Kwon. It is the result of the joint research project of the National Institute of Science and Technology Policy (NISTEP), the Research Institute for Economy, Trade and Industry (RIETI), and Hitotsubashi University, under the "Science for Science, Technology and Innovation Policy" program of the Ministry of Education, Culture, Sports, Science and Technology in Japan.

[15] Early work examined R&D spillovers at the industry level (e.g. Mohnen and Lepine, 1991; Audretsch and Feldman, 1996; Goto and Suzuki, 1989).

[16] Adams and Jaffe (1996) do analyse plant level productivity but focuses on the effects of internal R&D. The analysis of Griffith et al. (2009) for UK plants focuses on proximity effects but does not incorporate the role of R&D.

[17] An exception is Crespi et al. (2007), who examine data from UK Community Innovation Surveys for direct (self

investments has strongly suggested that 'vertical' spillovers through buyer-supplier relationships often is the key channel through with spillovers occur (e.g. Haskel et al., 2007; Görg and Strobl, 2001; Javorcik, 2004; Kugler, 2006). While knowledge and technology transfer in these relationships is often purposeful and embedded in intermediates, their value tends not to be fully reflected in the price of such intermediates, leading to ‘pecuniary spillovers (Hall et al., 2012; Crespi et al, 2007). Compared with ‘horizontal’ spillovers in technological proximity within narrowly defined industries, the absence of market rivalry provides greater incentives for productivity and growth enhancing knowledge exchange and spillovers (e.g. Bloom et al., 2013). Since suppliers and clients may be active in a variety of industries, these 'relational' spillovers are yet a different dimension of heterogeneity in spillover pools.

This chapter addresses these limitations in prior works. The analyses contribute an analysis of the various sources of R&D spillovers, which until now have not been considered simultaneously, and examine these relationships at the plant level. The chapter analyzes the effects of technologically, geographically, and relationally proximate private R&D stocks, as well as of technologically and geographically proximate public R&D stocks on TFP in an unbalanced panel of close to 20000 Japanese manufacturing plants, 1987-2007. The plant level data from the Census of Manufacturers are matched with information on R&D expenditures from the comprehensive Survey of R&D Activities in Japan covering virtually all R&D spending firms (and public research institutions). The R&D survey data, which are decomposed by field or industry of application, allow us to construct relevant R&D stocks weighted by technological proximity (e.g. Bloom et al., 2013), while the information on plant locations allows us to explore the role of geographic distance between firms and between firms and public research institutions in much more detail than in previous studies. Relationally proximate R&D stocks are calculated using input-output tables. Public R&D stocks are differentiated by science field, which can be mapped into technologies and industries reflecting their varying relevance for firms. This chapter estimates long (five year) difference models of plant TFP growth to reduce the influence of measurement errors and cyclical effects (e.g. Haskel et al., 2007; Branstetter, 2000). Also, gradual convergence in TFP is allowed for by estimating dynamic TFP growth models (e.g. Klette, 1996; Klette and Johanson, 1998; Lokshin et al., 2008), and distance effects are identified by estimating exponential decay parameters (e.g. Lychagin et al.,

assessed) evidence of incoming knowledge flows at the firm level. They find, among others, that supplier information positively affects TFP growth, but do not examine geographic or technological proximity.

2010; Duranton and Overman, 2005)). The simultaneous inclusion of multiple sources of spillovers, the detail on location and field of R&D, the long panel, and the uniquely large set of plants should allow more precise estimates of spillover effects and an assessment of their relative importance over time. This study contributes to the very limited literature on R&D and spillovers at the plant level.

This research is also motivated by the observation that Japan's total factor productivity growth has been declining since the mid-1980s (e.g. Fukao and Kwon, 2011), while at the same time R&D expenditures as a percentage of GDP have been steadily increasing to reach 3.8% in 2008, from 2.5% in 1980s. The discrepancy between the trends in R&D expenditures and TFP suggests that the aggregate returns to R&D have been falling. One possible explanation for this phenomenon may be a decline in R&D spillovers due to the exit (and potential relocation abroad) of sophisticated manufacturing plants of R&D intensive firms and the accompanied changing patterns of R&D agglomeration, which may have reduced the size and effectiveness of the relevant pool of R&D spillovers across firms. Prior studies suggest that exit rates of relatively productive plants operated by multi-plant (multinational) firms have been typically higher than the exit rates of single establishments (e.g. Fukao and Kwon, 2006; Kneller et al. 2012).

The remainder of the chapter is organized as follows. The next section describes the model, the particularities of the data and the empirical strategy followed. Section 3 presents the empirical results and section 4 concludes and discusses avenues for future research.

2. Model Setup and Data

This chapter conducts a plant-level panel analysis of total factor productivity, in which plant-level TFP will be related to firms' own R&D stock, private R&D stocks (the private spillover pool), public R&D stocks, and a set of plant-, firm- and industry-level controls. It is assumed that firm level R&D stocks are available to all the firms' plants and that R&D spillover occur between plants due to the R&D stock the plants have access to. This allows us to investigate the geographic dimension of R&D spillover in detail, taking into account the population of R&D conducting firms and the spatial and industry configuration of their plants.

This chapter adopts the standard knowledge stock augmented production function framework (e.g. Hall et al, 2012). The production function is defined at the plant-level generally as:

$$Q_{it} = f(L_{it}, K_{it}, M_{it})g(R_{it-1}, S_{it-1}, P_{it-1}, \mathrm{X}_{it})U_{it} \tag{3-1}$$

where: Q_{it}: Gross output of the plant; L_{it}, K_{it}, M_{it}: Inputs of plant i in year t; R_{it-1}: Firm-level R&D stock; S_{it-1}: Private R&D stock; P_{it-1}: Public R&D stock; X_{it}: a vector of other observable factors (control variables) affecting plant productivity; U_{it}: plant-year specific unobserved efficiency. Total factor productivity (TFP) is defined as:

$$TFP_{it} \equiv \frac{Q_{it}}{f(L_{it}, K_{it}, M_{it})} = g(R_{it-1}, S_{it-1}, P_{it-1}, \mathrm{X}_{it})U_{it} \tag{3-2}$$

R&D stocks are assumed to influence production with a one-year lag to reflect that the application of new knowledge and insights due to R&D takes time. If we adopt a log-linear specification for $g(R_{it-1}, S_{it-1}, P_{it-1})$ and allow $U_{it} = e^{\eta_i + u_{it}}$, where η_i is a plant specific fixed effect and u_{it} is a plant-year specific efficiency shock, we obtain:

$$\ln TFP_{it} = \alpha_R \ln R_{it-1} + \alpha_S \ln S_{it-1} + \alpha_P \ln P_{it-1} + \gamma' \mathrm{X}_{it} + \eta_i + u_{it} \tag{3-3}$$

and if we take a difference of the equation between two periods:

$$\Delta \ln TFP_{it} = \alpha_R \Delta \ln R_{it-1} + \alpha_S \Delta \ln S_{it-1} + \alpha_P \ln \Delta P_{it-1} + \gamma' \Delta \mathrm{X}_{it} + \Delta u_{it} \tag{3-4}$$

where the plant-specific efficiency parameter drops out. It is assumed that the change in plant-specific efficiency levels (Δu_{it}) is a function of past productivity relative to the industry mean, in order to allow for a gradual convergence in efficiency levels between firms (e.g. Lokshin et al., 2008). Klette (1996) and Griffith et al. (2009) have shown that the empirically observed persistent productivity differences between plants or firms require a model specification that allows for gradual convergence.[18] Specifically, this chapter models:

$$\varDelta u_{it} = u_{it} - u_{it-1} = \rho \ln TFP_{it*} + e_{it} \tag{3-5}$$

where $\ln TFP_{it*}$ is the level of TFP of plant i relative to the industry mean in the previous period. ρ is expected to fall within the interval [-1,0]. If ρ is zero there is no gradual convergence between leading firms and lagging firms; if ρ is –1 complete convergence materializes in one period. We assume that the error term e_{it} can be decomposed into four components, year-specific effects λ_t, industry-year specific technological opportunity or efficiency shocks μ_{st}(with s denoting industry), regional shocks ρ_r and measurement error ε_{it}:

$$e_{it} = \mu_{st} + \lambda_t + \rho_r + \varepsilon_{it} \tag{3-6}$$

18 Kneller et al. (2012) show that productivity catch up is an important phenomenon among Japanese manufacturing plants as well

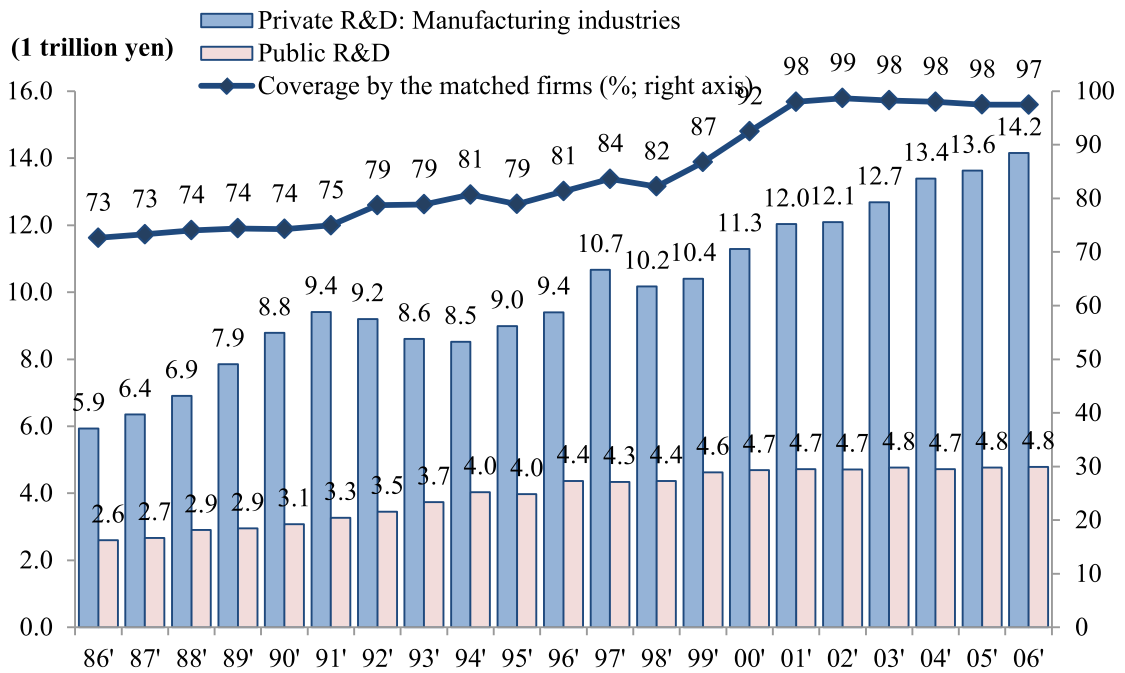
2.1. Data Sources and Sample

Plant level data from the Japanese *Census of Manufacturers* with information is matched on R&D expenditures from the yearly (comprehensive) *Survey of R&D Activities* in Japan, 1987-2007. The census has a comprehensive coverage of manufacturing plants with more than 4 employees. From 2001 onwards, information on plant level fixed capital investment has not been surveyed for plants with less than 30 employees, with the exception of the benchmark surveys organized every 5 years. The number of plants for which panel data on TFP can be calculated is roughly 40,000 yearly.

The Survey of R&D activities in Japan (hereafter R&D survey) is a comprehensive and mandatory survey of R&D performing firms and public research institutes and universities in Japan. It contains information on R&D expenditures, differentiated by field, for roughly 9,000 firms yearly and has a response rate greater than 90 percent. Large firms (with more than 1 billion Yen of capital) are always included in the survey; smaller firms are included in higher sampling rates if they are identified as R&D conducting firms in the previous survey. The information on R&D by field (30 fields are distinguished) is easily mapped into industries, and allows us to distinguish R&D expenditures relevant to 20 manufacturing industries. The response rate by research institutes and universities is close to 100 percent.

Each plant in the Census of Manufacturers is matched to a firm in the R&D survey for each year by using the information on firm name, address of headquarter, and stated capital. The matching between the surveys, however, is posed a number of challenges. Firm names are only recorded in the R&D survey from 2001 onwards and parent firm names are only provided on the plant records in the census from 1994 onwards. Firm identifiers in the R&D survey are not compatible between the years before and after 2001 because the identifiers for all firms were revised in 2001; only the R&D survey in 2001 includes both the old and new versions of firm identifiers. Because of the absence of common firm identifiers in the surveys, matching had to be done semi-manually (by firm name, address and capitalization). From 2001 onwards, more than 97.5 percent of reported R&D expenditures could be matched to firms and plants included in the census (Figure 3.1). The situation is more complicated for the years 1983-2000, for which R&D could be only matched to plants 1) that could be linked to the parent firm in 1994 or one of the later years, and 2) that belong to firms identified in the R&D survey of 2001. This caused the coverage rate to decline from 98 percent in 2001 to 92.5 percent in 2000, declining progressively further to 73 percent in 1983.

Figure 3.1 R&D expenditures and matching rate with census of manufacturers



The matching issues cause several problems. First, there is a difficulty ascertaining whether a plant belongs to a parent firm conducting R&D or not. Here all unmatched firms are excluded from the sample to avoid measurement error in R&D stocks at the firm level. Second, for some firms R&D series are incomplete. R&D stocks are calculated on the basis of the information available only if there was sufficient information to derive an R&D growth rate for a specific period. Firms that are included in the R&D survey multiple times reporting absence of R&D activities are included in the sample with zero R&D stock. Third, we require reliable estimates of private R&D spillover pools. Here estimates that are as accurate as possible can be obtained by 1) using the weights provided in the R&D survey to correct for non-response and arrive at an estimate of total R&D expenditures in Japan; 2) allocating the R&D (stocks) to locations and fields/industries for R&D conducting firms that could not be matched to the manufacturing census (and hence for which no geographic information on plants is available) on the basis of the location of the firm, rather than on the basis of the location of plants. The second correction may be a reasonable approximation as most of the unmatched firms are smaller enterprises for which the plant and administrative unit are collocated.

Using the above matching rules, an unbalanced panel of over 19000 plants is obtained, observed for a maximum of 20 years and a minimum of 5 years, during 1987-2007.[19] The five year minimum observation period is due to the fact that (five-year) long difference models will be estimated. About 57 percent of the plant observations, plants are owned by parent firms for which the absence of formal R&D could be confirmed. Zero R&D cases are not compatible with the specification in natural logarithms in (4) but provide important variation in the sample. The chapter deals with this in two ways: 1) a dummy for continuous engagement in, or absence of, R&D is included; 2) the value of one is added to the R&D stock before taking the logarithm, such that the continuous absence of R&D as zero growth is treated.

Table 3.1 Sample characteristics

Industries (R&D fields)	# of obs.		# of (unique) plants in sample		# of (unique) plants in Japan (%)	# of (unique) parent firms	Avg. # of plants per firm	Avg. parent R&D stock per plant (billion yen)	% of plants with positive parent R&D
	#	(%)	#	(%)					
Food products	5,048	(10.8)	1,961	(10.1)	(12.7)	1,032	1.9	7.3	42.8
Textile mill products	1,741	(3.7)	641	(3.3)	(10.5)	432	1.5	7.3	37.4
Pulp and paper products	1,838	(3.9)	660	(3.4)	(3.2)	365	1.8	2.6	32.6
Printing	1,270	(2.7)	489	(2.5)	(5.6)	332	1.5	4.1	15.7
Chemical fertilizers and industrial chemicals	2,049	(4.4)	786	(4.1)	(0.8)	519	1.5	17.6	61.0
Drugs and medicine	1,154	(2.5)	490	(2.5)	(0.5)	398	1.2	22.2	47.6
Miscellaneous chemicals	2,135	(4.6)	913	(4.7)	(1.1)	655	1.4	11.9	53.3
Petroleum and coal products	511	(1.1)	225	(1.2)	(0.3)	113	2.0	7.6	58.5
Rubber products	1,072	(2.3)	426	(2.2)	(1.4)	295	1.4	13.4	37.2
Ceramic, stone and clay products	2,969	(6.3)	1,187	(6.1)	(5.5)	669	1.8	5.7	41.4
Iron and steel	1,744	(3.7)	642	(3.3)	(2.6)	425	1.5	16.6	37.7
Non-ferrous metals and products	1,331	(2.8)	513	(2.6)	(1.7)	371	1.4	11.2	39.5
Fabricated metal products	4,196	(8.9)	1,818	(9.4)	(14.0)	1,271	1.4	3.8	31.3
General-purpose machinery	6,925	(14.8)	2,951	(15.2)	(14.1)	2,284	1.3	15.8	33.1
Home electronics	444	(0.9)	225	(1.2)	(1.9)	185	1.2	83.1	32.9
Electrical machinery	3,455	(7.4)	1,508	(7.8)	(6.8)	1,101	1.4	26.3	36.6
Info. .and com. electronics	3,585	(7.6)	1,714	(8.8)	(7.7)	1,247	1.4	56.9	31.5
Motor vehicles, parts and accessories	3,285	(7.0)	1,304	(6.7)	(5.1)	756	1.7	58.4	43.1
Other transportation equipment	724	(1.5)	289	(1.5)	(1.7)	235	1.2	36.5	39.5
Precision instruments and machinery	1,447	(3.1)	647	(3.3)	(2.7)	503	1.3	6.0	28.3
Total	46,923	(100.0)	19,389	(100.0)	(100.0)	13,188	1.5	19.4	38.2

Table 3.1 shows the distribution of plants over industries and compares this with the distribution of the population of Japanese manufacturing plants over industries. Plants in

[19] We take out the plant from the sample of year *t* if its parent firm is acquired by other firm in year *t* but that plant remains in the sample of other years.

technology intensive industries such as drugs and medicine and chemicals are overrepresented in the sample, but the difference with the distribution of all plants over industries is not generally pronounced. The 19389 unique plants are operated by 13188 firms, implying that on average there are 1.5 plant observations per firm in the sample. Parent firm R&D stocks are highest in the home electronics and information and telecommunication sectors, and lowest in pulp and paper and printing.

It is necessary to note that creating a sample of plants for which parent firms' R&D stocks can be calculated leads to various sample selection issues, with a natural oversampling of R&D conducting firms (although the majority of plants in the sample have no access to internal R&D), larger plants (post-2001), surviving plants (1987-1994), and surviving firms (1987-2001). Several sensitivity analyses will be conducted to examine potential selection bias.

2.2. Variables and Measurement

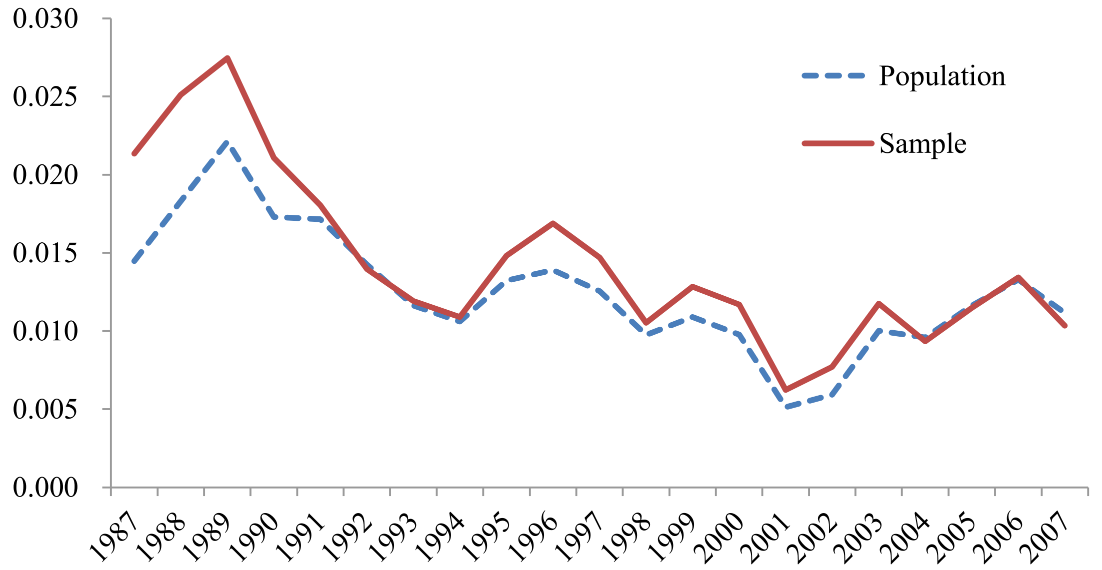
This chapter utilizes plant level TFP data from the Japan Industrial Productivity Database (JIP) 2010 (Fukao et al., 2008). TFP is measured using the index number method, following Good et al (1997):

$$\ln TFP_{fsit} = \left(\ln Q_{fsit} - \overline{\ln Q}_{st}\right) - \sum_{X=L,C,M} \frac{1}{2}\left(s^X_{fsit} + \bar{s}^X_{st}\right)\left(\ln X_{fsit} - \overline{\ln X}_{st}\right) + \sum_{j=1}^{t}\left(\overline{\ln Q}_j - \overline{\ln Q}_{j-1}\right) - \sum_{j=1}^{t}\sum_{X=L,C,M}\frac{1}{2}\left(\bar{s}^X_{fsij} + \bar{s}^X_{sj-1}\right)\left(\overline{\ln X}_s - \overline{\ln X}_{s-1}\right) \quad (3\text{-}7)$$

where $Q_{fsi,t}$ is the gross output of plant i of firm f in industry s in year t, $s^X_{,fsi,t}$ is the cost share of input X, and $X_{fsi,t}$ is the amount inputs of the plant. Three inputs, labor (L), capital (C), and intermediate input (M), are taken into account. Variables with upper bars denote the arithmetic mean of each variable over all plants in that industry s in year t. The JIP database provides index linked TFP estimates distinguishing 58 industries. The TFP indices express the plants' TFP as an index of the TFP level of a hypothetical representative plant in the industry (with an index of 1). One of the main advantages of the index number method is that it allows for heterogeneity in the production technology of individual firms, while other methods controlling for the endogeneity of inputs (e.g. Olley and Pakes, 1996; Levinsohn and Petrin, 2003) assume an identical production technology among firms within an industry (Van Biesebroeck, 2007; Aw et al., 2001).

Drawing on the JIP database, the five-year growth rate in TFP is calculated for the matched sample. The observations with the largest (top 1 percent) and lowest (bottom 1 percent) TFP growth are dropped to avoid a potentially strong influence of outliers. Figure 3.2 shows the 5-year moving average of the gross output weighted average TFP growth rate for the sample. The figure confirms that the rate of TFP growth has been decreasing over time, while there is a modest recovery in growth rates after 1999. The pattern of TFP growth in the sample closely follows the pattern of TFP growth in the population of Japanese plants.

Figure 3.2 Trends in TFP growth: sample plants and population of Japanese plants



2.3. R&D Stocks by Industry and Location

R&D stocks measured at the parent firm level can be separated by industry/field of application to arrive at R&D stocks of the firm per industry. This chapter utilizes a question in the R&D survey asking firms to allocate R&D expenditures by field, which easily maps into 20 industries. R&D stock of firm f in industry/field s is defined by:

$$K_{fst} = I_{fst} + (1 - \delta_s)K_{fst-1} \tag{3-8}$$

where I_{fst} is R&D investment of firm f for activities in industry s in year t and δ is a depreciation rate of the R&D stock. Industry-specific depreciation rates are used to reflect differences in the speed of obsolescence and technology life cycles. Industry specific depreciation rates are based on Japanese official surveys of "life-span" of technology conducted in 1986 and 2009 among R&D conducting firms[20] and vary between 8 (food industry) and 25 percent (precision instruments). To calculate initial R&D stocks (Hall and Oriani, 2006), industry-specific growth rates are similarly used, which are calculated from the R&D survey as average R&D growth rates per field in the 1980s**.** R&D investments are deflated using a deflator for private R&D from the JIP database, calculated from the price indices of the input factors for R&D expenditures for each industry; the deflator for public R&D is obtained from the White Paper on Science and Technology.

Firms can obtain technological knowledge not only from its own R&D investment but also from the other firm through mergers and acquisitions (M&A). However, since we do not use the information on M&A, we could not account the technological knowledge obtained through M&A into our R&D stock estimation. This implies that R&D stock of the firm disappeared through M&A is not included in R&D stock of any surviving firms.

Matching the field of firms' R&D with the industry of the firms' plants, R&D stocks across industries and space can be calculated, where it is assumed that the R&D stock in a field/industry is available to each same-industry plant of the firm. R&D stocks are mapped in geographic space by using the information on the location of the plant, where more than 1800 cities, wards, towns, and villages are distinguished.

2.4. Plant R&D Stocks

Plant R&D stocks are calculated as the R&D stock of the parent, and we assume that all parent R&D provides relevant productivity improving inputs to the plants. This implies that the firm R&D may generate common knowledge for all plants within the firm. Given that R&D at the firm level is often organized to benefit from scope economies (e.g. Henderson and Cockburn, 1996; Argyres and Silverman, 2004) and involves active knowledge transfer to business units and plants, this may be a suitable assumption.[21]

---

[20] See "White paper on Science and Technology" (1986, Science and Technology Agency) and "Survey on Research Activities of Private Corporations" (2009, National Institute of Science and Technology Policy).

[21] A technological proximity weighted parent R&D stock is also calculated, applying the weighting scheme for industries/fields outside the industry of the plant based on the technological proximity matrix used for R&D spillovers, but obtained weaker effects. As the co-occurrence of different technologies in the R&D portfolios of firms is often taken as an indicator of the potential for scope economies (Bloom et al. 2013; Breschi et al. 2003) this is

2.5. Private R&D stocks (spillover pools)

Private R&D stocks (spillover pools) are derived from the calculated parent firms' R&D stocks, while it is allowed for geographic decay in the effectiveness of spillovers. Technologically proximate R&D stocks are calculated based on the technological proximity between the R&D field/industry of the plant and the industry of other plants. The technologically relevant private R&D stock (spillover pool) is defined as the sum total of other firms' R&D assigned to their (nearest) plants in an industry, weighted by the technological relatedness between the industry of the plants and the industry of the focal plant:

$$S_{ifst}^{tech} = \sum_{f' \neq f} \sum_{s'} K_{f's't} T_{ss'} e^{\tau d_{if's't}} \tag{3-10}$$

where: $d_{if's't}$: Minimum geographic distance between plant i and the plant of firm f' in the field s' in year t; $T_{ss'}$: the technological proximity weight; $e^{\tau d_{if's't}}$: Weight for geographic proximity of plant i to R&D stock firm f' for field s'; τ: a decay parameter, with $\tau < 0$.

If firms operate multiple plants, the R&D stock is only counted once using the plant with the minimum distance to the focal plant, which avoids double counting of R&D.[22] This chapter models an exponential decay function in the effectiveness of spillovers with parameter τ to be estimated, in line with recent studies (e.g. Lychagin et al. 2010). Distance d is the distance between a pair of locations and is measured as the geo-distance between the center of cities, wards, towns, and villages. In order to correct for differences in the geographic areas covered by the regions, distance is the radius of the region if plants are located in the same region.

The technological relatedness measure is derived from patent data and based on Leten et al. (2007). The relatedness between technologies will be reflected in the intensity with which technologies in a field build on prior art in a different field. Patent citation data are available at the 4-digit IPC level. The IPC codes can subsequently be mapped onto industries using the industry-technology concordance table developed by Schmoch et al. (2003) in which each technology field is uniquely linked to its corresponding NACE two-digit industry. Appendix A

---

perhaps not surprising.

[22] One may argue that having multiple plants in the vicinity increases the likelihood of knowledge spillovers. However, assuming that each plant of a firm shares the same knowledge of this firm, at least it is consistent to assume that the largest spillover from the firm comes from the nearest plant of this firm, although in this way we may overestimate the inter-firm spillover for that from multi-plant firms. Seeking the most appropriate specification of knowledge spillover between multi-plants firms is one of the future research agenda.

shows the resulting technological relatedness coefficients (weights) between industries used in the analyses, with weights for the own industry normalized at 1.

Relationally proximate R&D stocks are measured by the R&D stocks of supplier and customer industries, identifying the importance of supplier and customer transactions from Input-Output tables (yearly between 1987 and 2007) for 52 JIP industries. The calculation of R&D stocks follows (10) but with $T_{ss'}$ substituted by supplier industry proximity weights $SUP_{ss'}$ and customer industry proximity weights $CUS_{ss'}$, with:

$$SUP_{ss't} = \frac{Q^*{}_{s'st}}{\sum_j Q^*{}_{jst}} \tag{3-12}$$

$$CUS_{ss't} = \frac{Q_{ss't}}{EX_{st}+Q_{st}} \tag{3-13}$$

where $Q^*{}_{s'st}$ denotes domestic sales of industry s' to industry s and EX_{st} denotes exports of industry s. In equation (3-12), $Q^*{}_{s'st}$ is the estimated output of industry s' sold to industry *s*. Since domestic sales in the input-output tables include domestic sales of imported goods, $Q^*{}_{s'st}$ is estimated by applying the following correction to the domestic sales data: $Q^*{}_{s'st} = Q_{s'st} * (\sum_s Q_{s'st})/(\sum_s Q_{s'st} + I_{s't})$, with I_{st} imports of industry s. Hence it is assumed that the imported goods of the industry are sold to other industries in proportion to total sales to these industries. It is necessary to note that industries s include services and other industries' sales to industry s', such that the sum of input shares for industry s' does not add up to 1. Weights for customer R&D stocks for industry s are the shares of sales by industry *s* to industry s' in total sales, with the latter including sales to non-manufacturing industries and exports. Yearly input-output tables provided by the JIP database are used, such that weights are varying by year. Appendix B and C show the average the input and output share weights for the industries in the analysis for the year 1990. Using these weights, then we define relevant R&D spillover pools from suppliers and customers, respectively, as follows:

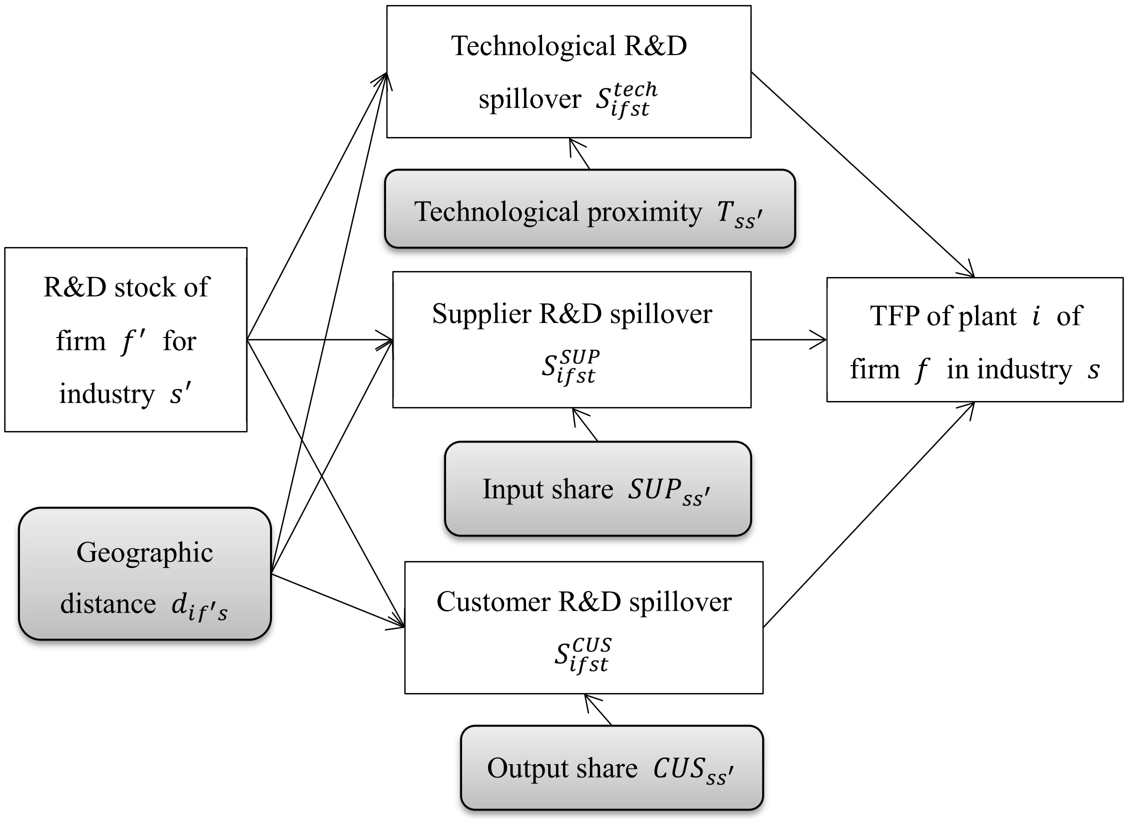
$$S_{ifst}^{SUP} = \sum_{f' \neq f} \sum_{s'} K_{f's't} SUP_{ss't} e^{\tau d_{if's't}} \tag{3-10b}$$

$$S_{ifst}^{CUS} = \sum_{f' \neq f} \sum_{s'} K_{f's't} CUS_{ss't} e^{\tau d_{if's't}} \tag{3-10c}$$

where $K_{f's't}$ is the R&D stock of firm f' for activities in industry s' at the beginning of year t.

