AN ABSTRACT OF THE DISSERTATION OF

Hisamitsu Saito for the degree of Doctor of Philosophy in Agricultural and Resource Economics presented on March 17, 2008.
Title: Three Essays on Firm Heterogeneity and Regional Development.

Abstract approved:

---

Munisamy Gopinath

The objective of this dissertation is to theoretically and empirically examine the role of firm heterogeneity in terms of productivity and skill-intensity in the agglomeration process and the effect of agglomeration on regional economic development.

In the first essay, I analyze the effect of trade liberalization on agglomeration of high- and low-productivity firms and the consequences for regional economic development. By extending a new-economic-geography model, I find that competition, domestic and international, disperses low-productivity firms to less-developed regions. Trading with advanced countries also appears to bring about dispersion of economic activity. However, attempts by less-developed regions to provide monetary incentives are less likely to attract high-productivity firms.

In the second essay, I empirically test the hypothesis that high-productivity (exporting) plants in Chile self-select to locate in large markets. Plants' raw productivity, i.e., productivity independent of agglomeration economies, is computed to obtain regional productivity-distribution measures. I find that high-productivity

(exporting) plants indeed locate in a region where other plants in the same industry agglomerate, industrial structure is diversified and market size is large. Finally, plants' self-selection outweighs the contribution of agglomeration economies in increasing a region's productivity.

In the third essay, I identify the mechanism by which human-capital spillovers occur at the plant-level and examine the relationship between spillovers and agglomeration of high skill-intensive plants in Chile. I employ plant-level production functions incorporating the absorptive capacity hypothesis, i.e., high skill-intensive plants benefit more from human-capital spillovers than others. Empirically, in 5 out of 8 manufacturing industries, the benefit from spillovers is larger in high skill-intensive plants. Plant entry and exit are also affected by spillovers resulting in regional skill disparities.

The results of the three essays reveal locational preferences of various types of firms. Policy options for economic development through increases in regional productivity include specializing in targeted industry, diversifying regional industrial structure, enlarging the market size and workforce education. The results of this dissertation help local governments to evaluate of the benefits from each policy option, which when compared with their knowledge of costs, aid in the selection of an effective policy to improve regional well-being.

Three Essays on Firm Heterogeneity and Regional Development

by
Hisamitsu Saito

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented March 17, 2008
Commencement June 2008

Doctor of Philosophy dissertation of Hisamitsu Saito presented on March 17, 2008.

APPROVED:

---

Major Professor, representing Agricultural and Resource Economics

---

Head of the Department of Agricultural and Resource Economics

---

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to my reader upon request.

---

Hisamitsu Saito, Author

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. I am deeply indebted to my supervisor Dr. M. Gopinath whose help, stimulating suggestions and encouragement helped me in all the time of research for and writing of this thesis.

My committee members, Dr. R. Brooks, Dr. S. Buccola, Dr. J. Wu and Dr. A. Tekin-Koru, supported me in my research work. I want to thank them for all their help, support, interest and valuable hints.

Also, I want to thank Dr. R. Echeverria, who helped me to obtain the Chilean data and provided helpful comments.

Finally, I would like to give my special thanks to my wife Yoko whose support enabled me to complete this work.

## CONTRIBUTION OF AUTHORS

Dr. Gopinath and Dr. Wu assisted in interpreting the results and writing of chapter 2. Dr. Gopinath also assisted in building the econometric model, interpreting the results and the writing of chapter 3 and chapter 4.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX TABLES

Three Essays on Firm Heterogeneity and Regional Development

## 1. Introduction

Economic inequality within a country has received increasing research attention in the past few decades. Some view the inequality as an opportunity to better understand factors causing regional economic growth and development, and the ensuing policy implications (Henderson et al., 2001). However, others see the uneven economic growth and development within a country as a consequence of myriad factors including globalization, skill-intensive technological change, and urban-biased public policies (Venables, 2005). Nevertheless, the sources and consequences of intra-country economic inequality are important in the context of developing as well as developed economies. In the former, the high expected-wages in urban areas cause labor in-migration, which leads to an excess supply of workers and the creation or expansion of slums housing the unemployed families. It is estimated that 78.2% of the urban population lives in slums in the least developed countries (UN-HABITAT, 2003). While urban areas of developing countries deal with resource stress and possible deterioration of public goods, rural areas continue to be challenged by out-migration of jobs and skilled workers.

In developed countries, dwindling rural population makes it difficult to maintain viable communities outside of metropolitan areas. For example, more than a half of the population is 65 or older in 12.7% of rural communities in Japan, which have experienced decline in population (Ministry of Land, Infrastructure and Transportation, 2007). Therefore, a better understanding of the spatial concentration or agglomeration of economic activities, the associated employment opportunities and migration patterns is required for public policies to address not only individual problems of urban and rural areas but also the overall objective of spatially-uniform economic development.

In regional economics, agglomeration or dispersion of firms is explained by centripetal and centrifugal forces (Fujita et al., 1999; Baldwin et al., 2003). The economic intuition behind the centripetal force, i.e., attraction to a core, is that transport costs to sell products and buy intermediate goods are saved in a large market. At the same time, as firms concentrate in a region, competition becomes severe, and congestion and commuting costs increase bringing about a centrifugal force, i.e., movement away from a core. In the presence of increasing-returns -to-scale technology, if the former positive effect exceeds the latter negative effect, agglomeration of economic activities occurs.

Previous studies in this emerging area of new economic geography (NEG) analyze agglomeration either theoretically or empirically at the industry level (Henderson, 1986; Fujita et al., 1999). The prototypical NEG model has two regions; two kinds of industries (agriculture and manufactures), and two primary factors of production ("farmers" and "workers"). Agriculture is assumed to be a constant-returns sector tied to land, and the monopolistically-competitive manufacturing sector with increasing-returns can be located in either region. Firms in manufactures produce differentiated products, but have the same production technologies. In equilibrium, all firms produce the same amount of output and receive the same price. Thus, the standard NEG models essentially do not consider heterogeneity among firms in the form of productivity or skill-intensity (Melitz, 2003; Yeaple, 2005).

Recent research shows that firm or plant heterogeneity is a key determinant of firm behavior including entry and exit in both domestic and international markets (Helpman, 2006). Jovanovic (1982) is one of the first authors to theoretically demonstrate that high-productivity firms are more likely to grow faster and survive longer than their low-productivity counterparts. Melitz (2003) also shows that trade-cost changes induce a country's high-productivity firms to either begin exporting or increase exports, while low-productivity firms are forced out of the domestic markets. Empirical analyses support such predictions. For example, Dunne et al. (1989) find that plant shutdown is negatively correlated with the size and age. More recently, Bernard and Jensen (1999) show that low-productivity firms pay lower wages, employ fewer workers and are more likely to stop exporting than their high-productivity counterparts. In terms of productivity growth, Cohen and Levinthal (1989) show that firms engaging in research and development (R&D) activity can benefit more from knowledge spillovers than others.

The evidence of firm heterogeneity and intra-country economic differences lead to a question on the type of firms locating in urban or rural regions. It is apparent that regions which attract high-productivity, large, and R&D intensive firms can grow faster than other regions of an economy. In a cross-country study, Baldwin and Okubo (2006) show that high-productivity firms are more likely to agglomerate in large countries than low-productivity firms. The result arises from the lower benefits of agglomeration to low-productivity firms when they face severe competition with their high-productivity counterparts. Hence, the average productivity in large countries is higher than that in small countries. However, few studies have explored the location decision of heterogeneous firms. Empirical research on firm heterogeneity in the spatial dimension is also limited with a few exceptions. For example, Syverson's (2004) investigation of spatial productivity differences in the U.S. ready-mix concrete industry finds that large markets have not only a higher average productivity but also a lower dispersion due to the truncation (from below) of the productivity distribution. The objective of this dissertation is to theoretically and empirically examine the role of firm heterogeneity in the agglomeration process and the latter's effect on regional economic development.

In the first of the three essays of this dissertation, I analyze the effect of trade liberalization on agglomeration and regional economic development. The globalization wave of the past few decades has made regional economies more dependent on international economic conditions. For example, northern Mexico has grown by attracting firms exporting to the U.S. market while southern Mexico's level of economic activity remains significantly lower than that in the northern regions. The impact of trade liberalization on the spatial distribution of economic activities has also been explored in the NEG literature (Krugman and Livas, 1996;

Behrens et al., 2003). For example, Krugman and Livas (1996) show that even if two regions are not identical at the beginning, as long as they face the same exporting and importing opportunities, trade liberalization will lead to an even distribution of firms between the two regions. Hanson (1997) finds that the regional wage structure in Mexico depends on the distance to the U.S. border after trade liberalization. At the same time, trade liberalization has been accused of increasing wage disparities, benefiting higher income groups more than the poor (Feliciano, 2001). Therefore, attracting high-wage jobs has become a major development strategy of many local governments. To understand how trade liberalization affects regional income inequality, I will examine the effect of trade liberalization on agglomeration process of heterogeneous firms by extending the NEG model of Ottaviano et al. (2002).

In the second essay, I empirically test the hypothesis derived from the first essay: high-productivity (exporting) firms self-select into large markets. To do so, I will employ plant-level data from Chile. As most regions in Chile have a coastline, the assumption in the first essay that every region faces the same level of trade costs will likely hold. However, several econometric problems arise when estimating plant-level production functions and productivity. For example, as high-productivity plants agglomerate in a region with large market to gain from agglomeration economies, plant's raw productivity, i.e. productivity independent of agglomeration economies, is likely correlated with variables measuring the regional market size (Henderson, 2003; Rosenthal and Strange, 2004). After obtaining plant's raw productivity by using Levinsohn and Petrin's (2003) approach, the self-selection hypothesis will be tested by regressing measures of regional productivity distribution, e.g., median, $10^{th}$, and $90^{th}$ percentiles, on indicators of the market size.

In the third essay, I examine the relationship between human-capital spillovers and agglomeration of high skill-intensive plants in Chile. Human-capital spillovers are considered to be an important source of economic growth (Lucas, 1988). Evidence of spillovers has often been found at the aggregate level by estimating wage hedonic equations (Acemoglu and Angrist, 2000; Moretti, 2004b; Rauch, 1993). However, the firm- or plant-level mechanism by which spillovers arise and yield benefits to recipients is not clear from the wage-hedonic approach. I will estimate a plant-level production function incorporating the absorptive capacity hypothesis of Cohen and Levinthal (1989) to identify the mechanism by which human-capital spillovers improve plants' productivity. In other words, I test whether or not plants employing high skill-intensive production can benefit more from spillovers than others. If the benefit from spillovers is limited to high-skill intensive plants, then they agglomerate in a high skill-intensive region, which will be empirically tested after estimating the plant-level production function.

Most studies on regional economic growth and development have not considered firm or plant heterogeneity. As a result, their policy suggestions do not account for the types of firms attracted to a region. The differences across firms within an industry are large and the impact on regional economy of attracting firms will vary depending on the type of firms. High-productivity firms engage in exporting, employ skilled workers with high wages and are more likely to survive longer than their low-productivity counterparts. Similarly, only firms engaging in R&D activity, a type of high-productivity firm, can benefit from R&D spillovers. The results of this dissertation show that the effect of economic development policies on regional economy depends on the type of firms attracted to a region. Suggestions on how to attract the desired type of firms to a region are also provided.

# 2. Heterogeneous Firms, Trade Liberalization and Agglomeration

## 2.1 Introduction

The effect of trade liberalization on economic growth has been at the core of international economics for decades. However, the recent debate on trade liberalization has centered around its effect on income inequality and poverty (Yale Global Forum, 2007). The domestic and international relocation of labor-intensive manufacturing jobs, associated with liberalized trade, and the consequences for low-income households have received much research attention (Feenstra and Hanson 1997; Galiani and Sanguinetti, 2003; Goldberg and Pavcnik, 2005). In some instances, trade liberalization has been accused of increasing wage disparities, benefiting higher income groups more than the poor (Feliciano, 2001).

Consideration of relocation decisions of different types of firms in the face of trade and globalization is important from an economic development perspective. Regions in a developing country that are beneficiaries in the relocation process tend to grow faster than others (Hanson, 1998). For example, wages in northern Mexico and coastal regions of China – home to a diverse set of manufacturing activities - are several times higher than that in southern Mexican regions and inland provinces of China (Hanson, 1997).[1] Thus, trade liberalization may influence the spatial distribution of economic activities within an economy, e.g., Ades and Glaeser (1995), where attracting high-wage jobs has become a major development strategy for many local and state governments. To understand how trade liberalization affects income inequality and poverty, it is essential to understand how the location patterns of different types of firms adjust with trade liberalization.

The spatial concentration of economic activity has become a major focus of economic research with the seminal contribution of Krugman (1991). The research has led to the development of the new economic geography (NEG), which has greatly increased our understanding of how regions can endogenously become differentiated into an industrialized "core" and an agricultural "periphery" (Fujita et al., 1999; Baldwin et al., 2003). The prototypical NEG model has two regions; two kinds of industries (agriculture and manufactures), and two primary factors of production ("farmers" and "workers"). Agriculture is assumed to be a constant-returns sector tied to land, and the monopolistically-competitive manufacturing sector with increasing-returns can be located in either region. Firms in manufactures produce differentiated products, but have the same production technologies. In equilibrium, all firms produce the same amount of output and receive the same price. Thus, the standard NEG models essentially do not consider firm heterogeneity (Melitz, 2003; Yeaple, 2005).

The impact of trade liberalization on the spatial distribution of economic activities has also been explored in the NEG literature (Krugman and Livas, 1996; Allonso-Villar, 1999). For example, Krugman and Livas (1996) show that even if two regions are not identical at the beginning, as long as they face the same exporting and importing opportunities, trade liberalization will lead to an even distribution of firms between the two regions. The reason is that firms must pay higher wages to urban residents to compensate for their high rents and commuting costs. When urban and rural areas face the same trade costs, benefits from locating in a city decline and firms relocate to rural areas. The enlargement of the European Union is a case in point (Behrens et al., 2003).[2] Mansori (2003) shows that increasing returns to scale in transport/trade costs can lead to agglomeration, a result opposite to that of Krugman and Livas (1996). Plauzie (2001) employs immobile workers instead of congestion costs as a dispersion force. Finally, Fujita et al. (1999) extend the model developed by Krugman and Livas (1996) to the case of two monopolistically competitive industries, where trade liberalization leads each region to specialize in one industry. These studies, however, have not considered the agglomeration process of heterogeneous firms with trade liberalization.

The few studies that consider location decisions of heterogeneous firms include Baldwin and Okubo (2006) and Syverson (2004). The former shows that the most productive firms are the first to relocate to a large country from a small country and the last to relocate from a large country, but does not address regional patterns of firm location within a country. The latter is an empirical study of U.S. ready-mixed concrete industry using plant-level data. Syverson (2004) finds that the distribution of productivity in a market shifts toward right, i.e., low-productivity firms disappear, as a market's demand density increases. Overall, there is very limited research on the relocation of different types of firms/jobs with trade liberalization.

The primary objectives of this study are 1) to explore the determinants of firms' location patterns in the presence of firm heterogeneity, in particular, differences in productivity and 2) to examine how heterogeneous firms' location patterns adjust with trade liberalization. For these purposes, we first extend the model of agglomeration developed by Ottaviano et al. (2002) to include different types of manufacturing firms in section 2.2. Then, in section 2.3, we use the model to explore how the degree of competition and trade costs affect the location choice of heterogeneous firms. In section 2.4, we extend the model to include a foreign market with identical trade costs to different regions in the home economy. We apply the extended model to examine firms' incentives to relocate within the home economy due to the trade liberalization. In section 2.5, we simulate the effects of regional governments' subsidies to attract firms to a region. Section 2.6 provides conclusions and policy implications.

## 2.2 The Model

Consider an economic space consisting of two regions within a country (henceforth, North and South, indexed *N* and *S*, respectively).[3] There are two sectors; sector *A* produces agricultural goods while the other sector produces manufactured goods. The manufacturing sector consists of two types of firms, i.e., low-productivity (*M*) and high-productivity (*H*) firms. The agricultural good is produced using labor (*L*) as the only input, whereas the manufactured goods are produced using both labor and human capital (*K*). The two types of firms in the manufacturing sector differ in the marginal labor requirement. There are three types of consumers in the two regions:

worker, industrialists, and entrepreneurs. Workers provide labor input for all sectors; industrialists provide human capital for low-productivity firms, and entrepreneurs provide human capital for high-productivity firms. Each worker provides one unit of labor, and each industrialist or entrepreneur provides one unit of human capital. We assume that the skilled industrialist or entrepreneur best describes low- or high-productivity firms. In other words, human capital is not assumed mobile between two types of firms but we assume it is mobile between regions. Each region is endowed with a given mass of workers, which is mobile between sectors but immobile between regions.[4]

Consider first the demand for goods produced by the two sectors. All consumers have identical preferences, which are defined by the quadratic utility function:

(1)
$$\begin{aligned} U = {} & \alpha\left(\int_0^n c_H(i)\,di + \int_0^n c_M(i)\,di\right) - \frac{\beta-\gamma}{2}\left(\int_0^n c_H^2(i)\,di + \int_0^n c_M^2(i)\,di\right) \\ & - \frac{\gamma}{2}\left[\left(\int_0^n c_H(i)\,di\right)^2 + \left(\int_0^n c_M(i)\,di\right)^2\right] - \delta\left(\int_0^n c_H(i)\,di\right)\left(\int_0^n c_M(i)\,di\right) + c_A, \end{aligned}$$

where $c_s(i)$ is the consumption of the *i*-th good produced by type *s* firms, where $s = H, M$ ; $c_A$ is the consumption of the agricultural good; *n* is the total mass of goods, which is assumed to be the same for low-productivity and high-productivity firms. Our focus is on the effect of competition between the two types of firms, and the latter assumption ensures that both types of firms face the same level of competition within respective sectors. Following Ottaviano et al. (2002), we assume that $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\delta \geq 0$, $\beta > \gamma$ and $\gamma \geq \delta$. Parameter $\gamma$

measures the degree of substitutability within *H* goods or *M* goods, while $\delta$ measures the degree of substitutability between *H* and *M* goods.[5] In previous studies, it is assumed that $\delta = \gamma$ because type *H* and *M* firms belong to the same industry.[6] However, goods produced by low-productivity firms may be weakly substitutable for goods produced by high-productivity firms and vice versa. For example, a computer sold online by IBM or Apple might be weak substitute for the one assembled by a local computer company. Therefore, we relax this assumption and allow $\delta$ to take a value other than $\gamma$.

Each consumer maximizes utility subject to a budget constraint:

(2) $$\max_{c_H(i), c_M(i), c_A} U \quad s.t. \int_0^n p_H(i) c_H(i) di + \int_0^n p_M(i) c_M(i) di + c_A = E$$

where, the agricultural good *A* is assumed to be a numeraire good, $p_s(i)$ is the price of the *i*-th good produced by type *s* firm, and *E* denotes consumer income which includes the endogenous wage and an exogenously determined endowment of agricultural goods.[7] The first order conditions are:

(3a) $$\alpha - (\beta - \gamma) c_H(i) - \gamma \int_0^n c_H(i) di - \delta \int_0^n c_M(i) di - \lambda p_H(i) = 0,$$

(3b) $$\alpha - (\beta - \gamma) c_M(i) - \gamma \int_0^n c_M(i) di - \delta \int_0^n c_H(i) di - \lambda p_M(i) = 0,$$

(3c) $$1 - \lambda = 0.$$

Integrating the equations (3a) and (3b) over *i*, we obtain:

(4a) $$\alpha n - (\beta - \gamma) \int_0^n c_H(i) di - \gamma n \int_0^n c_H(i) di - \delta n \int_0^n c_M(i) di - P_H = 0$$

(4b) $$\alpha n - (\beta - \gamma) \int_0^n c_M(i) di - \gamma n \int_0^n c_M(i) di - \delta n \int_0^n c_H(i) di - P_M = 0$$

where, $P_H = \int_0^n p_H(i)\,di$ and $P_M = \int_0^n p_M(i)\,di$. Solving (4a) and (4b) for $\int c_H(i)\,di$ and $\int c_M(i)\,di$ and substituting these into (3a) and (3b) gives us the demand functions for the goods produced by the two types of firms in the manufacturing sector:

(5a) $$c_H(i) = a - dp_H(i) + bP_H + cP_M$$

(5b) $$c_M(i) = a - dp_M(i) + bP_M + cP_H$$

where,

$$a = \alpha / \left[\beta - \gamma + n(\gamma + \delta)\right],$$

$$b = \left[\gamma(\beta - \gamma) + n(\gamma^2 - \delta^2)\right] \Big/ (\beta - \gamma)\left[(\beta - \gamma + \gamma n)^2 - n^2\delta^2\right],$$

$$c = \delta / \left[(\beta - \gamma + \gamma n)^2 - n^2\delta^2\right] \text{ and}$$

$$d = 1/(\beta - \gamma).$$

Given that $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\delta \geq 0$, $\beta > \gamma$ and $\gamma \geq \delta$, *a*, *b*, *c* and *d* all take positive values. The derivatives of *a*, *b, c* and *d* with respect to *n* and *γ* indicate how the level of competition between firms and the substitutability between goods affect the demand for a manufactured good. For instance, the derivatives of *a*, *b* and *c* with respect to *n* are negative, indicating that as the number of competitors increases, demand for the variety declines. Moreover, the derivative of *d* with respect to *γ* is positive, implying that demand for the variety decreases as substitutability between varieties increases. Finally, the demand for agricultural good is given by:

(5c) $$c_A = E - a\left(P_H + P_M\right) - b\left(P_H^2 + P_M^2\right) - 2cP_H P_M + d\left(\int p_H^2\left(i\right)di + \int p_M^2\left(i\right)di\right)$$

Substituting (5a)-(5c) into (1) and then simplifying it yields the indirect utility function:

(6) $$\begin{aligned} V = \frac{na^2}{d-bn-cn} - a\left(P_H + P_M\right) - \frac{b}{2}\left(P_H^2 + P_M^2\right) - cP_H P_M \\ + \frac{d}{2}\left(\int_0^n p_H^2(i)di + \int_0^n p_M^2(i)di\right) + E \end{aligned}$$

Now, consider the supply side of the markets. We extend the assumptions of Ottaviano et al. (2002) to the heterogeneous firms' setting for the specification of production technologies. Agriculture requires one unit of labor to produce one unit of output. With free trade in agriculture and zero transport costs, the choice of the agricultural good as the numeraire implies that the wage of labor is equal to one in both regions. Therefore, the supply of workers for the manufacturing sector is perfectly elastic as long as there is no shortage of workers in both regions. Each variety of manufactured good is produced by a firm, which is owned by the industrialists or entrepreneurs who provide human capital to the firm. The manufacturing firms use human capital as a fixed input and labor as a variable input, and the technology requires $k_s$ units of human capital to produce any amount of type $s$ goods. Thus, the total mass of type $s$ goods is $n = K_s / k_s$, where $K_s$ is total mass of human capital used for the production of type $s$ goods. Let $\lambda_H$ and $\lambda_M$ denote, respectively, the shares of the entrepreneurs and industrialists living in the North. The total mass of population in the two regions are

$M^N = \lambda_H K_H + \lambda_M K_M + 0.5L$ and $M^S = (1-\lambda_H) K_H + (1-\lambda_M) K_M + 0.5L$,

respectively. The problem for each firm is to choose the prices of the product in the two regions to maximize the profit:

$$\text{(7)} \qquad \begin{array}{c} \max\limits_{p_s^{rr}, p_s^{ro}} \pi_s^r = \left(p_s^{rr} - x_s\right) c_s^r M^r + \left(p_s^{ro} - t - x_s\right) c_s^r M^o, \\ s = H, M, \quad r, o = N, S, \quad r \neq o \end{array},$$

where $p_s^{rr}$ is the price of a type *s* good produced in region *r* and sold in region *r*, $p_s^{ro}$ is the price of a type *s* good produced in region *r* and sold in the other region, $c_s^r$ is the demand by a consumer living in region *r* for a type *s* good, *t* is the cost of transporting one unit of any variety from one region to the other and is assumed to account for all the impediments to trade, and $x_s$ is the amount of labor needed to produce a unit of a type *s* good. Since type *H* (*M*) firms are high (low) productivity, $x_H < x_M$.