Figure 3.3 illustrates our assumption on the effects of private R&D spillovers on TFP through the above three types of proximities: Technological proximity, supplier industry proximity and customer industry proximity. We assume that the R&D stock of a firm for activities of an industry affects TFP of the other firms' plant through three paths and define three spillover variables corresponding to each path. Since different weighting schemes are used to define each spillover variable, these effects can be empirically identified if these weighting schemes are sufficiently different each other.

Figure 3.3: Effects of Private R&D spillover on TFP



We include the parent plant's R&D, private R&D weighted by geographic distance and technological proximity, and private R&D weighted by geographic distance, supplier industry's share and customer industry's share as the explanatory variables of the TFP growth equation (3-2). Thus, we assume that the R&D investment of a firm affects not only the TFP of its own plants but also spills over and affects TFP of the other firms' plants which is proximate with respect to geographic, technological and/or relational dimensions. By estimating the parameters, we can evaluate the relative importance of these three proximity dimensions for R&D spillover effects.

2.6. Public R&D Stocks

Public R&D spillover pools derived from the R&D surveys have few measurement issues, as response rates are virtually 100 percent. This chapter differentiates public R&D by location based on the region (city, ward, town, village) of the research institute or university, and by industry/R&D field utilizing information on science fields with varying relevance for specific industries. The R&D stock of public research institution h in science field m is defined as:

$$A_{hmt} = E_{hmt} + (1 - \delta_A)A_{hmt-1} \quad (3\text{-}14)$$

where E_{hmt} is research expenditure of public research institution h in science field m in year t and δ_A is a depreciation rate of public R&D stock, which is set at 15 percent per year. Although the surveys do not include research expenditures by science field, they do contain information on the number of researchers by science field for each institution for each year. The public R&D expenditure E_{hmt} is estimated by multiplying total R&D expenditures with the share of the number of scientists in the field in the total number of scientists for each institution and year.

Second, a ‘relevant’ public R&D stock per industry/R&D field is estimated by using weights derived from a concordance matrix between science fields and industries. The weights are based on a study by Van Looy et al. (2004) examining citation frequencies on patent documents classified in different technology fields to Web of Science publications in each of the science fields. The concordance attaches to each scientific discipline probabilities that it is of relevance to each technology field (4-digit IPC fields). Applying this concordance to the public R&D expenditures per science field, the concordance matrix between IPC classes and industries due to Schmoch et al. (2003) is subsequently applied to arrive at public R&D stocks per industry. Appendix D shows the compound weights used to relate R&D stocks per science field to industries.

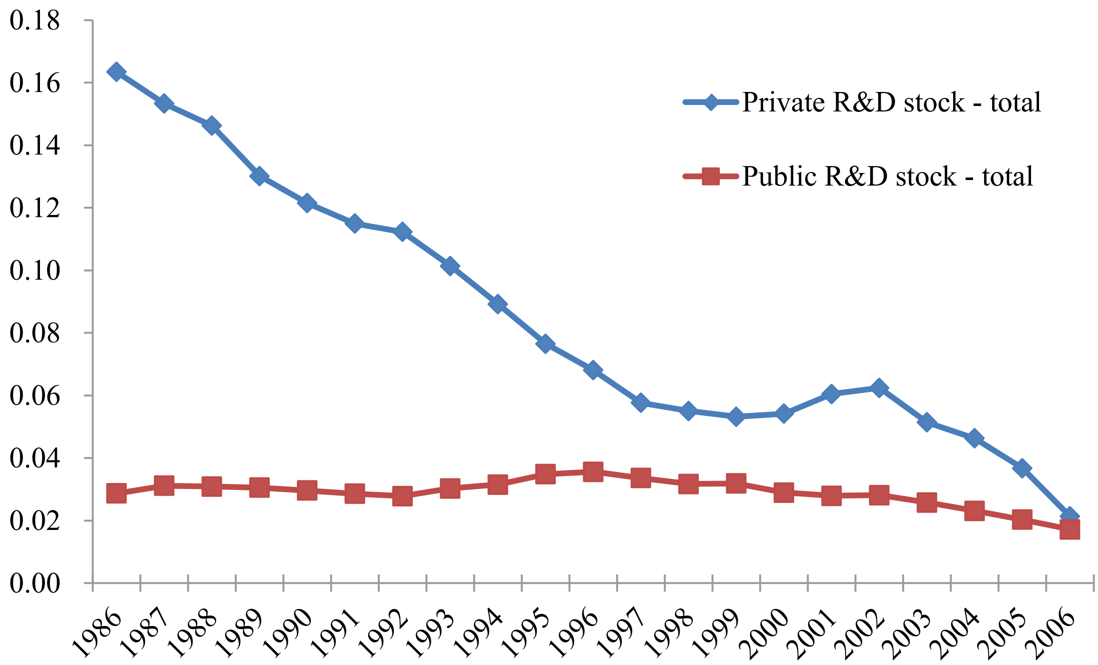
Using the above procedure, the technologically and geographically proximate public R&D stock is defined as:

$$P_{its} = \sum_h \sum_m A_{hmt} \tilde{T}_{sm} e^{\theta \tilde{d}_{ih}} \quad (3\text{-}15)$$

where: A_{hmt}: R&D stock of public institutes in location h for academic field m in year t; $\tilde{T}_{sm}$: The compound proximity weights between industry/R&D field s and science field m; $\tilde{d}_{ih}$: geographic distance between plant i and location h; θ: the geographic decay parameter, $\theta < 0$.

Figure 3.4 shows the 5-year moving average growth rates in the levels of public and private R&D stocks. Although growth rate of private R&D stock is higher than that of public R&D stock in all years between 1986 and 2006, this gap between them have been decreasing overtime. It indicates that private firms actively invest in R&D especially in1980's in comparison to university and public research institute. The recession in the early 1990's makes firms to reduce R&D investment and the growth rate of private R&D stock rapidly decreased, while the growth rate of public R&D stock gradually increased in that period. The growth in both public and private R&D shows a declining trend, as the increase in overall R&D investments (Figure 3.1) has slowed over time and had just exceeded deprecation rates in the most recent years.[23]

Figure 3.4 Growth rate in R&D stocks (5 year moving average)



2.7. Control Variables

The vector of time varying plant-specific characteristics X_{it} includes plant size

[23] This declining trend in private R&D stock in 1990's is consistent with the results in Development Bank of Japan (2005), "Investigation of the Seasonality and Profitability of R&D," Research Report No. 81.

(number of employees) and a dummy variable indicating whether the plant is active in multiple industries (at the 4 digit level).[24] In addition, it is controlled for parent firm size (number of employees) and the number of plants of the parent firm. On the one hand, increases in the number of a firm's plants may correlate with unmeasured firm-specific advantages. On the other hand a larger numbers of plants drawing on the same R&D pool may lead to reduced effective knowledge transfer (Adams and Jaffe, 1996). A set of year dummies λ_t and region (prefecture) dummies ρ_r is included. μ_{st} is modeled as a set of industry dummies μ_s in addition to the average TFP growth rate for all plants in the industry, $\widetilde{\ln tfp}_{st}$, which controls for industry-specific technological opportunity and demand shocks over time affecting TFP growth.

2.8. Specification

Equation (3-4) is estimated in its long difference form. Taking the first-order difference from the previous year, we control out the plant-specific fixed effects. Moreover, the long difference models, while sacrificing degrees of freedom, is a conservative estimation method to reduce the influence of measurement error and cyclical effects (e.g. Haskel et al, 2007; Branstetter, 2000). To strike a balance between degrees of freedom and reduction in measurement error, 5-year differences are taken starting from 1987, which leaves a maximum of exactly 4 non-overlapping long differenced observations (for plants observed over the entire period): 1987-1992, 1993-1997, 1998-2002 and 2003-2007. To facilitate interpretation of the descriptive statistics, the long difference is divided by 5 to arrive at annual average growth rates of TFP and R&D stocks during the 5-year periods. Since the geographic decay specification introduces nonlinearity in the TFP equation, equation (4) is estimated with nonlinear least squares. The distance decay parameters are estimated using a Taylor approximation.[25] Error

---

[24] Note that age effects are of no interest in differenced models, since the difference in age would be identical for all plants.

[25] Without approximation we would need to sum up over all R&D conducting firm-pairs and industries for each plant to arrive at an update of the distance parameter τ, which is computationally infeasible. Therefore the distance function is approximated by taking a H-order Taylor's expansion: $e^{\tau d_{if's't}} \cong \sum_{n=0}^{H} e^{\tau \bar{d}} (\tau)^n \frac{\left(d_{if's't}-\bar{d}\right)^n}{n!}$, such that the expression for the plant level technologically proximate R&D stock becomes:

terms are cluster-robust at the plant level.

Table 3.2 shows descriptive statistics of the variables and Table 3.3 contains the correlation matrix. The correlations between the (growth in) relationally proximate R&D stocks (buyers and suppliers) and the technologically proximate R&D stock are rather high at 0.66-0.78. This is mainly stemming from the correlation in same-industry R&D stocks, while correlations between stocks in other industries range between -0.04 and 0.12. Hence, the different measures of proximity do suggest rather different weightings for R&D stocks and the resulting spillovers potential.

Table 3.2 Descriptive statistics

	Mean	SD	Min	Median	Max
TFP	0.007	0.079	-1.409	0.006	1.025
PARENT R&D	0.023	0.055	-0.563	0.000	1.604
Tech-proximate PRIVATE R&D	0.040	0.038	-0.155	0.035	0.421
Supplier PRIVATE R&D	0.043	0.043	-0.168	0.036	0.237
Customer PRIVATE R&D	0.040	0.041	-0.751	0.033	0.420
PUBLIC R&D	0.030	0.008	0.002	0.030	0.072
Number of other plants of the parent firm	0.004	0.058	-1.099	0.000	1.099
Number of firm employees	-0.003	0.095	-2.290	-0.002	3.306
Number of plant employees)	-0.005	0.082	-2.297	-0.004	1.285
Multi-products (4 digit) plant dummy	-0.001	0.093	-1.000	0.000	1.000
Parent R&D stock > 0 (dummy)	0.435	0.485	0.000	0.000	1.000
Industry average TFP growth rate	0.006	0.019	-0.124	0.003	0.184
Prior TFP level relative to industry average	0.054	0.269	-1.529	0.036	1.383

Note: all variables are expressed as average 5-year differences, except for prior TFP

$$S_{ifst}^{tech} =\cong \sum_{n=0}^{H}\left[e^{\tau\bar{d}}(\tau)^n \sum_{f'\neq f}\sum_{s'}\left(K_{f's't}T_{ss'}\frac{\left(d_{if's't}-\bar{d}\right)^n}{n!}\right)\right]$$

The summation over f' and s' no longer depends on the distance decay parameter τ, and summation over H suffices. H is set conservatively at 50 and $\bar{d}$ is set at 1500 km (the midpoint of the smallest and largest possible distance).

Table 3.3 Correlation coefficients

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
[1]TFP	1.000												
[2]PARENT R&D	0.020	1.000											
[3]Tech-proximate PRIVATE R&D	0.071	0.086	1.000										
[4]Supplier PRIVATE R&D	0.076	0.103	0.612	1.000									
[5]Customer PRIVATE R&D	0.091	0.108	0.656	0.746	1.000								
[6]PUBLIC R&D	0.026	-0.021	0.065	0.213	0.100	1.000							
[7]Number of other plants of the parent firm	0.012	0.041	0.059	0.082	0.075	0.021	1.000						
[8]Number of firm employees	0.018	0.046	0.061	0.086	0.082	-0.057	0.297	1.000					
[9]Number of plant employees)	0.014	0.030	0.051	0.073	0.072	-0.070	-0.012	0.562	1.000				
[10]Multi-products (4 digit) plant dummy	-0.004	0.004	-0.001	0.001	0.005	0.000	-0.013	0.001	0.025	1.000			
[11]Parent R&D stock > 0 (dummy)	-0.017	0.451	-0.101	-0.099	-0.095	-0.077	-0.038	-0.059	-0.039	0.001	1.000		
[12]Industry average TFP growth rate	0.212	0.074	0.345	0.380	0.432	0.006	0.011	0.052	0.057	0.001	-0.045	1.000	
[13]Prior TFP level relative to industry average	-0.271	0.064	0.049	0.038	0.021	-0.004	0.000	-0.005	0.009	-0.010	0.128	-0.018	1.000

Note: all variables are expressed as 5-year differences, except for prior TFP

3. Empirical Results

Table 3.4 reports the estimation results. Model 1 only includes the technologically proximate R&D stock and the parent firm R&D stock. The coefficient on parent R&D suggests an elasticity of TFP with respect to R&D of 0.033 percent, which is within, but at the lower end, of the range estimated in Adams and Jaffe (1996) for plant level R&D effects.[26] The elasticity of the private R&D stock is higher – a common finding in R&D spillover studies- at 0.058, while spillover effects decay in distance, as the significant distance parameter suggests. The estimates on the past TFP level suggest that plants that are 1 percent more productive than the average TFP level in the industry have a 0.08 percent point smaller TFP growth rate, indicating that there is a modest gradual convergence in productivity. TFP growth of the plants is strongly influenced by opportunities and shocks captured by the average TFP growth in the industry, with an estimated elasticity of 0.89. Of the plant and firm control variables, only (growth in) the

[26] It should be noted that their specification was cross sectional, and one may expect smaller effects in a differenced model.

number of plants operated by the parent firm has a marginally significant positive effect on TFP.

In model 2 the dummy variable indicating continuous positive R&D is added. Both the dummy variable indicating positive R&D and the R&D stock are significant. The dummy variable suggests that R&D performing firms generate on average 0.5 percent points higher TFP growth (independent of variation in their R&D stocks). At the same time, the coefficient of the parent R&D stock declines to about 0.01. Model 3 adds the technologically proximate public R&D stock. The coefficient on public R&D, at 0.077 is larger than the coefficient on technologically proximate private R&D, demonstrating the importance of knowledge spillovers from public R&D. The estimates however do not suggest a significant geographic decay effect of public R&D spillovers. The addition of public R&D in model 3 does not materially affect the estimated coefficient on private R&D, which may indicate little overlap in the type of knowledge from technologically proximate private and public R&D.

In model 4 the relationally proximate R&D stocks of customers and suppliers are added. The relationally proximate R&D stock due to supplier linkages has a significant effect on TFP growth with elasticity of 0.031. The significant elasticity of customer R&D stocks is slightly smaller at 0.026. Meanwhile, the coefficient on the technologically proximate R&D stock reduces with the inclusion of the supplier and customer R&D stock variables, and at 0.035 is similar in magnitude as the elasticity of the supplier R&D stock. The estimated distance decay for private R&D spillovers becomes smaller overall, suggesting weaker proximity influences for relationally proximate R&D. Model 5 confirms this pattern: when we allow separate decay parameters for the three private R&D stocks, the decay parameter for technologically proximate R&D increases in strength whereas the model does not identify a distance decay effects for R&D spillovers from buyers and suppliers. For technologically proximate R&D spillovers, the decay function on the basis of model 5 is depicted in Figure 3.5. Spillover effects decline and become negligible at about 500 kilometers. This pattern is similar to the estimates reported in Lychagin et al. (2010) for US listed manufacturing firms based on inventor locations.

Model 6 presents the results of an alternative model with one parameter estimated for the (un-weighted) sum of the three types of private R&D. The estimated coefficient for this combined private R&D stock is close to 0.08 and larger than the estimated coefficient for technologically proximate R&D in models 1-3. This underscores that failure to take into account relational proximity may lead to an underestimation of R&D spillover effects. The estimate of the distance parameter for the combined private R&D stock is close to the parameter estimated in model 4.

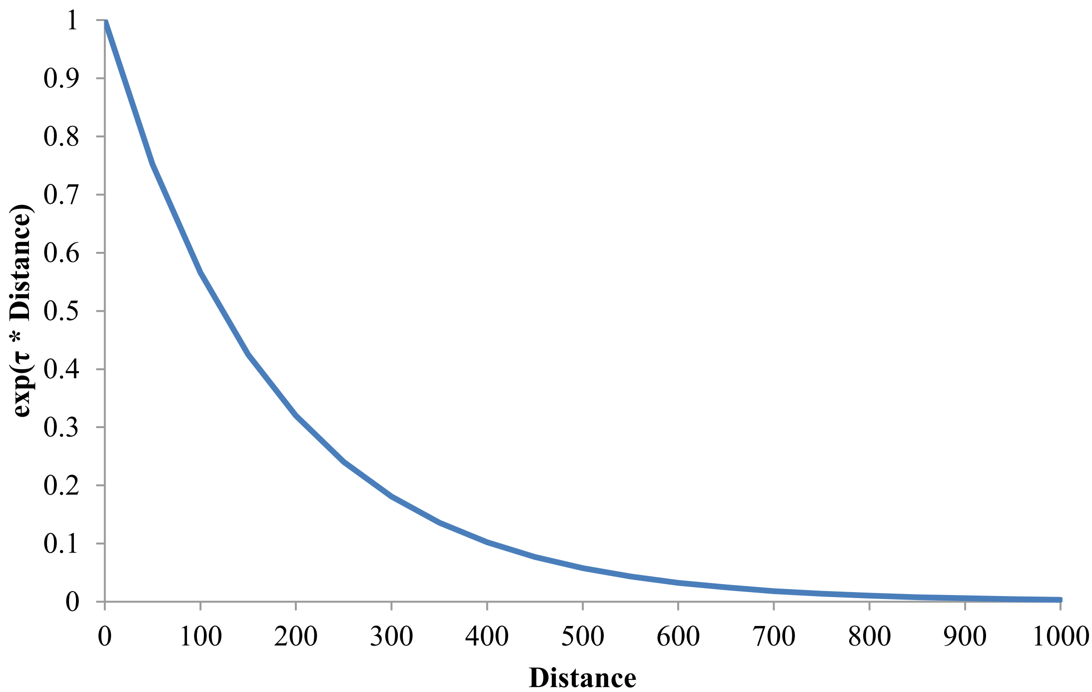
Table 3.4 Long Difference Analysis of Plant-level TFP (1987-2007)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Distance parameters:							
Tech-proximate PRIVATE R&D	-0.0040	-0.0038	-0.0040		-0.0057		-0.0058
	[0.0012]***	[0.0011]***	[0.0012]***		[0.0027]**		[0.0027]**
all PRIVATE R&D				-0.0018		-0.0017	
				[0.0008]**		[0.0010]*	
Supplier PRIVATE R&D					0.0000		0.0000
					[0.0027]		[0.0027]
Customer PRIVATE R&D					0.0000		0.0000
					[0.0037]		[0.0037]
PUBLIC R&D			0.0000	0.0000	0.0000	0.0000	
			[0.0024]	[0.0025]	[0.0024]	[0.0025]	
PUBLIC R&D (parent R&D>0)							0.0000
							[0.0020]
PUBLIC R&D (parent R&D=0)							-0.0060
							[0.0059]
R&D parameters:							
Parent R&D	0.0331	0.0097	0.0097	0.0096	0.0096	0.0096	0.0096
	[0.0036]***	[0.0043]**	[0.0043]**	[0.0043]**	[0.0043]**	[0.0043]**	[0.0043]**
Parent R&D stock > 0 (dummy)		0.0050	0.0050	0.0050	0.0050	0.0050	0.0034
		[0.0004]***	[0.0004]***	[0.0004]***	[0.0004]***	[0.0004]***	[0.0012]***
Tech-proximate PRIVATE R&D	0.0583	0.0600	0.0582	0.0392	0.0346		0.0347
	[0.0167]***	[0.0168]***	[0.0167]***	[0.0194]**	[0.0167]**		[0.0167]**
Supplier PRIVATE R&D				0.0311	0.0360		0.0364
				[0.0141]**	[0.0140]**		[0.0140]***
Customer PRIVATE R&D				0.0260	0.0260		0.0259
				[0.0131]**	[0.0131]**		[0.0130]**
all PRIVATE R&D						0.0775	
						[0.0180]***	
PUBLIC R&D			0.0766	0.0766	0.0832	0.0746	
			[0.0364]**	[0.0373]**	[0.0378]**	[0.0363]**	
PUBLIC R&D (parent R&D>0)							0.1211
							[0.0416]***
PUBLIC R&D (parent R&D=0)							0.0678
							[0.0356]*
Other parameters:							
Plant's relative prior TFP	-0.0792	-0.0802	-0.0802	-0.0803	-0.0803	-0.0802	-0.0803
	[0.0007]***	[0.0007]***	[0.0007]***	[0.0007]***	[0.0007]***	[0.0007]***	[0.0007]***
Industry average TFP growth	0.8917	0.8919	0.8971	0.8962	0.8966	0.8977	0.8970
	[0.0193]***	[0.0193]***	[0.0197]***	[0.0197]***	[0.0198]***	[0.0196]***	[0.0196]***
Number of other plants	0.0077	0.0087	0.0087	0.0087	0.0087	0.0087	0.0086
	[0.0053]	[0.0053]*	[0.0053]	[0.0053]	[0.0053]	[0.0053]	[0.0053]
Number of firm employees	-0.0008	0.0011	0.0012	0.0010	0.0010	0.0011	0.0010
	[0.0047]	[0.0047]	[0.0047]	[0.0047]	[0.0047]	[0.0047]	[0.0047]
Number of plant employees	-0.0040	-0.0032	-0.0031	-0.0033	-0.0033	-0.0032	-0.0032
	[0.0051]	[0.0051]	[0.0051]	[0.0051]	[0.0051]	[0.0051]	[0.0051]
Multi-products (4digit) plant	-0.0033	-0.0033	-0.0034	-0.0033	-0.0033	-0.0033	-0.0033
	[0.0029]	[0.0029]	[0.0029]	[0.0029]	[0.0029]	[0.0029]	[0.0029]
Constant	-0.0040	-0.0035	-0.0057	-0.0092	-0.0086	-0.0072	-0.0084
	[0.0073]	[0.0073]	[0.0073]	[0.0074]	[0.0074]	[0.0073]	[0.0073]
Industry dummies (JIP industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prefecture dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# observations	46,923	46,923	46,923	46,923	46,923	46,923	46,923
R-squared	0.1685	0.1696	0.1696	0.1697	0.1697	0.1696	0.1698
F statistic	9486.43***	9555.59***	9556.97***	9563.57***	9566.77***	9556.55***	9568.20***

* p < 0.1, ** p < 0.05, *** p < 0.01.

Prior studies have suggested that firms need to invest in internal R&D in order to benefit from academic research (e.g. Cassiman and Veugelers, 2006; Anselin et al., 1997; Belderbos et al., 2009), as firms need the absorptive capacity to screen, understand, and utilize the fruits of relevant scientific research (Cohen and Levinthal, 1990). In model 7, the effect of public R&D is separated into an effect for firms without formal R&D expenditures and an effect for firms with positive R&D. The results confirm that the presence of internal R&D increases the magnitude of public R&D spillovers: the elasticity increases to 0.12, while the coefficient for firms without internal R&D is only marginally significant (at 0.068). The difference between the two coefficients is statistically significant.

Figure 3.5 Decay in the effect of technologically proximate R&D spillovers as a function of distance



3.1. Sensitivity Analysis

In this subsection, the role of distance for public spillovers and the assumption that (private) R&D spillovers as a function of distance play out at the plant level are further explored. In an alternative specification, distance between the firms' R&D laboratories and between R&D laboratories and the location of public R&D institutions are examined. In particular for public spillovers, linkages may occur at the laboratory level and not necessarily at the plant level, while the R&D laboratories may not necessarily be located close to the firms' plants. It is derived as the location of R&D laboratories from published directories of R&D establishments in Japan. For R&D performing firms lacking laboratory location information, R&D is assigned to the location of headquarters – the safest option for these -mostly smaller- firms (e.g. Adams and Jaffe, 1996; Orlando, 2004). Results, however, did not show geographic decay effects in this specification either.

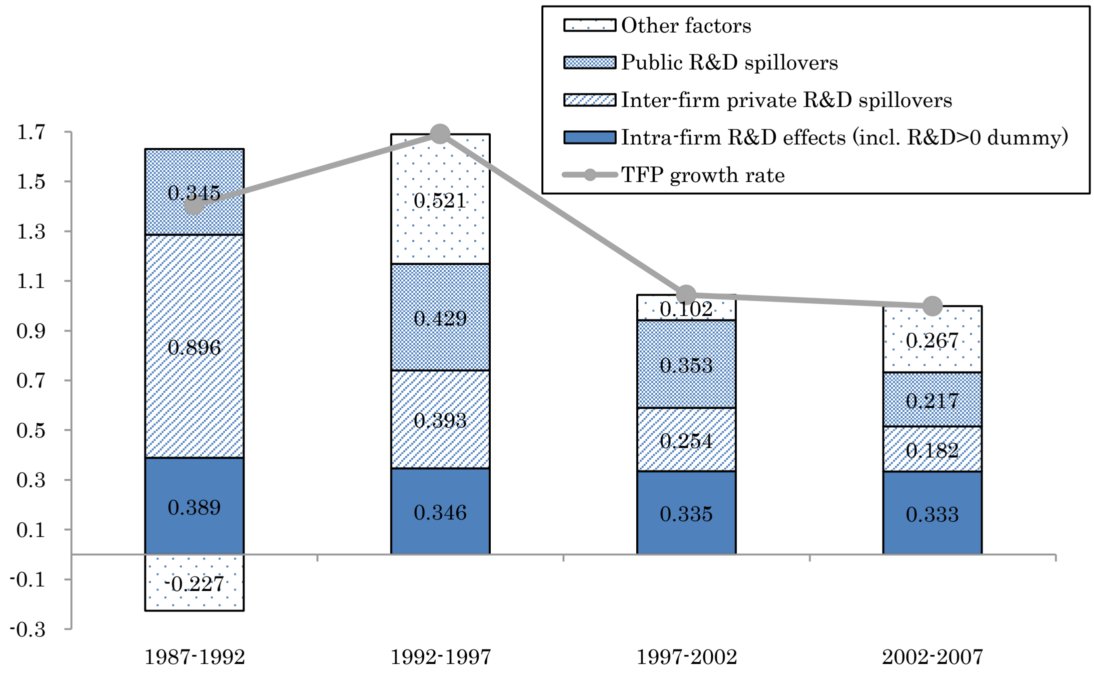
A number of additional sensitivity analyses are conducted, estimating model 6 on different samples. First, productivity models for the entire population of Japanese manufacturing plants (plants with TFP information; more than 230000 observations) are estimated to examine the robustness of the estimates. Here the unmatched plants are treated as zero R&D plants while including a separate dummy variable indicating that the plants lack R&D information. Second the model without smaller plants (leaving about 36000 observations) and on a balanced sample (limited to about 16000 observations) is estimated, to explore the implications of potential sample selection bias. All these models produced broadly similar results, with some exceptions. The distance effect for technologically proximate R&D proved difficult to identify in some of the models, while in two specifications only two of the individual effects of supplier, customer, and technologically proximate R&D were simultaneously estimated as significant. The chapter aims to further explore the robustness of the empirical model in future work.

3.2. Decomposition Analysis

Given the time dimension in the data and the changes over time in R&D investments and agglomeration, long term TFP growth effects can be decomposed into several factors: firms' internal R&D effects, private R&D spillover effects, and public R&D spillover effects. The results of the decomposition analysis based on model 7 are presented in Figure 3.6-Figure 3.9. The decomposition analysis is conducted for a balanced sample of close to 4200 plants. Keeping the sample of spillover receiving plants stable ensures that the decomposition is not influenced by period-on-period changes in the sample but highlights effects of the changing 'supply' of spillovers. The decomposition uses plants' gross output as weights. Figure 3.6 shows that declining R&D spillovers, in particular private R&D spillovers, play an important role in the decline in TFP growth over the years. The contribution of private R&D spillovers to TFP growth for the plants in the balanced sample reduced from 0.896 percent points in 1987-1992 to

0.182 percent points in 2002-2007. The contribution of public R&D spillovers also declined, but less so in relative and absolute terms. This is related to the more modest decline in the growth in public R&D and a changing composition of public R&D expenditures in the direction of life sciences with greater relevance for the private sector. The role of internal R&D remained relatively stable, although this is to an important extent due to the fact that R&D active firms record generally higher TFP growth than firms that are not engaged in R&D.

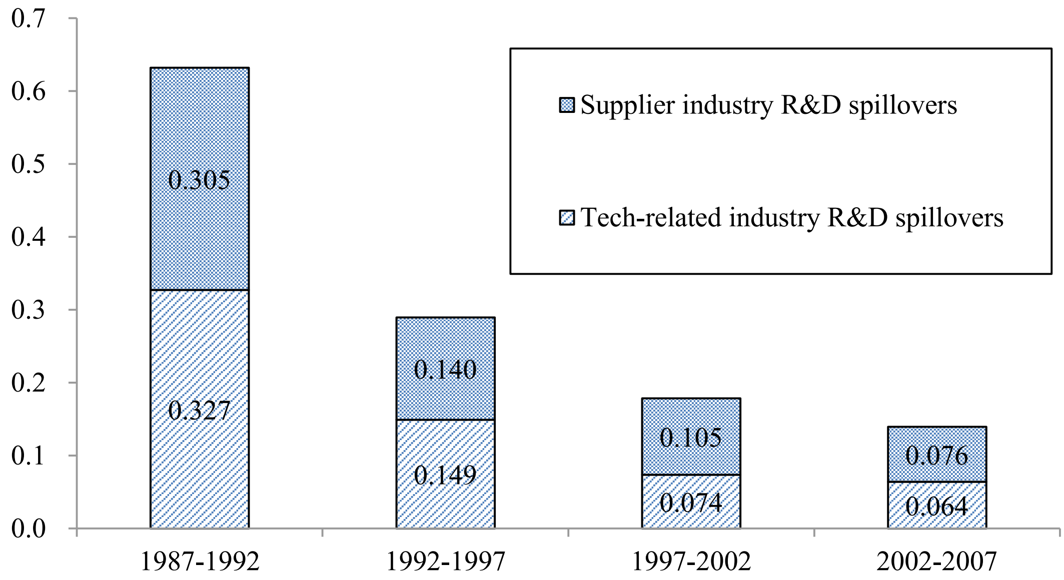
Figure 3.6 TFP Growth Composition: Intra-firm R&D vs. Private and Public Spillovers



Note: based on a balanced sample, 1987-2007

Further decomposition of the changing role of private R&D spillovers can be into the three types of spillovers: spillovers due to technological proximity, buyer effects, and supplier effects. Figure 3.7 shows that the technological proximity based spillovers and customer spillovers have declined most, while the decline in supplier spillovers has been more modest. These divergent effects arise because the share of procurement from (R&D intensive) local manufacturing industries has not decreased that much over time, while an increasing role of exports has reduced relational proximity to Japanese customer industries.

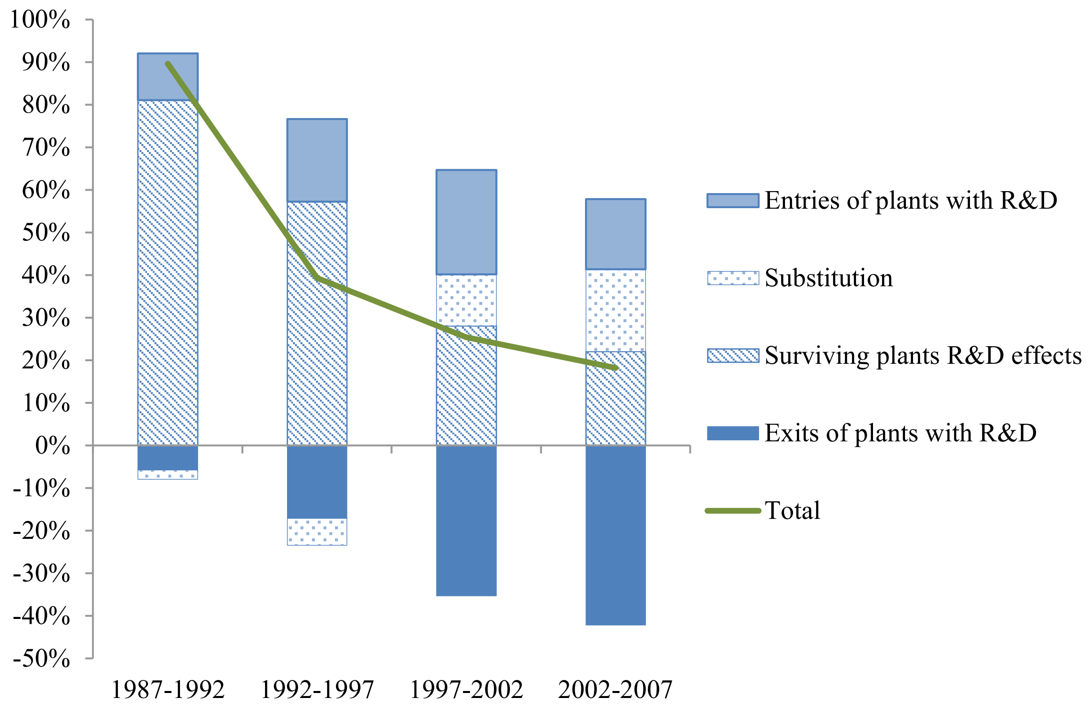
Figure 3.7 TFP Growth Composition: Effects of types of Private R&D spillovers



Note: based on a balanced sample, 1987-2007

Figure 3.8 decomposes private spillovers into effects due to the exit of R&D active plants, the entry of such plants, and the changing R&D stocks of surviving plants. The exit of R&D active plants reduces the R&D stock available to other plants and has a negative effect on TPF growth. However, if the parent firm operates multiple plants, the exit of one of its plants implies that another plant of the firm takes its place as 'minimum distance' plant providing R&D spillovers, such that there is a compensating 'plant substitution effect'. In such cases, net spillovers decline only to the extent that the exit increases average distance between plants. Similarly, if a firm opens up a new plant, this may increase the R&D stock available to plants in its proximity, but at the same time it displaces the R&D stock of the firm's plant that was previously located at minimum distance to these receiving plants. Hence, in case of entry there is a partially compensating negative substitution effect. This decomposition exercise shows that while the largest part of the decline in spillovers is due to a slowing down of R&D stock growth in surviving plants, increasing exit effects and reduced entry effects over time also play an important role. Figure 3.9 shows that most of the exits have taken place in the major industrial agglomerations in Japan around Tokyo and Kanagawa, Osaka, and Aichi (home of a large automobile cluster) during 1997-2007.

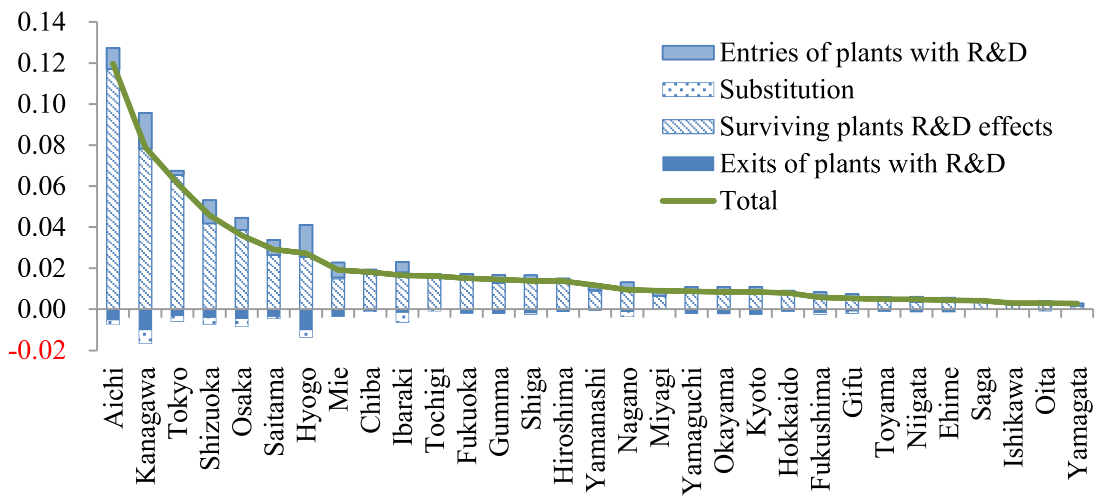
Figure 3.8 TFP Growth Composition: Effects of R&D Active Firms' Plant Entry and Exit



Note: based on a balanced sample, 1987-2007

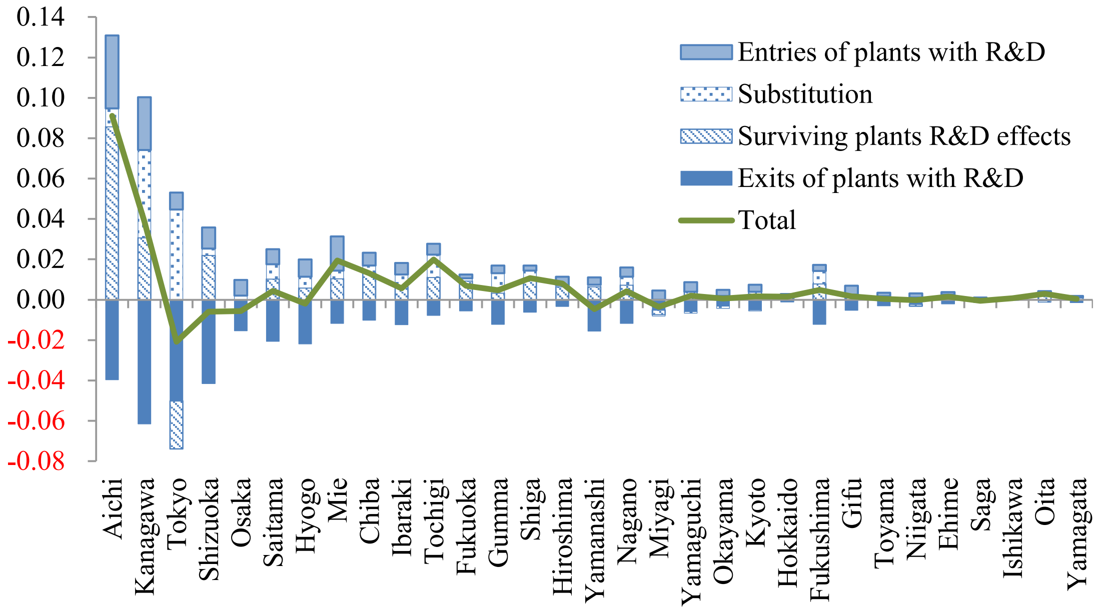
Figure 3.9 TFP Growth Composition: Effects Plant Entry and Exit by Prefecture

a. 1987-1997



Note: based on a balanced sample, 1987-1997

b. 1997-2007



Note: based on a balanced sample, 1997-2007

4. Conclusions

This chapter examined the effects of R&D spillovers as the main type of knowledge spillovers on total factor productivity in a large panel of Japanese manufacturing plants matched with R&D survey data. It simultaneously analyses the role of public (universities and research institutes) and private R&D spillovers, while examining effects due to 'relational' (supplier-customer) proximity as well as technological and geographic proximity. Analyses of this chapter confirm the importance of positive spillover effects from R&D by firms with plants in technologically related industries. The latter spillover effects are attenuated by distance and the estimates suggest that most spillover effects disappear beyond 500 kilometers. We also observe positive effects of public R&D spillovers, with the effects substantially larger for plants with access to internal R&D. It is not found the evidence that public R&D spillover effects are attenuated by distance. In addition to knowledge spillovers from technologically proximate plants, it is found the evidence that 'relational proximity' due to buyer and supplier linkages generates additional 'pecuniary' R&D spillovers of similar magnitude as the knowledge spillovers due to technological proximity. This chapter could not identify the role of geographic distance in these buyer and supplier spillovers.

This chapter concludes that public as well as private R&D spillovers matter for TFP growth, while relational proximity as well as technological proximity needs to be taken into account to arrive at representative estimates of the social effects of private R&D. Decomposition analysis shows that the contribution of private R&D spillovers to TFP growth has declined since the late 1990s. This is due to a declining growth in R&D stocks while another important factor is the exit of proximate plants operated by R&D intensive firms. A mildly declining contribution of public R&D spillovers is primarily due to a reduction in the growth of R&D by public research organization since the late 1990s. If we explore effects at the regional level, we observe that strong adverse exit effects occurred in particular in Japan's major industrial agglomerations such as Tokyo and Osaka.

These results help to explain the twin stylized facts of Japanese productivity growth: the exit of relatively productive plants and the declining TFP growth or surviving plants (Fukao and Kwon, 2006; Kneller et al., 2012). They suggest that these two trends may be causally related. The exit of plants by R&D intensive firms reduces the available R&D spillovers and hampers TFP growth of the surviving plants.

In future work, the research project aims to get a better understanding of the (absence of) distance effects in R&D spillovers. One reason for the lack of estimated distance effects for public R&D may be that public R&D spillovers occur most often through active collaboration across larger distances (Okamuro and Nishimura, 2013; Gittelman, 2007). We can explore these explanations by incorporating information available on research relationships between firms and universities. Additionally, previous studies show that the research collaboration between firms and universities have been increased from 1990's. This may accelerate knowledge transfer from university to private firm overtime. By using information on research relationships between firms and universities, we can also precisely test this hypothesis. Second, it aims to investigate the role of proximity effects in buyer-supplier relationships in more detail by utilizing data on the most important buyers and suppliers of individual Japanese firms. Third, it is planned to match the data with the Basic Surveys on Business Activities in Japan, which contain information on corporate relationships and foreign activities. Matching with the Basic Surveys allows bringing in controls on overseas R&D conducted/outsourced by the firms and the potentially resulting international transfers and knowledge spillovers (e.g. Branstetter, 2001; Griffith et al., 2008). It also allows analysis of potentially greater R&D spillovers for firms operating within business groups (Suzuki, 1993; Branstetter, 2000). Collectively, the remaining challenges for exploration of R&D spillover effects present a rich research agenda.

Appendix A. Technological proximity between industries

Spillovers sources (cited) / Focal industries (citing)	[04]	[05]	[06]	[07]	[08]	[09]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]
[04] Food products	1.00	.003	.006	.000	.125	.359	.041	.001	.000	.004	.001	.001	.001	.094	.021	.001	.003	.002	.000	.026	.026
[05] Textile mill products	.007	1.00	.045	.024	.631	.065	.104	.001	.002	.172	.007	.006	.023	.243	.026	.013	.033	.019	.005	.148	.114
[06] Pulp and paper products	.022	.073	1.00	.126	.415	.049	.089	.002	.000	.100	.003	.003	.043	.301	.009	.008	.190	.004	.001	.123	.083
[07] Printing	.000	.011	.042	1.00	.270	.021	.095	.000	.000	.028	.008	.011	.020	.085	.003	.003	.181	.002	.000	.087	.017
[08] Chemical fertilizers and industrial chemicals	.009	.020	.008	.015	1.00	.147	.050	.012	.004	.039	.007	.007	.005	.070	.005	.010	.032	.006	.001	.041	.027
[09] Drugs and medicine	.026	.002	.001	.001	.147	1.00	.013	.000	.000	.002	.000	.000	.000	.010	.001	.000	.005	.000	.000	.076	.001
[10] Miscellaneous chemicals	.031	.032	.012	.035	.488	.128	1.00	.020	.000	.038	.008	.007	.010	.093	.010	.006	.057	.014	.003	.055	.036
[11] Petroleum and coal products	.004	.004	.002	.001	.763	.031	.143	1.00	.000	.008	.006	.005	.014	.209	.003	.036	.074	.030	.004	.130	.014
[12] Rubber products	.000	.008	.001	.001	.400	.002	.006	.000	1.00	.008	.014	.011	.004	.030	.001	.005	.028	.064	.002	.050	.116
[13] Ceramic, stone and clay products	.003	.064	.026	.021	.439	.015	.047	.001	.001	1.00	.030	.027	.073	.225	.020	.022	.108	.032	.008	.112	.197
[14] Iron and steel	.001	.006	.002	.013	.248	.011	.028	.004	.007	.120	1.00	.580	.069	.410	.030	.059	.152	.036	.008	.065	.048
[15] Non-ferrous metals and products	.001	.009	.003	.030	.392	.020	.042	.004	.010	.187	1.00	1.00	.108	.486	.034	.111	.233	.052	.009	.097	.075
[16] Fabricated metal products	.001	.009	.012	.015	.066	.006	.016	.004	.000	.104	.025	.024	1.00	.259	.027	.050	.082	.081	.025	.070	.102
[17] General-purpose machinery	.010	.012	.008	.007	.114	.019	.018	.005	.001	.040	.019	.013	.033	1.00	.018	.020	.059	.078	.014	.082	.058
[18] Household appliances	.022	.015	.003	.004	.091	.012	.022	.001	.000	.039	.014	.010	.039	.188	1.00	.057	.121	.056	.004	.079	.106
[19] Electrical machinery	.000	.003	.001	.001	.080	.003	.004	.003	.000	.019	.013	.015	.026	.084	.022	1.00	.244	.082	.009	.127	.031
[20] Info.andcom. electronics	.000	.001	.003	.008	.024	.003	.005	.001	.000	.008	.003	.003	.005	.027	.005	.026	1.00	.010	.001	.068	.009
[21] Motor vehicles, parts and accessories	.000	.003	.001	.001	.028	.001	.008	.002	.003	.017	.004	.004	.029	.183	.012	.046	.055	1.00	.022	.076	.041
[22] Other transportation equipment	.000	.004	.001	.001	.032	.002	.012	.003	.000	.031	.006	.005	.064	.260	.008	.043	.041	.197	1.00	.060	.064
[23] Precision instruments and machinery	.003	.009	.004	.007	.070	.129	.011	.003	.001	.019	.003	.003	.009	.078	.007	.030	.151	.030	.003	1.00	.035
[24] Miscellaneous manufacturing	.011	.019	.009	.007	.180	.007	.024	.001	.008	.106	.007	.006	.042	.184	.034	.023	.076	.048	.009	.117	1.00

Source: calculations based on Leten et al. (2008)

Appendix B. Applied weights for relationally proximate (Supplier) R&D stocks

Focal industries (buyer) \ Spillover sources (supplier)		[04]	[05]	[06]	[07]	[08]	[09]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	Total
[04]	Food products	.120	.001	.015	.007	.006	.000	.002	.004	.000	.005	.000	.001	.018	.000	.000	.000	.000	.000	.000	.000	.181
[05]	Textile mill products	.003	.223	.009	.008	.034	.000	.009	.006	.003	.000	.000	.000	.002	.000	.000	.000	.001	.000	.000	.000	.298
[06]	Pulp and paper products	.003	.006	.275	.014	.018	.000	.012	.012	.001	.001	.000	.000	.001	.000	.000	.000	.001	.000	.000	.000	.344
[07]	Printing	.002	.001	.111	.081	.001	.000	.029	.002	.001	.000	.000	.002	.000	.000	.000	.000	.002	.000	.000	.000	.233
[08]	Chemical fertilizers and industrial chemicals	.003	.001	.005	.002	.339	.000	.007	.084	.001	.003	.000	.003	.005	.000	.000	.000	.001	.000	.000	.000	.454
[09]	Drugs and medicine	.012	.002	.033	.008	.071	.048	.013	.003	.002	.013	.000	.001	.013	.001	.000	.000	.002	.000	.000	.000	.222
[10]	Miscellaneous chemicals	.005	.001	.034	.012	.177	.001	.083	.005	.001	.006	.000	.004	.016	.001	.000	.000	.001	.000	.000	.000	.346
[11]	Petroleum and coal products	.001	.001	.000	.000	.002	.000	.003	.050	.000	.001	.000	.000	.002	.000	.000	.000	.000	.000	.000	.000	.060
[12]	Rubber products	.001	.017	.008	.002	.185	.000	.007	.005	.041	.001	.003	.001	.025	.000	.000	.000	.001	.000	.000	.000	.296
[13]	Ceramic, stone and clay products	.002	.003	.017	.003	.016	.000	.007	.022	.002	.090	.010	.003	.009	.003	.000	.000	.001	.000	.000	.000	.187
[14]	Iron and steel	.001	.001	.001	.001	.005	.000	.001	.029	.001	.007	.453	.006	.001	.001	.000	.000	.000	.000	.000	.000	.508
[15]	Non-ferrous metals and products	.001	.002	.004	.002	.013	.000	.003	.007	.000	.007	.002	.245	.002	.001	.000	.000	.001	.000	.000	.000	.289
[16]	Fabricated metal products	.002	.002	.004	.004	.002	.000	.008	.005	.002	.004	.192	.046	.062	.002	.000	.001	.004	.000	.000	.000	.342
[17]	General-purpose machinery	.001	.001	.003	.004	.001	.000	.005	.002	.011	.005	.073	.014	.034	.189	.000	.020	.022	.000	.000	.004	.391
[18]	Home electronics	.002	.003	.012	.014	.012	.000	.004	.002	.006	.003	.023	.022	.027	.021	.099	.033	.132	.000	.000	.002	.417
[19]	Electrical machinery	.002	.002	.011	.004	.007	.000	.005	.003	.006	.009	.039	.052	.025	.016	.000	.123	.028	.000	.000	.001	.334
[20]	Information and communication electronics	.003	.003	.012	.009	.008	.000	.005	.003	.004	.015	.004	.018	.016	.005	.001	.034	.256	.000	.000	.000	.396
[21]	Motor vehicles, parts and accessories	.001	.002	.003	.002	.002	.000	.007	.002	.015	.006	.030	.012	.007	.009	.005	.031	.005	.445	.000	.000	.583
[22]	Other transportation equipment	.001	.003	.002	.004	.002	.000	.013	.003	.014	.006	.092	.013	.028	.036	.003	.020	.008	.030	.189	.001	.470
[23]	Precision instruments and machinery	.001	.002	.010	.005	.004	.000	.003	.003	.005	.018	.011	.017	.016	.011	.000	.014	.066	.000	.000	.095	.284

Source: JIP database. Data are for 1990.

Appendix C. Applied weights for relationally proximate Buyer R&D stocks

	Spillover sources (buyer)	[04]	[05]	[06]	[07]	[08]	[09]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	Total
[04]	Food products	.120	.001	.001	.000	.001	.002	.001	.000	.000	.000	.000	.000	.001	.001	.000	.000	.001	.001	.000	.000	.132
[05]	Textile mill products	.006	.223	.005	.001	.001	.001	.001	.001	.005	.003	.001	.001	.002	.004	.002	.003	.007	.006	.001	.001	.275
[06]	Pulp and paper products	.067	.011	.275	.088	.007	.022	.026	.001	.003	.017	.002	.003	.006	.010	.009	.014	.029	.011	.001	.005	.607
[07]	Printing	.039	.012	.018	.081	.003	.007	.011	.001	.001	.004	.002	.002	.008	.017	.014	.005	.027	.010	.003	.003	.269
[08]	Chemical fertilizers and industrial chemicals	.020	.030	.014	.001	.339	.034	.100	.002	.049	.012	.008	.006	.003	.003	.006	.007	.015	.005	.001	.001	.655
[09]	Drugs and medicine	.002	.000	.000	.000	.000	.048	.002	.000	.000	.000	.000	.000	.000	.001	.000	.000	.001	.000	.000	.000	.056
[10]	Miscellaneous chemicals	.009	.013	.015	.030	.013	.011	.083	.006	.003	.009	.004	.002	.016	.020	.004	.007	.018	.038	.010	.002	.313
[11]	Petroleum and coal products	.011	.004	.008	.001	.073	.001	.002	.050	.001	.015	.041	.003	.005	.005	.001	.003	.004	.004	.001	.001	.236
[12]	Rubber products	.002	.008	.002	.002	.003	.003	.001	.000	.041	.004	.008	.000	.009	.098	.013	.021	.029	.179	.023	.006	.453
[13]	Ceramic, stone and clay products	.023	.001	.001	.000	.004	.009	.005	.001	.000	.090	.016	.004	.006	.017	.003	.011	.043	.027	.004	.008	.273
[14]	Iron and steel	.000	.000	.000	.000	.000	.000	.000	.000	.001	.004	.453	.001	.139	.107	.008	.022	.005	.059	.025	.002	.827
[15]	Non-ferrous metals and products	.005	.000	.000	.002	.006	.001	.004	.000	.001	.004	.018	.245	.107	.069	.026	.096	.067	.078	.011	.012	.751
[16]	Fabricated metal products	.051	.002	.001	.000	.004	.006	.008	.002	.006	.006	.001	.001	.062	.070	.014	.020	.025	.019	.011	.005	.312
[17]	General-purpose machinery	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	.001	.189	.005	.006	.004	.012	.007	.002	.227
[18]	Home electronics	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.001	.099	.000	.002	.026	.002	.000	.134
[19]	Electrical machinery	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.002	.053	.021	.123	.067	.102	.009	.005	.381
[20]	Information and communication electronics	.001	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.002	.023	.039	.012	.256	.007	.002	.008	.352
[21]	Motor vehicles, parts and accessories	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.445	.004	.000	.449
[22]	Other transportation equipment	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.189	.000	.189
[23]	Precision instruments and machinery	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.029	.004	.004	.002	.004	.002	.095	.140

Source: JIP database. Data are for 1990.

Appendix D: Applied weights in the science field - industry concordance

Focal industries (citing industries) \ Spillover sources (cited science fields)		Agriculture	Biology	Medicine	Nursing	Dentistry	Chemistry	Applied-Chemistry	Physics	Geology	Engineering	Electronics	Energy	Material Science	Mathematics	Education	Art-Literature-Society	Economics-Business-Management	History-Politics-Law	Philosophy
[04]	Food products	1.5	0.5	0.1	0.2	0.0	0.1	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
[05]	Textile mill products	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
[06]	Pulp and paper products	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
[07]	Printing	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
[08]	Chemical fertilizers and industrial chemicals	1.8	3.9	1.2	0.4	0.7	4.5	3.2	0.3	0.1	0.2	0.1	0.5	1.3	0.0	0.0	0.0	0.0	0.0	0.0
[09]	Drugs and medicine	3.4	15.6	5.8	2.3	2.1	7.0	3.2	0.3	0.1	0.2	0.3	0.4	0.3	0.0	0.1	0.2	0.0	0.0	0.0
[10]	Miscellaneous chemicals	0.2	0.1	0.0	0.0	0.0	0.2	0.5	0.1	0.0	0.0	0.1	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0
[11]	Petroleum and coal products	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
[12]	Rubber products	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.1	0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
[13]	Ceramic, stone and clay products	0.1	0.1	0.0	0.0	0.0	0.3	0.4	0.2	0.0	0.1	0.1	0.1	1.0	0.0	0.0	0.0	0.0	0.0	0.0
[14]	Iron and steel	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.0	0.1	0.2	0.1	0.9	0.0	0.0	0.0	0.0	0.0	0.0
[15]	Non-ferrous metals and products	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.0	0.1	0.2	0.1	0.9	0.0	0.0	0.0	0.0	0.0	0.0
[16]	Fabricated metal products	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0
[17]	General-purpose machinery	1.5	1.4	0.4	0.2	0.1	1.1	1.8	0.5	0.1	0.5	0.4	0.5	1.7	0.0	0.0	0.0	0.0	0.0	0.0
[18]	Home electronics	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
[19]	Electrical machinery	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.6	0.0	0.3	1.0	0.4	0.7	0.0	0.1	0.0	0.0	0.0	0.0
[20]	Information and communication electronics	0.1	0.4	0.2	0.1	0.1	0.9	0.4	2.5	0.2	1.2	12.5	0.8	2.0	0.3	2.2	0.1	0.3	0.0	0.0
[21]	Motor vehicles, parts and accessories	0.0	0.1	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
[22]	Other transportation equipment	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
[23]	Precision instruments and machinery	0.7	3.7	2.4	0.9	1.7	2.9	1.2	1.5	0.3	0.6	1.9	0.7	0.7	0.0	0.1	0.1	0.0	0.0	0.0
[24]	Miscellaneous manufacturing	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
[25]	Electricity and gas	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Source: Calculations based on Van Looy et al. (2004) and Schmoch et al. (2004)

Chapter 4 Effects of Regional Human Capital on Business Entry: a Comparison of Independent Start-ups and New Subsidiaries in Different Industries[27]

1. Introduction

The start-up of new businesses increases innovation and competition and creates local employment. Therefore, start-up activities have been encouraged and supported by various programs in many countries. However, even in Japan, where the start-up ratio[28] has been lower than the closure ratio[29] since the late 1980s, considerable efforts aimed at increasing the entry of start-ups have hitherto not met with much success (Okamuro and Kobayashi, 2006).

Business start-ups are important for both national and regional economies. In order to comprehensively consider the impact of business start-ups on the regional economy, it is appropriate to distinguish between new business entries of independent start-ups and subsidiaries of existing firms. The former type depends basically on the decision of the people living or working in the region regarding setting up independent businesses, which means the regional structure of human capital is expected to play a significant role. The latter type is based on decisions by the top management of existing firms, which could be located outside the region, regarding where to locate new subsidiaries. In this case, the regional level of demand and cost may be more important than the regional human capital. Bosma et al. (2008) investigate differences in the regional determinants of independent start-ups and new subsidiaries, focusing on agglomeration effects and comparing manufacturing and service sectors.

The effects of regional human capital on entry may differ considerably across sectors and industries. Industries differ in their sensitivity to regional supply and demand (market)

---

[27] This chapter is based on Ikeuchi and Okamuro (2011), co-authored with Hiroyuki Okamuro. Also, this study was supported from the Japan Society for the Promotion of Science (JSPS) under Grant-in-Aid for Scientific Research (A) (No. 20243018) for this study.

[28] The start-up ratio research to the number of new businesses in a given time. as a proportion of the total business stock.

[29] The closure ratio refers to the number of closing businesses in a given time period as a proportion of the total business stock

conditions as well as in the required levels and types of human capital. However, few studies have examined inter-industry differences of entry, apart from some studies comparing the manufacturing and service sectors. Okamuro (2008) compares the regional determinants of start-ups in high-tech versus low-tech industries in the manufacturing sector, finding that the agglomeration of specialized human capital and knowledge is important. In addition, Acs and Armington (2006) examine the differences in the regional determinants of entry among various sectors (manufacturing, retail trade, local market, distribution and business services), focusing on educational requirements and market segments.

However, in their analysis of the regional determinants of entry, these studies do not differentiate between independent start-ups and new subsidiaries of existing firms. Within the same sector, regional factors may differ between the types of start-ups. As mentioned earlier, this chapter may assume that the decisions on independent start-ups are mainly based on regional human capital, while the location of new subsidiaries is determined by considerations of demand and cost factors. Moreover, regional factors of start-ups may vary across sectors and industries, depending on whether we focus on independent start-ups or new subsidiaries. For example, the location choice of new subsidiaries would not necessarily depend on local demand conditions in manufacturing industries with wide, possibly foreign, markets, while it would be influenced by the human capital in the region in the case of knowledge-intensive services.

The aim of this chapter is, therefore, to investigate the impact of regional human capital on the start-up ratio using Japanese data at the prefecture level, differentiating between independent start-ups and new subsidiaries of existing firms. Moreover, it will be compared the effects of regional human capital on entry across different industries and sectors. Since these issues have not been explicitly explored by existing studies, this chapter makes a major contribution to the literature by analyzing them. One policy implication that could be derived from this study is that regional policies to activate business start-ups should recognize the differences between encouraging local entrepreneurship and attracting new subsidiaries. In addition, these differences may vary even within the service sector according to technological intensity.

The remainder of this chapter is organized as follows. Section 2 reviews the previous literature on regional variations of the start-up ratio. On the basis of the literature review, Section 3 provides hypotheses for the empirical analysis. In Section 4, a research framework is presented to capture the determinants of regional differences in the entry of independent start-ups and new subsidiaries. Section 5 provides the estimation results and discusses them. Finally, Section 6 provides concluding remarks.

2. Literature Review

Determinants of regional entry have been investigated since the 1990s in several countries using various kinds of regional variables, including demand factors, cost factors, business agglomeration, labor force structure, industry structure and some other factors (see, for example, Okamuro and Kobayashi, 2005, for a detailed survey of the relevant literature).

Several studies demonstrate that the start-up ratio is higher in regions with a higher level and growth of regional demand measured by the size and growth of population or income (e.g., Audretsch and Fritsch, 1994a; Davidsson et al., 1994; Reynolds et al., 1995; Acs and Armington, 2004). It is also empirically established that the start-up ratio is negatively correlated with the level of factor costs, especially with wage level (Gerlach and Wagner, 1994; Santarelli and Piergiovanni, 1995; Audretsch and Vivarelli, 1996). Moreover, several studies indicate that the start-up ratio is positively affected by agglomeration, measured by population or business density (Audretsch and Firtsch, 1994a; Davidsson et al., 1994; Acs and Armington, 2004). With regard to the industry structure, previous studies concur that a smaller share of the manufacturing sector and a larger share of the service sector have positive effects on the start-up ratio (Evans and Leighton, 1989; Reynolds et al., 1995; Egeln et al., 1997).

With regard to labor force structure, numerous studies focus on the effects of the qualitative and quantitative composition of regional labor force as well as the impact of the employment situation on the start-up ratio. The qualitative composition denotes the endowment of highly educated or skilled labor force (this issue will be addressed in detail later), while the quantitative composition is mainly measured by the age structure of the labor force (Evans and Leighton, 1989; Reynolds et al., 1995; Egeln et al., 1997). As for the effect of the unemployment ratio on the start-up ratio, there are contrasting views (the push hypothesis suggesting a positive impact and the pull hypothesis suggesting a negative impact), both of which find empirical support[30].

The impact of the qualitative composition of the regional labor force with regard to education, job experience and technical skills has attracted considerable attention from the human capital perspective. Several studies demonstrate that the ratio of white-collar to blue-collar workers (Keeble and Walker, 1994; Fotopoulos and Spence, 1999) and the

---

[30] For example, Evans and Leighton (1990) and Storey (1991) find evidence for the push hypothesis, while Reynolds et al. (1995) and Carree (2002) support the pull hypothesis.

proportions of college graduates (Guesnier, 1994; Armington and Acs, 2002; Acs and Armington, 2004) and the workforce in professional and managerial occupations (Guesnier, 1994; Hart and Gudgin, 1994) have positive effects on the start-up ratio.

Such a regional labor force structure proxies the agglomeration of human capital in the regions. Following Becker (1975), the previous literature distinguishes between the generic and specific components of human capital. Generic human capital is related to the general knowledge acquired by founders through formal education and professional experience. Specific human capital comprises capabilities that founders can directly apply to the entrepreneurial jobs in new businesses and that can be obtained through prior work experience in the same industry (industry-specific human capital) or through managerial and self-employment experience (entrepreneur-specific human capital) (Colombo and Grilli, 2005). Therefore, this chapter may expect that regions that show a higher ratio of highly educated labor force with rich professional, managerial or specific work experience would be characterized by larger agglomeration of human capital.

Why and how does such regional human capital positively affect the start-up of new businesses? Acs and Armington (2004, 2006) indicate the following three reasons. First, the agglomeration of highly educated and skilled labor force generates entrepreneurs with new ideas for creating new businesses (Glaeser et al., 1992). Second, it also promotes local knowledge spillovers, by which new start-ups are initiated and sustained (Reynolds et al., 1995). Third, it facilitates the founders of new firms to search for and hire skilled labor (Rauch, 1993).

A recent trend of research on regional variations of the start-up ratio is to differentiate between and compare start-up types, such as high- versus low-tech (Okamuro, 2008) and independent businesses versus new subsidiaries (Bosma et al., 2008). On the basis of micro data of start-ups in the Japanese manufacturing sector, Okamuro (2008) shows that regions characterized by agglomerations of highly educated and specialized human capital, as well as research institutes and high-tech industries, attract high-tech start-ups (those in high-tech industries), while a high unemployment ratio would draw only low-tech start-ups (push hypothesis).