The first order conditions of the profit maximization problem in equation (7), after indexing the variety-demand functions in (5a) and (5b) by regions, are:

$$\text{(8a)} \qquad a - 2dp_s^{rr} + bP_s^r + cP_{-s}^r + dx_s = 0,$$

$$\text{(8b)} \qquad a - 2dp_s^{ro} + bP_s^o + cP_{-s}^o + d\left(x_s + t_s\right) = 0,$$

where, the subscript "-*s*" denotes the other manufacturing sector (i.e., goods other than *s*). Solving (8a) and (8b) gives the prices set by type *s* firm located in region *r*:

$$\text{(9a)} \qquad p_s^{rr} = \frac{a + bP_s^r + cP_{-s}^r + dx_s}{2d}$$

(9b) $p_s^{ro} = p_s^{rr} + \frac{t}{2}$

Since $k_s$ units of human capital are required to set up a type *s* firm, each owner's share of profit in region *r* is $\pi_s^r / k_s$, and the corresponding utility for the owner is denoted by $V_s^r$. We assume that human capital is myopic and migrates to the region where it can get the highest utility.

To analyze the spatial distribution of manufacturing firms between the two regions, we define the concept of spatially stable equilibrium as follows.

Definition: A distribution of resources between the two regions is a spatially stable equilibrium if for each *s = H, M*:

(i) $\left(V_s^N - V_s^S\right) > 0$ and $\lambda_s = 1$,

(ii) $V_s^N - V_s^S = 0$ and $\frac{\partial\left(V_s^N - V_s^S\right)}{\partial \lambda_s} < 0$ and $\lambda_s \in [0,1]$, or

(iii) $\left(V_s^N - V_s^S\right) < 0$ and $\lambda_s = 0$.

If condition (i) or (iii) holds, then all type *s* firms fully agglomerate in one region. Such a distribution is a spatially stable equilibrium if the owners of the firms cannot achieve a higher utility level by moving to the other region. When condition (ii) holds, the owners of firms are indifferent between the two regions, and the firms could disperse between the two regions or be concentrated in only one of them. Such a distribution is a spatially stable equilibrium if relocation of a firm to the other region would result in lower utility for the owner.

**2.3 Spatial Equilibrium**

To analyze the equilibrium distribution of manufacturing firms between the two regions, we assume that the initial dispersion of human capital is symmetric between the two regions. Since both regions are initially identical, the relocation of the first firm is indeterminate. We assume that the first firm relocates to the North.

We first show how the degree of competition (*δ*) affects firm location and the spatial equilibrium in the two-sector setting. For analytical simplicity, we assume that each manufacturing firm requires one unit of human capital, i.e., $k_s = 1, \ \ s = H, L$. We normalize the unit of human capital so that $K_s = 1, \ \ s = H, L$.[8] Then, both types of manufacturing firms have the same mass of firms (i.e., $K_s / k_s = 1, \ \ s = H, L$). We further assume that $x_H = 0 < \theta_M = x_M$ without loss of generality. Under these assumptions, by substituting (5a), (5b), (7), (9a) and (9b) into (6) and taking a difference, the differences in the levels of utility for the entrepreneurs and industrialists can be derived as:

(10a) $$V_H^N - V_H^S = A_1 t\left(C_1 - t\right)\left(\lambda_H - \frac{1}{2}\right) + A_2 t\left(C_2 - t\right)\left(\lambda_M - \frac{1}{2}\right)$$

(10b) $$V_M^N - V_M^S = A_2 t\left(C_3 - t\right)\left(\lambda_H - \frac{1}{2}\right) + A_1 t\left(C_4 - t\right)\left(\lambda_M - \frac{1}{2}\right)$$

where, $A_1$, $A_2$, $C_1$, $C_2$, $C_3$ and $C_4$ are functions of parameters in the utility and profit functions and are presented in Appendix A. $A_1 \geq A_2 > 0$, $C_1 > C_4$ and $C_2 > C_3$.

Before discussing the spatial equilibrium, we need to ensure inter-regional trade, which may not occur if transport costs are prohibitive. Following Ottaviano et al. (2002), we set $t < t_{\max}$, where

$t_{\max} = \min\left\{t : p_s^{ro} - t - x_s \geq 0,\ \ r,o = N,S\ \ \ s = H,M,\ \ r \neq o\right\}$. To ensure $0 < t_{\max}$, we find that the productivity difference between firms should be bounded from above, $\theta_{M\max}$. The upper bound for transport cost and productivity difference makes $C_4$ positive. Finally, we set a lower limit for labor, $L_{\min}$, so that $t_{\max} > C_1$, i.e. every type of firms disperses at the initial point.[9] Then, we have the following result.

Result 1: There exists a $\delta^* \in [0,\gamma]$ such that for any $\delta \in \left(\delta^*, \gamma\right]$, $C_1 > C_2$, $C_3 > C_4$, $C_1 > C_3$ and $C_2 > C_4$ hold.

*Proof*: See Appendix B.

**Proposition 1. The equilibrium distribution of manufacturing firms:** When transport costs fall below a certain level, all high-productivity firms in the South will move to the North simultaneously. When the demand substitutability between *H* and *M* goods is sufficiently high, a fraction of low-productivity firms in the North will move to the South. However, with further reduction in transport costs, low-productivity firms in the South will gradually move back to the North and eventually all low-productivity firms will fully agglomerate in the North. When the substitutability between *H* and *M* goods is low, the agglomeration of low-productivity firms in the North begins without some firms moving to the South first.

*Proof*: See Appendix B.

Proposition 1 is illustrated in figure 2.1, where panel (a) and (b) correspond to the case of strong and weak substitutability between *H* and *M* goods, respectively. The thick (thin) line in both panels represents the share of the high-productivity (low-productivity) firms located in the North. As shown in the figure, regardless of the substitutability between *M* and *H* goods, all high-productivity firms in the South will move to the North simultaneously when transport costs fall below a certain threshold. In contrast, the agglomeration process of the low-productivity firms is more gradual and depends on the substitutability between *H* and *M* goods. When goods *H* and *L* are highly substitutable, a fraction of low-productivity firms in the North move to the South first. However, with weak substitutability between goods *H* and *L*, a decline in transport costs causes both types of firms to agglomerate in the North, although low-productivity firms' relocation to the North is not instantaneous.[10]

Agglomeration or dispersion of firms is determined by centripetal forces and centrifugal forces (Fujita et a., 1999; Baldwin et al., 2003). A primary centripetal force is the increased demand for manufacturing goods when firms are located together because of household preferences for variety (Ottaviano et al., 2002). However, competition between firms acts as a centrifugal force countering agglomeration. When transport costs are high, the latter outweighs the former, causing firms to spatially disperse. On the contrary, when transport costs are low, centripetal forces outweigh those that pull firms apart resulting in agglomeration

(Baldwin et al., 2003). In our case, competition between *H* and *M* goods works as another centrifugal force. In particular, when goods produced by high-productivity firms are strongly substitutable for goods produced by low-productivity firms (i.e., $\delta$ is large), competition becomes severe. Even though agglomeration of high-productivity firms in the North increases the demand for manufacturing goods, the severe competition reduces benefits of agglomeration. Since low-productivity firms are less competitive than high-productivity counterparts, low-productivity firms move to the South where high-productivity firms lose their competitiveness due to high transport costs (panel a). However, high-productivity firms become more competitive in the South as transport costs decline. As a result, the benefit of locating in the South falls for low-productivity firms, which begin their agglomeration in the North. When the competition between firms is weak, agglomeration of high-productivity firms benefits low-productivity firms by offering large demand for their goods as well.

Proposition 1 considers a general case where *H* and *M* goods can be weakly substitutable. A more restricted case where any manufacturing good regardless of the type faces the same substitutability is obtained by setting $\delta = \gamma$ (see footnote # 6). Location patterns of firms in this special case are summarized in the following corollary.

**Corollary 1.** When transport costs fall below a certain level, all high-productivity firms in the South will move to the North simultaneously, while a fraction of low-productivity firms in the North will move to the South. However, with further reduction in transport costs, some low-productivity firms in the South will move back to the North and eventually all firms will fully agglomerate in the North.

In the presence of productivity difference, we showed that competition between different types of firms works as a dispersion force. As competition becomes severe, i.e., $\delta$ is large, low-productivity firms tend to locate away from high-productivity counterparts. We next investigate how the equilibrium distribution will adjust with increasing trade and globalization.

**2.4 Trade Liberalization and Regional Structural Adjustments**

In this section, we analyze how trade liberalization affects regional structure by introducing a third region *F*, the rest of the world, into the agglomeration model with firm heterogeneity. For analytical purposes, we assume that identical firms exist in region *F*, which use one unit of human capital as a fixed input and labor as a variable input. The marginal labor requirement is $x_F = \theta_F$. In addition, to focus on the effect of trade liberalization, we set $\delta = \gamma$, which implies that goods are equally substitutable both within and across sectors. The region *F* is endowed with $K_F$ units of human capital and $L_F$ units of labor. Since the rest of the world is generally larger than any single economy, we assume that $K_F \geq 2$, which implies that the total mass of differentiated goods is larger in the foreign region *F* relative to the two home-country regions. Agricultural goods are freely traded between countries, which normalizes the wage rate to one in all three regions. It costs $t_F$ to trade one unit of differentiated goods between countries. $t_F$ includes both natural transport costs and artificial trade barriers and is assumed to be the same for the North and the South. With the access to the world market, the utility function becomes:

$$
\begin{aligned}
(11)\qquad U = \alpha\left(\int_0^n c_H(i)\,di + \int_0^n c_M(i)\,di + \int_0^{n_F} c_F(i)\,di\right) - \frac{\beta-\gamma}{2}\left(\int_0^n c_H^2(i)di\right. \\
\left. + \int_0^n c_M^2(i)\,di + \int_0^{n_F} c_F^2(i)\,di\right) - \frac{\gamma}{2}\left(\int_0^n c_H(i)\,di\right. \\
\left. + \int_0^n c_M(i)\,di + \int_0^{n_F} c_F(i)di\right)^2 + c_A,
\end{aligned}
$$

where, $n_F = K_F$. The demand function for each good can be derived in a similar way as equations (5a)-(5b) are derived. Given the additional region, the domestic firms' profit maximization problem is:

$$
\begin{aligned}
(12)\qquad \max_{p_l^{rr}, p_l^{ro}, p_l^{rF}} \pi_l^r = \left(p_l^{rr} - x_l\right)c_l^r M^r + \left(p_l^{ro} - t_l - x_l\right)c_l^s M^s + \left(p_l^{rF} - t_F - x_l\right)c_l^F M^F, \\
l = H, L, \quad r, o = N, S, \quad r \neq o,
\end{aligned}
$$

where, $M^F = K_F + L_F$. Likewise, foreign firms maximize profits by choosing prices for the three regions:

$$
\begin{aligned}
(13)\qquad \max_{p_F^{Fr}, p_F^{Fo}, p_F^{FF}} \pi_F = \left(p_F^{Fr} - t_F - x_F\right)c_F^r M^r + \left(p_F^{Fo} - t_F - x_F\right)c_F^s M^s \\
+ \left(p_F^{FF} - x_F\right)c_F^F M^F, \quad r, o = N, S, \quad r \neq o.
\end{aligned}
$$

As the mass of goods exported is exogenously determined in the utility function (11), we assume that every plant engages in exporting by setting the productivity range "*pd*" (See Appendix A). Prices for goods produced domestically and by foreign firms can be derived from the first-order conditions of these profit maximization problems. Given the prices, the indirect utility functions for industrialists and entrepreneurs can be derived by maximizing their utility subject to the budget constraint. The difference in utility for each group in the two regions is given by:

(14a) $$V_H^N - V_H^S = q_1(t_F, t)\left(\lambda_H - \frac{1}{2}\right) + q_2(t_F, t)\left(\lambda_L - \frac{1}{2}\right),$$

(14b) $$V_M^N - V_M^S = q_3(t_F, t)\left(\lambda_H - \frac{1}{2}\right) + q_4(t_F, t)\left(\lambda_L - \frac{1}{2}\right),$$

where, $q_i(t_F, t)$, $i = 1, ..., 4$, are linear in $t_F$ with a positive slope and quadratic in $t$ with a negative sign on $t^2$. Furthermore, $q_1(t_F, t) > q_2(t_F, t) > q_3(t_F, t) > q_4(t_F, t)$ holds for $t > 0$. Similar to the two-region setting and Ottaviano et al. (2002), there would be no trade among the three regions when $t_F$ and $t$ are high enough. Let $t_{2\max}$ and $t_{F\max}$ denote the highest levels of $t$ and $t_F$ that assure trade among the three regions. Implicit in the derivation of $t_{2\max}$ and $t_{F\max}$ are (i) $\theta_M$ and $\theta_F$ must be in the space "*pd*" shown in Appendix A, which assures that maximum transport costs are always positive, and (ii) $q_4(t_F, t) > 0$ for low domestic transport costs. Finally, we set a lower limit for labor, $L_{2\min}$, so that $q_1(t_F, t)$ is negative at $t_{2\max}$, which assures that both types of firms are dispersed when domestic transport costs are high.

From (14a) and (14b), $\partial\left(V_s^N - V_s^S\right)/\partial\lambda_k = q_i(t_F, t)$, $s, k = H, L$, $i = 1, ..., 4$. When $q_i(t_F, t) > 0$, agglomeration may occur because as human capital increases in one region, the utility level there will become higher than in the other region. As international transport costs decrease, e.g., from $t_{F2}$ to $t_{F1}$ in figure 2.2, $q_i\left(t_F, t\right)$, $i = 1,...,4$ shift downward and the range in which $q_i\left(t_F, t\right)$ takes a positive value shrinks. Therefore, international trade opportunity works against agglomeration.[11] All $q_i\left(t_F, t\right) = 0$ have one positive solution in terms of *t* in addition to zero implying that $q_3\left(t_F, t\right) + q_4\left(t_F, t\right) = 0$ has one positive solution. Let this solution be $t\left(t_F\right)$. Then, we have a following proposition.

**Proposition 2.**

When international transport costs are high, both types of firms agglomerate in the North for $t \in \left(t\left(0\right), t\left(t_{F\max}\right)\right)$. As international transport costs fall, low-productivity firms relocate to the South, while all high-productivity and some low-productivity firms remain in the North.

*Proof*: See Appendix B.

The explanation for this result is that large foreign markets become more important than domestic markets when international transport costs are low.

Low-productivity firms have to compete with the high-productivity firms if both are located in the North. With the increasing importance of foreign markets, low productivity firms are the first to relocate to the South. When the domestic transport cost *t* is greater than $t\left(t_{F\max}\right)$, low-productivity firms do not fully agglomerate in the North even when $t_F$ is high. On the other hand, if *t* is less than

$t(0)$, neither type of firms relocate from the North to the South. Figure 2.3 shows how the two types of firms change their location in response to a change in international transport costs for $t \in (t(0), t(t_{F\max}))$. As in figure 2.1, the thick (thin) line is the share of high-productivity (low-productivity) firms located in the North. As international transport costs decline, low-productivity firms relocate to the South and the difference in size between the North and the South shrinks with the latter specializing in low-productivity firms.

With regard to a change in productivity difference, we have the following proposition.

**Proposition 3.**

When the domestic productivity difference widens (i.e., $\theta_M$ increases), or when the productivity of foreign firms improves (i.e., $\theta_F$ declines), the level of international transport costs at which domestic low-productivity firms relocate increases. Moreover, when productivity of foreign firms improves, the number of firms relocating to the South also increases.

*Proof*: See Appendix B.

A narrowing gap in productivity between domestic firms fosters agglomeration of both types of firms. When productivity difference between domestic firms is large, low-productivity firms react to international trade opportunities at higher level of international transport costs because low-productivity firms are not competitive in the North. This result suggests that R&D subsidies for low-productivity firms will increase the regional economic disparity because as productivity gap narrows, trade liberalization will not cause the relocation of low-productivity firms to the South. On the contrary, R&D subsidies for high-productivity firms will lead to the opposite result.

Productivity improvement of foreign firms leads to dispersion. Alternatively, when productivity of foreign firms improves, competition in the domestic market becomes severe, which reduces the attractiveness of urban areas, especially for low-productivity firms. At the same time, prices in the South go down due to imports, which increase the utility level and attractiveness of the region. As a result, low-productivity firms move to the South, where they can hire human capital at lower costs. This result implies that trading with advanced countries narrows regional economic disparity because many low-productivity firms will relocate to the South.

To further illustrate the role of international trade, figure 2.4 depicts each region's export value based on a simulation under assumed parameter values.[12] The thick line is the export value of the North, which shows greater export value in this region relative to the South regardless of the level of international transport costs. When international transport costs approach zero, the export value of the North remains greater than the value of the South even though the former loses some exporting firms. This is because all high-productivity exporting firms remain in the North and only low-productivity exporting firms locate in the South, a result consistent with the export decision literature (Bernard and Jensen, 1999).

**2.5 A Policy Simulation: Role of Regional Government Subsidy**

In the previous section, the South is shown to attract only low-productivity firms with a fall in international transport costs. Moreover, the export value is always lower in the South than in the North. As a result, wages and incomes, and overall regional welfare including immobile workers' utility are likely to be lower in the South. The parallel to this situation in empirical studies is the case of rural areas, whose economic indicators lag behind their urban counterparts. For example, Glaeser and Mare (2001) report that wages are 33% higher in big cities in the United States. Local governments use various strategies to attract high-productivity firms, whose presence is strongly associated with regional economic development. These strategies include direct governmental subsidies through tax breaks or construction or improvement of local infrastructure. In this section, we use the model to simulate how local government subsidies affect the distribution of firms between the regions. For simplicity, the subsidy takes the form of a lump-sum payment to any firms located in the concerned region.

In the pre-subsidy regime, suppose many low-productivity firms ($1-\lambda_M = 0.79$) locate in the South due to the trade liberalization ($t_F = 0.025$) as shown in the figure 2.3. The subsidy is financed through a residency tax:

(15) $$\left(2-\lambda_H-\lambda_M+\frac{L}{2}\right)tax=\left(2-\lambda_H-\lambda_M\right)subsidy\,.$$

The differences in utility for the two groups in the home country are given by:

(16a) $$V_H^N-V_H^S=q_1\left(t_F,t\right)\left(\lambda_H-\frac{1}{2}\right)+q_2\left(t_F,t\right)\left(\lambda_M-\frac{1}{2}\right)-\frac{0.5L\times subsidy}{2-\lambda_H-\lambda_M+0.5L}\,,$$

(16b) $$V_M^N - V_M^S = q_3\left(t_F, t\right)\left(\lambda_H - \frac{1}{2}\right) + q_4\left(t_F, t\right)\left(\lambda_M - \frac{1}{2}\right) - \frac{0.5L \times subsidy}{2 - \lambda_H - \lambda_M + 0.5L}.$$

Figure 2.5 illustrates the reaction of high-productivity and low-productivity firms to such a subsidy.[13] For convenience, we assume that high-productivity firms first react to a subsidy even though the result does not change if low-productivity firms react first. As before, the thick (thin) line is the share of high-productivity (low-productivity) firms in the North. Since all high-productivity firms and some low-productivity firms locate in the North, $V_H^N - V_H^S > 0$ and $V_M^N - V_M^S = 0$ when subsidy is zero. When the amount of subsidy is small, only (16b) becomes negative and (16a) remains positive. Thus, only low-productivity firms react to a small subsidy from the South. However, once subsidy exceeds a threshold so that (16a) becomes negative, then all high-productivity firms react and relocate to the South, but, at the same time, many low-productivity firms relocate to the North to avoid competition from high-productivity firms. If the subsidy offered by the local government is large enough, the South would be able to attract all firms, including both high- and low-productivity firms.

Next, we consider a subsidy competition between North and South. It is likely that local government in the North also offers a subsidy to prevent firms from relocating to the South. It is clear from equations 16(a) and 16(b) that even when the South offers a large subsidy to attract high-productivity firms, the North can easily offset those incentives with a lower subsidy. Therefore, the North holds a distinct advantage in a subsidy competition. Hence, the subsidy instrument is less likely to attract high-productivity firms to rural areas for the purpose of economic development.[14]

## 2.6 Summary and Conclusion

In this study, we analyzed the agglomeration process in the presence of firm heterogeneity, i.e., productivity differences. We began with a closed-economy setting, where high- and low-productivity firms exhibit a different path in their location choices and the resulting spatial equilibrium. Similar to other studies, transport costs and the degree of competition are key factors determining firms' location choice and the agglomeration process in our model. However, we demonstrate that low-productivity firms can disperse with a fall in transport costs when facing severe competition with their high-productivity counterparts. With a reduction in transport costs, it gets less expensive to ship goods to other regions and the net benefit of agglomeration increases. When the net benefit increases enough to offset the cost of severe competition, low-productivity firms begin to agglomerate with high-productivity ones.