Bosma et al. (2008) is practically the first study that directly addresses and empirically investigates the determinants of location choice of subsidiaries in comparison to independent start-ups. Using a Dutch regional database, they find that localization economies positively affect independent business start-ups, while urbanization economies stimulate the entry of new subsidiaries. However, they do not explicitly consider the effects of regional human capital, although agglomeration economies also include general benefits such as access to a highly qualified labor pool.

Bosma et al. (2008) argue that the incentives for establishing a firm in a particular region are essentially different for independent start-ups and new subsidiaries. A founder of an independent firm will decide whether or not to start a business, comparing the expected utility of a business start-up with that of remaining an employee. In contrast, founders of new subsidiaries of established firms (especially of those in other regions) are often recruited from the core employees or managers of these firms working in other regions. Thus, established firms choose the best location for the new subsidiaries, considering several regional characteristics such as demand and cost factors. Thus, we expect that, among various regional factors, the regional level of human capital affects the entry of independent start-ups more strongly than that of new subsidiaries. In this sense, it is important to compare the effect of regional human capital between independent start-ups and new subsidiaries.

Moreover, policy measures to stimulate start-ups by inhabitants and to attract new subsidiaries, especially of firms located outside the region, may be quite different. Therefore, in the current study, different impacts of human capital on the entry of independent businesses and subsidiaries will be explored, founded on the basic models of Bosma et al. (2008) along with the concepts of Okamuro (2008).

Several previous studies compare the factors of regional entry in the manufacturing and service sectors (Audretsch and Fritsch, 1994a; Hart and Gudgin, 1994; Keeble and Walker, 1994; Audretsch and Vivarelli, 1996; Bosma et al., 2008) and in high- versus low-tech industries in the manufacturing sector (Nerlinger 1998; Okamuro 2008). Inter-industry comparison within the manufacturing or service sectors has not been conducted, except by Armington and Acs (2002) and Acs and Armington (2004), who performed sub-sample analyses in the service sector. In contrast to these studies, this research not only compares manufacturing and service sectors but also distinguishes between relatively high- and low-tech industries in the service sector. Thus, another contribution of this chapter is to compare different industries in the service sector, differentiating between independent start-ups and subsidiaries.

3. Hypotheses

Regional human capital might have different impacts on independent start-ups and new subsidiaries for the following reasons. On the one hand, the regional structure of human capital is expected to play a significant role for independent start-ups because they depend on the decisions of people living or working in the region[31]. On the other hand, location choices of new subsidiaries are based on the decisions by the top management of the existing firms, which could be located outside the region. In this case, the regional level of demand and cost may be more important than the regional human capital, because the heads of new subsidiaries often come from other regions, especially the headquarters.

As the measures of regional human capital, it is focused on the ratio of college graduates and the ratio of the workers in professional and technical occupations to the labor force. According to the above discussion on human capital based on Becker (1975), both measures relate to generic human capital.

With regard to the ratio of college graduates, previous studies such as Colombo and Grilli (2005) argue that highly educated workers are likely to be more capable ('capability theory') and earn higher incomes ('wealth effect') than others. They may have good ideas and projects for their new businesses and thus expect a high return from independent start-ups, on the one hand, and may be financially less constrained, on the other. Hence, they are more likely to start new businesses than others. However, previous studies for Japan demonstrate that highly educated persons are more reluctant to start their own businesses than others (e.g. Small and Medium Enterprise Agency, 2002), because, especially under the traditional Japanese employment system, they are aware of the high opportunity cost of quitting the current job to start a business.

Thus, the effect of high education on the regional entry of independent start-ups can be either positive or negative. Which of these contrasting effects is stronger than the other at the regional level is therefore an empirical question. Consequently, the following hypotheses with regard to the entry of independent start-ups are proposed.

Hypothesis 1a:

The agglomeration of college graduates *at the prefecture level has a positive impact on the entry rate of independent start-ups in the prefecture.*

Hypothesis 1b:

[31] Prior research shows that most founders start new businesses in their own region (Figueiredo et al., 2002; Stam, 2007).

The agglomeration of college graduates *at the prefecture level has a negative impact on the entry rate of independent start-ups in the prefecture.*

The top managers of the subsidiaries of existing firms are not necessarily recruited from among the inhabitants of the region where the subsidiaries are located. If the firms are located outside the region, the top managers of the subsidiaries are also often appointed from outside the region. Hence, in this regard, the availability of highly educated workers in the region is not expected to matter much for the entry of subsidiaries.

However, in the case of local firms, the top managers of new subsidiaries are usually recruited from among local workers or managers. Moreover, even firms located outside the region recruit (at least partly) local workers as the employees for their subsidiaries. In these cases, regions that can provide many highly educated and skilled workers would attract the entry of subsidiaries[32]. Therefore, a positive impact of college graduates is expected also on the entry of subsidiaries, which is formulated in the following hypothesis.

Hypothesis 2:

The agglomeration of college graduates *at the prefecture level has a positive impact on the entry rate of subsidiaries in the prefecture.*

According to the Japan Standard Occupation Classification, professional and technical occupations include various types of scientists and engineers, medical and health-care services, social welfare services, legal services and business support services (see footnote 33). People engaged in these occupations may not only have the potential to start new businesses (especially high-tech ventures or specialized service firms) by themselves, but also can be employed in new high-tech ventures or service firms. Moreover, they (especially those in legal and business support services) can also provide founders of new businesses with professional support. Thus, it is postulated that the agglomeration of the workers in professional and technical occupations positively affects the entry of independent start-ups.

With regard to subsidiaries, even if professional workers such as engineers are required, they can be recruited from other regions or brought from the headquarters or other subsidiaries.

[32] In this respect, location decisions on subsidiaries are similar to those on foreign plants by multinational companies. Coughlin and Segev (2000), for example, find positive effects of educational attainment on the location choice of multinational companies.

Moreover, unlike the de novo start-up of independent firms, the entry of new subsidiaries may not require local professional support. Therefore, any positive relationship between the regional structure of professional and technical workers and the entry of new subsidiaries is not expected.

Hypothesis 3:

The agglomeration of professional and technical workers *at the prefecture level has a positive impact only on the entry rate of independent start-ups in the prefecture.*

This chapter tests these hypotheses not only with a sample of all industries but also with sub-samples of manufacturing and service sectors. These sectors may differ in their sensitivity to regional supply and demand (market) conditions as well as in the required levels and types of human capital. Also it is examined whether or not the effects of regional human capital are different between high- and low-tech industries in the service sector. In high-tech (research-intensive) industries such as the information and communication industry, firms generally face a rapid technological development. To survive technological competition, entrepreneurs in high-tech industries require highly educated and skilled workers to a larger extent than those in low-tech industries.

4. Empirical Model and Data

In this chapter, the impact of various regional factors on the entry rate of independent start-ups and new subsidiaries is estimated for each industry sector in the sample. Relying on Bosma et al. (2008), the seemingly unrelated regression (SUR) method is employed, which assumes correlation between the error terms of two regression models, because variables affecting the entries of both independent businesses and subsidiaries might be omitted. By the SUR estimation procedure, regression models for both types of entries are simultaneously estimated, and asymptotically more efficient estimators (i.e. more efficient than the OLS estimator) can be obtained (Zellner, 1962, 1963). Moreover, as mentioned above, the same models for each industry sector are estimated in the sample and the results are compared.

Following Bosma et al. (2008), the following model is estimated:

$$\begin{cases} \ln SR^{Ind} = \alpha_0^{Ind} + \alpha_1^{Ind} H + x'\gamma^{Ind} + e^{Ind} \\ \ln SR^{Sub} = \alpha_0^{Sub} + \alpha_1^{Sub} H + x'\gamma^{Sub} + e^{Sub} \\ cor\left(e^{Ind}, e^{Sub}\right) = \rho \end{cases} \tag{4-1}$$

The dependent variables are the entry rates of independent establishment (SR^{Ind}) and of new subsidiaries (SR^{Sub}) in natural logarithms. Following Bosma et al. (2008), the variables of the workforce and the stock of existing establishments are used to measure and control for the effect of economic size in the regions. In other words, the ‘labor market approach’ is applied to independent start-ups and the ‘ecological approach’ to new subsidiaries (*cf.*, Audretsch and Fritsch, 1994b).

As the main subject of this chapter, the effects of regional human capital (*H*) on the entry rate of independent start-ups and new subsidiaries are examined. As the variables for regional human capital, the ratio of highly educated workforce (the ratio of college graduates) and the ratio of the workforce in professional and technical occupations are used. The other determinants of entry ($\mathbf{x}$) comprise demand and supply factors (population growth rate, wage rate and unemployment rate) and a measure of agglomeration economies.

4.1. Regional Entry in Japan

Pooled regional data at the prefecture level from four periods (1996–1999, 1999–2001, 2001–2004 and 2004–2006) is used. With 47 prefectures in Japan, at most 188 observations are available in the pooled sample. In general, a prefecture in Japan is, on average, smaller than a state in the US but larger than a county or city. Thus, it may be too large an area to represent the local (labor) market. However, appropriate data for a narrower regional classification level (i.e. of the ‘municipality’) cannot be obtained, which is the main reason for the use of the prefecture-level data.

Table 4.1 shows the definitions and the descriptive statistics of the variables used for regressions. Regional start-up data are obtained from the e-Stat Database of the Establishment and Enterprise Census. The number of regional independent start-ups in Japan from 1996 to 2006 is, on average, 4600 per prefecture, annually, which is more than the number of new subsidiaries (2400 on average). These numbers vary among regions significantly; the maximum number of regional start-ups is more than 60 times the minimum.

Table 4.1 Definitions of the variables and sample statistics

	No. of. obs.	Mean	S.D.	Min.	Max.
N^{Ind}= Number of independent start-ups per year (per 1000 establishments)					
Overall industry	188	4.60	5.73	0.63	44.75
Manufacturing sector	188	0.28	0.39	0.03	2.97
Service sector	188	3.85	4.89	0.51	38.38
Information & communication	94	0.08	0.30	0.01	2.61
Commerce & restaurants	94	1.81	1.98	0.27	13.42
N^{sub}= Number of new subsidiaries per year (per 1000 establishments)					
Overall industry	188	2.40	2.97	0.34	20.73
Manufacturing sector	188	0.11	0.15	0.01	0.99
Service sector	188	2.09	2.61	0.30	18.62
Information & communication	94	0.07	0.13	0.01	1.00
Commerce & restaurants	94	1.17	1.44	0.19	9.67
WF= Work force (1000 workforce)	188	1161.54	1347.17	228.67	8416.06
ES = Number of existing establishments (per 1000 establishments)					
Overall industry	188	130.41	126.09	27.91	759.21
Manufacturing sector	188	14.26	16.54	1.91	97.46
Service sector	188	100.98	99.83	22.28	595.23
Information & communication	94	1.22	2.75	0.18	18.83
Commerce & restaurants	94	54.27	51.34	12.37	299.27
SR^{Ind} = Entry rate of independent start-ups = 1000× N^{Ind}/WF					
Overall industry	188	4.16	1.87	2.00	14.16
Manufacturing sector	188	0.24	0.13	0.07	0.62
Service sector	188	3.45	1.59	1.66	12.72
Information & communication	94	0.04	0.04	0.01	0.34
Commerce & restaurants	94	1.75	0.84	0.88	7.29

Table 4.1 (continued)

	No. of. obs.	Mean	S.D.	Min.	Max.
SR^{sub} = Entry rate of new subsidiaries = 100× N^{sub}/ES					
Overall industry	188	1.70	0.68	0.58	3.45
Manufacturing sector	188	0.79	0.40	0.18	1.97
Service sector	188	1.92	0.75	0.64	3.83
Information & communication	94	5.93	2.52	1.99	13.69
Commerce & restaurants	94	1.95	0.72	0.82	3.79
College Grad = 100 ×college graduates/ workforce (in 2000)	188	12.23	3.74	7.18	24.19
Expert = 100×number of workers in professional and technical occupations/workforce	188	12.79	1.38	10.10	16.97
Wage = Wage rate (100 yen per hour)	188	2.06	0.27	1.55	2.93
PopGrowth = % growth between (t-4) and (t-1)	188	−0.05	1.07	−2.66	2.80
Unemp = Unemployment rate (%)	188	4.74	1.30	2.52	11.40
Localization = 1000×number of existing establishments/regional					
Overall industry	188	49.70	6.52	30.80	64.93
Manufacturing sector	188	5.15	2.16	2.07	12.09
Service sector	188	38.26	4.74	24.77	50.41
Information & communication	94	0.34	0.20	0.13	1.55
Commerce & restaurants	94	20.91	2.85	12.73	26.88

To control for the effects of regional economic size, the regional workforce and stock of establishments is used, obtained from the Establishment and Enterprise Census*,* as proxies for regional economic size. As shown in Table 4.1, the entry rate of independent start-ups is, on average, 4.16 per 1000 workers and ranges from 2.00 to 14.16 across prefectures, while that of new subsidiaries is, on average, 1.7% of the existing establishments and ranges from 0.58% to 3.45% across prefectures. Thus, although the entry rate in Japan is at the lowest level among OECD countries in recent years, the entry rates of both independent start-ups and new subsidiaries are significantly different among regions in Japan.

In addition, the ratio of the number of new subsidiaries to the total number of entries at the national level increased from 32.6% in the period 1996–1999 to 37.5% in the period 2004–2006. This ratio and its trend are almost the same as those in the Netherlands (Bosma et al. 2008). An increase in this ratio can also be observed at the prefecture level. It ranges from 15.3% to 41.5%

across prefectures in the period 1996–1999 and from 21.7% to 45.1% in the period 2004–2006.

The number of entries and the rate of entry also differ between industries. Table 4.1 shows the industrial composition of regional independent start-ups and new subsidiaries. The entry rates of both independent start-ups and new subsidiaries are higher in the service sector than in manufacturing. Within the service sector, they are relatively lower in the information and communication industry, compared to commercial establishments and the restaurants industry.

4.2. Independent Variables

In order to test the hypotheses on the relationship between regional human capital and regional new business start-ups, as mentioned in the previous section, the proportions of college graduates (*CollegeGrad*) and workers in professional and technical occupations[33] (*Expert*) to the entire workforce in each prefecture is used; these were obtained from the Population Census.[34] As shown in Table 6.1, the mean values of both *CollegeGrad* and *Expert* are approximately 12%–13%, while the regional variations of these variables are different. The proportion of college graduates ranges from 7.2% to 24.2% across regions (standard deviation is 3.7), while that of expert workers ranges from 10.1% to 17.0% (standard deviation is 1.4).

Several control variables are also included in the estimation models as additional determinants of regional start-up rate. The definitions and descriptive statistics of these variables are summarized in Table 4.1. Following Bosma et al. (2008), we included in the estimation models the population growth rate (*PopGrowth*), the natural logarithm of average wage rate (*Wage*) and the unemployment rate (*Unemp*) as demand and supply factors for regional entrepreneurship[35], and the ratio of establishment stock (*ES*) of each industry to the population as a measure of 'localization economy' (*Localization*).

---

[33] According to the Standard Occupation Classification of Japan, 'professional and technical occupations' include various types of scientists and engineers; medical and health-care services, such as doctors, pharmacists, and nurses; social welfare services; legal services, such as lawyers; business support services, such as accountants and management consultants; and teachers and artists.

[34] We use the ratios not but level of the regional human capital variables (the number of college graduates and professional and technical workers) since the dependent variables (start-up ratio) we used are not in level but ratio.

[35] The population growth rate and the unemployment rate are calculated from the *Population Census*, and the average wage rate from the *Basic Survey on Wage Structure* (*Wage Census*), at the prefecture level.

Also, it is considered that the population density of each prefecture should be included as a proxy for *urbanization economy* as another agglomeration factor, following Bosma et al. (2008). However, since the correlation between population density and *CollegeGrad* is very high (0.721), this proxy for urbanization economy is not included in the estimation models in order to avoid multicollinearity.

It is expected that the coefficients of the variables *PopGrowth* and *Localization* will be positive. For the independent start-up rate, the coefficient of the variable *Unemp* is expected to be positive according to the 'push hypothesis' and negative according to the 'pull hypothesis', while it is expected to be negative (also according to 'pull hypothesis') or insignificant for new subsidiaries.

The effect of the variable *Wage* on the entry rate of subsidiaries would be negative, since a higher wage rate implies a higher cost to hire employees for new subsidiaries from the local labor market. Further, the effect of the wage rate on the independent start-up rate would be negative since the regional wage rate should directly reflect a potential entrepreneur's opportunity cost of quitting the current job to start a new business.

However, the regional wage rate is highly correlated with the college graduates ratio (*CollegeGrad*) in the data set (see Table 4.2). Since the productivity of highly educated workers would be relatively high and the wage rate offered for such productive workers would also be high, it would not be able to distinguish the opportunity cost effects of new business start-ups for highly educated workers from the effects of wage rate. For this reason, the variable *Wage* is included only in the equation for the entry rate of new subsidiaries (SR^{Sub}) as a basic specification. Then the robustness of the estimation results of the basic models is checked by excluding *Wage* from the equation SR^{Sub}.

Finally, dummy variables for the periods 1999—2001, 2001—2004 and 2004—2006 (period dummies: the baseline reference is the period 1996-1999) are included in the estimation models in order to control for the time-variant differences of start-up ratio in each prefecture.

Table 4.2 Correlation coefficients of the variables

		InSR^Ind (1)	(2)	(3)	(4)	(5)
InSR^{Ind}						
Overall industry	(1)	1.000				
Manufacturing sector	(2)	0.748	1.000			
Service sector	(3)	0.996	0.699	1.000		
Information & communication	(4)	0.685	0.478	0.697	1.000	
Commerce & restaurants	(5)	0.969	0.592	0.978	0.604	1.000
InSR^{sub}						
Overall industry	(6)	0.716	0.568	0.708	0.675	0.642
Manufacturing sector	(7)	0.688	0.451	0.683	0.580	0.608
Service sector	(8)	0.682	0.590	0.669	0.650	0.619
Information & communication	(9)	0.782	0.596	0.770	0.531	0.696
Commerce & restaurants	(10)	0.665	0.500	0.659	0.645	0.536
CollegeGrad	(a)	−0.073	0.107	−0.063	0.327	−0.205
Expert	(b)	0.217	0.051	0.252	0.449	0.148
In *(wage)*	(c)	−0.208	0.124	−0.214	0.122	−0.448
PopGrowth	(d)	−0.123	0.021	−0.116	0.161	−0.313
Unemp	(e)	0.478	0.130	0.531	0.520	0.746
Localization (of own industry)	(f)	−0.120	0.335	0.027	0.562	0.006

		InSR^Sub (6)	(7)	(8)	(9)	(10)
InSR^{Sub}						
Overall industry	(6)	1.000				
Manufacturing sector	(7)	0.886	1.000			
Service sector	(8)	0.993	0.851	1.000		
Information & communication	(9)	0.875	0.726	0.874	1.000	
Commerce & restaurants	(10)	0.976	0.826	0.975	0.817	1.000

Table 4.2 (continued)

		lnSR^{Ind}				
		(6)	(7)	(8)	(9)	(10)
CollegeGrad	(a)	0.222	0.028	0.232	0.025	0.313
Expert	(b)	0.288	0.175	0.259	0.082	0.237
In *(wage)*	(c)	0.129	−0.026	0.155	−0.143	0.133
PopGrowth	(d)	0.008	−0.032	0.008	−0.171	0.104
Unemp	(e)	0.316	0.296	0.266	0.441	0.348
Localization (of own industry)	(f)	−0.454	−0.474	−0.410	−0.175	−0.585
		(a)	(b)	(c)	(d)	(e)
CollegeGrad	(a)	1.000				
Expert	(b)	0.699	1.000			
In *(wage)*	(c)	0.842	0.412	1.000		
PopGrowth	(d)	0.655	0.313	0.622	1.000	
Unemp	(e)	0.083	0.434	−0.108	−0.006	1.000
Localization	(f)					
Overall industry		−0.322	−0.384	−0.155	−0.139	−0.267
Manufacturing sector		0.165	--0.287	0.384	0.202	−0.441
Service sector		−0.356	0.224	−0.285	−0.192	−0.001
Information & communication		0.433	0.373	0.414	0.411	0.049
Commerce & restaurants		−0.480	−0.271	−0.361	−0.303	−0.009

5. Estimation Results

SUR estimation results of the overall industry (excluding the primary sector) are shown in Table 4.3 and Table 4.4. For each specification, the results of the equations for the entry rates of independent establishments and subsidiaries are shown in the first column (lnSR^{Ind}) and second column (lnSR^{Sub}) respectively, and the scores of the variation inflation factor (VIF) for each independent variable are also shown in the third column to check if a multicollinearity problem occurs[36].

[36] The VIF is a measure of the degree of multicollinearity of each independent variable in regression analysis. A common rule of thumb is the VIF of larger than 5 (or 10) as a sign of severe multicollinearity. However, some problems of this rule are also pointed out (O'Brien, 2007).

Table 4.3 SUR estimation results for overall industry

Specification	I(a)			I(b)		
Industry	Overall industry			Overall industry		
Dependent variable	InSR^{Ind}	InSR^{Sub}	VIF	InSR^{Ind}	InSR^{Sub}	VIF
Constant	−6.891***	−4.6***		−6.891***	−4.787***	
	[0.129]	[0.197]		[0.129]	[0.132]	
CollegeGrad	−0.039***	0.036***	8.3	−0.039***	0.026***	2.1
	[0.005]	[0.009]		[0.005]	[0.005]	
Expert	0.122***	–0.015	3.1	0.122***	−0.005	
	[0.013]	[0.016]		[0.013]	[0.014]	
In(*Wage Rate*)		−0.271	4.9			
		[0.212]				
Peroid dummy	Yes	Yes	1.3	Yes	Yes	1.2
N	188	188		188	188	
R squared	0.841	0.832		0.841	0.831	
Correlation between residuals of both equations:		0.016			0.046	
Breusch–Pagan test of independence:		0.0			0.4	

Specification	I(c)			I(d)		
Industry	Overall industry			Overall industry		
Dependent variable	InSR^{Ind}	InSR^{Sub}	VIF	InSR^{Ind}	InSR^{Sub}	VIF
Constant	−5.79***	−4.761***		−6.388***	−5.201***	
	[0.056]	[0.08]		[0.133]	[0.123]	
CollegeGrad	−0.008**	0.03***	3.6			
	[0.004]	[0.006]				
Expert				0.041***	0.003***	1.3
				[0.011]	[0.011]	
In(*Wage Rate*)		−0.196	3.6		0.426***	1.3
		[0.183]			[0.112]	
Period dummy	Yes	Yes	1.1	Yes	Yes	1.1

Table 4.3 (continued)

Specification	I(c)			I(d)		
Industry	Overall industry			Overall industry		
Dependent variable	InSR^{Ind}	InSR^{Sub}	VIF	InSR^{Ind}	InSR^{Sub}	VIF
N	188	188		188	188	
R squared	0.768	0.832		0.780	0.819	
Correlation between residuals of both equations:	−0.018			0.003		
Breusch–Pagan test of indepen-dence:	0.1			0.0		

Specification	I(e)			I(f)		
Industry	Overall industry			Overall industry		
Dependent variable	InSR^{Ind}	InSR^{Sub}	VIF	InSR^{Ind}	InSR^{Sub}	VIF
Constant	−5.791***	−4.83***		−6.388***	−5.116***	
	[0.056]	[0.047]		[0.133]	[0.125]	
CollegeGrad	−0.008**	0.025***	1.0			
	[0.004]	[0.003]				
Expert				0.041***	0.048***	
				[0.011]	[0.01]	1.1
Peroid dummy	Yes	Yes	1.0	Yes	Yes	1.1
N	188	188		188	188	
R squared	0.768	0.831		0.780	0.804	
Correlation between residuals of both equations:	0.023			−0.158		
Breusch–Pagan test of indepen-dence:	0.1			4.7**		

Notes:
Standard errors are shown in brackets.
*** $p < .01$, ** $p < .05$.
Sample periods: 1996–1999, 1999–2001, 2001–2004 and 2004–2006.

Table 4.4 SUR estimation results for overall industry (with all control variables)

Specification	I(g)			I(h)		
Industry	Overall industry		VIF	Overall industry		VIF
Dependent variable	InSR^{Ind}	InSR^{Sub}		InSR^{Ind}	InSR^{Sub}	
Constant	−6.984***	−3.825***		−6.984***	−3.901***	
	[0.121]	[0.204]		[0.121]	[0.172]	
CollegeGrad	−0.026***	0.02**		−0.026***	0.015**	
	[0.004]	[0.009]	9.8	[0.004]	[0.006]	4.0
Expert	0.56***	−0.024		0.056***	−0.019	
	[0.01]	[0.015]	3.7	[0.01]	[0.014]	3.1
In(*Wage Rate*)		−0.135				
		[0.192]	5.1			
PopGrowth	0.013	0.038**		0.013	0.037**	
	[0.011]	[0.016]	2.5	[0.011]	[0.016]	2.4
Unemp	0.12***	0.015		0.12***	0.016	
	[0.008]	[0.011]	1.7	[0.008]	[0.011]	1.7
Localization	0.004***	−0.012***		0.004***	−0.012***	
	[0.001]	[0.002]	1.4	[0.001]	[0.002]	1.3
Period dummy	Yes	Yes	2.1	Yes	Yes	1.9
N	188	188		188	188	
R squared	0.937	0.872		0.937	0.872	
Correlation between residuals of both equations:		−0.067			−0.044	
Breusch–Pagan test of independence:		0.9			0.4	

Notes:
Standard errors are shown in brackets.
*** $p < .01$, ** $p < .05$.
Sample periods: 1996–1999, 1999–2001, 2001–2004 and 2004–2006.

In specification I (a), in which the natural logarithm of the wage rate is included in the equation for the entry rate of subsidiaries, the VIF score of for *CollegeGrad* is 8.3. This indicates that the correlation between *CollegeGrad* and the other independent variable may cause multicollinearity. Thus, for a robustness check, other specifications are estimated in which the variables highly correlated with *CollegeGrad* are excluded. In specification I (b), I (e) and I (f), the wage rate is excluded. In specification I(c), I (d), I (e) and I (f), *CollegeGrad* and *Expert* are included interchangeably. Furthermore, results do not change even after including the other control variables in Table 4.4. The difference between specification I (g) and I (h) arises from the inclusion or exclusion of the wage rate.

For all specifications, the coefficient of *CollegeGrad* for independent start-ups is negative and significant at least at the 5% level; the coefficients of *Expert* for independent start-ups and *CollegeGrad* for new subsidiaries are positive and significant at the 1% level. These results support Hypothesis 1a, Hypothesis 2 and Hypothesis 3. According to these results, an agglomeration of a highly educated workforce attracts new subsidiaries, on the one hand, but decreases the independent start-up rate, on the other hand, while an agglomeration of workers in professional and technical occupations promotes regional entrepreneurship.

With regard to the effect of control variables, an increase in the wage level has an overall negative but not significant effect while the population growth and unemployment rate have positive impacts on both types of start-up rates. However, the effect of population growth is significant only for the entry rate of subsidiaries, while the effect of the unemployment rate is significant only for the ratio of independent start-ups. Similar to the results of Bosma et al. (2008), a significant and positive impact of localization economies only on independent start-ups is found.

5.1. Manufacturing and Service Industries

Table 4.5 and Table 4.6 show the estimation results for the manufacturing industry and the service industry respectively. There are some differences in the determinants of entry between the manufacturing and service sectors. In the manufacturing industry, the proportion of college graduates (*CollegeGrad*) positively affects independent start-ups, while this human capital has no significant effects on new subsidiaries; the coefficient of *Expert* for independent start-ups is negative and significant. These results support Hypothesis 1a but do not support Hypothesis 2 and Hypothesis 3. This implies that the agglomeration of highly educated workforce (rather than the professional and technical workforce) promotes regional entrepreneurship in the manufacturing sector, but regional human capital, contrary to natural expectations, do not influence the decision on the location of new subsidiaries.

In the service industry, the results are the same as those for the industry as a whole. The proportion of college graduates (*CollegeGrad*) has a negative and significant effect on the entry rate of independent establishments and a positive and significant effect on new subsidiaries at the 1% level. The proportion of professional and technical workers (*Expert*) has a positive and significant effect on the independent start-up rate at the 1% level. These results are consistent with Hypothesis 1b, Hypothesis 2 and Hypothesis 3. It implies that in the service sector an agglomeration of highly educated workforce attracts new subsidiaries, while workers in professional and technical occupations promote regional entrepreneurship.

Table 4.5 SUR estimation results for the manufacturing and service sectors

Specification	II(a)					II(b)				
Industry	Manufacturing		Service		VIF	Manufacturing		Service		VIF
Dependent variable	InSR^{Ind}	InSR^{Sub}	InSR^{Ind}	InSR^{Sub}		InSR^{Ind}	InSR^{Sub}	InSR^{Ind}	InSR^{Sub}	
Equation	(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)	
Constant	−8.209***	−5.611***	−7.257***	−4.4***		−8.209***	−5.656***	−7.257***	−4.44***	
	[0.226]	[0.335]	[0.139]	[0.173]		[0.226]	[0.248]	[0.139]	[0.123]	
	0.039***	0.004	−0.043***	0.034***		0.039***	0.001	−0.043***	0.032***	
CollegeGrad	[0.008]	[0.015]	[0.005]	[0.008]	8.3	[0.008]	[0.009]	[0.005]	[0.005]	2.1
	−0.092***	0.008	0.142***	−0.03**		−0.092***	0.011	0.142***	−0.028**	
Expert	[0.023]	[0.028]	[0.014]	[0.014]	3.1	[0.023]	[0.026]	[0.014]	[0.013]	2.3
		−0.065		−0.058						
In(*Wage Rate*)		[0.325]		[0.176]	4.9					
Period dummy	Yes	Yes	Yes	Yes	1.3	Yes	Yes	Yes	Yes	1.2
N	188	188	188	188		188	188	188	188	
R squared	0.693	0.654	0.813	0.848		0.693	0.653	0.813	0.848	
Correlation of the residuals with the equation:										
(i)	1.000	−0.503	0.331	−0.448		1.000	−0.502	0.331	−0.447	
(ii)	−0.503	1.000	0.129	0.648		−0.502	1.000	0.133	0.649	
(iii)	0.331	0.129	1.000	−0.006		0.331	0.133	1.000	0.001	
(iv)	−0.448	0.648	−0.006	1.000		−0.447	0.649	0.001	1.000	
Breusch–Pagan test of independence:		187.9***					188.1***			

Notes:
Standard errors are shown in brackets.
*** $p < .01$, ** $p < .05$.
Sample periods: 1996–1999, 1999–2001, 2001–2004 and 2004–2006.

Table 4.6 SUR estimation results for the manufacturing and service sectors with additional control variables

Specification	II(a)					II(b)				
Industry	Manufacturing		Service			Manufacturing		Service		
Dependent variable	$\text{In}SR^{Ind}$	$\text{In}SR^{Sub}$	$\text{In}SR^{Ind}$	$\text{In}SR^{Sub}$	VIF	$\text{In}SR^{Ind}$	$\text{In}SR^{Sub}$	$\text{In}SR^{Ind}$	$\text{In}SR^{Sub}$	VIF
Equation	(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)	
Constant	−8.125**	−5.319***	−7.04***	−4.296***		−8.125***	−5.504***	−7.04***	−4.399***	
	[0.226]	[0.319]	[0.085]	[0.171]		[0.226]	[0.239]	[0.085]	[0.123]	
	0.043***	−0.014	−0.029***	0.028***		0.043***	−0.024**	−0.029***	0.022***	
CollegeGrad	[0.011]	[0.016]	[0.004]	[0.009]	9.5	[0.011]	[0.012]	[0.004]	[0.006]	4.0
	−0.121***	0.003	0.063***	−0.028*		−0.121***	0.012	0.063***	−0.023	
Expert	[0.027]	[0.03]	[0.01]	[0.016]	3.7	[0.027]	[0.028]	[0.01]	[0.015]	3.1
		−0.269		−0.15						
In(*Wage Rate*)		[0.308]		[0.173]	5.0					
	0.008	0.131***	0.008	0.046***		0.008	0.129***	0.008	0.045***	
PopGrowth	[0.03]	[0.032]	[0.011]	[0.017]	2.4	[0.03]	[0.032]	[0.011]	[0.016]	2.4
	0.05**	0.022	0.135***	−0.001		0.05**	0.024	0.135***	0.000	
Unemp	[0.021]	[0.022]	[0.008]	[0.012]	1.7	[0.021]	[0.022]	[0.008]	[0.011]	1.7
Period dummy	Yes	Yes	Yes	Yes	1.9	Yes	Yes	Yes	Yes	1.7
N	188	188	188	188		188	188	188	188	
R squared	0.703	0.692	0.933	0.855		0.703	0.689	0.933	0.854	
Correlation of the residuals with the equation:										
(i)	1.000	−0.577	0.311	−0.480		1.000	−0.575	0.311	−0.479	
(ii)	−0.577	1.000	−0.015	0.628		−0.575	1.000	0.005	0.631	
(iii)	0.311	−0.015	1.000	−0.105		0.311	0.005	1.000	−0.084	
(iv)	−0.480	0.628	−0.105	1.000		−0.479	0.631	-0.084	1.000	
Breusch–Pagan test of independence:	200.4***					199.6***				

Notes:
Standard errors are shown in brackets.
*** $p < .01$, ** $p < .05$
Sample periods: 2001–2004 and 2004–2006, *$p < .10$.

5.2. High- and Low-Tech Service Industries

To check the robustness of the results for the service sector, it is focused on two sub-sectors with regard to technological intensity: information and communications and the commerce and restaurant subsectors. The R&D intensity of the information and communication industry is the highest (0.74%) in the service sector[37], according to the Input-Output Tables of 2005. In contrast, the R&D intensity of the commerce and restaurant industry is only 0.22%. Thus, the information and communication industry is regarded as a high-tech industry and the results of this industry are compared to the results of the commerce and restaurant industry[38].

Table 4.7, Table 4.8 and Table 4.9 shows the estimation results for these two industries in the service sector[39]. In both subsectors, the effects of the proportion of professional and technical workers (*Expert*) on independent start-ups are positive and significant at the 1% level. These results are consistent with Hypothesis 3. Human capital has a different effect on start-ups both in a high- and low-tech service: The coefficient of *CollegeGrad* is significantly negative for independent start-ups but significantly positive for new subsidiaries in the commerce and restaurant industry; while the coefficients of *CollegeGrad* are not significant for both types of start-ups in the information and communication industry. Thus, Hypothesis 1b and Hypothesis 2 are supported only in a relatively low-tech (commerce and restaurant) industry.

[37] The R&D intensity of a certain industry is defined as the ratio of its R&D expenditure to its total output.

[38] Other service industries include various industries with different levels of technology intensity, such as research institutes, postal service, medical service, education, social work, advertising, machine maintenance, amusement, barbers and laundries. Because of data limitations, it is not able to divide them in further detail. For that reason, this industry is excluded from detailed analysis to test Hypotheses 2a and 2b.

[39] Because of limitations of data, this analysis is restricted to two observation periods, 2001–2004 and 2004–2006.

Table 4.7 SUR estimation results for high- and low- tech service industries

Specification	II(a)					II(b)				
Industry	Info & communication		Commerce & restaurants			Info & communication		Commerce & restaurants		
Dependent variable	$\ln SR^{Ind}$	$\ln SR^{Sub}$	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF	$\ln SR^{Ind}$	$\ln SR^{Sub}$	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF
Equation	(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)	
Constant	−12.829***	−3.292***	−8.047***	−4.237***		−12.829***	−3.424***	−8.047***	−4.463***	
	[0.461]	[0.388]	[0.244]	[0.313]		[0.461]	[0.259]	[0.244]	[0.208]	
CollegeGard	0.007	0.007	−0.063***	0.048***	8.3	0.007	0.000	−0.063***	0.036***	2.1
	[0.016]	[0.018]	[0.008]	[0.014]		[0.016]	[0.009]	[0.008]	[0.007]	
Expert	0.161***	0.005	0.162***	−0.033	3.1	0.161***	0.012	0.162***	−0.021	2.3
	[0.044]	[0.029]	[0.024]	[0.023]		[0.044]	[0.025]	[0.024]	[0.02]	
In *(WageRate)*		−0.179		−0.307	4.9					
		[0.392]		[0.316]						
Period dummy	Yes	Yes	Yes	Yes	1.3	Yes	Yes	Yes	Yes	1.2

Specification	II(a)					II(b)				
Industry	Info & communication		Commerce & restaurants			Info & communication		Commerce & restaurants		
Dependent variable	$\ln SR^{Ind}$	$\ln SR^{Sub}$	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF	$\ln SR^{Ind}$	$\ln SR^{Sub}$	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF
Equation	(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)	
N	94	94	94	94		94	94	94	94	
R squared	0.535	0.739	0.724	0.776		0.535	0.737	0.724	0.775	
Correlation of residuals with the equation:										
(i)	1.000	0.003	0.369	0.200		1.000	0.003	0.369	0.197	
(ii)	0.003	1.000	0.219	0.432		0.003	1.000	0.231	0.439	
(iii)	0.369	0.219	1.000	0.050		0.369	0.231	1.000	0.077	
(iv)	0.200	0.432	0.050	1.000		0.197	0.439	0.077	1.000	
Breusch–Pagan test of independence:			38.8***				40.2***			

Notes:
Standard errors are shown in brackets.
***$p < .01$
Sample periods: 2001–2004 and 2004–2006.

Table 4.8 SUR estimation results for high- and low- tech service industries (with additional control variables)

Specification	III(a)					III(b)				
Industry	Info & communication		Commerce & restaurants			Info & communication		Commerce & restaurants		
Dependent variable	$\ln SR^{Ind}$	$\ln SR^{Sub}$	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF	$\ln SR^{Ind}$	$\ln SR^{Sub}$	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF
Equation	(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)	
Constant	−12.57***	−3.053***	−7.875***	−3.95***		−12.57***	−3.338***	−7.875***	−4.352***	
	[0.435]	[0.382]	[0.162]	[0.303]		[0.435]	[0.252]	[0.162]	[0.202]	
CollegeGard	0.001	0.021	−0.042***	0.045***	9.5	0.001	0.008	−0.042***	0.026***	4.0
	[0.021]	[0.018]	[0.008]	[0.015]		[0.021]	[0.012]	[0.008]	[0.01]	
Expert	0.106**	−0.038	0.074***	−0.048*	3.7	0.106**	−0.026	0.074***	-0.03	3.1
	[0.048]	[0.03]	[0.018]	[0.024]		[0.048]	[0.028]	[0.018]	[0.022]	
ln *(WageRate)*		−0.384		−0.541*	5.0					
		[0.385]		[0.303]						
PopGrowth	0.092*	0.012	0.002	0.061**	2.4	0.092*	0.008	0.002	0.056**	2.4
	[0.052]	[0.03]	[0.019]	[0.024]		[0.052]	[0.03]	[0.019]	[0.024]	
Unemp	0.113***	0.063***	0.147***	0.024	1.7	0.113***	0.065***	0.147***	0.028	1.7
	[0.036]	[0.021]	[0.014]	[0.017]		[0.036]	[0.021]	[0.014]	[0.017]	
Period dummy	Yes	Yes	Yes	Yes	1.9	Yes	Yes	Yes	Yes	1.7

Specification	III(a)					III(b)				
Industry	Info & communication		Commerce & restaurants			Info & communication		Commerce & restaurants		
Dependent variable	$\ln SR^{Ind}$	$\ln SR^{Sub}$	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF	$\ln SR^{Ind}$	$\ln SR^{Sub}$	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF
Equation	(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)	
N	94	94	94	94		94	94	94	94	
R squared	0.607	0.767	0.885	0.799		0.607	0.765	0.885	0.798	
Correlation of residuals with the equation:										
(i)	1.000	−0.131	0.168	0.095		1.000	−0.133	0.369	0.090	
(ii)	−0.131	1.000	−0.062	0.397		0.003	1.000	0.231	0.403	
(iii)	0.168	−0.062	1.000	−0.218		0.369	−0.027	1.000	−0.155	
(iv)	0.095	0.397	−0.218	1.000		0.197	0.403	0.077	1.000	
Breusch–Pagan test of independence:			24.7***					22.7***		

Notes:
Standard errors are shown in brackets.
***p< .01, **p< .05, *p< .10
Sample periods: 2001–2004 and 2004–2006.

Table 4.9 SUR estimation results for high- and low- tech service industries (controlled for localization economy)

Specification	III(a)						III(b)					
Industry	Info & communication			Commerce & restaurants			Info & communication			Commerce & restaurants		
Dependent variable	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF	$\ln SR^{Ind}$	$\ln SR^{Sub}$	VIF
Equation	(i)	(ii)		(iii)	(iv)		(i)	(ii)		(iii)	(iv)	
Constant	−12.483***	−3.218***		−8.19***	−3.51***		−12.483***	−3.363***		−8.192***	−3.692***	
	[0.272]	[0.375]		[0.201]	[0.305]		[0.272]	[0.245]		[0.201]	[0.235]	
CollegeGard	−0.014	0.019	10.0	−0.036***	0.024***	10.8	−0.014	0.012	4.5	−0.036***	0.014***	4.9
	[0.013]	[0.018]		[0.008]	[0.014]		[0.013]	[0.012]		[0.008]	[0.009]	
Expert	0.071**	−0.022	3.6	0.07***	−0.038*	3.5	0.071**	−0.016	2.9	0.07***	−0.021	2.9
	[0.03]	[0.03]		[0.017]	[0.022]		[0.03]	[0.027]		[0.017]	[0.02]	
In *(WageRate)*		−0.194	5.3		−0.261*	5.3						
		[0.379]			[0.279]							
PopGrowth	0.037*	0.025	2.6	0.006	0.051**	2.6	0.037*	0.023	2.6	0.006	0.048**	2.5
	[0.033]	[0.03]		[0.019]	[0.022]		[0.033]	[0.029]		[0.019]	[0.022]	
Unemp	0.108***	0.065***	1.5	0.144***	0.032	1.5	0.108***	0.066***	1.5	0.144***	0.034	1.5
	[0.023]	[0.021]		[0.013]	[0.016]		[0.023]	[0.021]		[0.013]	[0.015]	
Localization of own industry	1.609***	−0.445***	1.3	0.014***	−0.3**	1.6	1.608***	−0.45***	1.3	0.015***	−0.03***	1.6
	[0.132]	[0.12]		[0.006]	[0.007]		[0.132]	[0.118]		[0.006]	[0.007]	
Period dummy	Yes	Yes	1.4	Yes	Yes	1.6	Yes	Yes	1.4	Yes	Yes	1.6

Specification	III(a)						III(b)					
Industry	Info & communication		VIF	Commerce & restaurants		VIF	Info & communication		VIF	Commerce & restaurants		VIF
Dependent variable	$\ln SR^{Ind}$	$\ln SR^{Sub}$		$\ln SR^{Ind}$	$\ln SR^{Sub}$		$\ln SR^{Ind}$	$\ln SR^{Sub}$		$\ln SR^{Ind}$	$\ln SR^{Sub}$	
Equation	(i)	(ii)		(iii)	(iv)		(i)	(ii)		(iii)	(iv)	
N	94	94		94	94		94	94		94	94	
R squared	0.848	0.777		0.894	0.799		0.848	0.776		0.894	0.834	
Correlation of residuals with the equation:												
(i)	1.000	0.103		0.307	0.185		1.000	0.109		0.307	0.198	
(ii)	0.103	1.000		−0.049	0.430		0.109	1.000		−0.029	0.432	
(iii)	0.307	−0.049		1.000	−0.080		0.307	−0.029		1.000	−0.042	
(iv)	0.185	0.430		−0.080	1.000		0.198	0.432		−0.042	1.000	
Breusch–Pagan test of independence:				31.3***			31.5***					

Notes:
Standard errors are shown in brackets.
***p< .01, **p< .05, *p< .10
Sample periods: 2001–2004 and 2004–2006.

6. Conclusion

This chapter investigated the determinants of regional business entry distinguishing between independent start-ups and subsidiaries, with a special focus on the effects of regional human capital. This is the major contribution of this chapter. Another contribution to the literature is that it compares the determinants of regional entry between the manufacturing and service sectors as well as across subsectors within the service sector. For the empirical analyses, pooled data of 47 Japanese prefectures for four observation periods are used.

Table 4.10 Summary of empirical estimations

Type of human capital (independent variable)		College graduates			Professional and technical workers	
Types of start-up (dependent variable)		Independent start-ups		New subsidiaries	Independent start-ups	New subsidiaries
Hypothesis		1a	1b	2	3	
Expected sign		+	−	+	+	
Results	Overall industry		−	+	+	
	Manu-facturing sector	+			−	
	Service sector		−	+	+	
	High-tech service industry				+	
	Low-tech service industry		−	+	+	

Note: The + and − signs indicate significantly positive and negative relationships, respectively, while blank cells indicate insignificant relationships.

The estimation results of SUR indicate considerable differences in the impact of regional factors between independent start-ups and subsidiaries as well as among different industries. Table 4.10 summarizes the empirical tests of the hypotheses. First, the ratio of college graduates is correlated negatively with independent start-ups (Hypothesis 1b) and positively with the entry of subsidiaries (Hypothesis 2). Second, the ratio of professional and technical workers positively affects independent start-ups but not the entry of subsidiaries (Hypothesis 3). Third, the relationships between regional human capital and the entry of independent start-ups and new subsidiaries are different across industries. The determinants of entry differ not only between the manufacturing and service sectors but also within the service sector. Moreover, the differences of the determinants between the types of start-up vary across sectors.

The negative relationship between the ratios of college graduates and independent start-ups (for the overall industry and the service sector) is consistent with the previous empirical evidence for Japan (e.g., Small and Medium Enterprise Agency, 2002; Okamuro, 2008), but not with the results from other countries (Guesnier, 1994; Armington and Acs, 2002; Acs and Armington, 2004). This suggests that highly educated workers in Japan regard the opportunity cost of quitting their current job as considerably higher than the expected returns from start-ups their own business.

The estimation results for the manufacturing sector are not only different from, but almost contrasting to those for the service sector. For manufacturing, the ratio of college graduates is correlated positively with the ratio of independent start-ups. This may imply that high education (particularly in the natural sciences) is especially important for independent start-up in the manufacturing sector, or that highly educated workers in the manufacturing sector do not have high opportunity cost of self-employment, assuming that the founders of manufacturing firms come from the manufacturing sector.

Moreover, for the manufacturing sector, the ratio of college graduates has no significant effect on start-ups of new subsidiaries (which does not support Hypothesis 2) and that the ratio of professional and technical workers has a negative effect on independent start-ups (contrary to Hypothesis 3). The latter results are different from the previous studies (Guesnier, 1994; Hart and Gudgin, 1994) that find a positive relationship for the manufacturing sector. It may be attributed to the definition of professional and technical occupations in Japan that covers numerous skills in the service sector. It is also noteworthy that the results on the effect of unemployment ratio on the ratio of independent start-ups clearly support the push hypothesis, similar to the results of previous studies in Japan.

However, some limitations of the present study need to be addressed in future research. First, for the manufacturing sector, there are different results from the overall industry and the service sector. There might also be some heterogeneity among manufacturing industries, which means that a more detailed industry and occupation classification is needed in order to test the hypotheses more concretely. Second, although there are positive relationships between the regional structure of human capital and the ratio of independent start-ups, these relationships can be explained by two possibilities. One possibility is that entrepreneurs have accumulated their human capital within the regions, but another is the migration of high-skilled workers (e.g., Ritsila and Ovaskainen, 2001). Presently, on the basis of the data available in this study, it is unable to clarify these possibilities.

Despite these limitations, this study suggests that regional policies to activate business start-ups should recognize the differences between encouraging local entrepreneurship and attracting new subsidiaries. These differences may vary even within the service sector according to technological intensity (or innovativeness).

Chapter 5 R&D, Innovation, and Business Performance of Japanese Start-ups: A Comparison with Established Firms[40]

1. Introduction

Since J. A. Schumpeter, entrepreneurship and innovation have been regarded as major sources of economic growth. Several empirical studies confirm the contribution of innovation to productivity growth (e.g., Crépon et al. 1998; Griffith et al. 2006; OECD 2009) and to employment growth (Hall et al. 2008; Lachenmaier and Rottmann 2011) at the firm level. Moreover, Acs and Armington (2004) and Audretsch and Keilbach (2005) demonstrate that entrepreneurial activities measured as the start-up ratio are a key factor for regional economic growth and productivity.

Despite the importance of innovation activities in business start-ups, few studies have comprehensively compared these undertakings to equivalent ones in established firms. Several empirical studies estimate the determinants of R&D input and outcomes by focusing on start-ups (Kato et al. 2013) or SMEs (Hall et al. 2009). Okamuro et al. (2011) analyze the determinants of R&D cooperation of business start-ups with business partners or universities. Okamuro (2009) compares the determinants of the propensity to conduct R&D and the R&D intensity of start-ups and all SMEs in the manufacturing sector. Huergo and Jaumandreu (2004a) find a nonlinear relationship between firm age and the probability of introducing an innovation. However, to the best of the knowledge, few studies comprehensively compare the determinants of R&D intensity, innovation, and firm performance of start-ups and established firms. In order to understand the characteristics and impact of innovation activities in start-ups, we should focus not only on R&D input but also on innovation and its impact on firm performance in both start-ups and established firms.

[40] This chapter is based on Ikeuchi and Okamuro (2013), co-authored with Hiroyuki Okamuro, which was conducted as a research project of the National Institute of Science and Technology Policy (NISTEP) under the "Science for Science, Technology and Innovation Policy" program of the Ministry of Education, Culture, Sports, Science and Technology in Japan.

Moreover, especially in Japan, despite the growing policy interests in innovation[41], there is little empirical research that employs the national innovation surveys, except for a few studies, such as Kwon et al. (2008) and Isogawa et al. (2012). Thus, this chapter bridges these gaps by using comparable datasets from different surveys.

In sum, the empirical results suggest that 1) the effects of public financial support on R&D intensity are smaller for start-ups; 2) the effects of research cooperation with business partners or universities on innovation are larger for start-ups; and 3) the effects of product and process innovation on labor productivity (level and growth) are positive both for start-ups and established firms. These results imply that, in order to promote the innovation and growth of start-ups, we should provide them with more or better support to engage in research cooperation.

The remainder of this chapter is organized as follows: data and estimation models are explained in Sections 2 and 3. Subsequently, empirical results are presented in Section 4. Section 5 concludes this chapter.

2. Data

Based on the data sources, this chapter distinguishes start-ups from established firms as follows: The former are firms within two years of operation and the latter those with more than two years of operation.

Data on start-ups is obtained from an original questionnaire survey series for Japanese start-ups that were carried out annually from 2008 to 2011. The first wave of this survey targeted 14,401 start-ups in the manufacturing and the software industry in Japan incorporated between January 2007 and August 2008; it was compiled by Tokyo Shoko Research (TSR), a major credit investigation company in Japan and based on the Corporation Register. Since the sample may also include the firms that were established earlier but incorporated after January 2007, the “real” start-ups are extracted, that is, those that were established during 2007 and 2008, using the survey response. The first postal survey was conducted in 2008 and 1,514

[41] Since the mid-1990s, the Japanese government has intensively promoted R&D and innovation with the “Science and Technology Basic Plans.” Implementation of the science-based science and technology policy is a new and important agenda in the fourth plan starting in 2011.

responses were received, of which 1,060 were “real” start-ups[42].

Then follow-up surveys are carried out in the successive years for the respondents of the previous year’s survey until 2011. For the empirical analysis of this chapter, the respondent firms of the third survey in 2010 are extracted and incomplete responses and some outliers are excluded. Thus, the final dataset of start-ups comprises 894 firms less than 2 years of age at time of the initial survey in 2008. The data from the third survey wave (and not the first one) is used to obtain sufficient information on innovation and firm performance and to secure comparability with the dataset of established firms.

Comparable data of established firms (that comprises approximately 2,000 firms) were obtained from the Japanese National Innovation Survey 2009 (J-NIS 2009) conducted in 2009 by the National Institute of Science and Technology Policy (NISTEP), as official statistics carried out according to the Oslo Manual and the Community Innovation Survey 2010 (CIS2010) in the EU. The sample of the survey comprises the firms with more than ten employees and covers the entire manufacturing sector and most non-manufacturing sectors, including the software industry. In all, 15,871 firms were selected as the sample from the 331,037 firms in the list of the Establishment and Enterprise Census conducted in 2006 by the Statistics Bureau of the Ministry of Internal Affairs and Communications. Of 4,579 respondents, 1,993 firms could be classified as belonging to the manufacturing or the software industry. Excluding incomplete responses and some outliers in addition to young firms less than 2 years of age, final dataset of established firms comprises 1,517 firms that had at least 2 years of operation at time of the initial survey year, 2006.

Since the Japanese National Innovation Survey is a sampling survey, few start-up firms are included in this survey, e.g. in the sample only ten firms are less than 2 years of age and 23 firms are less than 5 years of age. Therefore, this survey itself should not be appropriate for detailed analysis of start-up firms and this is the reason why we use a comparable original special survey on start-up firms.

Table 5.1 shows the simple comparison between start-ups and established firms in the datasets: The former are 1) less likely to conduct R&D, but more R&D intensive on average; 2) less likely to cooperate with business partners, universities, or public research institutes, but more dependent on the information from competitors; 3) less likely to innovate; and 4) more likely to grow faster, but less productive and profitable.

[42] For further information on this survey, see Okamuro et al. (2011).

Table 5.2 shows the correlation matrix of the variables. It reveals that, while labor productivity is positively associated with product and process innovation, the correlation of the growth rate of labor productivity with product and process innovation is negligible. Productivity and profitability are positively correlated each other. Profitability is positively correlated with product innovation but negatively correlated with process innovation. R&D input is positively associated with productivity, profitability and product, and process innovation. Geographic factors, such as the expert ratio (the ratio of workers in professional and technical occupations[43] in the workforce) and the density of industry and university, are also positively correlated with R&D intensity.

3. Model

This chapter simultaneously examines the differences between start-up firms and established firms in the determinants of innovation input (R&D intensity) and output (introduction of new products and processes) and firm performance (productivity and profitability). For this purpose, a three-stage model proposed by Crepon et al. 1998 (see also OECD 2009) is employed in order to consider the selectivity and endogeneity issues. In the first stage, R&D intensity measured as the ratio of R&D expenditures per person (in natural logarithm) is determined. In the second stage, the relationship between innovation input (R&D intensity) and output is investigated, distinguishing between product and process innovation and considering the effect of R&D cooperation. In the third and final stage, the effects of innovation output on firm performance, measured as the level and growth rate of labor productivity and the positive profit dummy, is examined.

[43] According to the Standard Occupation Classification of Japan, 'professional and technical occupations' include various types of scientists and engineers; medical and health-care services, such as doctors, pharmacists, and nurses; social welfare services; legal services, such as lawyers; business support services, such as accountants and management consultants; and teachers and artists.

Table 5.1 Descriptive statistics

	Established firms (firm age ≥ 2)					Start-up firms (firm age < 2)				
Variables	n	Mean	S.D.	Min	Max	n	Mean	S.D.	Min	Max
Positive R&D (dummy)	1,283	0.461	0.499	0.000	1.000	880	0.308	0.462	0.000	1.000
R&D intensity (expenditure per person: 1mil. yen)	1,283	0.422	1.679	0.000	28.654	880	0.550	2.246	0.000	50.000
Log. of R&D intensity	591	-1.512	1.778	-7.378	3.355	271	-0.557	1.688	-6.765	3.912
Product innovation (dummy)	872	0.669	0.471	0.000	1.000	510	0.412	0.493	0.000	1.000
Process innovation (dummy)	872	0.429	0.495	0.000	1.000	510	0.161	0.368	0.000	1.000
Labor productivity (sales per person: 1 mil. yen)	674	36.228	45.841	0.000	458.652	223	17.030	32.669	0.000	360.000
Log. of labor productivity	674	3.211	0.879	0.000	6.130	223	2.288	1.020	0.000	5.889
Labor productivity growth rate	674	0.004	0.351	-2.244	3.714	223	0.120	0.880	-2.877	3.586
Positive profit (dummy)	743	0.709	0.454	0.000	1.000	247	0.543	0.499	0.000	1.000
Collaboration with business partners (dummy)	872	0.541	0.499	0.000	1.000	510	0.408	0.492	0.000	1.000
Collaboration with universities (dummy)	872	0.271	0.445	0.000	1.000	510	0.125	0.332	0.000	1.000
Information from competitor (dummy)	872	0.382	0.486	0.000	1.000	510	0.500	0.500	0.000	1.000
Employment size	1,517	321.937	1162.589	1.000	31595.000	894	11.892	42.296	1.000	620.000
Log. of employment size	1,517	4.342	1.927	0.000	10.361	894	1.404	1.134	0.000	6.430
Initial labor productivity (sales per person: 1 mil. yen)	1,517	31.585	44.660	0.000	671.597	894	15.407	30.763	0.000	600.000
Log. of initial labor productivity	1,517	3.007	0.966	0.000	6.511	894	2.186	1.056	0.000	6.399
Firm age	1,517	32.879	22.049	2.000	230.000	894	0.557	0.497	0.000	1.000
Affiliated firm dummy	1,517	0.405	0.491	0.000	1.000	894	0.219	0.414	0.000	1.000
Public financial support (dummy)	1,517	0.213	0.410	0.000	1.000	894	0.318	0.466	0.000	1.000
Expert ratio – city	1,517	0.141	0.033	0.059	0.247	894	0.153	0.037	0.064	0.247
Expert ratio – prefecture	1,517	0.140	0.019	0.111	0.171	894	0.146	0.020	0.111	0.171
Industry density – city	1,517	6.181	21.550	0.000	141.182	894	10.300	26.006	0.000	141.182
Industry density – prefecture	1,517	0.714	1.364	0.000	5.566	894	1.262	1.778	0.000	5.566
University density – city	1,517	0.029	0.083	0.000	0.707	894	0.040	0.097	0.000	0.707
University density – prefecture	1,517	0.008	0.010	0.000	0.028	894	0.011	0.011	0.000	0.028

Table 5.2 Correlation matrix of variables

A. Established firms (firm age ≥ 2 years-old)

		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]
[1]	Positive R&D (dummy)	1.000																				
[2]	Log of R&D intensity	.	1.000																			
[3]	Product innovation (dummy)	.332	.144	1.000																		
[4]	Process innovation (dummy)	.182	.005	.211	1.000																	
[5]	Log of labor productivity	.281	.333	.189	.094	1.000																
[6]	Labor productivity growth rate	-.001	.032	-.029	.056	.226	1.000															
[7]	Positive profit (dummy)	.052	-.002	.065	-.082	.142	.092	1.000														
[8]	Collaboration with business partners	.209	.068	.208	.198	.113	.013	-.028	1.000													
[9]	Collaboration with universities (dummy)	.278	.190	.221	.119	.119	-.014	.016	.271	1.000												
[10]	Information from competitor (dummy)	.005	.056	.092	.030	-.020	-.077	.017	.060	.042	1.000											
[11]	Log of employment size	.325	.070	.297	.179	.389	-.120	.095	.244	.301	-.047	1.000										
[12]	Log of initial labor productivity	.320	.325	.242	.136	.920	-.175	.127	.155	.172	-.035	.396	1.000									
[13]	Log. of firm age	.215	-.019	.234	.164	.190	-.126	.000	.130	.202	-.118	.510	.271	1.000								
[14]	Affiliated firm dummy	.133	.171	.132	.056	.340	-.033	.042	.155	.095	-.034	.457	.290	.130	1.000							
[15]	Public financial support (dummy)	.070	.090	-.019	.008	-.100	.070	-.072	.020	.238	.058	-.123	-.032	-.113	-.114	1.000						
[16]	Expert ratio – city	.125	.199	.053	-.050	-.002	-.003	.110	-.019	.110	.040	.049	.036	-.060	.005	.048	1.000					
[17]	Expert ratio – prefecture	.114	.194	.029	-.030	.044	.006	.105	-.007	.065	.013	.032	.077	-.029	.012	.018	.645	1.000				
[18]	Industry density – city	.005	.068	-.016	-.116	-.058	-.020	.101	-.066	-.042	.044	.050	-.052	-.056	-.005	-.064	.332	.318	1.000			
[19]	Industry density – prefecture	.014	.111	.009	-.084	-.010	-.022	.119	-.031	-.001	.069	-.005	-.028	-.071	-.029	-.056	.393	.490	.653	1.000		
[20]	University density – city	.105	.152	.023	-.009	.011	-.043	.055	.048	.110	.045	.103	.049	-.042	.056	.016	.586	.474	.402	.422	1.000	
[21]	University density – prefecture	.109	.199	.063	-.016	.069	-.041	.087	.018	.105	.020	.091	.102	-.008	.025	-.015	.597	.847	.418	.649	.574	1.000

Table 5.2 Correlation matrix of variables (cont.)

B. Start-up firms (firm age < 2 years-old)

		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]
[1]	Positive R&D (dummy)	1.000																				
[2]	Log of R&D intensity	.	1.000																			
[3]	Product innovation (dummy)	.265	.209	1.000																		
[4]	Process innovation (dummy)	.195	.122	.365	1.000																	
[5]	Log of labor productivity	.084	.011	.044	.037	1.000																
[6]	Labor productivity growth rate	.143	.152	.009	.026	.434	1.000															
[7]	Positive profit (dummy)	-.037	-.163	.008	-.114	.343	.105	1.000														
[8]	Collaboration with business partners	.142	.166	.370	.217	.086	.081	-.074	1.000													
[9]	Collaboration with universities (dummy)	.173	.196	.199	.192	.077	.148	.041	.142	1.000												
[10]	Information from competitor (dummy)	-.018	-.003	.010	-.081	.063	.014	.071	-.014	-.155	1.000											
[11]	Log of employment size	-.018	-.242	.135	.114	.226	.010	.045	.068	-.017	.000	1.000										
[12]	Log of initial labor productivity	-.073	-.038	.062	.064	.621	-.436	.276	.074	-.096	.017	.103	1.000									
[13]	Log. of firm age	.	.	.	.	.	.	.	.	.	.	.	.	.								
[14]	Affiliated firm dummy	-.065	.117	-.053	.002	.186	.022	.007	-.019	.007	-.060	.362	.114	.	1.000							
[15]	Public financial support (dummy)	.027	.000	.142	.077	.122	.154	-.008	-.026	.099	-.018	.062	.021	.	-.123	1.000						
[16]	Expert ratio – city	.091	.109	.009	-.055	-.037	-.038	-.025	-.042	.004	.096	-.103	.018	.	-.080	-.023	1.000					
[17]	Expert ratio – prefecture	.083	.150	.000	-.089	.001	.010	.005	-.046	-.026	.166	-.118	.037	.	-.097	-.001	.608	1.000				
[18]	Industry density – city	.003	.033	-.022	-.058	.004	.004	.113	-.129	-.033	.088	.069	-.002	.	.032	-.044	.366	.323	1.000			
[19]	Industry density – prefecture	-.019	.126	-.002	-.057	.058	.022	.077	-.096	-.032	.148	-.048	.017	.	-.031	-.040	.384	.571	.602	1.000		
[20]	University density – city	.070	.133	-.020	-.029	-.098	-.034	.035	-.028	-.021	.032	.013	-.007	.	.037	-.005	.572	.411	.488	.468	1.000	
[21]	University density – prefecture	.028	.215	.019	-.089	.014	.009	.003	-.070	-.027	.182	-.074	.047	.	-.057	-.012	.549	.862	.446	.764	.503	1.000

3.1. First stage: R&D intensity model

In the first stage of the model, the R&D intensity of firms, defined as R&D expenditures per employee, is determined by two equations employing the generalized Tobit model (Heckman, 1976, 1979): Firms decide at first whether or not they engage in R&D activity (the first equation) and then determine the level of R&D expenditures (the second equation). We use the same set of factors as explanatory variables for both equations, but estimate different sets of coefficients for each equation.

We focus on the differences between start-up and established firms with respect to the effects on R&D intensity of public financial support and local accessibility to research personnel. To examine the effects of public financial support, we utilize a dummy variable which takes the value of one if the firm obtains any public financial support from national or local government, and zero otherwise. We define the local accessibility to research personnel as the density of professional and technical workers which is calculated as the number of such workers per square kilometers.

Large body of literature argue that, due to information asymmetries, most start-up firms have difficulty in obtaining sufficient funds for investment from external capital markets (e.g., Honjo et al., 2014). Financial constraints might be essential particularly for R&D investment because of its intangible and inappropriable feature (Kamien and Schwartz, 1978) and high uncertainty (Carpenter and Petersen, 2002). Therefore, we expect that the effect of public financial support on R&D intensity of start-up firms is larger than that of established firms.