Extending the firm-heterogeneity model of agglomeration to an open-economy setting, we find that trade liberalization acts as a centrifugal force for low-productivity firms. That is, low-productivity firms locate away from their high-productivity counterparts, when both sets of firms face the same set of international trade opportunities. We also find that narrowing productivity differences between domestic firms favors agglomeration, while narrowing productivity differences between domestic and foreign firms favors dispersion.

Finally, we carry out two simulations in the open-economy setting: one for the export value in the two domestic regions and the other with a subsidy competition between these regions. Our results show that the gap in export values between the two regions narrows as international transport costs decrease, although it never disappears. When the less developed region with fewer firms attempts to lure firms away from the other region with a subsidy, it tends to attract low-productivity firms initially. However, when the subsidy becomes large enough to attract high-productivity firms, the other region can easily compensate those firms with a lower subsidy and avoid their relocation.

Our results show that trade liberalization is good for rural communities and under-developed regions. When facing the same set of international trade opportunities, less developed rural regions tend to gain less productive firms, which may only provide low-paying jobs relative to those located in more developed areas. An interesting extension of the model would explore agglomeration of heterogeneous firms when the two domestic regions face different level of international trade costs. Also, our results are based on a model that allows intra-regional mobility of human capital. However, trade liberalization may accompany foreign direct investment and outsourcing, and low-productivity firms may move to a foreign country rather than a less developed region. How globalization, with outsourcing and international movement of human capital, affects less developed regions remains an important topic for future research.

**2.7 Bibliography**

Ades, A. and E. Glaeser. (1995). "Trade and Circuses: Explaining Urban Giants." *Quarterly Journal of Economics*. 110: 195-227.

Allonso-Villar, O. (1999). "Spatial Distribution of Production and International Trade: A Note." *Regional Science and Urban Economics*. 29: 371–380.

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

Baldwin, R. and T. Okubo. (2006). "Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting." *Journal of Economic Geography*. 6: 323-346.

Behrens, K., Gaigne, C., Ottaviano, G. and J. F. Thisse. (2003). "Inter-regional and International Trade: Seventy Years After Ohlin." *CEPR Discussion Paper*. no. 4065.

Bernard, A. and J. B. Jensen. (1999). "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics*. 47: 1-25.

Coniglio, N. (2002). "Regional Integration and Migration: An Economic Geography Model with Heterogeneous Labour Force." *University of Glasgow Discussion Paper*.

Crozet, M. and P. Koenig. (2004). "EU Enlargement and the Internal Geography of Countries." *Journal of Comparative Economics*. 32: 265-279.

Dupont, V. and P. Martin. (2006). "Subsidies to Poor Regions and Inequalities: Some Unpleasant Arithmetic." *Journal of Economic Geography*. 6: 223-240.

Feenstra, R.C. and G.H. Hanson. (1997). "Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras." *Journal of International*

*Economics*. 42: 371-393.

Feliciano, Z.M. (2001). "Workers and Trade Liberalization: The Impact of Trade Reforms in Mexico on Wages and Employment." *Industrial and Labor Relations Review*. 55: 95-115.

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

Galiani, S. and P. Sanguinetti. (2003). "The Impact of Trade Liberalization on Wage Inequality: Evidence from Argentina." *Journal of Development Economics*. 72: 497-513.

Glaeser, E. and D. Mare. (2001). "Cities and Skills." *Journal of Labor Economics*. 19: 316-342.

Goldberg, P.K. and N. Pavcnik. (2005). "Trade, Wages, and the Political Economy of Trade Protection: Evidence from the Colombian Trade Reforms." *Journal of International Economics*. 66: 75-105.

Hanson, G.H. (1997). "Increasing Returns, Trade and the Regional Structure of Wages." *Economic Journal*. 107: 113-133.

Hanson, G.H. (1998). "Regional Adjustment to Trade Liberalization." *Regional Science and Urban Economics*. 28: 419-444.

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

Krugman, P. and R. Livas. (1996). "Trade Policy and the Third World Metropolis." *Journal of Development Economics*. 49: 137-150.

Mansori, K. (2003). "The Geographic Effects of Trade Liberalization with Increasing Returns in Transportation." *Journal of Regional Science*. 43: 249-268.

Melitz, M. (2003). “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity.” *Econometrica*. 71: 1695-1725.

Ottaviano, G., Tabuchi, T. and J. F. Thisse. (2002). “Agglomeration and Trade Revisited.” *International Economic Review*. 43: 409-435.

Overman, H. and L. A. Winters. (2006). “Trade and Economic Geography: The Impact of EEC Accession on the UK.” *CEPR Discussion Paper*. no. 5574.

Paluzie, E. (2001). “Trade Policy and Regional Inequalities.” *Papers in Regional Science*. 80: 67-85.

Redding, S. and A.J. Venables. (2004). “Economic Geography and International Inequality.” *Journal of International Economics*. 62: 53-82.

Syverson, C. (2004). “Market Structure and Productivity: A Concrete Example.” *Journal of Political Economy*. 112: 1181-1222.

Yale Global Forum. (2007). http://yaleglobal.yale.edu/globalization/.

Yeaple, S. (2005). “A Simple Model of Firm Heterogeneity, International Trade and Wages.” *Journal of International Economics*. 65: 1-20.

**2.8 Endnotes**

[1]In the context of international income inequality, see Redding and Venables (2004).

[2]The literature on EU enlargement is fairly extensive. See also Crozet and Koenig (2004) and Overman and Winters (2006).

[3]The following results can be recast in the context where a region can be regarded as a country.

[4]Support for the assumption that skilled labor is much more mobile than unskilled labor can be found in Coniglio (2002).

[5]To illustrate, consider the case where there are two high-productivity firms and two low-productivity firms. Then, the inverse demand functions from utility maximization are:

$$p_H(i) = \alpha - \beta c_H(i) - \gamma c_H(j) - \delta\left(c_M(i) + c_M(j)\right),\ i, j = 1,2 \ \ i \neq j$$
$$p_M(i) = \alpha - \beta c_M(i) - \gamma c_M(j) - \delta\left(c_H(i) + c_H(j)\right),\ i, j = 1,2 \ \ i \neq j$$

[6]When $\delta = \gamma$, the utility function reduces to:

$$U = \alpha\left(\int_0^n c_H(i)\,di + \int_0^n c_M(i)\,di\right) - \frac{\beta-\gamma}{2}\left(\int_0^n c_H^2(i)\,di + \int_0^n c_M^2(i)\,di\right) - \frac{\gamma}{2}\left(\int_0^n c_H(i)\,di + \int_0^n c_M(i)\,di\right)^2 + c_A.$$

[7]We assume that *E* is sufficiently large so that consumption of agricultural goods is positive.

[8]Suppose there are 1000 industrialists (entrepreneurs) in the economy. Then, our normalization measures industrialists (entrepreneurs) in one-thousand units.

[9]See Appendix B for the reason why $t_{\max} > C_1$ ensures the dispersion of firms at the initial point

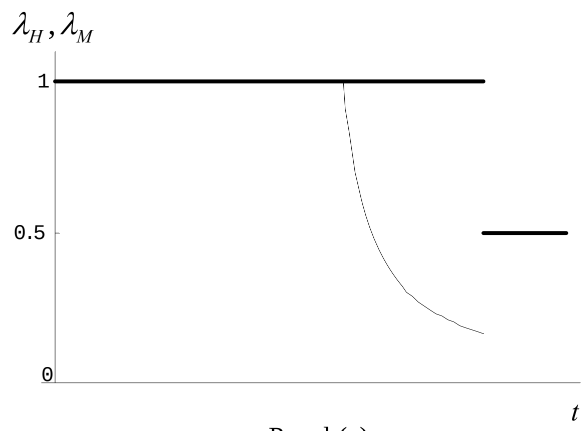
[10]At the extreme, i.e., when $\delta$ is very small, both high-productivity and low-productivity firms instantaneously move to the North.

[11]When *H* and *M* goods are not substitutable, i.e., $\delta = 0$, $q_i(t_F, t) = 0$ is not a function of $t_F$. This implies that international trade affects the firm location only when competition between *H* and *M* goods is not weak.

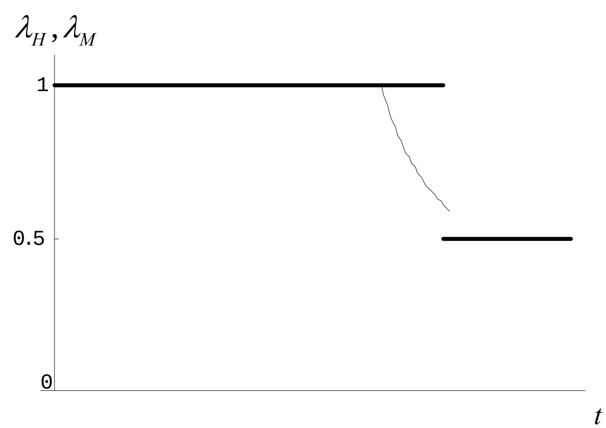
[12]Parameter values are $\alpha = 0.2$, $\beta = 1$, $\gamma = 0.5$, $Lab = 50$, $K_F = 5$, $Lab_F = 100$, $\theta_M = 0.005$, $\theta_F = 0.0025$ and $t = 0.03$.

[13]Parameter values used in this simulation are the same as those used for figure 2.4.

[14]In Dupont and Martin's (2006) study, with homogeneous firms, a subsidy attracts firms to underdeveloped areas, but, in some cases, it can be an income transfer from a poor to a rich area.




Panel (a)




Panel (b)

**Figure 2.1: Competition, Transport Costs and Agglomeration of Heterogeneous Firms**

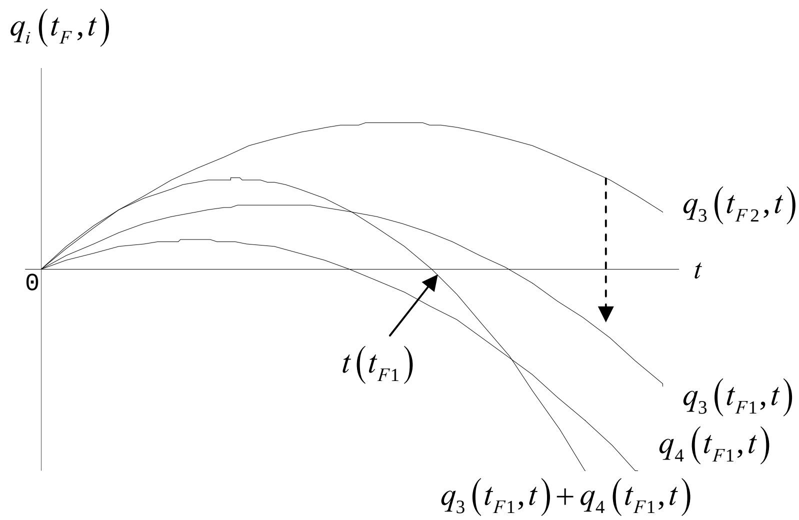



**Figure 2.2: International Trade Opportunity and Agglomeration**

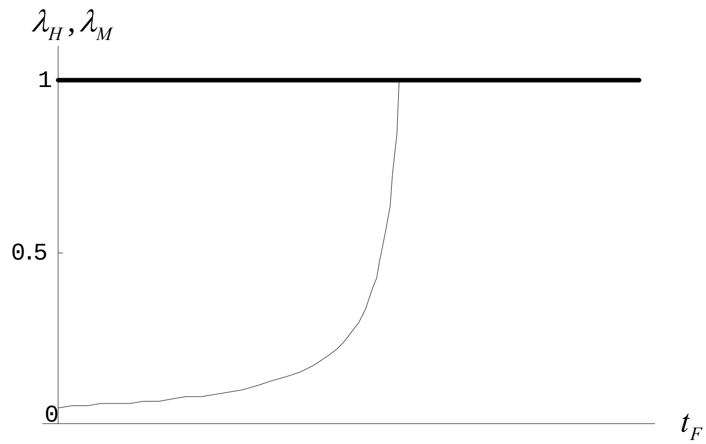



**Figure 2.3: International Transport Costs and Agglomeration of Heterogeneous Firms**

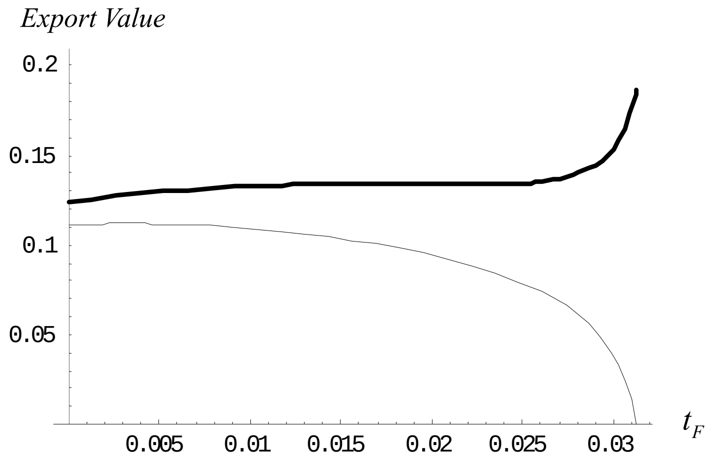



**Figure 2.4: Regional Export Values and International Transport Costs**

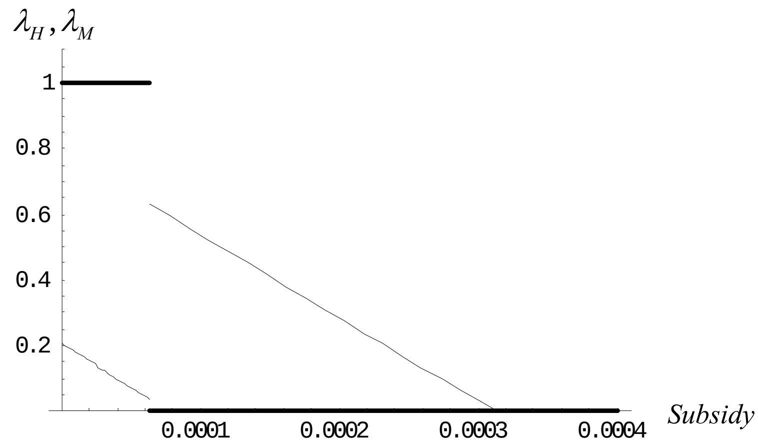



**Figure 2.5: Subsidy and Agglomeration of Heterogeneous Firms**

# 3. Plants' Self-Selection, Agglomeration Economies and Regional Productivity in Chile

## 3.1 Introduction

Observed differences among firms within an industry – employment and output size, capital and skill intensity, and productivity – are large (Bernard and Jensen, 1999; Tybout, 2000). Such heterogeneity, observed in the spatial dimension, has become a centerpiece of the emerging literature on new economic geography (Krugman, 1991; Fujita et al., 1999; Baldwin et al., 2002). For instance, Henderson et al. (2001) find that rural regions' output, employment, and wages have persistently lagged behind that of their urban counterparts. Focusing on the sources of spatial inequalities, Rice et al. (2006) attribute most of the urban-rural income difference to that in productivity between the two regions. Hence, the question why we observe large inter-regional productivity differences takes on significance in policymakers' search for solutions to spatial inequality in economic development.

Since regional productivity is inherently related to the inhabitants, the new-economic-geography literature has investigated firms' location decision focusing on demand- or supply-side factors (Fujita et al., 1999; Baldwin et al., 2002). Both economic factors contribute to the agglomeration of firms in urban areas, which has received much research attention. The forces of agglomeration include transport costs, market competition, availability of skilled labor and intermediates, congestion, and knowledge or informational spillovers. When forces pulling firms toward a center or core (centripetal) more than offset those that pull them apart (centrifugal), agglomeration occurs. The resulting spatial pattern determines firms' productivity and profitability causing economic differences among regions. Empirical evidence on each of the agglomeration forces begins to accumulate (Rosenthal and Strange, 2004). For instance, the urban economics literature shows the existence of localization and urbanization economies, i.e., externalities, using aggregate as well as plant-level data (Henderson,1986 and 2003; Ciccone and Hall, 1996; Holmes and Stevens, 2002).[1]

Agglomeration economies, however, are not the only reason for regional productivity differences. A key component of the agglomeration process is the self-selection of heterogeneous firms to locate in markets with specific characteristics. In analyzing firms' location decisions in the presence of heterogeneity, Baldwin and Okubo (2006) show that high-productivity firms self-select into large markets or countries. The reason for the self-selection is the larger benefits from agglomeration to high-productivity firms relative to their low-productivity counterparts, when their products are highly substitutable. Hence, average productivity level in large markets is higher than that in small markets. On the latter theme, Syverson (2004) investigates the spatial productivity differences in the U.S. ready-mix concrete industry, where plants produce a homogeneous product. In large markets, Syverson (2004) finds not only a higher average productivity but also a lower dispersion due to the truncation (from below) of the productivity distribution.[2]

The objective of this study is to assess the relative contribution of plants' self-selection and agglomeration economies to regional productivity. For this purpose, we focus on plants in the Chilean food manufacturing industry. Indeed Chile and, in particular, the food industry provide an excellent setting to study self-selection for several reasons. Most of Chile's 13 regions have a coastline. Thus, the assumption of the first essay that every region faces the same level of trade costs will likely hold. Also, since over 90% of the plants are single-plant firms, we are less likely to face measurement errors in modeling location decision relative to the case with multi-plant firms (Pavcnik, 2002). Next, the Santiago Metropolitan Region accounts for 40 percent of the population, suggesting significant spatial concentration of economic activity. In contrast, the food industry appears to be more dispersed than any other industry in the Chilean manufacturing sector. In addition, the food firms constitute the largest manufacturing industry and the major source of employment in Chile. Finally, processed foods are major exports, which allows us to explore the role of self-selection of exporting plants in augmenting regional productivity.

We first measure plant-level productivity controlling for self-selection and a simultaneity bias. The former is discussed above, but the latter is due to Olley and Pakes (1996), who argue for the possible simultaneity between a plant's productivity and input levels (Levinsohn and Petrin, 2003). The resulting productivity estimates and their regional distributional measures - median, $10^{th}$, and $90^{th}$ percentile values – allow us to explore spatial productivity differences emphasizing the roles of self-selection and agglomeration economies.

### 3.2 Theoretical Basis and Empirical Methodology

Our objective is to identify the relative contribution of plants' self-selection and agglomeration economies to a region's productivity level. For this purpose, we estimate a plant-level production function, which corrects for (i) simultaneity between input quantities and productivity, and (ii) plants' self-selection into large markets. For the latter, we explore two possible cases: high-productivity and high-productivity exporting plants, both of which are likely to self-select into large markets. Then, we obtain a distribution of plant productivity by region for each of the two cases of self-selection and derive the median, $10^{th}$, and $90^{th}$ percentile values. These distributional measures serve as dependent variables in a tobit model to assess the effects of self-selection and agglomeration economies on regional productivity. We employ a tobit model to allow for the possibility that some regions may have a degenerate productivity distribution.

We first outline the broader simultaneity problem. Most studies of plant-level productivity begin with the estimation of a first-order approximation of a production function (Olley and Pakes, 1996; Pavcnik, 2002; Henderson, 2003; Levinsohn and Petrin, 2003; Syverson, 2004). Here, productivity is often represented as a parameter in the production function employed. Otherwise, productivity would be omitted from the production function, but become a part of the random disturbance. In the latter case, any simultaneity between conventional input levels and productivity would be ignored.[3] Then, least squares estimates of the production function, with productivity as an unidentified part of the disturbance term, are biased. Hence, the above plant-level studies estimate the productivity parameter as either a fixed plant- and region-specific or a random, time-varying coefficient (Olley and Pakes, 1996; Pavcnik, 2002; Henderson, 2003; Levinsohn and Petrin, 2003). The seminal contribution of Olley and Pakes (1996) shows conditions under which an investment proxy controls for the simultaneity between input levels and productivity. Pavcnik (2002) applies Olley and Pakes' (1996) technique to Chilean manufacturing plants, but Levinsohn and Petrin (2003) show that the investment proxy to control for the above simultaneity is valid only when plants report non-zero investments. In several plant-level databases from U.S., U.K., Turkey, Chile and others, the zero-investment problem is fairly large. Levinsohn and Petrin (2003) propose the use of an intermediate-input proxy to correct for the simultaneity between conventional input levels and productivity.

Self-selection creates a similar simultaneity problem if externalities such as localization and urbanization economies are included in the set of independent variables used in the production function (Henderson, 2003). That is, if high-productivity (exporting) plants self-select into large markets as in Baldwin and Okubo (2006) or Syverson (2004), then productivity and externality variables are likely correlated. Again, least squares estimates of the production function are biased if productivity is not directly represented, but included in the disturbance term. With an explicit productivity parameter, most plant-level studies have employed a two-step procedure involving semi-parametric or series approximation techniques to correct for either self-selection or simultaneity, but not both.