In addition to financial resources, sufficient specific human capital is needed to conduct R&D projects, embodied in engineers, researchers, and/or scientists. Therefore, accessibility to R&D personnel can also affect R&D intensity. For start-up firms, it is difficult to procure R&D personnel from the distant labor market areas. This indicates that a start-up firm relies more on local labor market than an established firm does. Therefore, we expect that the effect of local accessibility to R&D personnel on R&D intensity is larger for start-up firms than for established firms.

In addition, we control for the effects of firm size and age, the differences between affiliated and independent firms, industry-specific effects, and the density of businesses and universities in the municipality and prefecture where the firms' headquarters are located.

3.2. Second Stage: Innovation Model

Firms generate new products and processes as innovation outputs. In this regard, it is distinguished between product innovation (the generation of new or significantly improved products) and process innovation (the implementation of new or significantly improved production method)[44].

As the determinants of innovations, the predicted values of R&D intensity in the first stage are a main variable. In addition, Robin and Schubert (2013) have recently found a positive effect of cooperation with public research institutes on the probability of introducing product innovation but no effect on process innovation. As shown in Belderbos et al. (2004), supplier and customer firms and competitors might be also important as collaboration partners and external knowledge sources. Therefore, first, the cooperation with universities and firms is distinguished with supplier/customer relationships. Second, the effects of external knowledge from competitors are examined by utilizing a survey question on the importance of competitors as information sources in R&D (innovation) activity. Then the difference in the magnitude of the effects of those cooperation and external knowledge from competitors on innovation between start-ups and established firms is examined.

Because of short experiences in their business, start-up firms tend to have smaller network and lower reputation than established firms do. It should be a difficult task for start-up firms to find appropriate research partners, and even if they find any, it is difficult for them to successfully make a contract with research partners. Therefore, the marginal benefits of collaboration with external organization and external knowledge is expected to be larger for start-up firms than established firms. Also, due to scarce internal resources of start-up firms, it is more difficult for them to complete R&D projects and commercialize the outcomes by themselves. For start-up firms, a collaboration with a large firm helps to commercialize R&D outcomes and a collaboration with a university helps to utilize basic research and scientific knowledge. Also, because start-up firms are free from innovators' dilemma as compared to

---

[44] According to Oslo Manual (OECD 2005), process innovation covers not only the implementation of a new or significantly improved production methods but also that of new or significantly improved delivery methods and techniques, equipment, and software in ancillary support activities. Since the survey for start-ups did not consider the latter two types of process innovation, only the implementation of a new production method is regarded as process innovation.

established firms, it takes advantage in smoothly utilizing the collaborative partners' specialty.

3.3. Third Stage: Performance Model

Finally, to validate the measurement of the indicators for innovations and to access the differences in an economic impact of innovations between start-ups and established firms, the effects of product and process innovation on firm economic performance are estimated, such as the levels or growth rates of labor productivity and profitability. Since start-up firms may have less complementarity assets necessary to commercialize innovation compared to established firm, the effects of innovations on firm economic performance may differ between start-ups and established firms. As the proxy for productivity, labor productivity is employed. Since the dataset of startups does not consist of physical capital accumulation and the input of materials, it is not able to measure the total factor productivity and also not control for capital intensity or intermediate inputs. Instead, several control variables are included: initial employment size, age, affiliated firm dummy, and initial labor productivity level. Choice of the proxy for profitability is also limited because of a lack of detailed financial information. A dummy variable is used, which takes the value of one, if the firm's (operating) profit is positive[45].

Product and process innovation may be complimentary. However, a marginally strong correlation between these two types of innovations (0.306 as shown in Table 5.2) might make it difficult to identify the effects of these two types of innovations. To explore the relevant specification, several approaches are examined: First, the predicted probability that the firm introduces either the product or process innovation is used as an explanatory variable. Second, the predicted probabilities of product innovation and process innovation are used, alternately or independently, as explanatory variables. Third, the predicted probability of product innovation only, process innovation only, and product and process innovations together is used as explanatory variables.

[45] For the start-ups, it is not able to identify the firms' answers to the profitability question based on which kind of profit.

4. Results

4.1. First stage: R&D intensity model

Table 3.3 shows the estimation results of the generalized Tobit model for R&D intensity. For each specification, the first column shows the coefficients of the probit model in which the dependent variable is a dummy variable for R&D conducting firms, and the second column reports the coefficients of linear model of the level of R&D intensity. In addition, in the last row, the correlation coefficients of the residuals of two equations are reported for each specification. The results show the positive effects of initial labor productivity on both the selection equation and R&D intensity and the positive effects of employment size and firm age on only R&D intensity. Affiliated firms conduct R&D investment at a higher probability, but their R&D intensity is lower than that of independent firms. Public financial support and the expert ratio in local labor market increase the probability of R&D investment and the R&D intensity of firms (see Figure 5.1 and Figure 5.2). The geographic agglomeration of industry and university have no effects on either the selection or the intensity of R&D. Interestingly, the effects of public support on both the selection and intensity of R&D are significantly smaller for start-ups than established firms, while there is no significant difference in the effects of the expert ratio between these groups.

Table 5.3 First stage results for R&D intensity

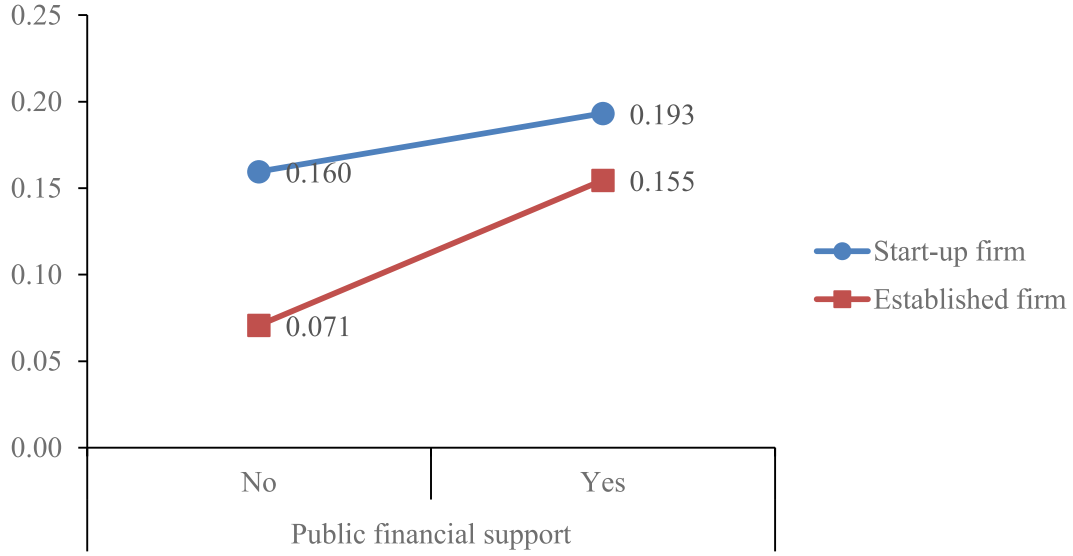
(Generalized tobit model - ML estimation)

Dependent variable: positive R&D dummy and log of R&D per employee

	[1]		[2]		[3]		[4]		[5]	
Dependent variable	R&D>0	R&D int.	R&D>0	R&D int.	R&D>0	R&D int.	R&D>0	R&D int.	R&D>0	R&D int.
Initial labor productivity	0.405***	0.121***	0.428***	0.125***	0.429***	0.127***	0.435***	0.128***	0.439***	0.128***
	[0.081]	[0.034]	[0.080]	[0.034]	[0.079]	[0.034]	[0.080]	[0.034]	[0.080]	[0.034]
Initial employment size	-0.072	0.188***	-0.077	0.190***	-0.073	0.197***	-0.078	0.196***	-0.073	0.196***
	[0.053]	[0.024]	[0.053]	[0.024]	[0.053]	[0.024]	[0.053]	[0.024]	[0.053]	[0.024]
Age	-0.137**	0.010	0.143	0.108***	0.143	0.114***	0.149*	0.117***	0.154*	0.113***
	[0.058]	[0.026]	[0.088]	[0.041]	[0.088]	[0.041]	[0.088]	[0.041]	[0.089]	[0.042]
Affiliated (dummy)	0.523***	-0.152**	0.548***	-0.154**	0.542***	-0.160**	0.535***	-0.161**	0.526***	-0.163**
	[0.150]	[0.071]	[0.147]	[0.071]	[0.147]	[0.071]	[0.148]	[0.071]	[0.148]	[0.071]
Public financial support (dummy)	0.504***	0.203***	0.498***	0.208***	0.695***	0.404***	0.685***	0.404***	0.694***	0.407***
	[0.136]	[0.067]	[0.135]	[0.067]	[0.161]	[0.092]	[0.160]	[0.093]	[0.161]	[0.093]
Expert ratio – city	5.414**	3.429***	5.235**	3.390***	4.987**	3.226***	7.009**	3.860**	7.575***	3.882**
	[2.441]	[1.201]	[2.427]	[1.208]	[2.407]	[1.210]	[2.823]	[1.541]	[2.883]	[1.600]
Expert ratio – prefecture	1.477	7.886**	1.641	7.947***	1.901	8.007***	-0.455	8.028**	0.537	3.637
	[5.963]	[3.069]	[5.916]	[3.083]	[5.919]	[3.090]	[6.597]	[3.497]	[7.248]	[3.998]
Industry density – city	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.002	-0.001
	[0.004]	[0.002]	[0.004]	[0.002]	[0.004]	[0.002]	[0.004]	[0.002]	[0.005]	[0.002]
Industry density – prefecture	-0.080	-0.038	-0.088	-0.039	-0.084	-0.037	-0.089	-0.036	-0.135	-0.059
	[0.077]	[0.037]	[0.076]	[0.037]	[0.076]	[0.037]	[0.076]	[0.038]	[0.091]	[0.051]
Univ. density – city	0.126	0.559	0.281	0.594	0.269	0.609	0.340	0.629	0.119	0.579
	[0.929]	[0.433]	[0.941]	[0.437]	[0.933]	[0.440]	[0.919]	[0.441]	[1.028]	[0.619]
Univ. density – prefecture	26.206*	-7.254	23.433*	-8.149	22.781	-8.113	22.114	-8.223	22.364	5.084
	[14.033]	[6.940]	[13.923]	[6.984]	[13.891]	[6.996]	[13.828]	[7.010]	[16.087]	[8.937]
Start-up (dummy)			1.219***	0.402***	1.388***	0.549***	1.291	0.796*	1.819	-0.642
			[0.271]	[0.125]	[0.286]	[0.137]	[0.965]	[0.456]	[1.585]	[0.774]
Start-up x Public financial support					-0.505*	-0.401***	-0.510*	-0.402***	-0.512*	-0.404***
					[0.291]	[0.136]	[0.293]	[0.136]	[0.294]	[0.136]
Start-up x Expert ratio – city							-5.823	-1.519	-6.541	-1.604
							[4.542]	[2.159]	[4.976]	[2.408]
Start-up x Expert ratio – prefecture							6.816	-0.106	3.418	11.498*
							[7.860]	[3.849]	[12.400]	[6.269]
Start-up x Industry density – city									-0.004	0.002
									[0.007]	[0.003]
Start-up x Industry density – prefecture									0.093	0.072
									[0.141]	[0.068]
Start-up x Univ. density – city									0.544	0.053
									[2.039]	[0.880]
Start-up x Univ. density – prefecture									1.272	-34.268**
									[27.853]	[13.965]
Constant	-4.999***	-2.773***	-5.958***	-3.105***	-6.002***	-3.194***	-5.961***	-3.293***	-6.236***	-2.736***
	[0.933]	[0.408]	[0.977]	[0.421]	[0.983]	[0.424]	[1.020]	[0.458]	[1.093]	[0.517]
Industry dummies (2 digit)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	2,163		2,163		2,163		2,163		2,163	
# of firms no R&D	1301		1301		1301		1301		1301	
Chi-squared (statistics)	328.2231		347.5391		356.9189		355.8323		357.3385	
Chi-squared (p-value)	0.000		0.000		0.000		0.000		0.000	
Correlation between errors	0.543		0.533		0.528		0.526		0.551	

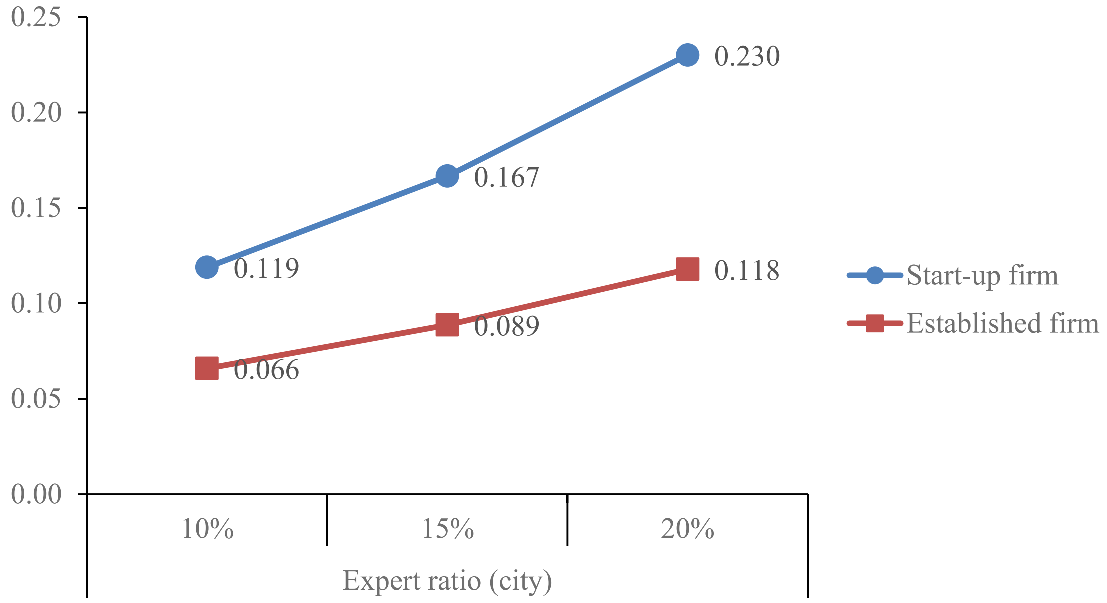
Notes: Robust standard errors are in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Figure 5.1: Marginal effects of public financial support on R&D intensity



Notes: The vertical axis is the predicted value of R&D expenditure (in 1 million yen) per person. The predicted values are calculated from the estimation results of column [4] in Table 3 at the mean values of the remaining covariates.

Figure 5.2: Marginal effects of expert ratio in city on R&D intensity



Notes: The vertical axis is the predicted value of R&D expenditure (in 1 million yen) per person. The predicted values are calculated from the estimation result of column [4] in Table 3 at the mean values of the remaining covariates.

4.2. Second Stage: Innovation Model

Table 5.4 shows the second stage results of the bivariate probit model for product and process innovation. For each specification, the coefficients of the product innovation equation and those of the process innovation equation are reported in the first column and the second column, respectively. The effects of predicted R&D intensity are significantly positive on product innovation (see Figure 5.3) but not on process innovation (see Figure 5.4). There are positive effects of collaboration with business partners (see Figure 5.5 and Figure 5.6) and universities (see Figure 5.7 and Figure 5.8) both on product and process innovation while the information from competitors affects only product innovation (Figure 5.9 and Figure 5.10). Firm size has positive effects, but firm age has no effect. Affiliated firms have a lower probability of product innovation but there is no significant difference in the probability of process innovation between affiliated and independent firms. There are several significant differences in the effects of collaboration with partner firms and universities and in information from competitors on innovation between start-ups and established firms: the positive effects of collaboration with business partners (supplier and client) and universities on product innovation are greater in start-ups than in established firms, while the effect of information from competitors on product innovation is lower in start-ups than in established firms. Collaborations with universities also increase the probability of process innovation more in start-ups than in established firms. As the same as in the first stage of the R&D intensity model, there is not any significant effects of geographic agglomeration factors on innovations.

4.3. Third Stage: Performance Model

Table 5.5-Table 5.7 reports the third stage results of the firm performance model with three different dependent variables: the level of labor productivity in Table 5.5, the growth rate of labor productivity in Table 5.6, and profitability in Table 5.7. While the models shown in first five columns of Table 5.5 and Table 5.6 estimate the common coefficients for start-ups and established firms, the models in the successive five columns (6-10), include the interaction terms of these innovation indicators with start-up firm dummy. In those tables, the last two columns examine the direct effects of R&D intensity on productivity.

Table 5.4 Second stage results for product and process innovation

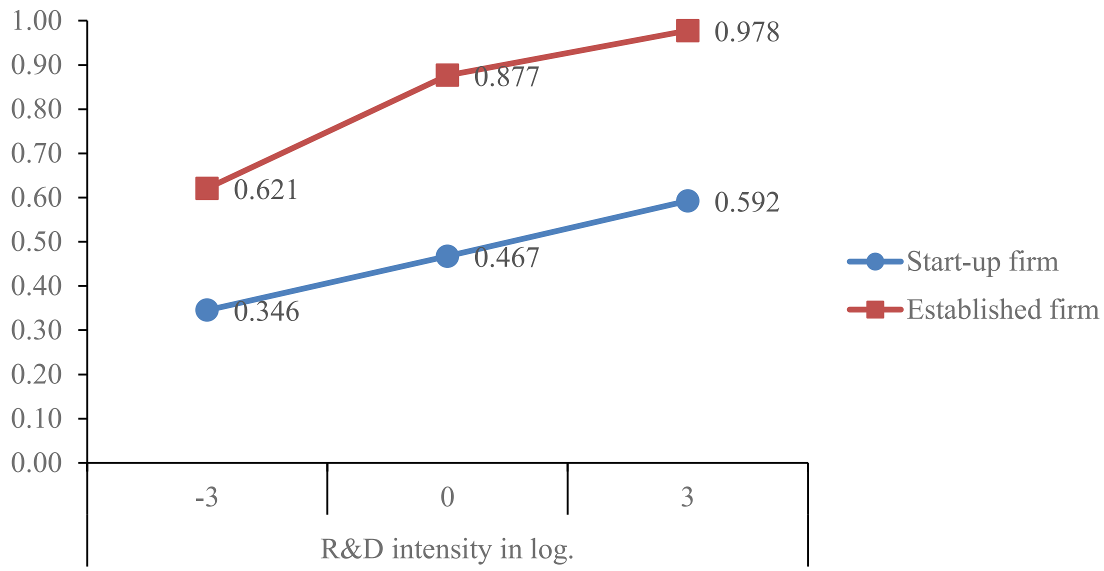
(bivariate probit model - ML estimation)

Dependent variables: Dummy variables indicating the introduction of product innovation and process innovation

	[1]		[2]		[3]	
Dependent variable	Product	Process	Product	Process	Product	Process
Predicted R&D intensity	0.169***	-0.057	0.186**	-0.001	0.283***	0.039
	[0.065]	[0.067]	[0.082]	[0.086]	[0.091]	[0.089]
Collaboration with business partners (dummy)	0.565***	0.443***	0.567***	0.433***	0.355***	0.375***
	[0.076]	[0.079]	[0.076]	[0.079]	[0.098]	[0.095]
Collaboration with universities (dummy)	0.467***	0.236**	0.467***	0.231**	0.336***	0.109
	[0.105]	[0.095]	[0.105]	[0.095]	[0.124]	[0.111]
Information from competitors (dummy)	0.213***	0.024	0.207***	0.023	0.344***	0.086
	[0.077]	[0.078]	[0.077]	[0.078]	[0.101]	[0.093]
Initial employment size	0.150***	0.081***	0.153***	0.089***	0.167***	0.097***
	[0.029]	[0.028]	[0.029]	[0.029]	[0.030]	[0.030]
Age	0.074	0.062	0.066	0.052	0.065	0.054
	[0.050]	[0.048]	[0.051]	[0.050]	[0.051]	[0.052]
Affiliated (dummy)	-0.198**	-0.104	-0.206**	-0.141	-0.235**	-0.164
	[0.098]	[0.098]	[0.102]	[0.103]	[0.106]	[0.106]
Start-up (dummy)	-0.069	-0.200	-0.104	-0.253	-0.635**	-0.460
	[0.163]	[0.174]	[0.172]	[0.184]	[0.289]	[0.308]
Start-up x Predicted R&D intensity					-0.178*	-0.050
					[0.094]	[0.101]
Start-up x Collaboration with business partners					0.568***	0.192
					[0.157]	[0.175]
Start-up x Collaboration with universities					0.412*	0.473**
					[0.229]	[0.224]
Start-up x Information from competitors					-0.277*	-0.137
					[0.160]	[0.177]
Industry density – city			-0.001	-0.004*	-0.001	-0.004
			[0.002]	[0.002]	[0.002]	[0.003]
Industry density – prefecture			0.016	0.075	0.020	0.074
			[0.044]	[0.048]	[0.045]	[0.048]
Univ. density – city			-0.754	0.315	-0.795	0.320
			[0.486]	[0.461]	[0.506]	[0.540]
Univ. density – prefecture			2.538	-9.868	2.825	-9.821
			[6.380]	[6.329]	[6.571]	[6.581]
Constant	-0.342	-1.287***	-0.271	-1.074***	0.039	-0.940**
	[0.324]	[0.335]	[0.391]	[0.406]	[0.400]	[0.407]
Industry dummies (2 digit)	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	1,382		1,382		1,382	
Chi-squared (statistics)	446.021		456.576		470.224	
Chi-squared (p-value)	0.000		0.000		0.000	
Correlation between errors	0.367		0.366		0.360	

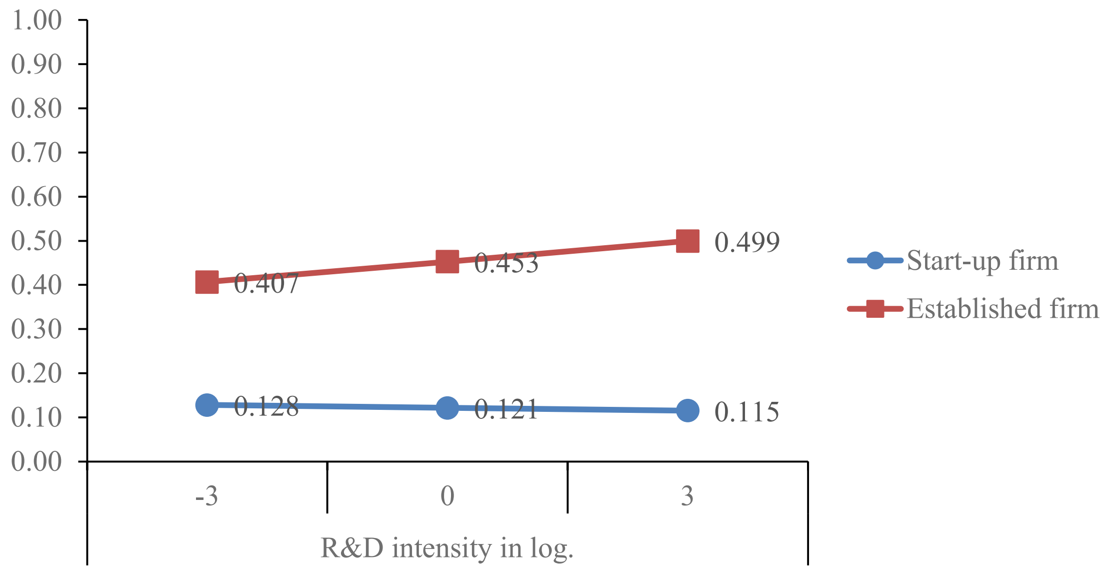
Notes: Robust standard errors are in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Figure 5.3: Marginal effects of R&D intensity on product innovation



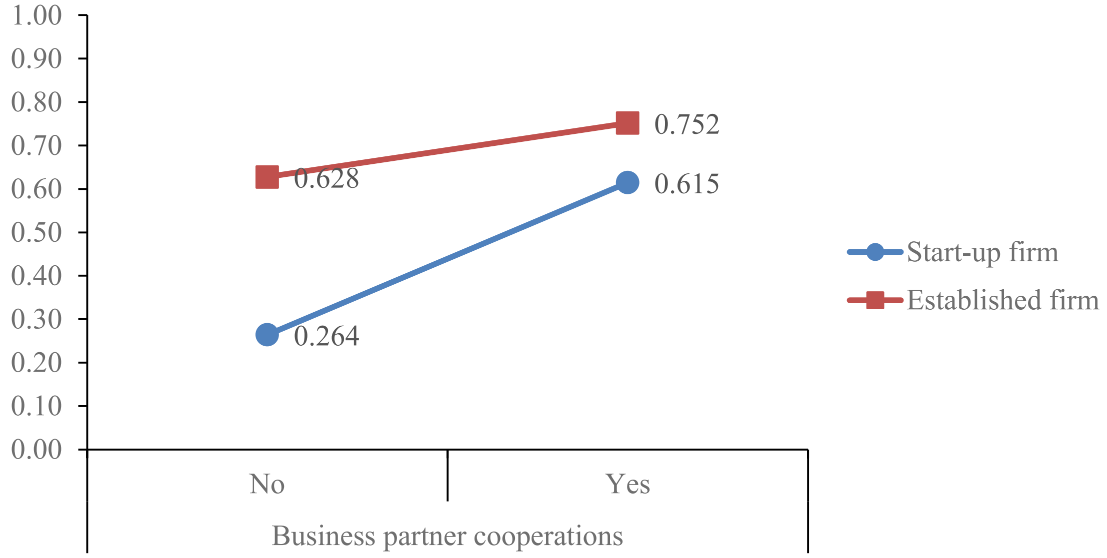
Notes: The vertical axis is the predicted probability to have a product innovation. The predicted values are calculated from the estimation result of column [3] in Table 4 at the mean values of the remaining covariates.

Figure 5.4: Marginal effects of R&D intensity on process innovation



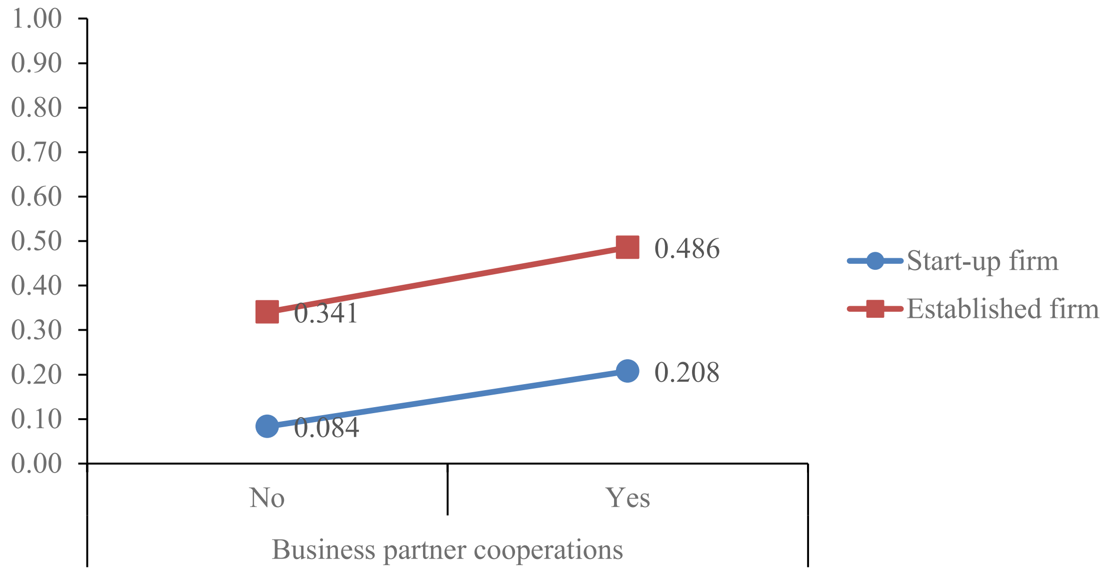
Notes: Vertical axis is the predicted probability to have a process innovation. Predicted values are calculated from the estimation result of column [3] in Table 4 at the mean values of the remaining covariates.

Figure 5.5: Marginal effects of business partner cooperation on product innovation



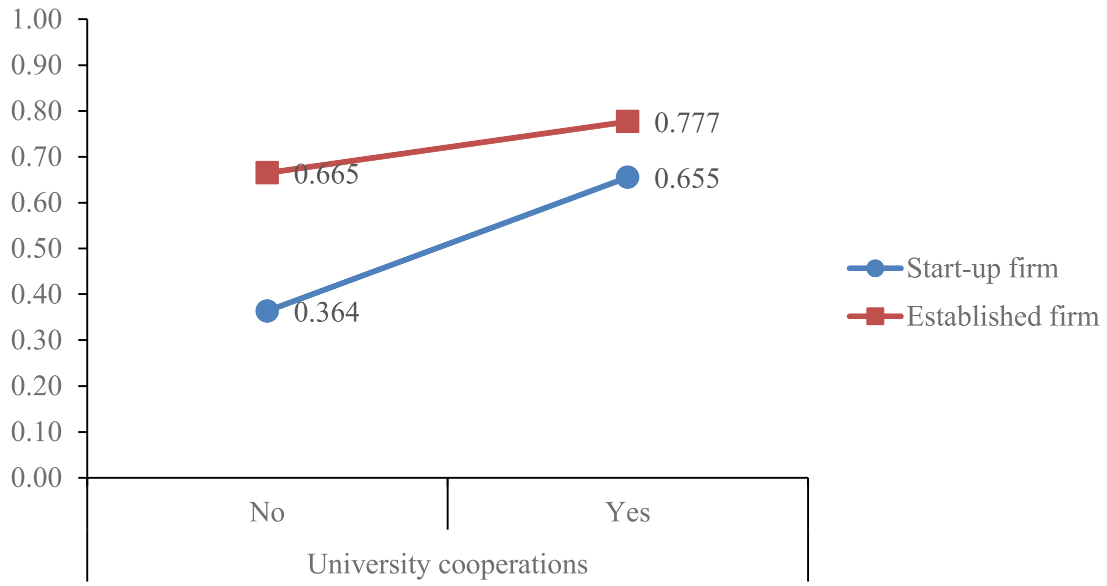
Notes: The vertical axis is the predicted probability to have a product innovation. The predicted values are calculated from the estimation result of column [3] in Table 4 at the mean values of the remaining covariates.

Figure 5.6: Marginal effects of business partner cooperation on process innovation



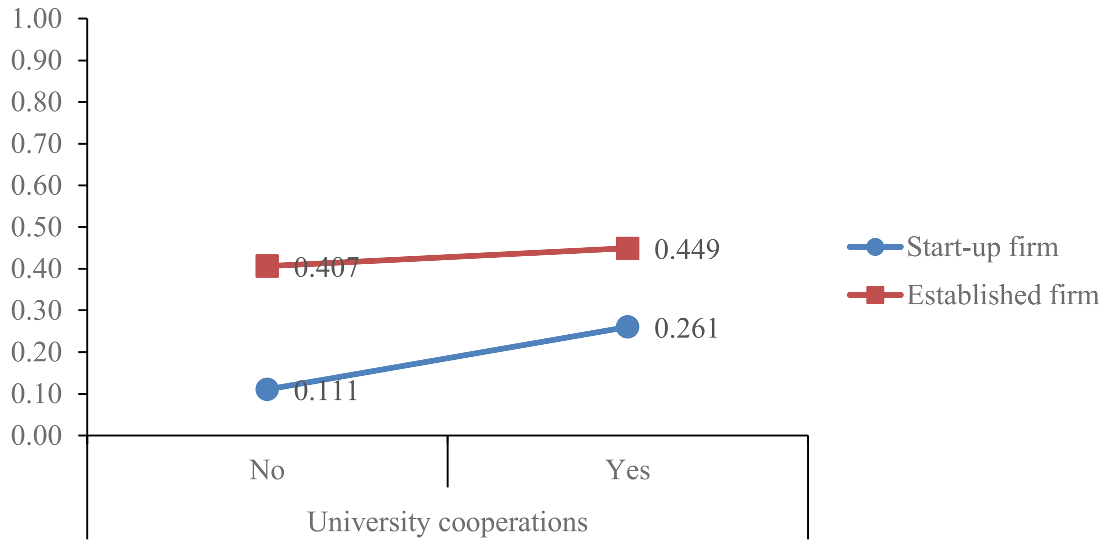
Notes: The vertical axis is the predicted probability to have a process innovation. The predicted values are calculated from the estimation result of column [3] in Table 4 at the mean values of the remaining covariates.

Figure 5.7: Marginal effects of university cooperation on product innovation



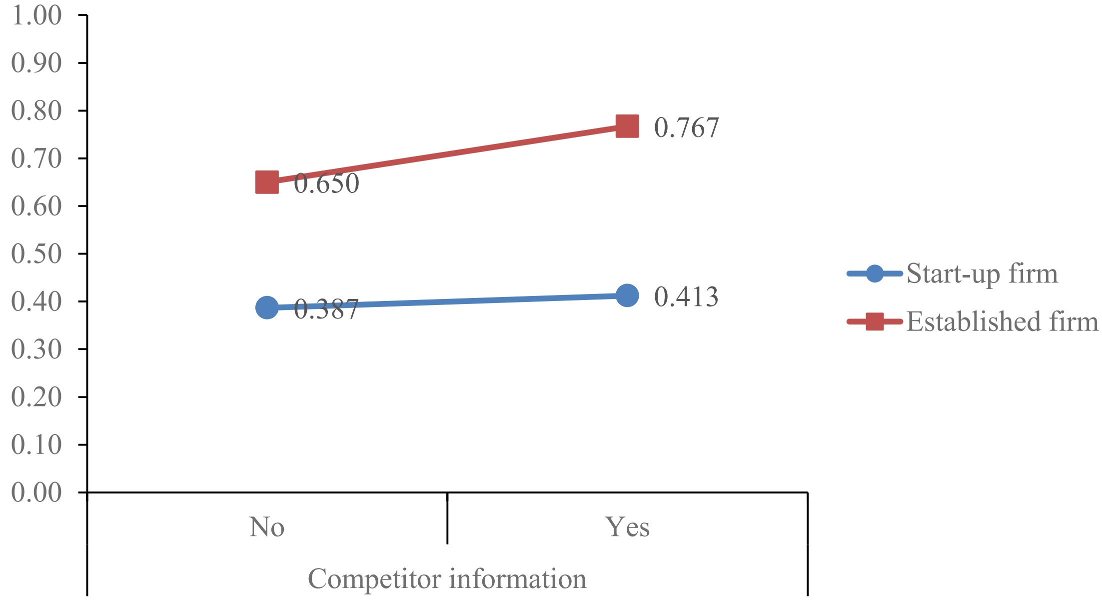
Notes: the vertical axis is the predicted probability to have a product innovation. The predicted values are calculated from the estimation result of column [3] in Table 4 at the mean values of the remaining covariates.

Figure 5.8: Marginal effects of university cooperation on process innovation



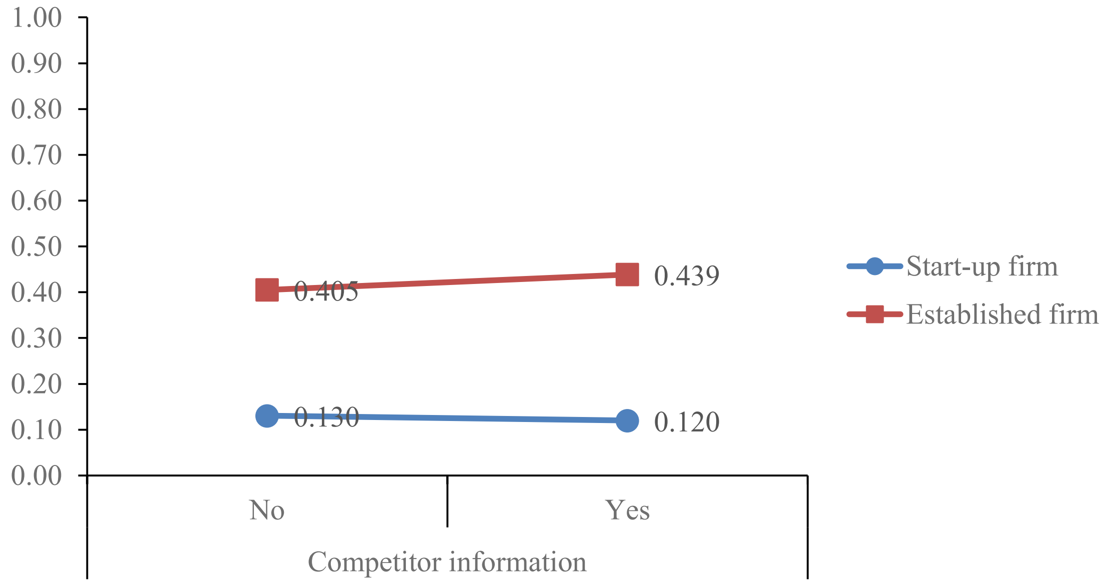
Notes: The vertical axis is the predicted probability to have a process innovation. The predicted values are calculated from the estimation result of column [3] in Table 4 at the mean values of the remaining covariates.

Figure 5.9: Marginal effects of competitor information on product innovation



Notes: The vertical axis is the predicted probability to have a product innovation. The predicted values are calculated from the estimation result of column [3] in Table 4 at the mean values of the remaining covariates.

Figure 5.10: Marginal effects of competitor information on process innovation



Notes: The vertical axis is the predicted probability to have a process innovation. The predicted values are calculated from the estimation result of column [3] in Table 4 at the mean values of the remaining covariates.

The results in column [1] to [3] in Table 5.5 show that positive effects of product and process innovation on the level of labor productivity, controlling for effects of scale economy and affiliated firms. When we jointly include product and process innovation in the specification [4] and [5] of Table 5.5, however, the coefficient of process innovation turn negative. The effects of process innovation on productivity are also controversial in the literature. On the one hand, OECD (2009) consistently reports the significantly negative coefficients of process innovation on productivity of 18 countries, while the coefficients of product innovation are jointly estimated as positive. On the other hand, Griffith et al. (2006) report the significantly positive effects of process innovation and product innovation, using capital investment intensity only as an instrumental variable for process innovation[46].

There can be found the negative coefficients of the interaction terms between the start-up firm dummy and product and process innovations. These imply that the effects of product or process innovation are smaller in start-ups than in established firms. In column [11] and [12], there are also the significant effects of predicted R&D intensity on productivity. These imply that the innovation indicators might not capture the whole effects of R&D.

[46] Hall et al. (2009) confirms that the effect of process innovation on productivity is estimated as significantly positive only when they instrument it by capital investment intensity and do not include capital investment intensity in the productivity equation; otherwise, it is estimated as negative or positive but not as significant.

Table 5.5 Third stage results for performance (1): Level of labor productivity (linear model - OLS estimation)

Dependent variable: Log. of labor productivity

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Product or process innovation (predicted probability)	1.088***					1.670***					0.447**	0.284
	[0.221]					[0.309]					[0.201]	[0.287]
Product innovation (predicted probability)		1.119***		1.560***			1.510***		2.263***			
		[0.193]		[0.285]			[0.254]		[0.359]			
Process innovation (predicted probability)			0.865***	-0.927**				0.956***	-1.340***			
			[0.292]	[0.421]				[0.301]	[0.438]			
Product innovation only (predicted probability)					1.224***					2.077***		
					[0.354]					[0.552]		
Process innovation only (predicted probability)					-2.426***					-3.108***		
					[0.931]					[1.179]		
Product and process innovation (predicted probability)					0.560*					0.502		
					[0.294]					[0.365]		
Start-up (dummy)						0.214	-0.054	-0.096	-0.111	0.200	-1.078***	-1.248***
						[0.242]	[0.204]	[0.161]	[0.203]	[0.382]	[0.133]	[0.299]
Start-up x Product or process innovation						-0.891**						0.242
						[0.397]						[0.346]
Start-up x Product innovation							-0.606*		0.123			
							[0.360]		[0.477]			
Start-up x Process innovation								-0.430	-2.006***			
								[0.487]	[0.694]			
Start-up x Product innovation only										-0.779		
										[0.788]		
Start-up x Process innovation only										-8.436***		
										[2.448]		
Start-up x Product and process innovation										-0.409		
										[0.600]		
Predicted R&D intensity											0.707***	0.720***
											[0.051]	[0.054]
Start-up x Predicted R&D intensity												-0.021
												[0.078]
Initial employment size	0.091***	0.081***	0.120***	0.088***	0.086***	0.067**	0.058**	0.117***	0.065**	0.072***	0.114***	0.120***
	[0.028]	[0.028]	[0.028]	[0.028]	[0.028]	[0.028]	[0.028]	[0.028]	[0.028]	[0.027]	[0.024]	[0.024]
Age	0.011	0.019	0.011	0.038	0.050*	-0.063	-0.062	-0.032	-0.067*	-0.067*	-0.129***	-0.124***
	[0.029]	[0.029]	[0.029]	[0.029]	[0.030]	[0.038]	[0.038]	[0.038]	[0.038]	[0.038]	[0.035]	[0.037]
Affiliated (dummy)	0.363***	0.362***	0.363***	0.339***	0.337***	0.359***	0.357***	0.357***	0.309***	0.278***	-0.077	-0.085
	[0.067]	[0.067]	[0.068]	[0.067]	[0.067]	[0.067]	[0.067]	[0.068]	[0.066]	[0.067]	[0.066]	[0.064]
Constant	1.379***	1.462***	1.694***	1.457***	1.676***	1.302***	1.576***	1.812***	1.603***	1.912***	4.692***	4.808***
	[0.165]	[0.151]	[0.153]	[0.150]	[0.208]	[0.209]	[0.184]	[0.188]	[0.182]	[0.329]	[0.306]	[0.337]
Industry dummies (2 digit)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	897	897	897	897	897	897	897	897	897	897	897	897
F-test (statistics)	16.5	17.3	16.1	16.8	18.4	17.3	18.0	15.6	17.6	18.1	29.9	28.2
F-test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R-squared	0.387	0.392	0.374	0.396	0.398	0.393	0.399	0.376	0.413	0.422	0.524	0.525

Notes: Robust standard errors are in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Table 5.6 shows the estimation results for the growth rate of labor productivity rather than the level of labor productivity, as in Table 5.5. In general, there are not large differences in the results on the effects of process innovation and interaction terms between start-ups and product and/or process innovations. The results in column [1] to [3] in Figure 5.11 show the positive effects of product and process innovation on the labor productivity growth. There are also no significant coefficients of the interaction terms between the start-up firm dummy and product and process innovations in column [6] to [8] in Figure 5.11. These imply that the effects of product or process innovation are positive and not significantly different in start-ups and in established firms.

But in column [4] in Table 5.6 there is no significant coefficient when we jointly include product and process innovation, and in column [5] there is a significant positive coefficient only on joint introduction of product and process innovations. These results indicate the strong complementarity of product and process innovation. Moreover, the results in column [10] indicate that this complementarity works more in start-ups than in established firms. In particular, the result indicates that, for start-ups, labor productivity growth rate falls when they introduce process innovation but not product innovation.

The first six columns in Table 5.7 show the estimation results of profitability equation without control variables, and the last four columns of this table display the results with control variables. The results without control variables have almost the same implications as the results for labor productivity growth: the positive and significant effects of product and process innovation, when they are not distinguished (column [1]) or included independently (column [2] and [3]); but no significant coefficients when they are jointly included (column [4]) and when they complement each other (column [5]). There is no significant difference between start-ups and established firms in the effects of innovation on profitability (column [6]). However, these significant results disappear when we add one of the control variables (column [7] to [10]): firm age, size, or initial labor productivity. Since a dummy variable is used and a continuous variable is not used for profitability, the data may not have sufficient variation to identify these effects.

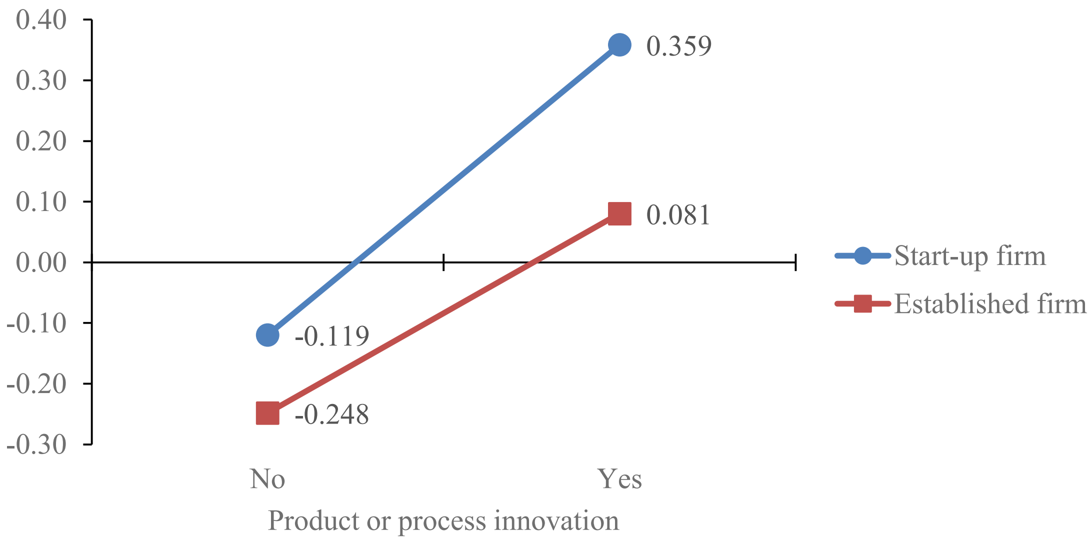
Table 5.6 Third stage results for performance (2): Growth rate of labor productivity (linear model - OLS estimation)

Dependent variable: Growth rate of labor produtivity

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Product or process innovation (predicted probability)	0.417**					0.329*					0.421**	0.099
	[0.172]					[0.178]					[0.174]	[0.184]
Product innovation (predicted probability)		0.369**		0.210			0.261*		0.060			
		[0.153]		[0.220]			[0.142]		[0.208]			
Process innovation (predicted probability)			0.567***	0.328				0.426**	0.335			
			[0.193]	[0.270]				[0.175]	[0.270]			
Product innovation only (predicted probability)					0.252					0.277		
					[0.270]					[0.326]		
Process innovation only (predicted probability)					0.522					0.080		
					[0.570]					[0.726]		
Product and process innovation (predicted probability)					0.547***					0.266		
					[0.191]					[0.229]		
Start-up (dummy)						-0.150	-0.204	-0.159	-0.196	0.151		-0.738***
						[0.163]	[0.141]	[0.111]	[0.142]	[0.260]		[0.234]
Start-up x Product or process innovation						0.149						0.414
						[0.276]						[0.283]
Start-up x Product innovation							0.251		0.157			
							[0.255]		[0.349]			
Start-up x Process innovation								0.556	0.328			
								[0.416]	[0.579]			
Start-up x Product innovation only										-0.670		
										[0.590]		
Start-up x Process innovation only										-3.885**		
										[1.941]		
Start-up x Product and process innovation										1.221*		
										[0.626]		
Predicted R&D intensity											-0.006	0.078**
											[0.032]	[0.035]
Start-up x Predicted R&D intensity												-0.213***
												[0.071]
Initial employment size	-0.001	-0.001	0.000	-0.003	-0.003	0.001	0.002	0.000	0.000	0.003	-0.001	0.011
	[0.018]	[0.018]	[0.017]	[0.018]	[0.018]	[0.017]	[0.017]	[0.017]	[0.017]	[0.017]	[0.018]	[0.018]
Age	-0.021	-0.017	-0.027	-0.024	-0.025	-0.032	-0.031	-0.030	-0.030	-0.033	-0.022	-0.036
	[0.022]	[0.022]	[0.023]	[0.022]	[0.022]	[0.025]	[0.025]	[0.024]	[0.025]	[0.025]	[0.022]	[0.025]
Affiliated (dummy)	0.087**	0.086**	0.094**	0.093**	0.093**	0.086**	0.084**	0.094**	0.093**	0.075*	0.090*	0.044
	[0.043]	[0.043]	[0.043]	[0.043]	[0.043]	[0.043]	[0.043]	[0.044]	[0.044]	[0.044]	[0.046]	[0.044]
Initial labor productivity	-0.212***	-0.214***	-0.206***	-0.211***	-0.210***	-0.212***	-0.214***	-0.202***	-0.206***	-0.211***	-0.210***	-0.207***
	[0.034]	[0.034]	[0.034]	[0.034]	[0.034]	[0.034]	[0.034]	[0.034]	[0.034]	[0.034]	[0.035]	[0.035]
Constant	0.384***	0.438***	0.460***	0.436***	0.407***	0.481***	0.549***	0.525***	0.534***	0.535**	0.361**	0.816***
	[0.108]	[0.098]	[0.096]	[0.099]	[0.134]	[0.141]	[0.127]	[0.126]	[0.130]	[0.216]	[0.169]	[0.261]
Industry dummies (2 digit)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	897	897	897	897	897	897	897	897	897	897	897	897
F-test (statistics)	2.9	2.9	2.9	2.9	2.8	2.8	2.8	2.8	2.7	2.6	2.8	2.7
F-test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R-squared	0.131	0.131	0.131	0.132	0.132	0.133	0.134	0.135	0.136	0.146	0.131	0.150

Notes: Robust standard errors are in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Figure 5.11: Marginal effects of product/process innovation on labor productivity growth



Notes: The vertical axis is the predicted growth rate of labor productivity. The predicted values are calculated from the estimation result of column [6] in Table 6 at the mean values of the remaining covariates.

Table 5.7 Third stage results for performance (3): Profitability (probit model - ML estimation)

Dependent variable: Positive profit dummy

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Product or process innovation (predicted probability)	0.824***					0.584*	0.398	0.248	-0.127	0.258
	[0.229]					[0.355]	[0.283]	[0.246]	[0.330]	[0.249]
Product innovation (predicted probability)		0.731***		0.461						
		[0.214]		[0.430]						
Process innovation (predicted probability)			1.004***	0.441						
			[0.304]	[0.609]						
Product innovation only (predicted probability)					0.774					
					[0.515]					
Process innovation only (predicted probability)					1.751					
					[1.298]					
Product and process innovation (predicted probability)					0.916***					
					[0.318]					
Start-up (dummy)						-0.150				
						[0.328]				
Start-up x Predicted product or process innovation						-0.433				
						[0.502]				
Affiliated (dummy)	0.143	0.143	0.168*	0.148	0.150	0.097	0.127	0.095	-0.033	-0.012
	[0.092]	[0.092]	[0.090]	[0.093]	[0.093]	[0.094]	[0.092]	[0.095]	[0.103]	[0.096]
Age							0.089**			
							[0.035]			
Initial profitability (positive profit dummy)								1.000***		
								[0.112]		
Initial employment size									0.136***	
									[0.034]	
Initial labor productivity										0.308***
										[0.054]
Constant	-0.365*	-0.252	-0.171	-0.244	-0.443*	-0.088	-0.259	-0.613***	-0.204	-0.676***
	[0.215]	[0.200]	[0.189]	[0.200]	[0.266]	[0.303]	[0.220]	[0.219]	[0.219]	[0.224]
Industry dummies (2 digit)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	990	990	990	990	990	990	990	979	990	990
F-test (statistics)	52.2	51.0	49.8	51.3	52.7	63.6	56.8	124.5	67.3	81.3
F-test (p-value)	0.002	0.002	0.003	0.003	0.003	0.000	0.001	0.000	0.000	0.000

Notes: Robust standard errors are in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

4.4. Robustness checks

University collaboration maybe endogenous with regard to product and process innovation. Table 5.8 shows the results of bivariate probit model of product and process innovation with predicted probability of university collaboration as an explanatory variable, in which the density of university in the region is used as the instrument for the collaboration with a university. Compared to the results shown in Table 5.4, the estimated coefficients on university collaboration are smaller. However, while the effects of university collaboration on established firms' product and process innovation are no longer significant, the coefficient of the interaction term between university collaboration and start-up firms is still significantly positive. This suggests that the positive effect of university collaboration on start-up firms' product innovation is robust.

We perform another robustness checks. Since the data sources are different between start-up firms and established firms, data generating processes might also be different between them. Therefore, we separate the sample into start-up firms and established firms and estimate the same models for each group. Table 5.9, Table 5.10 and Table 5.11 show the estimation results of the first stage, the second stage, and the third stage model, respectively. The difference between start-up firms and established firms in the effects of public financial support on R&D intensity, in the effects of collaboration on innovations, and in the effects of product innovation on labor productivity is almost similar to the results of the pooled sample with interaction terms with start-up firm dummy variable shown in Table 5.3, Table 5.4 and Table 5.5.

Finally, the definition of established firms in this chapter (operating for more than 2 years) may be arbitrary. Specifically, it may be problematic that the group of established firms also includes several start-ups (for example, those operating for more than 2 years but still less than 5 years). We may focus on older, really established firms. In this sense, we check the sensitivity of the estimation results to this definition of established firms and confirmed that the results do not significantly change even if we redefine established firms as those operating for 10 years or more, 20 years or more, or 30 years or more.

5. Conclusion

In this chapter, the differences between start-ups and established firms are empirically examined with respect to determinants of R&D and innovation and the relationship between innovation and firm performance using a comprehensive datasets derived from two surveys on innovation activities in Japanese private firms in the last years of the first decade of the new century; one is the survey of start-ups and another is the Japanese national innovation survey. Empirical results suggest that 1) the local labor pooling of research relevant workforces (professional and technical occupations) positively relates to the R&D intensity of the firms located in neighborhood, 2) the effects of public financial support on R&D intensity are generally positive but smaller for start-ups, 3) the effects of research cooperation with business partners or universities on innovation are generally positive but larger for start-ups, and 4) the effects of product and process innovation on labor productivity (level and growth) are positive both for start-ups and established firms.

Among the control variables, we found interesting results on the effects of difference between affiliated firm and independent firm (affiliated firm dummy) on R&D, innovation and performance: In comparison to independent firms, affiliated firm is more likely to conduct R&D but their R&D intensity is lower (Table 5.3) and less likely to innovate their product (Table 5.4) while the level and growth rate of labor productivity is higher (Table 5.5). These mixed results may indicate that firms within business group take some specific R&D and innovation strategy.

However, this study has several limitations: First, an appropriate correction for the reported standard errors is needed. Second, the correction for endogeneity in public subsidies and R&D cooperation should be examined. Third, differences in intensity, magnitude, or quality of innovations between firms are ignored. Fourth, since we do not identify survival effects and aging effects in the empirical model, our estimates of the differences between start-ups and established firms are mix of both effects.

Despite these limitations, empirical results imply that in order to promote innovation and growth of start-ups, we should provide more or better support for start-ups to engage in research cooperation with both business partners and universities, rather than the financial support. In general, start-up firms have scarce internal knowledge and R&D stock compared to established or mature firms, despite their greater incentives for innovation; and they rely heavily on external knowledge and research collaboration with others. These findings indicate that governments can accelerate innovation and productivity growth more efficiently by promoting research collaborations between start-up firms and universities and between start-ups and their business partners, rather than by increasing public financial supports for start-ups.

Table 5.8 Robustness check for the second stage model with additional instruments for university collaboration

(bivariate probit model - ML estimation)

Dependent variables: Dummy variables indicating the introduction of product innovation and process innovation

	[1]		[2]	
Dependent variable	Product	Process	Product	Process
Predicted R&D intensity	0.201***	-0.021	0.336***	0.006
	[0.077]	[0.080]	[0.091]	[0.090]
Collaboration with business partners (dummy)	0.619***	0.479***	0.412***	0.401***
	[0.075]	[0.078]	[0.096]	[0.093]
Predicted probability to collaborate with universities	-0.017	0.016	-0.288	-0.044
	[0.420]	[0.383]	[0.423]	[0.396]
Information from competitors (dummy)	0.198***	0.013	0.344***	0.098
	[0.076]	[0.078]	[0.100]	[0.093]
Initial employment size	0.166***	0.095***	0.168***	0.094**
	[0.035]	[0.034]	[0.037]	[0.036]
Age	0.077	0.057	0.095*	0.066
	[0.050]	[0.048]	[0.051]	[0.051]
Affiliated (dummy)	-0.229**	-0.133	-0.260**	-0.139
	[0.104]	[0.105]	[0.107]	[0.109]
Start-up (dummy)	-0.072	-0.239	-0.928***	-0.381
	[0.163]	[0.173]	[0.341]	[0.369]
Start-up x Predicted R&D intensity			-0.279***	-0.047
			[0.103]	[0.111]
Start-up x Collaboration with business partners			0.558***	0.230
			[0.155]	[0.173]
Start-up x Predicted prob. to collaborate with universities			1.646*	0.479
			[0.922]	[0.970]
Start-up x Information from competitors			-0.334**	-0.263
			[0.158]	[0.174]
Industry density – city	-0.001	-0.004	-0.001	-0.004
	[0.002]	[0.002]	[0.002]	[0.003]
Industry density – prefecture	0.011	0.034	0.017	0.036
	[0.033]	[0.037]	[0.034]	[0.038]
Constant	-0.252	-1.185***	0.191	-1.098***
	[0.346]	[0.360]	[0.376]	[0.387]
Industry dummies (2 digit)	Yes	Yes	Yes	Yes
# of observations	1,383		1,383	
Chi-squared (statistics)	450.1029		453.54	
Chi-squared (p-value)	0.000		0.000	
Correlation between errors	0.379		0.377	

Notes: Robust standard errors are in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Instrumental variables for collaboration with universities are university density at city and prefecture level.

Table 5.9 Robustness check for the first stage model with separate specification

(Generalized tobit model - ML estimation)

	Established firms		Start-up firms	
Dependent variable	R&D>0	R&D int.	R&D>0	R&D int.
Initial labor productivity	0.982***	0.345***	-0.028	-0.092*
	[0.116]	[0.050]	[0.093]	[0.049]
Initial employment size	-0.011	0.231***	-0.503***	0.039
	[0.057]	[0.031]	[0.116]	[0.047]
Age	0.018	0.052		
	[0.093]	[0.045]		
Affiliated (dummy)	0.341**	-0.181**	0.797***	-0.194
	[0.164]	[0.091]	[0.276]	[0.127]
Public financial support (dummy)	0.701***	0.452***	0.259	-0.001
	[0.162]	[0.097]	[0.226]	[0.100]
Expert ratio – city	8.852***	3.879**	-1.495	1.699
	[2.873]	[1.688]	[3.781]	[1.856]
Expert ratio – prefecture	-0.276	3.259	-2.108	14.781***
	[6.986]	[4.164]	[10.124]	[4.909]
Industry density – city	0.001	-0.001	-0.002	0.001
	[0.005]	[0.003]	[0.006]	[0.002]
Industry density – prefecture	-0.156*	-0.072	-0.049	0.030
	[0.094]	[0.055]	[0.119]	[0.054]
Univ. density – city	0.141	0.553	1.188	0.975
	[1.021]	[0.653]	[1.760]	[0.632]
Univ. density – prefecture	12.316	4.218	42.554*	-31.219***
	[15.611]	[9.416]	[25.392]	[11.572]
Constant	-7.440***	-3.194***	-1.175	-2.521***
	[1.221]	[0.557]	[1.653]	[0.626]
Industry dummies (2 digit)	Yes	Yes	Yes	Yes
# of observations	1,283		881	
# of firms no R&D	692		609	
Chi-squared (statistics)	271.0883		.	
Chi-squared (p-value)	0.000		.	
Correlation between errors	0.587		0.240	

Notes: Robust standard errors are in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Table 5.10 Robustness check for the second stage model with separate specification

(bivariate probit model - ML estimation)

	Established firms		Startup firms	
Dependent variable	Product innovation	Process innovation	Product innovation	Process innovation
Predicted R&D intensity	0.185***	-0.011	0.902*	0.301
	[0.058]	[0.055]	[0.484]	[0.583]
Collaboration with business partners (dummy)	0.353***	0.386***	0.948***	0.641***
	[0.100]	[0.096]	[0.130]	[0.161]
Collaboration with universities (dummy)	0.364***	0.122	0.729***	0.650***
	[0.125]	[0.111]	[0.200]	[0.207]
Information from competitors (dummy)	0.353***	0.087	0.084	-0.130
	[0.103]	[0.095]	[0.130]	[0.162]
Initial employment size	0.118***	0.087**	0.652***	0.306
	[0.036]	[0.035]	[0.244]	[0.295]
Age	0.122**	0.073		
	[0.053]	[0.053]		
Affiliated (dummy)	-0.036	-0.095	-1.083***	-0.592
	[0.118]	[0.110]	[0.404]	[0.483]
Industry density – city	-0.002	-0.006	0.003	0.001
	[0.003]	[0.004]	[0.004]	[0.005]
Industry density – prefecture	-0.017	0.052	0.095	0.108
	[0.061]	[0.060]	[0.070]	[0.087]
Univ. density – city	-0.718	0.429	-1.555*	-0.123
	[0.676]	[0.646]	[0.904]	[1.134]
Univ. density – prefecture	8.010	-5.374	-30.200	-24.747
	[7.881]	[7.130]	[19.825]	[24.140]
Constant	-0.229	-1.076***	0.420	-1.102
	[0.335]	[0.331]	[0.823]	[0.988]
Industry dummies (2 digit)	Yes	Yes	Yes	Yes
# of observations	872		511	
Chi-squared (statistics)	233.2894		147.3032	
Chi-squared (p-value)	0.000		0.000	
Correlation between errors	0.270		0.590	

Notes: Robust standard errors are in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Table 5.11 Robustness check for the third stage model with separate specification

Dependent variable: Level of labor productivity (linear model - OLS estimation)

	[1] Established firms	[2] Startup firms	[3] Established firms	[4] Startup firms	[5] Established firms	[6] Startup firms
Product or process innovation (predicted probability)	1.894***	0.513*				
	[0.328]	[0.308]				
Product innovation (predicted probability)			2.026***	0.518*		
			[0.262]	[0.304]		
Process innovation (predicted probability)					-0.044	0.958*
					[0.337]	[0.535]
Initial employment size	0.069**	0.116*	0.041	0.116*	0.160***	0.120**
	[0.031]	[0.062]	[0.030]	[0.062]	[0.030]	[0.061]
Age	-0.080**		-0.096**		-0.018	
	[0.038]		[0.038]		[0.039]	
Affiliated (dummy)	0.313***	0.128	0.288***	0.125	0.331***	0.129
	[0.070]	[0.184]	[0.069]	[0.184]	[0.071]	[0.184]
Constant	1.129***	1.650***	1.397***	1.659***	1.990***	1.735***
	[0.233]	[0.291]	[0.197]	[0.288]	[0.220]	[0.279]
Industry dummies (2 digit)	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	674	223	674	223	674	223
R-squared	0.359	0.204	0.389	0.205	0.322	0.205

Notes: Robust standard errors are in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Chapter 6 Conclusion

This thesis investigated several sources of the agglomeration economies in each chapter. In Chapter 2, the effects of transportation costs were examined. Combining a spatial demand function derived from the theoretical literature on the NEG (e.g., Krugman 1980; Fujita et al. 1999) with a production function, a revenue function was proposed, which would capture the effects of transportation costs on a firm's revenue. Since such spatial effects are generated by the transportation costs on a firm's own products and its intermediate goods, the suggested revenue function makes it possible to relate the geographic agglomeration economy to the transportation costs, which was not conducted in previous empirical studies. The proposed model was empirically examined using the regional panel data of the manufacturing sector in Japan. The estimation results of the revenue function show significant, robust, and positive transportation costs for the manufacturing products. Moreover, consistent GMM estimation results present evidence of positive transportation costs for the outputs of both the manufacturing and primary and service sectors. These outcomes indicate that the efficiency of the manufacturing firms depends on their access to the markets and the intermediate goods supply.

Chapter 3 examined the effects of R&D spillovers on TFP, with a large panel dataset of Japanese manufacturing plants matched with R&D survey data. This chapter simultaneously analyzed the role of public (universities and research institutes) and private R&D spillovers, while examining the effects due to 'relational' (supplier-customer) proximity, as well as technological and geographic proximity. My analysis confirms the importance of positive spillover effects from R&D by the firms with plants in technologically related industries. The latter spillover effects are attenuated by distance, and my estimates suggest that most spillover effects disappear beyond 500 km. Positive impacts of public R&D spillovers are also observed, with the effects substantially larger for the plants with access to internal R&D. However, no evidence shows that public R&D spillover effects are attenuated by distance.