In the following, we extend the Levinsohn and Petrin (2003) and Henderson (2003) approaches to estimate a plant-level (Cobb-Douglas) production function, which controls for simultaneity between conventional inputs and productivity, and possible self-selection of high-productivity (exporting) plants. In particular, we (i) directly account for agglomeration economies on plant output, as in Henderson (1986, 2003) and Nakamura (1985) and (ii) incorporate productivity as a random, plant-, region- and time-specific coefficient:[4]

$$(1) \qquad \begin{aligned} \ln Y_{irt} = \beta_0 + \beta_1 \ln S_{rt} + \beta_2 \ln L^s_{irt} + \beta_3 \ln L^u_{irt} + \beta_4 \ln K_{irt} \\ + \beta_5 \ln M_{irt} + \beta_6 \ln E_{irt} + \omega_{irt} + \varepsilon_{irt} \end{aligned},$$

where, *i*, *r* and *t* denote plant, region and time, respectively. The variables $L^s$, $L^u$, *K*, *M*, *E* and *S* respectively represent the number of skilled workers, the number of unskilled workers, capital stock, material (intermediates), energy, and the market characteristics. The parameter $\omega$ denotes the productivity of the plant, and $\varepsilon$ is an *i.i.d.* disturbance term. If we do not directly include $\omega$ in the production function, the resulting disturbance term would be $\xi_{irt} = \omega_{irt} + \varepsilon_{irt}$ and $\text{cov}[\xi_{irt}, M_{irt}] \neq 0$. Hence, the estimate of $\beta_3$ would be biased if we ignored simultaneity between productivity and materials.[5] Similarly, if $\text{cov}[\xi_{irt}, S_{rt}] \neq 0$, the estimate of $\beta_1$, a measure of externality effects, would be biased.[6]

As noted earlier, we extend Levisohn and Petrin's (2003) approach to control for both simultaneity bias and self-selection by using a material-input proxy. That is, a plant determines the material input level given the productivity realization and capital stock as follows:[7]

(2) $$\ln M_{irt} = m\left(\omega_{irt}, \ln K_{irt}\right),$$

where $m(\cdot)$ is a monotonic function in the level of productivity. Hence, $m(\cdot)$ is invertible and plant's productivity can be written as:

(3) $$\omega_{irt} = \omega\left(\ln M_{irt}, \ln K_{irt}\right).$$

Then, the production function is rewritten as:

(4) $$\ln Y_{irt} = \beta_1 \ln S_{rt} + \beta_2 \ln L_{irt}^{s} + \beta_3 \ln L_{irt}^{u} + \beta_6 \ln E_{irt} + \varphi\left(\ln M_{irt}, \ln K_{irt}\right) + \varepsilon_{irt},$$

where, $\varphi(\cdot)$ is a function of capital and material inputs:

(5) $$\varphi\left(\ln M_{irt}, \ln K_{irt}\right) = \beta_0 + \beta_4 \ln K_{irt} + \beta_5 \ln M_{irt} + \omega\left(\ln M_{irt}, \ln K_{irt}\right).$$

The first stage of Levinsohn and Petrin's (2003) procedure is a least squares estimation of equation (4) by approximating $\varphi(\cdot)$ with a third-order polynomial series estimator:[8]

(6) $$\begin{aligned}\ln Y_{irt} &= \beta_1 \ln S_{rt} + \beta_2 \ln L_{irt}^{s} + \beta_3 \ln L_{irt}^{u} + \beta_6 \ln E_{irt} \\ &\quad + \sum_{p=0}^{3} \sum_{q=0}^{3-p} \theta_{pq} \ln M_{irt}^{p} \ln K_{irt}^{q} + \varepsilon_{irt}\end{aligned}.$$

Since $\varepsilon$ is *i.i.d*. and assumed to be independent of $L^s$, $L^u$ and *E*, the OLS estimates of $\beta_1$, $\beta_2$, $\beta_3$ and $\beta_6$ are consistent but $\beta_4$ and $\beta_5$ are not identified at this stage.

For consistent estimation of $\beta_4$ and $\beta_5$, we assume that a plant's productivity follows a first-order Markov process as in Levinsohn and Petrin (2003):

(7) $$\begin{aligned}\beta_0 + \omega_{irt+1} &= \beta_0 + E\left[\omega_{irt+1} \mid \omega_{irt}\right] + \eta_{irt+1} \\ &= g\left(\omega_{irt}\right) + \eta_{irt+1}\end{aligned},$$

where $\eta_{irt+1}$ is considered to be an innovation. Substituting equation (7) into the production function in equation (1) for $t+1$, we obtain:

(8)
$$\begin{aligned}\ln Y_{irt+1} &= \beta_1 \ln S_{rt+1} + \beta_2 \ln L^s_{irt+1} + \beta_3 \ln L^u_{irt+1} + \beta_4 \ln K_{irt+1} \\ &+\beta_5 \ln M_{irt+1} + \beta_6 \ln E_{irt+1} + g\left(\omega_{irt}\right) + \varepsilon^*_{irt+1}\end{aligned},$$

where $\varepsilon^*_{irt+1} = \eta_{irt+1} + \varepsilon_{irt+1}$. As noted prior to equation (2), capital at period $t+1$, $K_{t+1}$, is uncorrelated with $\varepsilon^*_{t+1}$ because investment decision is assumed to be made at period *t*:

(9)
$$E\left[\varepsilon^*_{t+1} \ln K_{t+1}\right] = 0 .$$

From equation (5), $\omega_{irt}$ can be rewritten as:

(10)
$$\begin{aligned}\beta_0 + \omega_{irt} &= \hat{\varphi}\left(\cdot\right) - \beta_4 \ln K_{irt} - \beta_5 \ln M_{irt} \\ &= \sum_{p=0}^{3} \sum_{q=0}^{3-p} \hat{\theta}_{pq} \ln M^p_{irt} \ln K^q_{irt} - \beta_4 \ln K_{irt} - \beta_5 \ln M_{irt}\end{aligned}.$$

Using a third-order approximation for $g\left(\cdot\right)$, we estimate the following equation by a nonlinear instrumental variables (NIV) estimator,

(11)
$$\begin{aligned}Z_{irt+1} = \beta_4 \ln K_{irt+1} + \beta_5 \ln M_{irt+1} + \sum_{p=0}^{3} \delta_p \left(\hat{\varphi}\left(\cdot\right) - \beta_4 \ln K_{irt} - \beta_5 \ln M_{irt}\right)^p \\ +\varepsilon^*_{irt+1}\end{aligned},$$

where $Z_{irt+1} = \ln Y_{irt+1} - \hat{\beta}_1 \ln S_{rt+1} - \hat{\beta}_2 \ln L^s_{irt+1} - \hat{\beta}_3 \ln L^u_{irt+1} - \hat{\beta}_6 \ln E_{irt+1}$. Since $\ln M_{irt+1}$ is likely to be correlated with $\varepsilon^*_{irt+1}$, we use an NIV procedure to obtain consistent estimates of $\beta_4$ and $\beta_5$. The instruments used in the estimation of equation (11) are $\hat{\varphi}\left(\ln M_{irt}, \ln K_{irt}\right)$ from the first stage, $\ln K_{irt+1}$, $\ln K_{irt}$, $\ln M_{irt}$ and $\ln E_{irt}$.

Given the multiple steps in the estimation, it is difficult to compute the standard error of parameter estimates and hence, we use bootstrapping procedures (Levinsohn and Petrin, 2003). First, plants are drawn from the original sample with replacement and with equal probability until the number of plant-year observations in the new sample (closely) equals the number of plant-year observations in the original sample. When a plant is drawn, all observations for that plant are included in the new sample. The square root of the variance of estimated parameters across samples is used to compute the standard error.[9]

Finally, the raw productivity of plant *i* in region *r* at period *t*, controlling for agglomeration economies, is derived as follows:

$$
\begin{aligned}
(12) \qquad prod_{irt} &= e^{\omega_{irt}} \\
&= \exp\left(\ln Y_{irt} - \hat{\beta}_1 \ln S_{rt} - \hat{\beta}_2 \ln L^{s}_{irt} - \hat{\beta}_3 \ln L^{u}_{irt} - \hat{\beta}_4 \ln K_{irt} \right.\\
&\qquad \left. - \hat{\beta}_5 \ln M_{irt} - \hat{\beta}_6 \ln E_{irt}\right).
\end{aligned}
$$

Note that equation (12) derives a plant's raw productivity by subtracting the effect of agglomeration economies (*S*) in the right-hand side. For every time period, we can then obtain the median, 10$^{th}$, and 90$^{th}$ percentile values of plant productivity for each region after dropping provinces with less than 10 plants as in Syverson (2004).[10] The 10$^{th}$ and 90$^{th}$ percentile values are used to avoid misrepresenting productivity when some plants realize large positive or negative shocks to their productivity (Syverson, 2004).

Recall that our estimation of the plant-level production function is flexible enough to account for self-selection, but does neither test nor identify the process by which plants self-select into certain markets. Given the distributional measures of regional productivity, we first test the hypothesis that high-productivity plants agglomerate in large markets by estimating the following equation:

(13) $$z_{prt} = \alpha_0 + \alpha_1 \ln S_{rt} + \alpha_2 X_{rt} + \nu_{prt}, \quad p = 10, 50, 90,$$

where $z_{prt}$ is the *p*-th percentile value of the raw productivity distribution of region *r* at period *t*. As noted before, $S_{rt}$ represents *r*-th market's characteristics in period *t*, while $X_{rt}$ is a set of control variables such as regional characteristics. We recognize that $z_{prt}$ may take value zero in some regions, e.g., some regions may not have any plants, as is the case in our application to Chile. Estimating equation (13) with positive $z_{prt}$ alone would yield biased estimates. Hence, we use a tobit model, where the observable variable $z_{prt}$ equals the latent variable only when the latter takes a positive value:

(14) $$z_{prt} = \begin{cases} z^*_{prt} & if \quad z^*_{prt} > 0 \\ 0 & if \quad z^*_{prt} \leq 0 \end{cases},$$

where $z^*_{prt}$ is the latent variable, which takes the following form:

(15) $$z^*_{prt} = \alpha_0 + \alpha_1 \ln S_{rt} + \alpha_2 X_{rt} + \nu_{prt} \quad \nu_{prt} \sim N\left(0, \sigma^2\right).$$

Equations (13)-(15) to test self-selection are applied to the two cases: all and exporting plants. In the latter, we use only exporting plants to compute regional (raw) productivity measures.[11]

The estimation of the tobit model leads to an evaluation of the relative contribution of self-selection and agglomeration economies to regional productivity. To do so, consider the following representation of *p*-th percentile regional productivity following the production function in equation (1) and the productivity representation in endnote 4:

(16) $$v_p = S^{\beta_1} \exp\left(\omega_p\right) = S^{\beta_1} z_p,$$

where $z_p$ is the *p*-th percentile value of regional raw productivity distribution and $\omega_p$ is the corresponding plant productivity. Then, the elasticity of regional productivity level with respect to market characteristics, *S*, is

(17) $$\frac{d \ln v_p}{d \ln S} = \beta_1 + \frac{1}{E\left[z_p \mid \bar{x}\right]} \frac{\partial E\left[z_p \mid \bar{x}\right]}{\partial \ln S},$$

where $\partial E\left[z_p \mid \bar{x}\right] / \partial \ln S$ is the marginal impact on the expected *p*-th percentile value of the raw regional productivity distribution evaluated at the sample mean. The first term on the right-hand-side of equation (17) is the increase in regional productivity associated with agglomeration economies, while the second term is attributed to self-selection, i.e., the extent to which market characteristics raise the *p*-th percentile productivity of a region.

### 3.3 Data and Variables

For estimating plant and regional productivity and identifying the latter's relationship to self-selection, we use data from the Chilean Manufacturing Census from 1998 to 2003 collected by National Statistics Institute (INE). Plants with at least 10 employees are surveyed in the Census, with information on plant location at three administrative levels: macro-region, province and county. We avoid using the macro-region and county as the geographical unit since the former is too aggregated (13 macro-regions) and the latter has too many counties without manufacturing plants. Hence, we use each of the 51 provinces as a geographical unit (region) to assure enough variation in data.

We focus on the food industry (ISIC 31) because it is the largest manufacturing industry in Chile (Roberts and Tybout, 1997). It accounts for 32.4 and 41.3 percent of manufacturing sectors' real GDP and real exports, respectively in 2005 (Central Bank of Chile). In addition, plants in this industry are dispersed around the country, providing an important source of employment to most of the 51 regions (figure 3.1). In contrast, most plants in other manufacturing industries concentrate in Santiago Metropolitan Region (table C.1 in the Appendix C). Moreover, we do not have enough degrees of freedom to estimate the tobit model for other industries. For example, we have less than 10 uncensored observations in two other industries that appear as dispersed as the food industry: glass (ISIC 36) and basic metals (ISIC 37).[12]

Six variables are constructed to estimate the plant-level production function: skilled labor, unskilled labor, materials, energy, capital and output.[13] The unit of labor is the number of workers, while other variables are in Chilean Peso. We use deflators (base year 1996) from the Chilean National Account to convert the latter variables into constant Chilean Peso. Three types of externalities are considered in urban economics: localization economies, urbanization economies, and demand-driven scale economies (Henderson, 2003; Lall, Shalizi, and Deichmann, 2004). As noted earlier, localization economies arise if firms or plants belonging to the same industry concentrate in a region. Urbanization externalities relate to spatial concentration of plants from all industries. Furthermore, greater demand in large markets makes it possible for plants to adopt efficient production processes. All three externalities are considered to be centripetal forces of agglomeration (Eberts and McMillen, 1999). We measure these three externalities respectively by the employment share of the food industry in the manufacturing sector ($LOC_{rt}$), the Herfindahl-Hirschman Index based on employment by industry ($HHI_{rt}$), and regional population ($SCL_{rt}$).[14] The first two variables are constructed from the Chilean Manufacturing Census, while data on regional population is obtained from Population Projections, INE. Note that a higher $HHI_{rt}$ implies less industrial diversification (greater specialization).

For the provincial characteristics, $X_{rt}$ in equation (13), we use (i) employment in agricultural and fishery sectors ($AG_r$) from Housing and Population Census, (ii) human capital index ($HC_{rt}$) from The Regional Competitiveness Report, Chilean Ministry of Economy (CME) and (iii) distance from each provincial capital to Santiago Province ($DIST_r$), which is the economic center of Chile.[15] Great-circle distance is used to measure distance in kilometers. Since agricultural products are major material inputs in the food industry, we expect that plants favor a location where they have an easy access to intermediates. If high-productivity plants are human capital intensive, they may prefer a region with higher human capital. Since province is an administrative geographical unit, economic interaction may not complete within a province. Rather than use regional population or GDP, Head and Mayer (2004) use "market potential" constructed by inter-regional trade flows. Because data on regional trade flows are not available, we include the distance to Santiago Province to capture market potential.

In the case where we focus on only exporting plants, the share of manufacturing employment engaged in exporting plants in total manufacturing employment of a region ($EXP_{rt}$) is also included in $X_{rt}$. This is based on the positive spillover effect from other exporting firms on a firm's export decision and export values (Aitken et al., 1997; Malmberg et al., 2000). Hence, exporters might prefer to locate near other exporters. Table 3.1 provides descriptive statistics on each of the above variables used in our estimation.

## 3.4 Results

<u>Plant-Level Production Function Estimates</u>: The estimated plant-level production function is:[16]

(18) $$\begin{aligned}\ln Y_{irt} = \beta_0 + \beta_{11} \ln LOC_{irt} + \beta_{12} \ln HHI_{rt} + \beta_{13} \ln SCL_{rt} + \beta_2 \ln L^s_{irt} + \beta_3 \ln L^u_{irt} \\ + \beta_4 \ln K_{irt} + \beta_5 \ln M_{irt} + \beta_6 \ln E_{irt} + \omega_{irt} + \varepsilon_{irt}.\end{aligned}$$

Table 3.2 shows the parameter estimates of equation (18). The largest coefficient is that on material, suggesting that the food industry is material-intensive. Capital has the second largest elasticity, while labor and energy elasticities are similar. Standard errors from bootstrapping suggest that the coefficients on all conventional inputs are statistically significant at the 1 percent level. Returns to scale, which equals the sum of coefficients on conventional inputs, is estimated to be 0.96 and the 95 percent confidence interval includes the value 1. Hence, the technology in food processing appears to exhibit constant returns to scale. With regard to externalities in production, only one of the three coefficients, that on demand-driven scale economies, is statistically significant with the expected positive sign. Although the coefficient on urbanization economies (*HHI*) has the expected negative sign, it is not statistically significant.

Regional Productivity and A Heteroskedastic Tobit Model: Given the parameter estimates of equation (18), we compute raw, plant-level productivity using equation (12). We rewrite equation (15) in estimable form as:

(19) $$z^*_{prt} = \alpha_0 + \alpha_{11} \ln LOC_{rt} + \alpha_{12} \ln HHI_{rt} + \alpha_{13} \ln SCL_{rt} + \alpha_{21} \ln AG_r + \alpha_{22} \ln HC_{rt} + \alpha_{23} \ln DIST_r + \nu_{prt}.$$

For the case of exporting plants, equation (15) takes the following form:

(20) $$z^*_{prt} = \alpha_0 + \alpha_{11} \ln LOC_{rt} + \alpha_{12} \ln HHI_{rt} + \alpha_{13} \ln SCL_{rt} + \alpha_{21} \ln AG_r + \alpha_{22} \ln HC_{rt} + \alpha_{23} \ln DIST_r + \alpha_{24} \ln EXP_{rt} + \nu_{prt}.$$

The expected sign of $\alpha_{11}$ and $\alpha_{13}$ ($\alpha_{12}$) is positive (negative), while the control variables except the distance are expected to take a positive sign.

Although we have captured most regional characteristics in equations (19) and (20), there may still exist some unobserved heterogeneity among regions

(Hanson, 2001). Unfortunately, we cannot add a dummy for each region (51 provinces) given the limited degrees of freedom. Moreover, the fixed effects tobit model cannot be implemented when all observations in a region are censored as in a couple of cases of our sample (Greene, 2006). Note also that variables $AG_r$ and $DIST_r$ are time invariant in our case. Therefore, we use macro-regional (12) and time (5) dummies to capture any left-over heterogeneity.[17]

We first estimate tobit models assuming homoskedastic disturbances. To assess the effect of explanatory variables on actual regional productivity measures ($z_{prt}$), marginal effects evaluated at the sample mean are presented in columns (1) to (3) of tables 3 and 4.[18] Standard errors are computed using the delta method (Greene, 2003). Although the initial version of the tobit model shows expected externality effects on regional productivity, the maximum likelihood estimates of equations (19) and (20) are not consistent in the presence of heteroskedasticity (Greene, 2003). Heteroskedasticity matters because the variance of the regional productivity distribution may be inversely related to market size as in Syverson (2004). Moreover, the number of plants per province is positively related to the market size, leading to more accurate measures (less variable) of productivity distribution ($z_{prt}$) in large markets. Hence, we correct for heteroskedasticity of the following form in the tobit model:[19]

(21) $$\sigma_i^2 = \sigma^2 \left[ 1 + \exp\left( \gamma' x_i \right) \right],$$

where, $x_i$ is $i$-th observation of all explanatory variables except the intercept and regional and time dummies in the self-selection equations (19) and (20). Likelihood ratio (LR) statistic for the hypothesis that $\gamma = 0$, i.e., homoskedastic disturbances in the $10^{th}$, $50^{th}$ (median) and $90^{th}$ percentiles equations, is 59.91, 81.53 and 13.27, respectively, for the self-selection model with all plants. Corresponding LR statistic for the case of exporting plants is 37.33, 42.55 and 137.71 respectively. Since the calculated LR statistic is larger than the critical value from a chi-squared distribution with six (12.59) and seven (14.07) degrees of freedom for all and exporting plants, respectively, the null hypothesis is rejected in every equation.

We derive marginal effects which account for the heteroskedastic specification of the tobit model as in Greene (2003). The expected value of the *p*-th percentile regional productivity is:

$$(22) \qquad E\left[z_p \mid \bar{x}\right] = \Phi\left(\frac{\bar{x}'\beta}{\bar{\sigma}}\right)\bar{x}'\beta + \bar{\sigma}\phi\left(\frac{\bar{x}'\beta}{\bar{\sigma}}\right),$$

where, $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution and probability density function of a standard normal variable and $\bar{\sigma}$ is a squared error of variance from (21) evaluated at sample mean. Given $\bar{\sigma}^2$ in equation (21), taking derivative of (22) with respect to the *j*-th explanatory variable yields:

$$(23) \qquad \frac{\partial E\left[z_p \mid \bar{x}\right]}{\partial x_j} = \Phi\left(\frac{\bar{x}'\beta}{\bar{\sigma}}\right)\beta_j + \phi\left(\frac{\bar{x}'\beta}{\bar{\sigma}}\right)\frac{\bar{\sigma}\gamma_j \exp\left(\gamma'\bar{x}\right)}{2\sqrt{1+\exp\left(\gamma'\bar{x}\right)}}.$$

Marginal effects corrected for heteroskedasticity are presented in columns (4) to (6) of tables 3 and 4.[20] Note that the marginal effects are very similar to the coefficients in tables 3.7 and 3.8 because $\Phi(\cdot)$ and $\phi(\cdot)$ evaluated at the sample mean approach one and zero, respectively.[21] The latter might be due to the proportion of zero observations in the estimation of the tobit model, 7.5 and 37.5 percent in equations (19) and (20), respectively.

With all plants (table 3.3), the heteroskedastic-tobit model shows that most coefficients on externality variables - *LOC*, *HHI* and *SCL* - have the expected sign. Although only 4 out of the possible 9 externality coefficients are statistically significant in table 3.3 (columns 4-6), a pattern emerges on the impact of externalities on the median, left- and right-tail of the regional productivity distribution. The coefficients on *LOC* and *SCL* are significantly positive and relatively large in the regression of 90$^{th}$ percentile productivity (column 6), highlighting the importance of agglomeration of the food industry and large markets (Syverson, 2004). Industrial diversification appears less important in extending the right-tail of a region's productivity distribution. Similarly, the 10$^{th}$ percentile productivity is positively impacted by the agglomeration of the food and other industries (*LOC* and *HHI*). Surprisingly, the median productivity of a region is not affected by externality variables. The explanation of the latter result likely appears in the sign and significance of the control variables. In particular, the coefficient on human capital is significant only in the median productivity regression. That is, increases in educational attainment are related to a region's median (or average) productivity, but the variance is likely explained by strong input-output linkage with other food processing plants or knowledge spillovers from other plants in the same industry (e.g., *LOC*). With regard to the other control variables, the coefficient on agricultural employment (*AG*) is significant in the $10^{th}$ and $50^{th}$ percentile regression. The parameter estimate on distance to Santiago, where we had hoped to capture the effect of market potential, takes an unexpected positive sign. The result on distance may reflect the spatial pattern of land rents with Santiago Province as the economic center of Chile. The lower the land rent, i.e., the farther from Santiago, the greater is the cost-saving, which might attract plants.

For the case of exporting plants only, the results of the heteroskedastic tobit model (columns 4-6, table 3.4) show that 7 out of 9 coefficients on externality variables take the expected sign with statistical significance. They also show a large coefficient on externality variables in the $90^{th}$ percentile productivity regression (column 6) relative to that of the median or $10^{th}$ percentile productivity models (column 5 or 4). The differential impact of agglomeration economies on median and variance of regional productivity in table 3.3 disappears in the case of regressions with only exporting plants for the following reason: since most exporting plants have higher productivity relative to non-exporters (Aitken et al, 1997; Bernard and Jensen, 1999), this is essentially a microscopic view of high-productivity plants (column 6) in table 3.3. With regard to other control variables, we find a positive and statistically significant coefficient on *EXP* for each of the three regional productivity models with only exporting plants. This suggests that exporting plants are attracted to a region where the information and knowledge associated with exporting are readily available. The coefficient on agricultural employment takes an unexpected sign in the median productivity regression (column 5, table 3.4), which might imply that exporting plants are less agricultural-input intensive relative to other plants. The positive effect of distance persists for exporting plants as well.