'Relational proximity' due to buyer and supplier linkages generates additional 'pecuniary' R&D spillovers of a similar magnitude as the knowledge spillovers due to technological proximity. However, the role of geographic distance in these buyer and supplier spillovers cannot be identified. The chapter concludes that both public and private R&D spillovers matter for TFP growth, while relational and technological proximity should be considered to arrive at representative estimates of the social effects of private R&D.

Decomposition analysis shows that the contribution of private R&D spillovers to TFP growth has declined since the late 1990s. This is due to a declining growth in R&D stocks, while another important factor is the exit of proximate plants operated by R&D-intensive firms. A mildly decreasing contribution of public R&D spillovers is primarily due to a reduced growth of R&D by public research organizations since the late 1990s. Exploring the effects at the regional level, we observe strong, adverse exit effects occurring particularly in Japan's major industrial agglomerations such as Tokyo and Osaka. The results help explain the two stylized facts of the Japanese productivity growth: the exit of relatively productive plants and the declining TFP growth of surviving plants (Fukao and Kwon 2006; Kneller et al. 2012). They suggest that these two trends may be causally related. The exit of plants owned by R&D-intensive firms reduces the available R&D spillovers and hampers the TFP growth of the surviving plants.

Chapter 4 investigated the determinants of regional business entry, distinguishing between independent start-ups and subsidiaries, with a special focus on the effects of regional human capital. It is this chapter's major contribution to the literature. Another one is the comparison of the determinants of regional entry between the manufacturing and service sectors, as well as across subsectors within the service sector. For the empirical analyses, pooled data of 47 Japanese prefectures for four observation periods were used. The estimation results of SUR indicate considerable differences in the impact of regional factors between independent start-ups and subsidiaries, as well as among different industries.

First, the ratio of college graduates is correlated negatively with independent start-ups (Hypothesis 1b) and positively with the entry of subsidiaries (Hypothesis 2). Second, the ratio of professional and technical workers positively affects independent start-ups but not the entry of subsidiaries (Hypothesis 3). Third, the relationships between regional human capital and the entry of independent start-ups and new subsidiaries are different across industries.

The determinants of entry differ not only between the manufacturing and service sectors, but also within the service sector. Moreover, the differences of the determinants between the start-up types vary across sectors. The negative relationship between the ratios of college graduates and independent start-ups (for the overall industry and the service sector) is consistent with the previous empirical evidence for Japan (e.g., Small and Medium Enterprise Agency 2002; Okamuro 2008), but not with the results from other countries (Guesnier 1994; Armington and Acs 2002; Acs and Armington 2004). This suggests that highly educated workers in Japan regard the opportunity cost of quitting their current jobs as considerably higher than the expected returns from start-up their own business.

The estimation results for the manufacturing sector are not only different from, but almost contrasting those for the service sector. For manufacturing, the ratio of college graduates is positively correlated to that of independent start-ups. This may imply that a high educational level (particularly in the natural sciences) is especially important for independent start-ups in the manufacturing sector or that highly educated workers in the manufacturing sector do not have a high opportunity cost of self-employment, assuming that the founders of manufacturing firms come from the manufacturing sector. Moreover, for the manufacturing sector, the ratio of college graduates has no significant effect on start-ups of new subsidiaries, while the ratio of professional and technical workers has a negative impact on independent start-ups.

The latter results are different from those of the previous studies (Guesnier 1994; Hart and Gudgin 1994) that found a positive relationship for the manufacturing sector. It may be attributed to the definitions of professional and technical occupations in Japan that cover numerous skills in the service sector. It is also noteworthy that this study's results regarding the effect of the unemployment ratio on the ratio of independent start-ups clearly support the push hypothesis, similar to the findings of previous research in Japan. This study suggests that regional policies to activate business start-ups should recognize the differences between encouraging local entrepreneurship and attracting new subsidiaries. These differences may vary even within the service sector, according to technological intensity (or innovativeness).

In Chapter 5, the differences between start-ups and established firms were empirically examined, with respect to the determinants of R&D and innovation and the relationship between innovation and firm performance. A comprehensive dataset was used, derived from two surveys on innovation activities in Japanese private firms; one was an original survey of start-ups and the other was the Japanese National Innovation Survey. The empirical results show that 1) the local labor pooling of the research-relevant workforce (professional and technical occupations) positively relates to the R&D intensity of the firms located in the neighborhood, 2) the effects of public financial support on R&D intensity are generally positive but smaller for start-ups, 3) the impacts of research cooperation with business partners or universities on innovation are generally positive but larger for start-ups, and 4) the effects of product and process innovation on labor productivity (level and growth) are positive for both start-ups and established firms.

The contribution of this thesis to the literature is its empirical examination of the mechanisms of regional economic growth, using the data from Japan in recent years. In sum, the results suggest the importance of Marshall's (1920) three sources of the agglomeration economies (input sharing, knowledge spillovers, and labor pooling) in enhancing regional productivity growth, business start-ups, or innovation activities. However, some limitations of this thesis should be addressed in future research. First, the three sources of agglomeration economies were examined separately. The relative importance of the sources should be different, may also vary across industries, and have been changing over time.

Second, to investigate effective regional and location policies, additional empirical analyses of the dynamics of firm and labor supply locations are needed. From the results of Chapter 2, only comparative statics can be performed. From the dynamic perspective, other essential aspects should be considered, including relocation cost, entry cost, or time lags. They can be analyzed only in dynamic models of location choice.

Third, this study ignores the export and import activities and roles of trade hubs, such as harbors, airports, or train stations. Since the distance from such trade hubs should affect both the transportation costs and efficiency of firms, it is necessary to control for such effects.

Fourth, the lack of estimated distance effects for public R&D (Chapter 3) may be because public R&D spillovers occur most often through active collaboration across longer distances (Gittelman 2007; Okamuro and Nishimura 2013). This explanation can be explored by incorporating the available information on research relationships between firms and universities.

Fifth, this thesis disregards the effects of overseas R&D conducted or outsourced by the firms and the potential, international knowledge transfers and spillovers (e.g., Branstetter 2001; Griffith et al. 2008).

Finally, the positive relationships between the regional structure of human capital and the ratio of independent start-ups (Chapter 4) may have two explanations. One reason is the entrepreneurs' accumulation of human capital within the region; the other is the migration of high-skilled workers (e.g., Ritsila and Ovaskainen 2001). On the basis of the data available in this study, it is not currently possible to distinguish between these effects.

Bibliography

Acs, Z. J., and Armington, C. (2004a). Employment Growth and Entrepreneurial Activity in Cities. *Regional Studies* 38: 911–927.

Acs, Z. J., and Armington, C. (2004b). The Impact of Geographic Differences in Human Capital on Service Firm Formation Ratio. *Journal of Urban Economics* 56: 244–278.

Acs, Z. J., and Armington, C. (2006). Entrepreneurship, Geography, and American Economic Growth. New York: Cambridge University Press.

Adams, J. (1990). Fundamental Stocks of Knowledge and Productivity Growth. *Journal of Political Economy* 98: 673–702.

Adams, J. D., and Jaffe, A. B. (1996). Bounding the Effects of R&D: An Investigation Using Matched Establishment-firm Data. *Rand Journal of Economics* 27: 700–721.

Aldieri, L., and Cincera, M. (2009). Geographic and Technological R&D Spillovers within the triad: micro evidence from US patents. *Journal of Technology Transfer* 34(2): 196–211.

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*. Cambridge: Harvard University Press.

Alvarez, I., and Molero, J. (2005). Technology and the generation of international knowledge spillovers: An application to Spanish manufacturing firms. *Research Policy* 34: 1440–1452.

Anselin, L., Varga, A., and Acs, Z. (1997). Local Geographic Spillovers between University Research and High Technology Innovations. *Journal of Urban Economics* 42: 422–448.

Arellano, M., and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* 58: 277–297.

Argyres, N., and Silverman, B. (2004). R&D, Organization Structure, and the Development of Corporate Technological Knowledge. *Strategic Management Journal* 25: 929–958.

Armington, C., and Acs, Z. J., 2002. The Determinants of Regional Variation in New firm Formation. *Regional Studies* 36: 373–386.

Audretsch, D. B., and Fritsch, M. (1994a). The Geography of Firm Births in Germany. *Regional Studies*. 28 (4): 359–365.

Audretsch, D. B., and Fritsch, M. (1994b). On the Measurement of Entry Rates’, *Empirica* 21: 105–113.

Audretsch, D. B., and Vivarelli, M. (1996). Determinants of New-firm Start-ups in Italy. *Empirica* 23 (1): 91–105.

Audretsch, D. B., and Feldman, P. (1996). R&D Spillovers and the Geography of Innovation and Production. *American Economic Review* 86: 630–640.

Audretsch, D. B., and Keilbach, M. (2005). Entrepreneurship Capital and Regional Growth. *Annals of Regional Science* 39 (3): 457–469.

Aw, B. Y., Chen, X., and Roberts, M. J. (2001). Firm-level Evidence on Productivity Differentials and Turnover in Taiwanese Manufacturing. *Journal of Development Economics* 66: 51–86.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G., and Robert-Nicoud, F. (2003). Economic Geography and Public Policy: Princeton University Press.

Becker, G. S. (1975). *Human Capital*. New York: National Bureau of Economic Research.

Belderbos, R., Carree, M., and Lokshin, B. (2004). Cooperative R&D and Firm Performance. *Research Policy* 33 (10): 1477–1492.

Belderbos, R., Ikeuchi, K., Fukao, K., Kim, Y. G., and Kwon, H. U. (2013). Plant Productivity Dynamics and Private and Public R&D Spillovers: Technological, Geographic and Relational Proximity. *Center for Economic Institutions Working Paper Series*, (2013-5).

Belderbos, R., Leten, B. and Suzuki, S. (2009). Does Academic Excellence Attract Foreign R&D? *Global COE Hi-Stat Discussion Paper Series*, Institute for Economic Research, Hitotsubashi University.

Bloom, N., Schankerman, M. and van Reenen, J. (2013). Identifying Technology Spillovers and Product Market Rivalry. *Econometrica* 81(4): 1347–1393.

Blundell, R., and Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* 87: 115–143.

Bosma, N., Van Stel, A., and Suddle, K. (2008). The Geography of New Firm Formation: Evidence from Independent Start-ups and New Subsidiaries in the Netherlands. *International Entrepreneurship and Management Journal* 4 (2): 129–146.

Brakman, S., Garretsen, H., Gorter, J., Horst, A. V. D., and Schramm, M. (2009). New Economic Geography, Empirics, and Regional Policy. *CPB Special Publication* 56, CPB Netherlands Bureau for Economic Policy Analysis.

Branstetter, L. (2000). Vertical Keiretsu and Knowledge Spillovers in Japanese Manufacturing: An Empirical Assessment. *Journal of the Japanese and International Economies* 14(2): 73–104.

Branstetter, L. (2001). Are Knowledge Spillovers International or Intra-national in Scope? Microeconometric evidence from Japan and the United States. *Journal of International Economics* 53: 53–79.

Brulhart, M., and Mathys, N. A. (2008). Sectorial Agglomeration Economies in a Panel of European Regions. *Regional Science and Urban Economics* 38(4): 348–362.

Breschi, S., Lissoni, F. and Malerba, F., 2003. Knowledge-relatedness in Firm Technological Diversification. *Research Policy* 32: 69–87.

Cassiman, B., and Veugelers, R. (2006). In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition. *Management Science* 52(1): 68–82.

Carpenter, R. E., and Petersen, B. C. (2002). Capital Market Imperfections, High-tech Investment, and New Equity Financing. *Economic Journal* 112: F54–F72.

Carree, M. A. (2002). Does Unemployment Affect the Number of Establishments? A Regional Analysis for US States. *Regional Studies* 36 (2): 389–398.

Carree, M., Lokshin, B., and Belderbos, R. (2006). Complementarity in R&D Cooperation Strategies. *Review of Industrial Organization* 28 (4): 401–426.

Cefis, E., and Marsili, O. (2006). Survivor: The Role of Innovation in Firms' Survival. *Research Policy* 35 (5): 626–641.

Ciccone, A. (2002). Agglomeration Effects in Europe. *European Economic Review* 46: 213–227.

Ciccone, A., and Hall, R. E. (1996). Productivity and the Density of Economic Activity. *American Economic Review* 86: 54–70.

Cohen, W. M., and Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly* 35(1) Special Issue: Technology, Organizations, and Innovation: 128–152.

Colombo, M.G., and Grilli, L. (2005). Founders' Human Capital and the Growth of New Technology-based Firms: a Competence-based View. *Research Policy* 34: 795–816.

Combes, P., Duranton, G., Gobillon, L., and Roux, S. (2008). Estimating Agglomeration Economies with History, Geology, and Worker Effects. *CEPR Discussion Paper* No. DP6728.

Coughlin, C. C., and Segev, E. (2000). Location Determinants of New Foreign-owned Manufacturing Plants. *Journal of Regional Science* 40 (2): 323–351.

Crespi, G., Criscuolo, C., Haskel, J., and Slaughter, M. (2007). Productivity Growth, Knowledge Flows and Spillovers. *CEP Discussion Papers*.

Crépon, B., Duguet, E., and Mairesse, J. (1998). Research, Innovation and Productivity: An Econometric Analysis at the Firm Level. *NBER Working Papers* No.6696.

Crozet, M. (2004). Do Migrants Follow Market Potentials? An Estimation of a New Economic Geography Model. *Journal of Economic Geography* 4: 439–458.

Davis, D. R, and Weinstein, D. E. (2008). A Search for Multiple Equilibria in Urban Industrial Structure. *Journal of Regional Science* 48(1): 29–65.

Davidsson, P., Lindmark, L., and Olofsson, C. (1994). New Firm Formation and Regional Development in Sweden. *Regional Studies* 28: 395–410.

Dixit, A. K., and Stiglitz J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review* 67(3): 297–308.

Duranton, G. and Overman, H. G. (2005). Testing for Localization Using Micro-Geographic Analysis. *Review of Economic Studies*, 72(4): 1077–1106.

Egeln, J., Licht, G., and Steil, F. (1997). Firm Foundations and the Role of Financial Constraints. *Small Business Economics* 9: 137–150.

Ellison, G., and Glaeser, E. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy* 105: 889–927.

Ellison, G., Glaeser, E. L., and Kerr, W. R. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review* 100 (3): 1195–1213.

Evans, D. S., and Leighton, L. S. (1989). The Determinants of Changes in U.S. Self-employment, 1968–1987. *Small Business Economics* 1 (2): 111–119.

Evans, D. S., and Leighton, L. S. (1990). Small Business Formation by Unemployed and Employed Workers. *Small Business Economics* 2 (4): 319–330.

Figueiredo, O., Guimaraes, P., and Woodward, D. (2002). Home-field Advantage: Location Decisions of Portuguese Entrepreneurs. *Journal of Urban Economics* 52: 341–361.

Fosfuri, A., and Ronde T. (2004). High-tech Clusters, Technology Spillovers, and Trade Secret Laws. *International Journal of Industrial Organization* 22: 45–65.

Fotopoulos, G., and Spence, N. (1999). Spatial Variations in New Manufacturing Plant Openings: Some Empirical Evidence from Greece. *Regional Studies* 33: 219–229.

Fujita, M., Krugman, P., and Venables, A. (1999). *The Spatial Economy: Cities, Regions and International Trade*. MIT Press.

Fujita, M., and Thisse, J. F. (2002). *Economics of Agglomeration*. Cambridge University.

Fukao, K. and Kwon, H. U. (2011). The Key Drivers of Future Growth in Japan, presentation prepared for the CCJ Growth Strategy Task Force White Paper, June 10.

Fukao, K, Kim, Y. G. and Kwon. H. U. (2008). *Plant Turnover and TFP Dynamics in Japanese Manufacturing*. In A. Heshmati and J.-D. Lee (eds.), *Micro-Evidence for Dynamics of Industrial Evolution*, Nova Science Publishers, Inc., Ch. 3, 23–59.

Fukao, K. and Kwon, H. U. (2006). Why Did Japan's TFP Growth Slowdown in the Lost Decade? An Empirical Analysis Based on Firm–level Data of Manufacturing Firms. *Japanese Economic Review* 57(2): 195–228.

Furman, J., Kyle, M., Cockburn, I., and Henderson, R. (2005). Public and Private Spillovers, Location and the Productivity of Pharmaceutical Research. *Annales d'Economie et de Statistique*, 79/80: 165–188.

Gerlach, K. and Wagner, J. (1994). Regional Differences in Small Firm Entry in Manufacturing Industries: Lower Saxony, 1979–1991. *Entrepreneurship and Regional Development* 6: 30–80.

Gittelman, M. (2007). Does Geography Matter for Science–Based Firms? Epistemic Communities and the Geography of Research and Patenting in Biotechnology. *Organization Science* 18 (4): 724–741.

Glaeser, E. L. (1999). *Learning in Cities. Journal of Urban Economics* 46(2): 254–277.

Glaeser, E. L., Kallal, H., Scheinkman, J. A., and Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy* 100 (6): 1126–1152.

Goldstein, G. S., and Gronberg, T. J. (1984). Economies of scope and economies of agglomeration. *Journal of Urban Economics* 16(1): 91–104.

Good, D. H., Nadiri, M. I. and Sickles, R. C. (1997). Index Number and Factor Demand Approaches to the Estimation of Productivity. In M. H. Pesaran and P. Schmidt (eds.), *Handbook of Applied Econometrics: Vol.2. Microeconometrics*, Basil Blackwell, 14–80.

Görg, H. and Strobl, E. (2001). Multinational Companies and Productivity Spillovers: A Meta–Analysis. *Economic Journal* 111(475): F723–F739.

Goto, A. and Suzuki, K. (1989). R&D Capital, Rate of Return on R&D Investment and Spillover of R&D in Japanese Manufacturing Industries. *Review of Economics and Statistics* 71: 555–564.

Griffith, R., Redding, S. and Simpson, H. (2009). Technological Catch–up and Geographic Proximity. *Journal of Regional Science* 49: 689–720.

Griffith, R., Harrison, R. and van Reenen, J. (2008). How Special is the Special Relationship? Using the Impact of US R&D Spillovers on UK Firms as a Test of Technology Sourcing. *American Economic Review* 96(5): 1859–1875.

Griffith, R., Huergo, E., Mairesse, J., and Peters, B. (2006). Innovation and Productivity across Four European Countries. *Oxford Review of Economic Policy* 22 (4): 483–498.

Griliches, Z. (1992). The Search for R&D Spillovers. *Scandinavian Journal of Economics* 94: 29–47.

Guesnier, B. (1994). Regional Variations in New Firm Formation in France. *Regional Studies* 28 (4): 347–358.

Hall, B. H, Lotti, F., and Mairesse, J. (2008). Employment, Innovation, and Productivity: Evidence from Italian Microdata. *Industrial and Corporate Change* 17 (4): 813–839.

Hall, B. H, Lotti, F., and Mairesse, J. (2009). Innovation and Productivity in SMEs: Empirical Evidence for Italy. *Small Business Economics* 33 (1): 13–33.

Hall, B. H., Mairesse, J. and Mohnen, P. (2012). Measuring the Returns to R&D. In B. Hall and N. Rosenberg (eds.), *Handbooks in Economics: Economics of Innovation Volume 2*. North–Holland. 1034–1074.

Hall, B. H. and Oriani R. (2006). Does the Market Value R&D Investment by European Firms? Evidence from a Panel of Manufacturing Firms in France, Germany, and Italy. *International Journal of Industrial Organization* 5: 971–993.

Haneda, S., and Ito, K. (2013). Modes of international activities and the innovativeness of firms: an empirical analysis based on the Japanese National Innovation Survey for 2009. *Economics of Innovation and New Technology*. 23 (8): 758–779.

Hanson, G. H. (2005). Market Potential, Increasing Returns, and Geographic Concentration. *Journal of International Economics* 67(1): 1−24.

Hart, M. and Gudgin, G. (1994). Spatial Variations in New Firm Formation in the Republic of Ireland, 1980–1990. *Regional Studies* 28 (4): 367–380.

Harrison, R., Jaumandreu, J., Mairesse, J., and Peters, B. (2008). Does Innovation Stimulate Employment? A Firm–Level Analysis Using Comparable Micro–Data from Four European Countries. *NBER Working Papers* No. 14216.

Haskel, J. E., Slaughter, M. J., Jones, C., Keller, W., Markusen, J., Nickell, S., Petrin, A., and van Reenen, J. (2005). Global Engagement and the Innovation Activities of Firms. *NBER Working Papers* No. 11479.

Haskel, J., Pereira, S. and Slaughter, M. (2007). Does Inward Foreign Investment Boost the Productivity of Domestic Firms? *The Review of Economics and Statistics* 89(3): 482–496.

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement* 5 (4): 475–492.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica* 47 (1): 153–161.

Helsley, R. W., and Strange, W. C. (2002). Innovation and Input Sharing. *Journal of Urban Economics* 51(1): 25–45.

Henderson, J. V. (2007). Understanding Knowledge Spillovers. *Regional Science and Urban Economics* 37: 497−508.

Henderson, R. and Cockburn, I. (1996). Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery. *Rand Journal of Economics* 27: 31–59.

Honjo, Y., Kato, M., and Okamuro, H. (2014). R&D financing of start-up firms: How much does founders' human capital matter? *Small Business Economics*, 42 (2): 207–220.

Huergo, E., and Jaumandreu, J. (2004a). How Does Probability of Innovation Change with Firm Age ? *Small Business Economics* 22: 193–207.

Huergo, E., and Jaumandreu, J. (2004b). Firms' Age, Process Innovation and Productivity Growth. *International Journal of Industrial Organization* 22 (4): 541–559.

Ikeuchi, K., and Okamuro, H. (2011). Effects of regional human capital structure on business entry: a comparison of independent start-ups and new subsidiaries in different industries. In M. Raposo, D. Smallbone, B. Karoly, and L. Hortoványi (Eds.), *Entrepreneurship, Growth and Economic Development: Frontiers in European Entrepreneurship Research* (pp. 118–153). Edward Elgar Publishing Limited.

Ikeuchi, K., Fukao, K., Belderbos, R., Kwon, H. U., and Kim, Y. G. (2013). Plant Location and the Effects of Private and Public R&D Spillovers: an Analysis of the Productivity Effects of R&D Spillovers via Technological, Geographic and Relational Proximity (*Kou-jou ri-tchi to min-kan/kou-teki R&D spillover kou-ka: gi-jutsu-teki/chi-ri-teki/kan-kei-teki kin-setsu-sei wo tsu-ji-ta spillover no sei-san-sei-kou-ka no bun-seki* in Japanese). *NISTEP Discussion Paper* (93).

Ikeuchi, K., and Okamuro, H. (2013). R&D, Innovation, and Business Performance of Japanese Start-ups: A Comparison with Established Firms. *NISTEP Discussion Paper* (104).

Isogawa, D., Nishikawa, K., and Ohashi, H. (2012). New–to–Market Product Innovation and Firm Performance: Evidence from a Firm–Level Innovation Survey in Japan. *RIETI Discussion Paper Series* No. 12–E–077.

Jaffe, A. B. (1989). Real Effects of Academic Research. *American Economic Review* 79: 957–970.

Jaffe, A. B., Trajtenberg, M. and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics* 108: 577–598.

Javorcik, B. S. (2004). Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages. *American Economic Review* 94(3): 605–627.

Kain, J. F. (1962). The Journey-to-Work as a Determinant of Residential Location. *Papers of the Regional Science Association* 9(1): 137–160.

Kamien, M. I., and Schwartz, N. L. (1978). Self-financing of an R&D Project. *American Economic Review* 68: 252–261.

Kato, M., Okamuro, H., and Honjo, Y. (2015). Does Founders' Human Capital Matter for Innovation? Evidence from Japanese Start–Ups. *Journal of Small Business Management*. 53 (1): 114–128.

Keeble, D. and Walker, S. (1994). New Firms, Small Firms, and Dead Firms: Spatial Patterns and Determinants in the United Kingdom. *Regional Studies* 28 (4): 411–427.

Klette, T. J. (1996). R&D, Scope Economics, and Plant Performance. *Rand Journal of Economics* 27: 502–522.

Klette, T. J., and Griliches, Z. (1996). The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous. *Journal of Applied Econometrics* 11 (4): 343–361.

Klette, T. J. and Johansen, F. (1998). Accumulation of R&D Capital and Dynamic Firm Performance: A Not–so–fixed Effect Model. *Annals of Economics and Statistics* 49/50: 389–419.

Kneller. R., McGowan, D., Inui, T. and Matsuura, T. (2012). Globalization, Multinationals and Productivity in Japan's Lost Decade. *Journal of the Japanese and International Economies* 26: 110–112.

Krugman, P. R. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review* 70: 950−959.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy* 99(3): 483–499.

Kugler, M. (2006). Spillovers from Foreign Direct Investment: Within or Between Industries? *Journal of Development Economics* 80: 444–477.

Kwon, H. U., Fukao, K., and Kim, Y. G. (2008). Innovation and Productivity Growth: An Empirical Analysis Based on Micro Data of the Basic Survey of Business Structure and Activity and the Japanese National Innovation Survey. *Global COE Hi–Stat Discussion Paper Series* No. gd08–002.

Lachenmaier, S., and Rottmann, H. (2011). Effects of Innovation on Employment: A Dynamic Panel Analysis. *International Journal of Industrial Organization* 29 (2): 210–220.

Leten, B., Belderbos, R. and Van Looy, B. (2007). Technological Diversification, Coherence, and Performance of Firms. *Journal of Product Innovation Management* 24 (6): 567–579.

Levinsohn, J., and Melitz, M. (2002). Productivity in a Differentiated Products Market Equilibrium. Mimeo.

Levinsohn, J., and Petrin, A. (2003). Estimating Production Functions using Inputs to Control for Unobservables. *Review of Economic Studies*, 70: 317–341.

Lokshin, B., Belderbos, R. and Carree, M. (2008). The Productivity Effects of Internal and External R&D: Evidence from a Dynamic Panel Data Model. *Oxford Bulletin of Economics and Statistics* 70(3): 399–413.

Lychagin, S., Pinkse, J., Slade, M. E., and van Reenen, J. (2010). Spillovers in Space: Does Geography Matter? *NBER Working Paper* No. 16188.

Mairesse, J., and Pierre, M. (2002). Accounting for Innovation and Measuring Innovativeness: An Illustrative Framework and an Application." *American Economic Review* 92 (2): 226–230.

Mairesse, J., and Pierre, M. (2010). Using Innovations Surveys for Econometric Analysis. *NBER Working Papers* No. 15857.

Mairesse, J., and Mulkay, B. (2008). An Exploration of Local R&D Spillovers in France. *NBER Working Paper*, National Bureau of Economic Research.

Marshall, A. (1920). *Principles of Economics*. London: MacMillan.

Mion, G. (2004). Spatial Externalities and Empirical Analysis: The Case of Italy. *Journal of Urban Economics* 56(1): 97−118, July.

Mohnen, P., and Lepine, N. (1991). R&D, R&D Spillovers and Payments for Technology: Canadian Evidence. *Structural Change and Economic Dynamics* 2(1): 213–28.

Mohnen, P., Mairesse, J., and Dagenais M. (2006). Innovativity: A Comparison Across Seven European Countries. *Economics of Innovation and New Technology* 15 (4–5): 391–413.

Monjon, S., and Waelbroeck, P. (2003). Assessing Spillovers from Universities to Firms: Evidence from French Firm–level Data. *International Journal of Industrial Organization* 21: 1255−1270.

Mori, T., Nishikimi, K., and Smith, T. E. (2005). A Divergence Statistic for Industrial Localization. *Review of Economics and Statistics* 87(4): 635–651.

Nakajima, K., Saito, Y. U., and Uesugi, I. (2012). Measuring economic localization: Evidence from Japanese firm-level data. *Journal of the Japanese and International Economies* 26(2): 201–220.

Nerlinger, E. A. (1998). Standorte und Entwicklung Junger Innovativer Unternehmen: Empirische Ergebnisse für West–Deutschland, Baden–Baden: Nomos.

O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors, *Quality and Quantity* 41 (5): 673–690.

OECD (2005). *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*. 3rd edition. OECD Publishing.

OECD (2009). Innovation in Firms: A Microeconomic Perspective. OECD Publishing.

Okamuro, H. (2009). Determinants of R&D Activities by Start–up Firms: Evidence from Japan. In *Small Business: Innovation, Problems and Strategy* edited by J. E. Michaels and L. F. Piraro, 27–44. New York: Nova Science Publishers.

Okamuro, H., Kato, M., and Honjo, Y. (2011). Determinants of R&D Cooperation in Japanese Start–Ups. *Research Policy* 40 (5): 728–738.

Okamuro, H. (2008). How Different are the Regional Factors of High–tech and Low–tech Start–ups? Evidence from Japanese Manufacturing Industries. *International Entrepreneurship and Management Journal* 4 (2): 199–215.

Okamuro, H., and Kobayashi, N. (2005). Determinants of Regional Variations in the Start–up Ratio: Evidence from Japan. *COE/RES Discussion Paper Series* No. 115. Hitotsubashi University.

Okamuro, H., and Kobayashi, N. (2006). The Impact of Regional Factors on the Start–up Ratio in Japan. *Journal of Small Business Management* 44 (2): 310–313.

Okamuro, H., and Nishimura, J. (2013). Impact of University Intellectual Property Policy on the Performance of University–industry Research Collaboration. *Journal of Technology Transfer* 38(3): 273–301.

Olley, S., and Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* 64: 1263–1297.

Orlando, M. (2004). Measuring Spillovers from Industrial R&D: On the Importance of Geographic and Technology Proximity. *Rand Journal of Economics* 35: 777–786.

Ornaghi, C. (2006). Spillovers in Product and Process Innovation: Evidence from Manufacturing Firms. *International Journal of Industrial Organization* 24: 349−380.

Pons, J., Paluzie, E., Silvestre, J., and. Tirado, D. A. (2007). Testing the New Economic Geography: Migrations and Industrial Agglomerations In Spain. *Journal of Regional Science* 47 (2):289−313.

Quigley, J. M. (1998). Urban Diversity and Economic Growth. *Journal of Economic Perspectives*, 12(2), 127–138.

Rauch, J. E. (1993). Productivity Gains from Geographic Concentration of Human Capital: Next Term Evidence from the Cities. *Journal of Urban Economics* 34 (3): 380–400.

Redding, S. J. (2010). The Empirics of New Economic Geography. *Journal of Regional Science* 50: 297−311.

Robin, S., and Schubert, T. (2013). Cooperation with Public Research Institutions and Success in Innovation: Evidence from France and Germany. *Research Policy* 42 (1): 149–166.

Reynolds, P. D., Miller, B., and Maki, R. (1995). Explaining Regional Variation in Business Births and Deaths: U.S. 1976–88. *Small Business Economics* 7 (5): 389–407.

Ritsila, J., and Ovaskainen, M. (2001). Migration and Regional Centralization of Human Capital. *Applied Economics* 33 (3): 317–325.

Rosenthal, S. S., and Strange, W. C. (2001). The Determinants of Agglomeration. *Journal of Urban Economics* 50: 191–229.

Schmoch, U., Laville, F., Patel, P., and Frietsch, R. (2003). Linking Technology Areas to Industrial Sectors. *Final Report to the European Commission*, DG Research.

Suzuki, K. (1993). R&D Spillovers and Technology Transfer among and within Vertical Keiretsu Groups: Evidence from the Japanese Electrical Machinery Industry. *International Journal of Industrial Organization* 11 (4): 573–591.

Thünen, J. H. von (1826). *The Isolated State*. Oxford: Pergammon Press.

Tveteras, R., and Battese, G. E. (2006). Agglomeration Externalities, Productivity, and Technical Inefficiency. *Journal of Regional Science* 46 (4): 605−625.

Van Biesebroeck, J. (2007). Robustness of Productivity Estimates. *Journal of Industrial Economics* 55(3): 529–569.

Van Looy, B., Tijssen, R. J. W., Callaert, J, Van Leeuwen, T., and Debackere, K. (2004). European Science in Industrial Relevant Research Areas: Development of an Indicator–based Bibliometric Methodology for Performance Analyses of Countries and Research Organizations. *Report for the European Commission* (DG Research).

Vernon, R. (1960). *Metropolis 1985: An Interpretation of the Findings of the New York Metropolitan Region Study*. Harvard University Press.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* 57 (298): 348–368.

Zellner, A. (1963). Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results. *Journal of the American Statistical Association* 58 (304): 977–992.