In general, the estimates of the heteroskedastic-tobit model suggest that high-productivity (exporting) plants self-select into large markets, where other plants in food industry agglomerate. Self-selection effects appear stronger in the case of exporting plants relative to those of all plants suggesting that the former is more sensitive to market characteristics.

Self-Selection Versus Agglomeration Economies in Regional Productivity: Based on the parameter estimates of the tobit model, we now compare the effect of plants' self-selection relative to that of agglomeration economies on regional productivity distribution.

Recall that equation (17) decomposed the elasticity of *p*-th percentile regional productivity with respect to market characteristics into that attributable to (i) agglomeration economies and (ii) self-selection. Although equation (16) considered all three externality effects on the production function, we restrict our attention to only the significant effect of the demand-driven scale economies (*SCL*) in table 3.2. Therefore, the estimate of (i) is the same as the coefficient on ln*SCL*, 0.020, from the production function estimates of table 3.2. Moreover, this estimate of (i) remains the same for all regressions of productivity distributional measures in both cases: all and exporting plants. For (ii), we compute the elasticity of productivity with respect to *LOC*, *HHI* and *SCL* (and the control variables) following the second term in the right-hand side of equation (17), i.e., marginal effect divided by the expected value of the productivity indicator. The latter results are presented in the column labeled “Elasticity” in table 3.5. Note that most elasticities from the self-selection component are larger (in absolute value) than 0.020, the elasticity with respect to agglomeration economies. Hence, the effects of self-selection outweigh that of the agglomeration economies. Note also that every elasticity is larger in the case of exporting plants than in the case of all plants, which is due to high-productivity of exporters relative to non-exporters both in the literature and in our sample (Aitken et al., 1997; Bernard and Jensen, 1999).

We also compute the change in regional productivity due to a one-standard deviation increase in *LOC*, *HHI* and *SCL* from their mean. The column titled “change in productivity” shows that market size (*SCL*) increases regional productivity level the most followed by food industry agglomeration (*LOC*) for all plants. The one-standard deviation increase in the presence of other exporters (*EXP*) also brings about a relatively large increase in regional productivity, when considering only exporting plants.

### 3.5 Summary and Conclusions

In this study, we tested whether high-productivity (exporting) plants self-select into large markets. In addition, we assessed the relative contribution of plants’ self-selection and agglomeration economies to a region’s productivity level.

For this purpose, we chose plants from the Chilean food industry, which is not only a key source of employment, GDP, and exports but also well dispersed among the Chilean provinces. We first estimated a production function using plant-level data from 1998 to 2003 while correcting for two sources of bias: simultaneity and self-selection problem. Simultaneity arises when plants realizing a large positive productivity shock respond by using more conventional inputs. If productivity is not explicitly included in the estimation of the production function, the estimates suffer from the lack of independence between the disturbance term and conventional inputs. Following prior studies, we used a material-input proxy to model plant productivity in the estimation of the production function without the simultaneity bias. We include three types of externalities in the production function: localization, urbanization and demand-driven scale economies. By modeling productivity as a random, time-, region- and plant-specific parameter, we again avoid possible simultaneity between agglomeration economies and productivity, i.e., the self-selection problem.

The estimates of the production function showed that the food industry is material intensive and exhibited constant returns to scale. Among the three externalities, only the demand-driven scale economies has a significantly positive effect on production. Plant-level productivity by region allowed us to derive measures characterizing regional productivity distribution: median, $10^{th}$ and $90^{th}$ percentile values. Each of these distributional measures is then regressed on the three types of agglomeration economies controlling for regional characteristics. We employ a heteroskedastic-tobit model since some regions did not have any food manufacturing plants. We found that high-productivity (exporting) plants locate in a region where other plants in food industry agglomerate, industrial structure is diversified and market size is large. However, exporters are more sensitive to market characteristics than other plants. Finally, we compared the relative importance of plants' self-selection and agglomeration economies to a region's productivity level. Our results suggest that plants' self-selection outweighs the contribution of externalities in increasing a region's productivity.

Policies to increase regional productivity include specializing in targeted industry, diversifying regional industrial structure, enlarging the market size and encouraging agricultural production. However, costs to implement each policy should be considered. Our results provide information to local governments about the benefits or gains from each policy option, which when compared with costs, aid in the selection of an effective policy to improve regional productivity.

**3.6 Bibliography**

Aitken, B., Hanson, G.H. and A. Harrison. (1997). "Spillovers, Foreign Investment, and Export Behavior." *Journal of International Economics*. 43:103-132.

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

Baldwin, R. and T. Okubo. (2006). "Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting." *Journal of Economic Geography*. 6: 323-346.

Bernard, A. and J. B. Jensen. (1999). "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics.* 47:1-25.

Campbell, J. and H. Hopenhayn. (2005). "Market Size Matters." *Journal of Industrial Economics*. 53:1-25.

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

Eberts, R. and D. McMillen. (1999). "Agglomeration Economies and Urban Public infrastructure." in P. Cheshire and E. S. Mills, eds., *Handbook of Regional and Urban Economics*, Volume 3. Elsevier. Amsterdam.

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

Greene, W. (2003). *Econometric Analysis 5$^{th}$ Edition*. Prentice Hall. Upper Saddle River, NJ.

Greene, W. (2006). "Censored Data and Truncated Distributions." in T. Mills and K. Patterson, eds., *Palgrave Handbook of Econometrics Volume 1; Theoretical Econometrics*. Palgrave Macmillan. London.

Hanson, G.H. (2001). “Scale Economies and the Geographic Concentration of Industry.” *Journal of Economic Geography*. 1:255-276.

Head, K. and T. Mayer. (2004). “Market Potential and the Location of Japanese Investment in the European Union.” *Review of Economics and Statistics*. 86: 959-972.

Henderson, J. V. (1986). “Efficiency of Resource Usage and City Size.” *Journal of Urban Economics*. 19:47-70.

Henderson, J. V. (2003).“Marshall”s Scale Economies.”*Journal of Urban Economics*. 53:1-28.

Henderson, J. V., Z. Shalizi, and A. J. Venables (2001). “Geography and Development.” *Journal of Economic Geography*. 1: 81-105.

Holmes, T. and J. Stevens. (2002). “Geographic Concentration and Establishment Scale.” *Review of Economics and Statistics*. 84: 682-690.

Lall, S., Shalizi, Z. and U. Deichmann. (2004). “Agglomeration Economies and Productivity in Indian Industry.” *Journal of Development Economics*. 73: 643-673.

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

Liu, L. (1991). “Entry-Exit and Productivity Changes: An Empirical Analysis of Efficiency Frontiers.” Ph.D. Thesis, University of Michigan.

Malmberg, A., Malmberg, B. and P. Lundequist. (2000). “Agglomeration and Firm Performance: Economies of Scale, Localisation, and Urbanisation among Swedish Export Firms.” *Environmental and Planning A*. 32: 305-321.

Nakamura, R. (1985). “Agglomeration Economies in Urban Manufacturing

Industries: A Case of Japanese Cities." *Journal of Urban Economics*. 17: 108-124.

Olley, S. and A. Pakes. (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*. 64: 1263-1298.

Pavcnik, N. (2002). "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants." *Review of Economic Studies*. 69: 245-76.

Rice, P., Venables, A. J. and E. Patacchini (2006). "Spatial Determinants of Productivity: Analysis for the Regions of Great Britain." *Regional Science and Urban Economics*. 36: 727-752.

Roberts, M. and J. Tybout. (1997). "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs." *American Economic Review*. 87: 545-564.

Rosenthal, S. and W. Strange. (2004). "Evidence on the Nature and Sources of Agglomeration Economies." in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economics*, volume 4. Elsevier. Amsterdam.

Syverson, C. (2004). "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy*. 112: 1181-1222.

Tybout, J. (2000). "Manufacturing Firms in Developing Countries: How Well They Do and Why?" *Journal of Economic Literature*. 38: 11-44

**3.7 Endnotes**

[1]Localization (urbanization) economies are externalities due to the concentration of firms in the same (any/all) industry. Henderson's (2003) estimation of a plant-level production function for high-tech and machinery industries finds positive localization economies for the former, but none for the latter. See also Nakamura (1985) and Lall, Shalizi, and Deichmann (2004).

[2]See also Campbell and Hopenhayn (2005), who find that retail firms' sales and employment are positively correlated with that of their location (city size).

[3]Alternatively, the marginal product of an input increases with a positive productivity shock, which results in a change in the allocation and optimal level of inputs.

[4]The underlying production function takes the form:
$Y_{irt} = g\left(\omega_{irt}, S_{rt}\right) f\left(L^{s}_{irt}, L^{u}_{irt}, K_{irt}, M_{irt}, E_{irt}\right)$, where $g\left(\omega_{irt}, S_{rt}\right) = e^{\omega_{irt}} S_{rt}^{\beta_1}$.

[5]Estimates of parameters on other input variables such as labor and energy are also biased for the same reason.

[6]Even if $\operatorname{cov}\left[\xi_{irt}, S_{irt}\right] = 0$, note that the estimation of equation would yield biased estimates of $\beta_3$, which, in turn, would bias our estimates of $\omega$ derived later in this section. In that sense, our estimation of equation (1) can handle both the presence and absence of self-selection.

[7]Implicit in the proxy is a sequencing of plants' decision within a time period. Given a capital stock, plants first realize their productivity level and then, choose the level of conventional inputs and the type of externalities they would like to associate with (e.g., locational attributes, *S*).

[8]Levinsohn and Petrin (2003) also employ a third-order polynomial series estimator, but the results are very similar to that from the non-parametric estimation.

[9]For details on bootstrapping procedures, refer to Levinsohn and Petrin (2003).

[10]We test the sensitivity of our results to this cutoff value, i.e., 10 plants per region. When setting the cutoff value set to 3, the number of observation increases while accuracy of productivity (distribution) measures deteriorates. Nevertheless, the re-estimation with the lower cut-off value yields externality effects that are qualitatively and quantitatively similar to those reported in the latter sections. Minor quantitative differences are only observed in the case of control variables.

[11]Since the number of exporting plants is less than 10 for many provinces, the cut-off

value to exclude a region from the self-selection model is set at five.

[12]Although plants in wood industry (ISIC 33) disperse around a country and the number of uncensored observations is not small, we omit this industry from the analysis because location of the industry heavily depends on forest coverage.

[13]Refer to Liu (1991) for details on constructing plant-level variables.

[14]The Herfindahl-Hirschman Index is computed as $\sum_j S_{jrt}^2$ where, $S_{jrt}$ is the employment share of the 2-digit manufacturing industry $j$ in province $r$ at period $t$.

[15]Data on agricultural and fishery employment is available for only 2002 and data on human capital index is available only at the macro-regional level for 1999, 2001 and 2003.

[16]Since we subtract $i$-th (own) plant's employment from numerator and denominator, $LOC_{irt}$ is plant specific.

[17]Since region 11 had fewer number of observations, all of which are censored, we use one regional dummy to represent provinces in regions 11 and 12. The underlying assumption here is that regional characteristics do not differ much for neighboring two regions. In the case of exporting plants, region 3 (9) is combined with region 4 (10). The results do not change much if region 11 is combined with region 10 in the former case and region 3 (9) is combined with region 2(8) in the latter case.

[18]Parameter estimates are presented in columns (1) to (3) of tables C.2 and C.3 in the Appendix C.

[19]We attempted other forms, $\sigma_i^2 = \sigma^2 \exp(\gamma' x_i)$ and $\sigma_i^2 = \sigma^2 \left[1 + (\gamma' x_i)^2\right]$, but the log-likelihood value is lower relative to that of the specification in (21).

[20]Corresponding parameter estimates are presented in columns (4) to (6) of tables C.2 and C.3 in the Appendix C.

[21]We also evaluated $\Phi(\cdot)$ and $\phi(\cdot)$ at the sample median, but they again approached one and zero, respectively.

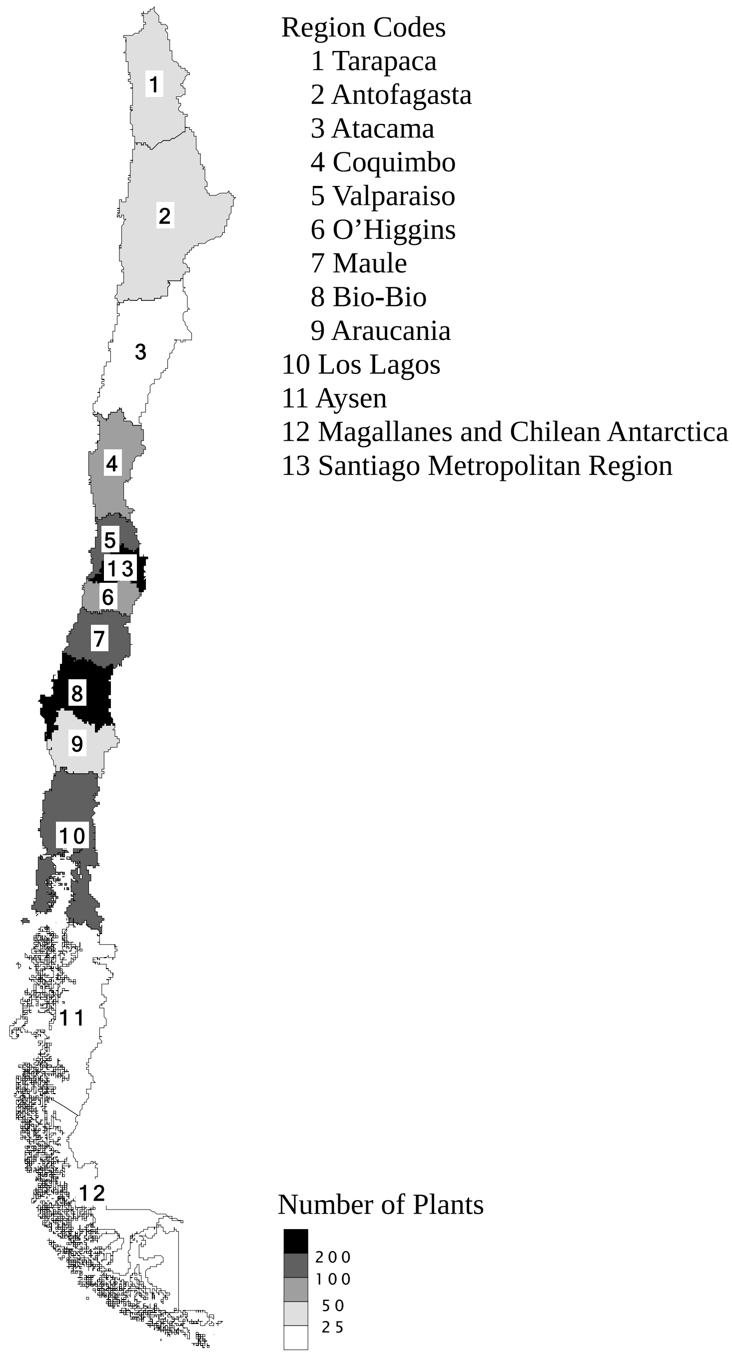



**Figure 3.1: Regional Plant Distribution of Food Industry in Chile**

**Table 3.1: Descriptive Statistics on Chilean Processed Food Industry 1998-2003 Average**

| Variable | Unit | Mean | Std. Deviation |
|---|---|---|---|
| Output (*Y/1,000,000*) | Constant peso | 3.77 | 14.70 |
| Skilled worker ($L^s$) | Number | 21.55 | 56.26 |
| Unskilled worker ($L^u$) | Number | 50.38 | 120.65 |
| Material (*M/1,000,000*) | Constant peso | 2.04 | 6.28 |
| Energy (*E/1,000*) | Constant peso | 78.02 | 338.18 |
| Capital (*K/1,000,000*) | Constant peso | 5.31 | 135.56 |
| Food industry's share (*LOC*) | Index, 0-1 | 0.46 | 0.29 |
| Herfindahl-Hirschman Index (*HHI*) | Index, 0-1 | 0.50 | 0.24 |
| Province population (*SCL/1,000*) | Number | 303.36 | 665.30 |
| Agricultural employment (*AG/1,000*) | Number | 11.42 | 10.92 |
| Human capital index (*HC*) | Index, 0-1 | 0.49 | 0.14 |
| Distance to Santiago (*DIST*) | Kilometer | 768.80 | 787.83 |
| Exporters' share (*EXP*) | Index, 0-1 | 0.52 | 0.25 |

Source: INE, Chilean Manufacturing Census, Various Years.
INE, Housing and Population Census, 2002.
INE, Population Projections
CME, The Regional Competitiveness Report, 2003.

**Table 3.2: Parameter Estimates of the Plant-Level Production Function**

| Variable | Estimate | Variable | Estimate |
|---|---|---|---|
| ln*LOC* | -0.003 | ln*M* | 0.567*** |
| | (0.012) | | (0.102) |
| ln*HHI* | -0.001 | ln*E* | 0.074*** |
| | (0.032) | | (0.010) |
| ln*SCL* | 0.020** | ln*K* | 0.170*** |
| | (0.010) | | (0.049) |
| ln$L^s$ | 0.069*** | Returns to Scale | 0.960 |
| | (0.009) | ($\beta_2+\beta_3+\beta_4+\beta_5+\beta_6$) | (0.064) |
| ln$L^u$ | 0.079*** | | |
| | (0.010) | | |

Note: *, ** and *** indicate statistical significance at the 10%, 5% and 1%, respectively. Value in parenthesis is the bootstrapped standard error based on 200 iterations. Dependent variable is the log of output.

**Table 3.3: Marginal Effects from the Self-Selection Model: All Plants**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Percentile | 10 | 50 | 90 | 10 | 50 | 90 |
| ln*LOC* | 1.910*** | 2.400*** | 5.531*** | 2.193*** | 0.451 | 6.400*** |
| | (0.339) | (0.261) | (0.755) | (0.587) | (0.459) | (0.934) |
| ln*HHI* | -0.442 | -1.302** | -3.554** | -1.228* | -0.615 | -2.319 |
| | (0.579) | (0.567) | (1.617) | (0.650) | (0.606) | (1.710) |
| ln*SCL* | 1.319*** | 1.194*** | 3.455*** | 0.312 | -0.208 | 3.746*** |
| | (0.324) | (0.337) | (0.976) | (0.248) | (0.317) | (1.016) |
| ln*AG* | 1.021*** | 1.681*** | 0.851 | 0.396* | 0.572* | -0.234 |
| | (0.338) | (0.352) | (1.012) | (0.220) | (0.299) | (1.017) |
| ln*HC* | 2.112 | 5.644*** | 3.556 | -2.430 | 2.738* | 8.406 |
| | (1.638) | (1.660) | (4.825) | (1.502) | (1.609) | (5.284) |
| ln*DIST* | 1.035*** | 1.445*** | 3.652*** | 0.040 | 0.270 | 2.973*** |
| | (0.292) | (0.303) | (0.878) | (0.244) | (0.303) | (0.952) |
| Heteroskedasticity | No | No | No | Yes | Yes | Yes |
| OBS. | 173 | 173 | 173 | 173 | 173 | 173 |

Note: *, ** and *** indicate statistical significance at the 10%, 5% and 1%, respectively. Value in parenthesis is the standard error. Dependent variable is the measure of regional productivity distribution and the row labeled ‘Percentile’ indicates which measure is used.

**Table 3.4: Marginal effects from the Self-Selection Model: Exporting Plants**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Percentile | 10 | 50 | 90 | 10 | 50 | 90 |
| ln*LOC* | 5.245*** | 8.123*** | 23.132*** | 3.165** | 9.127*** | 8.845*** |
| | (1.211) | (1.350) | (5.552) | (1.531) | (2.324) | (3.100) |
| ln*HHI* | -2.633 | -1.271 | -23.210*** | -5.235** | -4.468 | 3.933 |
| | (1.928) | (1.990) | (8.184) | (2.353) | (2.825) | (6.317) |
| ln*SCL* | 3.958*** | 7.751*** | 4.541 | 3.379*** | 8.489*** | 10.988*** |
| | (1.030) | (1.130) | (5.645) | (0.759) | (0.733) | (2.841) |
| ln*AG* | -0.415 | -2.116 | 3.786 | 3.133*** | -2.257** | 9.782*** |
| | (1.297) | (1.345) | (6.488) | (0.785) | (1.092) | (3.991) |
| ln*HC* | 3.400 | 12.232** | 9.020 | 3.425 | -5.499 | 6.094 |
| | (5.419) | (5.949) | (27.708) | (5.148) | (5.589) | (14.443) |
| ln*DIST* | 1.810** | 4.240*** | 2.383 | 2.078*** | 4.866*** | 9.221*** |
| | (0.909) | (0.971) | (4.902) | (0.681) | (0.750) | (2.286) |
| ln*EXP* | 4.848*** | 4.861*** | 13.217*** | 8.937*** | 5.901*** | 10.346*** |
| | (1.899) | (0.914) | (5.089) | (2.301) | (1.601) | (1.896) |
| Heteroskedasticity | No | No | No | Yes | Yes | Yes |
| OBS. | 160 | 160 | 160 | 160 | 160 | 160 |

Note: *, ** and *** indicate statistical significance at the 10%, 5% and 1%, respectively. Value in parenthesis is the standard error. Dependent variable is the measure of regional productivity distribution and the row labeled 'Percentile' indicates which measure is used.

**Table 3.5: Agglomeration Economies and Change in Productivity Level due to Increase in Explanatory Variables by One Standard Deviation**

| | All plants | | | | | |
|---|---|---|---|---|---|---|
| Percentile | 10 | | 50 | | 90 | |
| Variable | Elasticity | Change in productivity (%) | Elasticity | Change in productivity (%) | Elasticity | Change in productivity (%) |
| Externalities (SCL) | 0.020* | 4.481 | 0.020* | 4.481 | 0.020* | 4.481 |
| LOC | 0.229* | 14.541 | 0.030 | 1.919 | 0.241* | 15.312 |
| HHI | -0.128* | -6.053 | -0.041 | -1.946 | -0.087 | -4.121 |
| SCL | 0.033 | 7.158 | -0.014 | -3.054 | 0.141* | 30.958 |
| AG | 0.041* | 3.959 | 0.038* | 3.664 | -0.009 | -0.843 |
| HC | -0.254 | -7.206 | 0.183* | 5.208 | 0.317 | 8.991 |
| DIST | 0.004 | 0.426 | 0.018 | 1.852 | 0.112* | 11.478 |

| | Exporting plants | | | | | |
|---|---|---|---|---|---|---|
| Percentile | 10 | | 50 | | 90 | |
| Variable | Elasticity | Change in productivity (%) | Elasticity | Change in productivity (%) | Elasticity | Change in productivity (%) |
| Externalities (SCL) | 0.020* | 4.481 | 0.020* | 4.481 | 0.020* | 4.481 |
| LOC | 0.339* | 21.510 | 0.496* | 31.468 | 0.256* | 16.242 |
| HHI | -0.560* | -26.439 | -0.243 | -11.447 | 0.114 | 5.367 |
| SCL | 0.362* | 79.324 | 0.461* | 101.093 | 0.318* | 69.690 |
| AG | 0.335* | 32.088 | -0.123* | -11.723 | 0.283* | 27.064 |
| HC | 0.367 | 10.409 | -0.299 | -8.477 | 0.176 | 5.003 |
| DIST | 0.222* | 22.799 | 0.264* | 27.075 | 0.267* | 27.327 |
| EXP | 0.957* | 45.310 | 0.320* | 15.175 | 0.299* | 14.170 |

Note: * indicates that corresponding marginal effect is statistically significant at least at the 10% level.

# 4. Human-Capital Spillovers, Absorptive Capacity and Agglomeration of High Skill-Intensive Production in Chilean Manufacturing

## 4.1 Introduction

The accumulation of knowledge, a non-rival and partially non-excludable good, is a key determinant of economic growth of nations (Barro and Sala-i-Martin, 2003; Lucas, 1988; Romer, 1990). Not surprisingly, the empirical literature has found that the emergence of multi-purpose technologies, e.g. computers, and the increased globalization have raised the knowledge- or skill-intensity of production processes (Berman et al., 1994; Bernard and Jensen, 1997; Feenstra and Hanson, 1997). However, the growth and trade literature often assume that the resulting economic gains apply uniformly across regions of a country. In fact, wide disparities exist in the level and types of economic activity among regions within a country (Henderson et al., 2001). In light of endogenous growth models, the spatial concentration of high skill-intensive production, e.g. Silicon Valley in the United States or Bangalore in India, and the consequences, especially the wage enhancement effects, have received substantial empirical attention (Acemoglu and Angrist, 2000; Glaeser and Mare, 2001; Moretti, 2004b; Rauch, 1993).

In this study, we focus on factors causing spatial concentration or agglomeration of high skill-intensive production in the context of a developing economy. In particular, we focus on the partial non-excludability of knowledge, i.e., human-capital spillovers in the production process, as an agglomerating force of high skill-intensive production. Evidence of human-capital spillovers as an important source of economic growth has been found at the national and industry levels, but the majority of studies rely on a wage hedonic equation to examine the existence of spillovers (Moretti, 2004c). An exception is Moretti (2004a), whose production-function estimates show that U.S. manufacturing plants' productivity increases by 0.6-0.7 percent as the share of college graduates in a region increases by one percentage point. Nevertheless, issues such as the mechanism by which spillovers arise and benefit plants, and whether or not every plant can equally benefit from spillovers have received limited attention in the literature. Furthermore, the existence or expectations of such spillovers may impact plants' location and exit choices, which can explain concentration of high skill-intensive plants in particular regions.

We draw on the absorptive capacity hypothesis of Cohen and Levinthal (1989) to identify the mechanism by which human-capital spillovers improve plants' productivity. The absorptive capacity hypothesis argues that firms must invest in research and development in order to access new knowledge created by either other firms or the public sector. Firm-level evidence on the absorptive capacity hypothesis can be found in Arora and Gambardella (1994) and Cockburn and Henderson (1998). We extend this idea to the case of skilled workers, meaning such workers' presence is a key channel by which knowledge is transmitted across firms or plants.

In addition to the testing the absorptive capacity hypothesis, we explore the role of spillovers in plants' location choice and exit. Audretsch and Feldman (1996) show that high skill-intensive industries are likely to agglomerate because they tend to benefit more from spillovers than low skill-intensive industries. Similarly, high skill-intensive plants are likely to be attracted to a high skill-intensive region because plants with low skill-intensity, i.e. low absorptive capacity, may not benefit from spillovers. As high skill-intensive plants are attracted to the concerned region, regional skill intensity will also rise, which further increases plants' productivity leading to the agglomeration of high skill-intensive plants (Carlton, 1983; Crozet et al., 2004; Head et al., 1999). Complementing the location-choice analysis is an examination of plant exit as in Dunne et al. (1989) and Olley and Pakes (1996). For instance, Bernard and Jensen (2007) find that the higher a plant's productivity, the lower the probability of exit. To the extent that spillovers augment productivity, they also lower plant death. Therefore, spillovers improve the productivity and lower the exit probability of high skill-intensive incumbents, and attract high skill-intensive entrants to a location, which together can explain the agglomeration of high skill-intensive production.

Chile provides an interesting case study of agglomeration of high skill-intensive production. Figure 4.1 shows average of plants' skill intensity, i.e., share of skilled workers in total employment, by province from 1998 to 2003. It is clear that relative demand for skilled workers is not uniform across provinces and high skill-intensive plants concentrate around northern and central provinces, suggesting that benefits from human-capital spillovers are likely to be large in those provinces. Plants' location choice and exit should additionally explain this uneven distribution of high skill-intensive production because of substantial entry and exit in Chile.[1] Figure 4.2 shows the kernel density estimates for entrants' skill intensity in provinces with above and below median provincial skill intensity. It is apparent that high (low) skill-intensive entrants are more likely to locate in provinces with above (below) median provincial skill intensity. Alternatively, high skill-intensive provinces appear to attract high skill-intensive plants. Figure 4.3 shows the kernel density estimates for exiting and all plants' skill intensity in provinces with above median provincial skill intensity. Although, we do not observe large density differences between exiting and all plants, it can be expected that high skill-intensive plants are less likely to exit due to gains from spillovers in high skill-intensive provinces.

In the following, we first detail the estimation of plant-level production functions with and without the absorptive capacity hypothesis. Then, we evaluate the role of productivity, especially that part arising from spillovers, in plants' location choice and exit. Finally, we numerically evaluate the above three effects of spillovers – productivity, location choice and exit – to illustrate the agglomeration of high skill-intensive production in particular regions.

### 4.2 Plant-Level Production Function with Human-Capital Spillovers

To test the existence of human-capital spillovers and identify the channel through which spillovers affect plant-level productivity, we consider the following Cobb-Douglas production function:

(1) $$\ln Y_{irt} = \beta_0 + \beta_1 \ln L^s_{irt} + \beta_2 \ln L^u_{irt} + \beta_3 \ln M_{irt} + \beta_4 \ln E_{irt} + \beta_5 \ln K_{irt} + \ln A_{irt} + \omega_{irt} + \varepsilon_{irt},$$

where $Y_{irt}$, $L^s_{irt}$, $L^u_{irt}$, $M_{irt}$, $E_{irt}$ and $K_{irt}$ represent output, skilled worker, unskilled worker, material, energy and capital of plant *i* in the *r*-th region at period *t*. The term $\omega_{irt}$ denotes plant-, time- and location-specific productivity shock and $\varepsilon_{irt}$ is an *i.i.d.* disturbance term (Levinsohn and Petrin, 2003). Following the absorptive capacity hypothesis, we measure plant-specific regional productivity-enhancement effects, $A_{irt}$, by:[2]

(2) $$\ln A_{irt} = \beta_6 \ln G_{irt} + \beta_7 \eta_{irt-1} S_{irt},$$

where $G_{irt}$ and $S_{irt}$ respectively measure agglomeration economies and knowledge resources accessible to a plant, both of which are defined in the following. The latter, $S_{irt}$, is scaled by our measure of the absorptive capacity, $\eta_{irt-1}$, which is the one-period lagged share of skilled workers to total workers of plant *i* in the *r*-th region. Our specification of absorptive capacity is consistent with the idea that plants must develop the ability to utilize knowledge over time in order to benefit from spillovers (Cohen and Levinthal, 1989). If $\beta_7$ is positive and significant, the absorptive-capacity hypothesis is supported and high skill-intensive plants gain more from spillovers than others.

The new economic-geography literature has well documented the effect of agglomeration economies, $G_{irt}$, i.e., plant's productivity increases by locating near other plants arising from ease of input- and output-market access (Baldwin et al., 2003; Fujita et al., 1999). Therefore, $\beta_6$ is expected to take a positive sign, but may take a negative sign in the case that congestion costs exceed the effect of agglomeration economies. Following Crozet et al. (2004), we define:

$$(3) \qquad G_{irt} = \left(N_{irt} - 1\right) + \sum_{j \neq r} \left( \frac{N_{ijt}}{d_{rj}} \right),$$

where $N_{irt}$ is the number of plants in the *i*-th plant's industry belonging to region *r* at time *t*. We subtract one from $N_{irt}$ so as to not include the own plant's effect. Economic interaction among plants may not be complete within a region since it is an administrative geographical unit, e.g. state or province. So, in equation (3) we include $N_{ijt}$, the number of firms in the *i*-th plant's industry in region $j \neq r$ weighted by $d_{rj}$, the great-circle distance between region *r* (base) and *j* in kilometers. Note that the second term on the right hand side of equation (3) allows agglomeration effects to decay when the distance between base and other regions increases (Lall et al., 2004).

We refer to $S_{irt}$, in equation (2), as regional skill intensity, which is measured by the share of skilled workers in total workforce (Moretti, 2004a):

$$(4) \qquad S_{irt} = \frac{\left(L_{rt}^{s} - L_{irt}^{s}\right) + \sum_{j \neq r} \left(L_{jt}^{s} / d_{rj}\right)}{\left(L_{rt}^{s} - L_{irt}^{s}\right) + \sum_{j \neq r} \left(L_{jt}^{s} / d_{rj}\right) + \left(L_{rt}^{u} - L_{irt}^{u}\right) + \sum_{j \neq r} \left(L_{jt}^{u} / d_{rj}\right)},$$

where $L_{rt}^{s}$ and $L_{rt}^{u}$ are respectively the number of skilled and unskilled workers in the *r*-th region at time *t*. Again, we subtract own plant's skilled (skilled and unskilled) workers when computing the numerator (denominator). The terms $L_{jt}^{s}$

and $L_{jt}^{u}$ denote the number of skilled and unskilled workers in region $j \neq r$. The effect of skilled workers in any other region ( $j \neq r$ ) on plants in the *r*-th region decays when distance between the two regions increases (Bottazzi and Peri, 2003; Jaffe et al., 1993). Note $S_{irt}$ approaches one (zero) as more skilled (unskilled) workers are employed in the surrounding regions.

By substituting (2) into (1):

$$\ln Y_{irt} = \beta_0 + \beta_1 \ln L_{irt}^{s} + \beta_2 \ln L_{irt}^{u} + \beta_3 \ln M_{irt} + \beta_4 \ln E_{irt} + \beta_5 \ln K_{irt} + \beta_6 \ln G_{irt} + \beta_7 \eta_{irt-1} S_{irt} + \omega_{irt} + \varepsilon_{irt}. \tag{5}$$

A plant's productivity shock is not observable to us, but it may contain two components that pose problems in the consistent estimation of the parameters of equation (5) (Levinsohn and Petrin, 2003; Olley and Pakes, 1996). Suppose $\omega_{irt}$ consists of a plant's productivity shock, $\chi_{it}$, and unobserved regional shocks such as infrastructure and climate, $\tau_{rt}$, both of which contribute to changes in a plant's productivity. The former, $\chi_{it}$, is likely correlated with conventional input use. For example, a plant realizing a positive productivity shock may respond by using more workers or material (intermediate) inputs to increase the output, i.e., $E\left(\chi_{it}, L_{irt}^{u}\right) \neq 0$. Additionally, plants might be attracted to a region with positive unobserved regional shocks, which leads to an endogeneity problem between unobserved regional characteristics, $\tau_{rt}$, and observed regional characteristics, $G_{irt}$ and $S_{irt}$ (Rosenthal and Strange, 2004). Thus, ordinary least squares (OLS) estimation of equation (5) where the productivity shock, $\omega_{irt}$, is added to the *i.i.d.* disturbance, $\varepsilon_{irt}$, would result in inconsistent parameter estimates of the production function.

To correct these two problems, we employ the Levinsohn and Petrin's (2003) approach. Given plants' predetermined capital stock and absorptive capacity, plants first realize their productivity level and then, choose the level of conventional inputs.[3] For this purpose, we assume that material input use is a monotonic function of productivity. Then, we can invert the latter function and express the productivity shock as a function of material input, capital stock and plant's skill intensity in the last period:

(6) $$\omega_{irt} = \omega\left(\ln M_{irt}, \ln K_{irt}, \eta_{irt-1}\right).$$

By substituting (6) into (5):

(7) $$\begin{aligned} \ln Y_{irt} = \beta_1 \ln L^s_{irt} + \beta_2 \ln L^u_{irt} + \beta_4 \ln E_{irt} + \beta_6 \ln G_{irt} + \beta_7 \eta_{irt-1} S_{irt} \\ + \varphi\left(\ln M_{irt}, \ln K_{irt}, \eta_{irt-1}\right) + \varepsilon_{irt}, \end{aligned}$$

where:

(8) $$\varphi\left(\ln M_{irt}, \ln K_{irt}, \eta_{irt-1}\right) = \beta_0 + \beta_3 \ln M_{irt} + \beta_5 \ln K_{irt} + \omega\left(\ln M_{irt}, \ln K_{irt}, \eta_{irt-1}\right).$$

Since $\varepsilon_{irt}$ is assumed to be independent of $L^s$, $L^u$, *E*, *G* and *S*, we can consistently estimate $\beta_1$, $\beta_2$, $\beta_4$, $\beta_6$ and $\beta_7$ by approximating $\varphi(\cdot)$ with a third-order polynomial and using OLS.

Following Levinsohn and Petrin (2003), we proceed to a second stage, where the productivity shock follows a first-order Markov process, to identify $\beta_3$ and $\beta_5$.

That is, production function, equation (5), at period $t+1$ is written as:

(9)
$$\ln Y_{irt+1} = \beta_1 \ln L^s_{irt+1} + \beta_2 \ln L^u_{irt+1} + \beta_3 \ln M_{irt+1} + \beta_4 \ln E_{irt+1} + \beta_5 \ln K_{irt+1} + \beta_6 \ln G_{irt+1} + \beta_7 \eta_{irt} S_{irt+1} + g\left(\omega_{irt}\right) + \varepsilon_{irt+1} + \xi_{irt+1}.$$

where, $\xi_{irt+1}$ is an innovation. From equation (8), we can express plant's productivity as follows:

(10)
$$\beta_0 + \omega_{irt} = \hat{\varphi}\left(\cdot\right) - \beta_3 \ln M_{irt} - \beta_5 \ln K_{irt}.$$

Given $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_4$, $\hat{\beta}_6$, $\hat{\beta}_7$ and $\omega_{irt}$ from the equation (7), we can consistently estimate $\beta_3$ and $\beta_5$ by approximating $g\left(\cdot\right)$ with a third-order polynomial and using nonlinear instrumental variables. Instrumental variables estimator is preferred because ($\varepsilon_{irt+1} + \xi_{irt+1}$) is correlated with $\ln M_{irt+1}$. We use $\hat{\varphi}\left(\ln M_{irt}, \ln K_{irt}\right)$, $\ln K_{irt+1}$, $\ln K_{irt}$, $\ln M_{irt}$ and $\ln E_{irt}$ as instruments. Given the multiple steps in the estimation, the standard error of parameter estimates is obtained using bootstrapping procedures.[4]

### 4.3 Plants' Location Choice and Exit

Having illustrated a technological framework to capture human-capital spillovers in a plant's productivity, we now turn our attention to the role of spillovers in plants' location choice and exit. Plants' entry and exit are often analyzed as a dynamic process as in Jovanovic (1982), Hopenhayn (1992) and, more recently, in Melitz (2003). For instance, a potential entrant in period *t* makes the entry decision by comparing expected discounted profits and sunk entry costs, where the latter serves as the initial investment to set up a plant. After incurring sunk costs, each plant realizes the period *t* productivity, which steers the plant to a region where the expected discounted profit is the highest among alternative regions (1,…, *R*). Let $v_t(X_{it}, Z_{rt})$ be a value function, i.e., expected discounted profits, of entrant *i* in period *t*, where, $X_{it}$ and $Z_{rt}$ are vectors of plant and regional characteristics, respectively. Then, the *i*-th plant's location decision is expressed as:

(11) $$r = \arg\max_r \{v_t(X_{it}, Z_{rt})\}.$$

In analyzing the location of Japanese investment in the European Union, Head and Mayer (2004) derive a plant's profit, analogous to the value function in equation (11), in the *r*-th region as:

(12) $$\pi(X_{it}, Z_{rt}) = \pi(prod_{irt}, MP_{rt}, w_{rt}),$$

where $Z_{rt} = \{w_{rt}, MP_{rt}\}$ respectively represents factor prices and market potential in the region *r*, which quantifies the demand for goods produced by plants in region *r*. Plant *i*'s productivity in period *t*, $prod_{irt}$, is the primary plant-specific characteristic in the profit function. Head and Mayer (2004) show that $\partial\pi/\partial prod_{irt} > 0$, $\partial\pi/\partial MP_{rt} > 0$ and $\partial\pi/\partial w_{rt} < 0$. That is, profit is higher in a region where productivity-enhancement effects are large, the demand for goods is large and factor prices are low, respectively. Drawing on Head and Mayer (2004), we assume that the value function takes the following linear specification of the profit function (12):[5]

(13) $$v_t(X_{it}, Z_{rt}) = \alpha_1 \ln prod_{irt} + \alpha_2 \ln MP_{rt} + \alpha_3 \ln w_{rt} + \mu_{irt},$$

where $\mu_{irt}$, the unobserved advantage of region *r* for the *i*-th plant, follows a type I extreme value distribution. From equation (11), the probability that plant *i* chooses the region *r* is (McFadden, 1984):

(14) $$\text{Prob}_i\left(r\right)=\frac{\exp\left(\alpha_1 \ln prod_{irt}+\alpha_2 \ln MP_{rt}+\alpha_3 \ln w_{rt}\right)}{\sum_{j=1}^{R}\exp\left(\alpha_1 \ln prod_{ijt}+\alpha_2 \ln MP_{jt}+\alpha_3 \ln w_{jt}\right)}.$$

Given the probability, parameters of equation (14) can be estimated by maximizing the following log likelihood:

(15) $$\ln L=\sum_{i=1}^{n}\sum_{j=1}^{R}\delta_{ij}\ln \text{Prob}_i\left(j\right),$$

where, $\delta_{ij}=1$ if plant *i* locates in the region *j* and $\delta_{ij}=0$ otherwise.

When an entrant begins production, it becomes an incumbent and decides, prior to observing the next period's productivity shock, whether to continue or stop production. Employing the value function in equation (13), the Bellman equation can be written as:

(16) $$v_t\left(X_{it},Z_{rt}\right)=\pi\left(X_{it},Z_{rt}\right)+\max\left\{F,\rho E\left[v_{t+1}\left(X_{it+1},Z_{rt+1}\right)\mid I_t\right]\right\},$$

where, *F* is the opportunity cost of continuing the operation, $\rho$ is the discount factor and $I_t$ is the information set at period *t*. Let $D_{irt}$ be a dummy variable that takes value one when plant *i* exits from the market in the following period. Then, equation (16) implies the following:[6]

(17) $$D_{irt}=\begin{cases}0 & if \quad F\le\rho E\left[v\left(X_{it+1},Z_{rt+1}\right)\mid I_t\right]\\ 1 & if \quad F>\rho E\left[v\left(X_{it+1},Z_{rt+1}\right)\mid I_t\right].\end{cases}$$

Our initial specification of the $t+1$ period's expected value of the plant is obtained by augmenting equation (13):

(18) $$E\left[v_{t+1}\left(X_{it+1}, Z_{rt+1}\right) | I_t\right] = \gamma_1 \ln prod_{irt} + \gamma_2 X_{it} + \gamma_3 \ln MP_{rt} + \gamma_4 \ln w_{rt} + \kappa_{irt},$$

where $\kappa_{irt}$, the unobserved shocks to the expected value function, follows a logistic distribution and $X_{it}$ includes additional plant-specific characteristics. Then, the parameters of equation (18) can be estimated via a logit model.

### 4.4 Chilean Data

We use data from the *Chilean Manufacturing Census* from 1998 to 2003 collected by National Statistics Institute (INE). Plants with at least 10 employees are surveyed in the Census, with information on plant location at three administrative levels: macro-region, province and county. We avoid using the macro-region and county as the geographical unit since the former is too aggregated (13 macro-regions) and the latter has too many counties without manufacturing plants. Hence, we use each of the 51 provinces as a geographical unit (region) to assure enough variation in data. Summary statistics on each of the variables used in this study are presented in table 4.1.

Data for Production Function: Six variables are constructed to estimate the plant-level production function: skilled worker, unskilled worker, materials, energy, capital and output.[7] The unit of worker is the number of workers, while other variables are in Chilean Peso. We use deflators (base year 1996) from the *Chilean National Accounts* to convert the latter variables into constant Chilean Peso. As noted earlier, the number of plants in the same industry in a province is used to proxy for agglomeration economies, $G_{irt}$. Figure 4.1 shows that several adjacent provinces tend to have similar skill intensities, and hence, the potential for neighboring province's economic activity to affect a given province's plants. Equation (3) accounts for the latter effect by including the number of plants in other provinces weighted by inter-provincial distance. As noted in equation (4), provincial skill intensity or stock of knowledge resources available to a plant, $S_{irt}$, is measured by the share of skilled worker in total workforce as in Moretti (2004a). Skilled worker includes executives, administrators and white-collar production workers.

Data for Location Decision: Plants which show up in data in year *t* and do not appear between 1998 and $t-1$ are defined as entrants. Since only plants with 10 or more employees are reported in data, small plants which employed less than 10 workers in previous years may reappear in the following periods because they now employ more than 10 workers. To minimize such selection bias, we created an additional sample where observations of plants with less than 20 employees are excluded. The conditional logit model for location choice, equation (14), is estimated for both these samples.

Provincial wages are obtained by dividing the total wage bill of all plants in a province by their total number of workers.[8] Market potential is the distance-weighted average of provincial population:

(19) $$MP_{ir} = Pop_{ir} + \sum_{j \neq r} \left( \frac{Pop_{ij}}{d_{rj}} \right),$$

where, $Pop_{ij}$ denotes population in the province *j* from *Population Projections* (Head and Mayer, 2004). A number of controls are included in equation (13) to minimize specification bias. For instance, higher wages may reflect workers' higher average educational attainment in a province, e.g. a greater share of college and university graduates in provincial population (Head et al., 1999). To account for the latter in equation (13), we include the percent of workers who have completed a university degree (17 or more years of education), which is taken from the *2002 Housing and Population Census*. Also, we use the provincial population density as a proxy for rents as in Guimaraes et al. (2000).

<u>Data for Exit Decision</u>: Plants which report data until year *t* but do not show up after $t+1$ are defined as exits. Again, observations of plants with less than 20 employees are excluded from our sample. The logit model of exit is estimated with and without the latter criterion.

A plant's size, age, productivity and capital-labor ratio are used as explanatory variables in the logit model for the exit decision. Size is measured by plant's total employment. Age takes value one when the plant first appears in the database and increases by one as it continues production in every period. Large capital stocks imply high future returns and therefore, the plant continues the operation even if plant's current productivity is low (Olley and Pakes, 1996). We also include two dummy variables in the logit model: the first one takes value one if the plant receives foreign direct investment and zero otherwise; the second dummy takes value one if the plant engages in export and zero otherwise. The inclusion of the two dummy variables in the logit model of exit, equation (18), is motivated by previous research, e.g. Bernard and Jensen (2007), where a positive relationship is found between plant's productivity (size) and the probability of survival. Exporting plants are less likely to exit relative to non-exporters and the decision to shutdown plants receiving foreign direct investment is made by a multinational firm based on the global strategy rather than an individual plant's characteristics (Bernard and Jensen, 2007; Melitz, 2003).

### 4.5 Econometric Results

In this section, we first discuss the estimates of the plant-level production with and without the assumption of absorptive capacity. After identifying the appropriate specification of the production function based on a statistical test, we report results from the plants' location choice and exit models. Finally, we combine the above results to discuss the provincial concentration of high skill-intensive production.

Production Function Estimates: Table 4.2 shows the parameter estimates of the production function in equation (7) for each of the 8 Chilean manufacturing industries.[9] Most parameters on conventional inputs –skilled and unskilled workers, materials, energy and capital- are positive and statistically significant. Some coefficients on energy (ISIC 35 and 37) and capital (ISIC 33, 34, 37 and 38) are not significant. Similar to plant-level studies of other countries, e.g. Henderson (2003), Chilean manufacturing plants show high intensity of material inputs in production.

Agglomeration economies, the coefficients of $\ln G$ in table 4.2, are found to be positive and statistically significant for food, textile, and machinery industries (ISIC 31, 32 and 38, respectively). That is, plants' productivity in these industries improves by locating near other plants in the same industry, i.e. localization economies (Rosenthal and Strange, 2004). Our specification of absorptive capacity appears to well fit the data since 5 of 8 parameters on $\eta S$ are positive and statistically significant in table 4.2. This result is consistent with the argument that high skill-intensity is necessary for a plant to utilize new knowledge created by other plants in the industry presumably through the interaction among skilled workers, i.e. absorptive capacity (Cockburn and Henderson, 2003; Cohen and Levinthal, 1989). Alternatively, we find evidence that the presence of high-skilled workers serves as an important channel through which knowledge is transmitted across plants. Multiplying these parameters by plants' average skill intensity, we find marginal effects of spillover in the range of 0.18 and 0.74, whose mid-point is similar to that in Moretti (2004a). Unlike the latter study**,** our results suggest that plants with high skill-intensity, regardless of industry classifications as high- or low-tech, benefit from human-capital spillovers.

Appendix D describes an alternative specification of the production function, equation (7), where every plant, regardless of the skill intensity, benefits from spillovers. Our specification (*J*) test shows that the data better fit equation (7) and the results reported in table 4.2 are statistically preferred over those in the table D.1 in the Appendix D.

Plants' Location Choice Model: For the conditional logit model of plants' location choice, we augment the linear specification of the value function in equation (13) as:

$$v_t\left(X_{it}, Z_{rt}\right) = \alpha_{11}\ln AG_{irt} + \alpha_{12}\ln AS_{irt} + \alpha_2 \ln MP_{rt} + \alpha_{31}\ln w_{rt} + \alpha_{32}\ln POPDEN_{rt} + \alpha_{33}UNIV_r + d_r + \mu_{irt}, \quad (20)$$

where *POPDEN*, *UNIV* and $d_r$ denote population density, share of university graduates in provincial population and macro-region dummy, respectively. The index *t* here represents the year when the plant entered the market. Correlation between error terms, which is a violation of the independence from irrelevant alternatives assumption (IIA), will bias the estimates of equation (20). To minimize such correlation arising from unobserved regional characteristics, we include macro-region dummies (13) in equation (20) as in Head et al. (1999). Finally, $\ln prod$ in equation (13) is replaced by $\ln AG$ and $\ln AS$, which are defined using the estimates of the production function in table 4.2 as follows:

$$\ln AG_{irt} = \hat{\beta}_6 \ln G_{irt}, \quad (21)$$

$$\ln AS_{irt} = \hat{\beta}_7 \eta_{irt} S_{irt}. \quad (22)$$

Equation (21) measures a plant's productivity gain from agglomeration economies, while the effect of human-capital spillovers in equation (22) on plants' location choice requires additional explanation. Since entrants do not engage in production at period $t-1$, they cannot benefit from human-capital spillovers at period *t* due to the lack of absorptive capacity. However, entrants with a large initial investment will employ high skill-intensive technology because capital has been found to be a complement to skilled workers (Pavcnik, 2003). Hence, it is likely that such entrants have expectations, as in equation (16), of potential gains from human-capital spillovers in period $t+1$ and beyond, especially in high skill-intensive provinces. Thus, equation (22) measures the expected productivity gain from human-capital spillovers given current information on entrant's and provincial skill intensity. Since $\ln AS$ embodies expectations, the productivity coefficient in equation (13), i.e., $\alpha_1$, is allowed to vary for $\ln AG$ and $\ln AS$ in equation (20).[10]

The estimates of the conditional logit model of plants' location choice are presented for the two alternative samples in table 4.3. Column (1) reports results with the cut-off of plants employing 20 or more workers. Here, we find that the parameters on market potential and the two sources of productivity gain are positive and significant. These results are consistent with those in Head and Mayer (2004) and Crozet et al. (2004), who find positive and significant effects of market potential and agglomeration economies on firms' location choice. Our estimates suggest that plants tend to locate where demand for their goods and the possibility of productivity improvement, via agglomeration economies and spillovers, are large. Note that the parameter on $\ln AG$ is several times larger than that on $\ln AS$. The latter result may arise if plants regarded actual productivity gain ($\ln AG$) more important than expected productivity gains ($\ln AS$). With regard to control variables, the parameter on the share of university graduates alone is positive and significant suggesting that plants prefer a province where educated workers are relatively abundant. The coefficients on *POPDEN* and *w* are not statistically significant in column (1) of table 4.3.

To test the sensitivity of results, we expand our sample to include all entrants recorded during the sample period, including those that may have had less than 10 employees in the previous period. Then, we estimate the conditional logit model of plants' location choice for this larger sample. The estimates are shown in column (2) of table 4.3. With the exception of the parameter on population density, results in column (2) are quantitatively and qualitatively similar to those reported in column (1). Hence, large markets, productivity-improvement via agglomeration economies and spillovers, and abundance of human-capital are important factors attracting plants to a location.

Logit Model of Plant Exit: To investigate plant exit, we expand on equation (18), the expected value of plant *i* in period $t+1$, as follows:

$$\begin{aligned} E\left[v_{t+1}\left(X_{it+1}, Z_{rt+1}\right) \mid I_t\right] &= \gamma_0 + \gamma_1 \ln prod_{irt} + \gamma_{21} age_{irt} + \gamma_{22} \ln size_{irt} \\ &+ \gamma_{23} \ln K_{irt} / L_{irt} + \gamma_{24} EXP_{irt} + \gamma_{25} FDI_{irt} + \gamma_3 \ln MP_{rt} + \gamma_{41} \ln w_{rt} \\ &+ \gamma_{42} \ln POPDEN_{rt} + \gamma_{43} UNIV_r + d_t + d_r + d_i + \kappa_{irt}, \end{aligned} \tag{23}$$

where, $X_{it} = \left\{age_{irt}, size_{irt}, K_{irt} / L_{irt}, EXP_{irt}, FDI_{irt}\right\}$ are plant-specific characteristics – plant's age, size, capital-labor ratio, exporter and foreign-direct-investment dummy – in addition to the productivity. Additional controls include $d_t$ and $d_i$, which denote dummy variables for year and industry, respectively. Finally, we measure *prod* as a residual from the estimates of the production function in table 4.2:

$$\ln prod_{irt} = \ln Y_{irt} - \hat{\beta}_1 \ln L^s_{irt} - \hat{\beta}_2 \ln L^u_{irt} - \hat{\beta}_3 \ln M_{irt} - \hat{\beta}_4 \ln E_{irt} - \hat{\beta}_5 \ln K_{irt}. \tag{24}$$

The estimates of the logit model of plant exit are presented in table 4.4. Again, we report two sets of estimates: one each for the sample with and without the cut-off of plants employing 20 or more workers, columns (1) and (2), respectively.

The variables for which a significant and negative coefficient is obtained include productivity, age, capital-labor ratio and population density. For only one variable, wages, we obtain a positive and significant coefficient. All other parameter estimates of the logit model of plant exit are not statistically significant. Our results suggest that high-productivity plants are less likely to exit relative to others, a result consistent with Olley and Pakes (1996) and others. Similarly, plants that are older, have a higher capital-labor ratio and engage in export activity are less likely to exit (Bernard and Jensen, 2007). The parameter on population density implies a lower probability of exit in densely populated provinces, which may be due to urbanization economies - positive externalities from city size (Guimaraes et al., 2000). Higher factor prices, e.g. wage, increase the likelihood of plant exit, all else constant.[11]

Applying the logit model to the bigger sample (without the 20 or more employees cut-off), we obtain similar results, which are reported in column (2) of table 4.4. The qualitative similarity of the results in columns (1) and (2) is striking, but some quantitative differences exist, e.g. age, capital-labor ratio. Other than magnitudinal differences between the two specifications, the parameter on plant's size becomes negative and significant in column (2), which implies that larger plants are less likely to exit.[12]

### 4.6 Effect of Regional Skill Intensity on Plants' Productivity, Location Choice and Exit

Our results from the previous section have identified that high skill-intensive plants can improve their productivity by absorbing knowledge from similar plants in a province. Moreover, plants are attracted to high skill-intensive provinces so as to benefit from human-capital spillovers. At the same time, lower productivity increases the probability of plant exit. The combination of these results suggests that a province can accumulate high skill-intensive production by attracting high skill-intensive entrants, lowering the probability of exit for high skill-intensive plants and improving the productivity of high-skill incumbents. In the following, we examine this accumulation process by deriving the effect of provincial skill intensity, $S_{irt}$, on plants' productivity, location choice and exit.

In table 4.5, first consider the third column labeled 'Productivity' which presents the elasticity of plant's productivity with respect to provincial skill intensity for each Chilean manufacturing industry. Since the effect of provincial skill intensity varies according to a plant's skill intensity, elasticities are evaluated at three different levels, i.e., mean skill intensity, mean minus one standard deviation (low) and mean plus one standard deviation (high). Note that the obtained productivity elasticities show substantial variation within and across industries. Plants in chemical industries (ISIC 35) have larger productivity elasticities relative to other industries. The textile industry (ISIC 32) appears to have the lowest productivity elasticities with respect to provincial skill intensity. In the aggregate (last three rows of table 4.5), doubling a region's skill intensity increases average productivity of low, medium, and high skill-intensive plants by 2.8, 11.2 and 20 percent, respectively. While low skill-intensive plants gain less from spillovers, additional benefits to high skill-intensive plants are likely reflected in their entry and exit decisions.

From the conditional logit model of location choice, the elasticity of the probability of choosing province *r* with respect to provincial productivity enhancement effects is obtained as:

$$(25)\qquad \frac{\partial \ln \mathrm{Prob}_i\left(r\right)}{\partial \ln AS_{irt}} = \left(1 - \mathrm{Prob}_i\left(r\right)\right)\alpha_{12}.$$

Since there are 51 alternatives, province *r* is chosen with a probability of 2 percent on average. Using the productivity elasticity with respect to provincial skill intensity obtained above, we can compute the effect of doubling provincial skill intensity on the probability of choosing province *r*. The results are presented in the fourth column of table 4.5 labeled ‘Entry’. Again, substantial differences in location probabilities are obtained across and within industries. On average, doubling a province’s skill intensity increases the probability that it is chosen by high skill-intensive plants (for locational purposes) by 34.9 percent. In contrast, the probability that the concerned province will be chosen by low skill-intensive plants increases by only 4.9 percent. Thus, high skill-intensive plants prefer high skill-intensive provinces, but low skill-intensive plants do not show locational preferences in terms of regional skill intensity.

Similarly, the elasticity of the probability of exit with respect to a plant’s productivity is obtained as:

$$(26)\qquad \frac{\partial \ln \mathrm{Prob}\left(D_{irt} = 1\right)}{\partial \ln prod_{irt}} = \left(1 - \mathrm{Prob}\left(D_{irt} = 1\right)\right)\gamma_1.$$

The average probability of exit is 8.4 percent.[13] Given the estimate of $\gamma_1$ from table 4.4 and the productivity elasticity with respect to provincial skill intensity, we derive the effect of doubling provincial skill intensity on the probability of plant exit. The results are shown in the fifth column of table 4.5 labeled 'Exit'. Note that the exit probabilities (elasticities) are smaller than those of entry in absolute value. The chemical industry (ISIC 35) appears to have lower probabilities of exit relative to other industries. Focusing again on the average (last three rows of table 4.5), the probability of exit falls by 3.2 percent for high skill-intensive plants and by 0.4 percent for low skill-intensive plants when the provincial skill intensity is doubled.

Together, provincial skill intensity's contribution to (i) the productivity of high skill-intensive incumbents, (ii) attracting high skill-intensive entrants, and (iii) lowering the exit probability of high skill-intensive plants, lead to agglomeration of high skill-intensive production. Figures 4.2 and 4.3 indicate that high skill-intensive entrants are more likely to choose high skill-intensive provinces. However, the exit probability is almost identical for high and low skill-intensive plants in high skill-intensive provinces. Our numerical results are consistent with the patterns in figures 4.2 and 4.3. In terms of relative importance, channel (ii) appears more important than (iii) for the agglomeration of high skill-intensive production. The combination of these three forces explains provincial skill disparities illustrated in figure 4.1.

**4.7 Conclusion**

In this study, we examined the concentration of high skill-intensive production in particular regions by focusing on human-capital spillovers. Recent studies of endogenous growth suggest such spillovers are an important source of economic growth, but the mechanism by which knowledge spreads across firms or plants has received limited empirical attention. To address this gap in the literature, we extend the absorptive capacity hypothesis, i.e., own investment in skills is required over time to utilize knowledge created by other firms, to a plant-level setting. We anticipate that high skill-intensive plants benefit more from spillovers than their low-skill counterparts, which leads to concentration of high skill-intensive plants in high skill-intensive regions.

To test the absorptive capacity hypothesis, we first estimated a plant-level production function incorporating human-capital spillovers. We draw on recent approaches to estimating plant-level production functions, which account for the correlation between unobserved plant's productivity and conventional input use. Our estimates of the production function show statistical support for the absorptive capacity hypothesis in 5 of 8 Chilean manufacturing industries. High skill-intensive plants, regardless of designations such as low- or high-tech industries, benefit from spillovers. In other words, plants' heterogeneity in terms of skill intensity matters for human-capital spillovers.

Using a conditional logit model, we then showed that plants are attracted to a region where (i) output demand is large, (ii) they can improve productivity through agglomeration economies and human-capital spillovers and (iii) human-capital is abundant. Complementary results were found in the examination of plant exit, where higher plant productivity lowered the probability of exit. Based on the above results, we numerically evaluate the contributions of regional skill intensity to enhancing plant's productivity, attracting plants and lowering the probability of plant exit. We found that the effect of regional skill intensity substantially varies within and across industries. High (low) skill-intensive plants benefit more (less) from spillovers and are more likely to be attracted to and less likely to exit from high skill-intensive regions.

The agglomeration of high skill-intensive production contributes to regional economic development through human-capital spillovers. Our results show that educating local population is necessary in the context of regional development since it raises the regional skill intensity, the source of human-capital spillover, and at the same time, it increases the supply of human-capital which helps attract plants to the region. However, the supply of skilled workers is assumed fixed in this study although inter-regional migration is an important source of labor supply in Chile. Consideration of the inter-regional migration decision of skilled workers should add depth to future analysis of high-skill agglomeration.

**4.8 Bibliography**

Acemoglu, D. and J. Angrist. (2000). "How Large are Human-Capital Externalities? Evidence from Compulsory Schooling Laws." *NBER Macroeconomics Annual*. 15: 9-59.

Arora, A. and A. Gambardella. (1994). "Evaluating Technological Information and Utilizing It: Scientific Knowledge, Technological Capability, and External Linkages in Biotechnology." *Journal of Economic Behavior and Organization*. 24: 91-114.

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

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

Barro, R and X. Sala-i-Martin. (2003). *Economic Growth*. MIT Press. Cambridge, MA.

Berman, E., Bound, J. and Z. Griliches. (1994). "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures." *Quarterly Journal of Economics*. 109: 367-397.

Bernard, A. and J. B. Jensen. (1997). "Exporters, Skill Upgrading, and the Wage Gap." *Journal of International Economics*. 42: 3-31.

Bernard, A. and J. B. Jensen. (2007). "Firm Structure, Multinationals, and Manufacturing Plant Deaths." *Review of Economics and Statistics*. 89: 193-204.

Bottazi, L. and G. Peri. (2003). 'Innovation and Spillovers in Regions: Evidence from European Patent Data.' *European Economic Review*. 47: 687-710.

Carlton, D. (1983). “The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables.” *Review of Economics and Statistics*. 65: 440-449.

Cockburn, I. and R. Henderson. (1998). “Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery.” *Journal of Industrial Economics*. 46: 157-182.

Cohen, W. and D. Levinthal. (1989). “Innovation and Learning: The Two Faces of R&D.” *Economic Journal*. 99: 569-596.

Crozet, M., Mayer, T. and J. Mucchielli. (2004). “How Do Firms Agglomerate? A Study of FDI in France.” *Regional Science and Urban Economics*. 34: 27-54.

Dunne, T., Roberts, M. and L. Samuelson. (1989). “The Growth and Failure of U. S. Manufacturing Plants.” *Quarterly Journal of Economics*. 104: 671-698.

Feenstra, R. and G.H. Hanson. (1997). “Foreign Direct Investment and Relative Wages: Evidence from Mexico’s Maquiladoras.” *Journal of International Economics*. 42: 371-393.

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

Glaeser, E. and D. Mare. (2001). “Cities and Skills.” *Journal of Labor Economics*. 19: 316- 342.

Guimaraes, P., Figueiredo, O. and D. Woodward. (2000). “Agglomeration and the Location of Foreign Direct Investment in Portugal.” *Journal of Urban Economics*. 47: 115-135.

Head, K. and T. Mayer. (2004). “Market Potential and the Location of Japanese Investment in the European Union.” *Review of Economics and Statistics*. 86: 959-972.

Head, K., Ries, J. and D. Swenson. (1999). “Attracting Foreign Manufacturing: Investment Promotion and Agglomeration.” *Regional Science and Urban Economics*. 29: 197-218.

Henderson, J. V. (2003). “Marshall’s Scale Economies.” *Journal of Urban Economics*. 53: 1-28.

Henderson, J. V., Z. Shalizi, and A. J. Venables (2001). “Geography and Development.” *Journal of Economic Geography*. 1: 81-105.

Hopenhayn, H. (1992). “Entry, Exit, and firm Dynamics in Long Run Equilibrium.” *Econometrica*. 60: 1127-1150.

Jaffe, A., Trajtenberg, M. and R. Henderson. (1993). “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.” *Quarterly Journal of Economics*. 63: 577-598.

Jovanovic, B. (1982). “Selection and the Evolution of Industry.” *Econometrica*. 50: 649-670.

Lall, S., Shalizi, Z. and U. Deichmann. (2004). “Agglomeration Economies and Productivity in Indian Industry.” *Journal of Development Economics*. 73: 643-673.

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

Liu, L. (1991). “Entry-Exit and Productivity Changes: An Empirical Analysis of Efficiency Frontiers.” Ph.D. Thesis, University of Michigan.

Lucas, R.E. (1988). “On the Mechanics of Economic Development.” *Journal of Monetary Economics*. 22: 3-42.

McFadden, D. (1984). “Econometric Analysis of Qualitative Response Models.” in Z.

Griliches and M. Intriligator, eds., *Handbook of Econometrics, Volume 2*. Elsevier. Amsterdam.

Melitz, M. (2003). “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity.” *Eonometrica*. 71: 1695-1725.

Moretti, E. (2004a). “Workers’ Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions.” *American Economic Review*. 94: 656-690.

Moretti, E. (2004b). “Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Section Data.” *Journal of Econometrics*. 121: 175-212.

Moretti, E. (2004c). “Human-Capital Externalities in Cities.” in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economics, Volume 4*. Elsevier. Amsterdam.

Pavcnik, N. (2003). “What Explains Skill Upgrading in Less Developed Countries?” *Journal of Development Economics*. 71: 311-328.

Rauch, J. E. (1993). “Productivity Gains from Geographic Concentration of Human-Capital: Evidence from the Cities.” *Journal of Urban Economics*. 34: 380-400.

Romer, P. (1990). “Endogenous Technological Change.” *Journal of Political Economy*. 98: S71-102

Rosenthal, S. and W. Strange. (2004). “Evidence on the Nature and Sources of Agglomeration Economies.” in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economics, Volume 4*. Elsevier. Amsterdam.

**4.9 Endnotes**

[1]During the sample period, 1998-2003, 2488 plants are recorded as entrants, while 2225 plants appear to have discontinued production. About 2875 plants reported production accounts (survived) for all six years.

[2]In the Appendix D, we estimate the conventional specification in which every plant, regardless of the skill intensity, can benefit from human-capital spillovers. That is, a production function with and without $\eta_{irt-1}$ in the right-hand side of equation (2) is estimated and a J test is carried out to choose between these two nonnested models.

[3]Due to other plants' entry, exit and employment decisions, $G_{irt}$ and $S_{irt}$ are not assumed to be predetermined.

[4]For details on bootstrapping procedures, refer to Levinsohn and Petrin (2003).

[5]We assume that plant characteristics matter only through the productivity term. The reasoning here is that plant characteristics do not vary across regions and are canceled out in the estimation of equation (14).

[6]Melitz (2003), following Hopenhayn (1992), suggests that these heterogeneous firms face bad shocks in each period forcing a proportion of them to exit in every period.

[7]Refer to Liu (1991) for details on constructing plant-level variables.

[8]In provinces where no plants locate, we use the average wage across other provinces in the same macro-region.

[9]Some may argue that this result reflects heterogeneity across industries at finer industry classification not across plants. For example, all wine-makers could have share of skilled workers near 1, while vegetable processors do not employ any skilled workers. We compute plant's skill intensity at 4-ditit level. Although the mean shows some variations across industries, the standard deviation centers around 0.3 for most industries. In other words, there is enough variation in plant's skill intensity even at 4-digit level.

[10]The null hypothesis that two parameters are equal, i.e., $\alpha_{11} = \alpha_{12}$, is rejected by a LR test at the 1 percent level.

[11]Note that the nonsignificant effect of wages applies to plants' location choice, but to the extent that higher wages are not offset by higher productivity, plants are likely

forced to exit high skill-intensive provinces (Moretti, 2004a).

[12]We suspect that the effect of plant size is not linear. When the first and second-order terms of $\ln size$ are included, the coefficient on $\ln size$ becomes negative and significant, and that on the square of $\ln size$ turns positive and significant. In other words, as plants become large, the probability of exit declines at decreasing rate.

[13]Of the 10579 plant-year observations, $D_{irt}$ takes a value one for 884 plant-year observations.

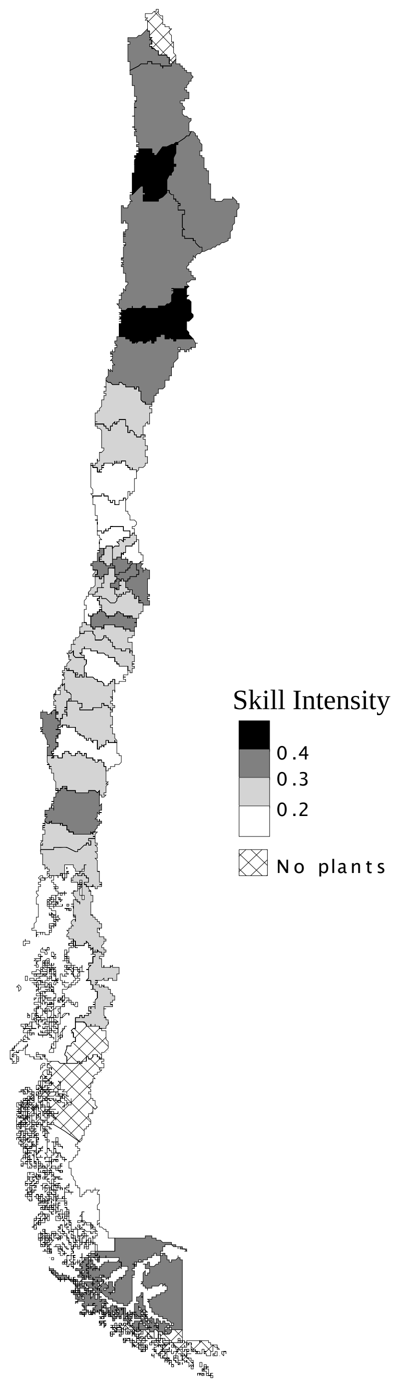



**Figure 4.1: Average of Plants' Skill Intensity in Chilean Provinces**

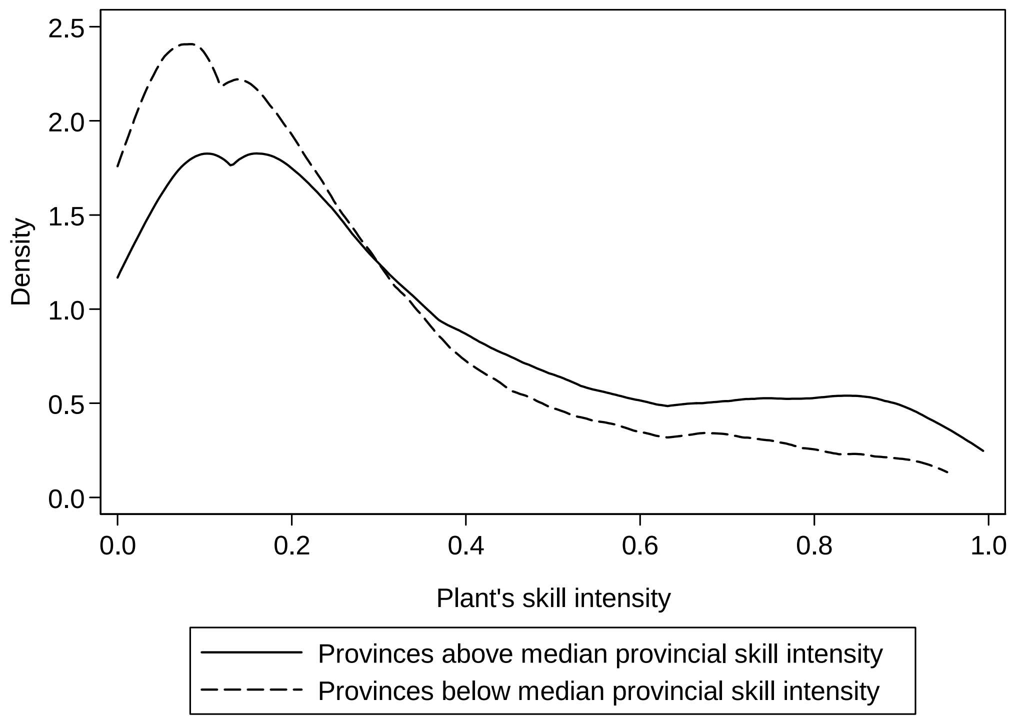



**Figure 4.2: Kernel Density Estimates for Entrants' Skill Intensity in Provinces with above and below Median Provincial Skill Intensity**

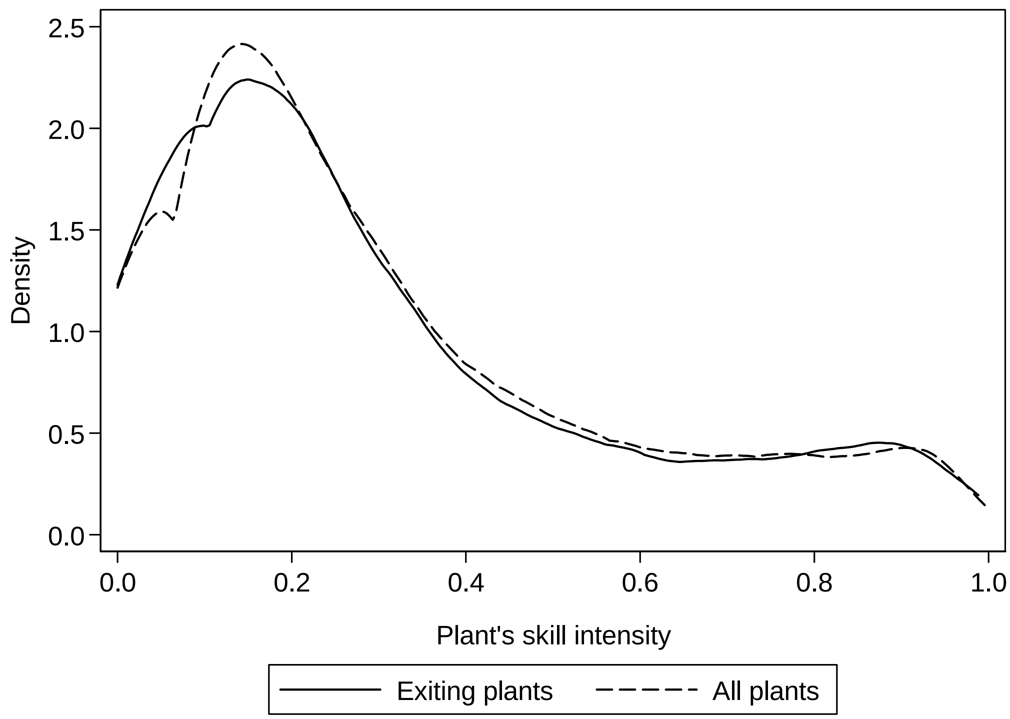



**Figure 4.3: Kernel Density Estimates for Exiting and All Plants' Skill Intensity in Provinces with above Median Provincial Skill Intensity**

**Table 4.1: Summary Statistics**

| Variable | Unit | Mean | Std. Deviation |
|---|---|---|---|
| Output (*Y/1,000,000*) | Constant peso | 4.21 | 21.39 |
| Skilled worker ( $L^s$ ) | Number | 24.77 | 70.51 |
| Unskilled worker ( $L^u$ ) | Number | 51.91 | 104.99 |
| Material (*M/1,000,000*) | Constant peso | 2.43 | 15.10 |
| Energy (*E/1,000,000*) | Constant peso | 0.12 | 0.87 |
| Capital (*K/1,000,000*) | Constant peso | 4.03 | 92.21 |
| Share of skilled worker in a plant ($\eta$) | Index, 0-1 | 0.32 | 0.25 |
| Share of skilled worker in a province (*S*) | Index, 0-1 | 0.39 | 0.12 |
| Agglomeration (*G*) | Number | 14.33 | 50.18 |
| Market potential (*MP/1,000,000*) | Number | 0.37 | 0.68 |
| Population density (*POPDEN*) | Number / $km^2$ | 93.97 | 330.44 |
| Wage rate (*w/1,000*) | Constant peso | 5.05 | 2.88 |
| Share of university graduates (*UNIV*) | Percent | 2.79 | 1.35 |
| Plant's age (*AGE*) | Number | 2.69 | 1.42 |

Source: INE, Chilean Manufacturing Census, Various Years.
INE, Housing and Population Census, 2002.
INE, Population Projections.

**Table 4.2: Parameter Estimates of the Production Function**

| | ISIC | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
| $\ln L^s$ | 0.069*** | 0.127*** | 0.133*** | 0.145*** | 0.123*** | 0.075*** | 0.115** | 0.144*** |
| | (0.009) | (0.014) | (0.023) | (0.021) | (0.025) | (0.026) | (0.051) | (0.013) |
| $\ln L^u$ | 0.084*** | 0.115*** | 0.116*** | 0.050*** | 0.087*** | 0.118*** | 0.047* | 0.125*** |
| | (0.010) | (0.011) | (0.024) | (0.016) | (0.018) | (0.024) | (0.029) | (0.012) |
| $\ln M$ | 0.555*** | 0.523*** | 0.520*** | 0.717*** | 0.724*** | 0.673*** | 0.520*** | 0.660*** |
| | (0.084) | (0.073) | (0.102) | (0.097) | (0.103) | (0.071) | (0.166) | (0.093) |
| $\ln E$ | 0.076*** | 0.058*** | 0.097*** | 0.054*** | -0.012 | 0.110*** | 0.032 | 0.074*** |
| | (0.011) | (0.014) | (0.016) | (0.018) | (0.017) | (0.030) | (0.029) | (0.012) |
| $\ln K$ | 0.125** | 0.138* | 0.008 | 0.045 | 0.225** | 0.056* | 0.215 | 0.047 |
| | (0.045) | (0.053) | (0.056) | (0.064) | (0.084) | (0.062) | (0.180) | (0.055) |
| $\ln G$ | 0.022*** | 0.036*** | 0.007 | 0.004 | -0.008 | -0.023 | 0.016 | 0.017*** |
| | (0.006) | (0.007) | (0.013) | (0.010) | (0.013) | (0.017) | (0.019) | (0.006) |
| $\eta S$ | 0.706*** | 0.243 | 0.792 | 0.956** | 2.307*** | 0.450 | 1.178** | 0.559** |
| | (0.195) | (0.467) | (0.672) | (0.551) | (0.458) | (0.395) | (0.540) | (0.274) |
| OBS. | 4504 | 1991 | 1318 | 923 | 1860 | 583 | 253 | 2489 |

Note: *, ** and *** indicate statistical significance at the 10%, 5% and 1%, respectively. Value in parenthesis is the bootstrapped standard error based on 200 iterations. Dependent variable is the log of output.
ISIC Code: 31 Food Processing, 32 Textiles, 33 Wood, 34 Paper, 35 Chemicals, 36 Glass, 37 Basic Metals, 38 Machinery.

**Table 4.3: Parameter Estimates of the Plants' Location Choice Model**

| Variable | (1) | (2) |
|---|---|---|
| ln*AG* | 6.371*** | 7.172*** |
| | (1.168) | (0.739) |
| ln*AS* | 1.782*** | 1.522*** |
| | (0.573) | (0.383) |
| ln*MP* | 0.907*** | 1.103*** |
| | (0.156) | (0.108) |
| ln*POPDEN* | 0.048 | 0.101* |
| | (0.088) | (0.060) |
| ln*w* | 0.331 | 0.157 |
| | (0.219) | (0.145) |
| *UNIV* | 0.215*** | 0.149*** |
| | (0.085) | (0.058) |
| Number of plants | 970 | 2434 |
| 20 or more employees only | Yes | No |
| Log Likelihood | -2436 | -5858 |

Note: * and *** indicate statistical significance at the 10% and 1%, respectively. Value in parenthesis is the standard error. Unit of the observations is province and the dependent variable is a dummy which takes value one if the entrant locates in the province.

**Table 4.4: Estimation Results for Plants' Exit Model**

| Variable | (1) | (2) |
|---|---|---|
| Productivity | -0.174** | -0.167*** |
| | (0.089) | (0.066) |
| Age | -0.220*** | -0.180*** |
| | (0.040) | (0.028) |
| Size | 0.066 | -0.084*** |
| | (0.046) | (0.031) |
| Capital labor ratio | -0.119*** | -0.063*** |
| | (0.026) | (0.020) |
| Export dummy | -0.244*** | -0.193** |
| | (0.092) | (0.083) |
| FDI dummy | 0.000 | 0.000 |
| | (0.002) | (0.002) |
| ln*MP* | 0.080 | 0.069 |
| | (0.201) | (0.163) |
| ln*POPDEN* | -0.199* | -0.154* |
| | (0.105) | (0.083) |
| Ln*w* | 0.532** | 0.374** |
| | (0.234) | (0.169) |
| *UNIV* | 0.058 | 0.030 |
| | (0.104) | (0.084) |
| OBS. | 10579 | 16345 |
| 20 or more employees only | Yes | No |
| Log Likelihood | -2956 | -4898 |

Note: *, ** and *** indicate statistical significance at the 10%, 5% and 1%, respectively. Unit of the observations is plant and the dependent variable is a dummy which takes value one if the plant stops its operation in the next year.

**Table 4.5: Elasticity of Productivity, Entry and Exit with respect to Provincial Skill Intensity**

| ISIC | Plant's skill intensity ($\eta$) | Elasticity of Productivity | Elasticity of Entry | Elasticity of Exit |
|---|---|---|---|---|
| 31 | Low | 0.022 | 0.039 | -0.004 |
| | Mean | 0.088 | 0.154 | -0.014 |
| | High | 0.157 | 0.274 | -0.025 |
| 32 | Low | 0.008 | 0.013 | -0.001 |
| | Mean | 0.030 | 0.053 | -0.005 |
| | High | 0.054 | 0.094 | -0.009 |
| 33 | Low | 0.025 | 0.043 | -0.004 |
| | Mean | 0.099 | 0.173 | -0.016 |
| | High | 0.176 | 0.308 | -0.028 |
| 34 | Low | 0.030 | 0.052 | -0.005 |
| | Mean | 0.119 | 0.209 | -0.019 |
| | High | 0.213 | 0.371 | -0.034 |
| 35 | Low | 0.072 | 0.126 | -0.011 |
| | Mean | 0.288 | 0.503 | -0.046 |
| | High | 0.513 | 0.896 | -0.082 |
| 36 | Low | 0.014 | 0.025 | -0.002 |
| | Mean | 0.056 | 0.098 | -0.009 |
| | High | 0.100 | 0.175 | -0.016 |
| 37 | Low | 0.037 | 0.064 | -0.006 |
| | Mean | 0.147 | 0.257 | -0.023 |
| | High | 0.262 | 0.457 | -0.042 |
| 38 | Low | 0.017 | 0.031 | -0.003 |
| | Mean | 0.070 | 0.122 | -0.011 |
| | High | 0.124 | 0.217 | -0.020 |
| Average | Low | 0.028 | 0.049 | -0.004 |
| | Mean | 0.112 | 0.196 | -0.018 |
| | High | 0.200 | 0.349 | -0.032 |

Note: Low, Mean and High correspond to mean minus one standard deviation, mean and mean plus one standard deviation of plant’s skill intensity, respectively.

## 5. Conclusion

The location of firms has been a principal research topic in regional and international economics with the object of mitigating economic inequality within a country. While large physical and performance differences across firms or plants have been observed in recent studies, their locational preferences have received limited attention. The contribution of this dissertation is to examine physical, performance and locational differences of firms as sources of regional economic development. More specifically, the objective of this dissertation is to theoretically and empirically examine the role of firm heterogeneity in the agglomeration process and the latter's effect on regional economic development.

In the first essay, I analyzed the effect of trade liberalization on heterogeneous firms' location. The first result suggests that competition between high- and low- productivity of firms works as dispersion force. In other words, low-productivity firms locate away from high-productivity firms because the former is less competitive in a region where the latter agglomerates. This result holds true even when trade liberalization, with identical trade costs across regions, works as a dispersion force. Moreover, the relocation is a more likely equilibrium when productivity gap between high- and low- productivity firms widens or by trading with advanced countries. Finally, a subsidy by less-developed regions is not effective to attract high-productivity firms because urban areas have the distinct advantage in such competition. The results imply that trade liberalization is effective in creating employment in less-developed regions, but the effect might be limited because only low-productivity firms are attracted to the regions. International competitiveness, the widening of productivity gap between domestic and foreign firms, may also help in achieving spatially-uniform economic development.

In the second essay, I empirically tested the theoretical results obtained in the first essay with plant-level data from Chilean manufacturing. Levinsohn and Petrin's (2003) approach to estimating plant-level production functions was employed to correct for (i) possible simultaneity between productivity and conventional inputs and (ii) plants' self-selection to locate in specific markets. Plant's raw productivity, i.e. productivity independent of agglomeration economies, was computed to obtain measures of regional productivity-distribution, e.g., median, $10^{th}$ and $90^{th}$ percentiles. Then, a heteroskedastic-tobit model related the latter to agglomeration economies. Results show that high-productivity (exporting) plants locate in a region where other plants in the same industry agglomerate, industrial structure is diversified and market size is large. Hence, regional productivity can be improved by specializing in targeted industry, diversifying regional industrial structure and enlarging the market size, but the magnitude of improvements vary among these options. Finally, the results suggest that plants' self-selection outweighs the contribution of agglomeration economies in increasing a region's productivity.

In the third essay, I studied the effect of human capital spillovers on the agglomeration of high skill-intensive plants. First, the absorptive capacity hypothesis was empirically tested to identify a mechanism for knowledge spillovers. The hypothesis is supported in 5 out of 8 Chilean manufacturing industries. That is, high skill-intensive plants, regardless of industry designations as high- or low-tech, benefit more from spillovers than their low skill counterparts. Moreover, productivity, especially that part attributable to human-capital spillovers, significantly affects plants' location choice and exit. In other words, human-capital spillovers lower the exit probability of high skill-intensive incumbents and attract high skill-intensive entrants. Results imply that educating local population is effective in the context of regional development since it increases not only the regional skill intensity, the source of human-capital spillover, but also the supply of human-capital, which helps attract plants to the region.

The results obtained in the above essays reveal locational preferences of various types of firms. For instance, high-productivity firms prefer agglomerated regions, while low-productivity firms tend to locate away from high-productivity firms. Also, high skill-intensive firms are more likely to locate in a region where human capital is abundant than their low skill-intensive counterparts. Policy options for promoting spatially-uniform economic development through increases in regional productivity include specializing in targeted industry, diversifying regional industrial structure, enlarging the market size and workforce education. However, local governments should consider the cost of implementing each policy. The results of this dissertation help local governments' evaluation of the benefits or gains from each policy option, which when compared with their knowledge of costs, aid in the selection of an effective policy to improve regional well-being.

**Bibliography**

Acemoglu, D. and J. Angrist. (2000). “How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws.” *NBER Macroeconomics Annual*. 15: 9-59.

Ades, A. and E. Glaeser. (1995). “Trade and Circuses: Explaining Urban Giants.” *Quarterly Journal of Economics*. 110: 195-227.

Aitken, B., Hanson, G.H. and A. Harrison. (1997). “Spillovers, Foreign Investment, And Export Behavior.” *Journal of International Economics*. 43:103-132.

Allonso-Villar, O. (1999). “Spatial Distribution of Production and International Trade: A Note.” *Regional Science and Urban Economics*. 29: 371–380.

Arora, A. and A. Gambardella. (1994). “Evaluating Technological Information and Utilizing It: Scientific Knowledge, Technological Capability, and External Linkages in Biotechnology.” *Journal of Economic Behavior and Organization*. 24: 91-114.

Audretsch, D. and M. Feldman. (1996). “R&D Spillovers and the Geography of Innovation and Production.” *American Economic Review*. 86: 630-640.

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

Baldwin, R. and T. Okubo. (2006). “Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting.” *Journal of Economic Geography*. 6: 323-346.

Barro, R and X. Sala-i-Martin. (2003). *Economic Growth*. MIT Press. Cambridge, MA.

Behrens, K., Gaigne, C., Ottaviano, G. and J. F. Thisse. (2003). "Inter-regional and International Trade: Seventy Years After Ohlin." *CEPR Discussion Paper*. no. 4065.

Berman, E., Bound, J. and Z. Griliches. (1994). "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures." *Quarterly Journal of Economics*. 109: 367-397.

Bernard, A. and J. B. Jensen. (1997). "Exporters, Skill Upgrading, and the Wage Gap." *Journal of International Economics*. 42: 3-31.

Bernard, A. and J. B. Jensen. (1999). "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics*. 47: 1-25.

Bernard, A. and J. B. Jensen. (2007). "Firm Structure, Multinationals, and Manufacturing Plant Deaths." *Review of Economics and Statistics*. 89: 193-204.

Bottazi, L. and G. Peri. (2003). 'Innovation and Spillovers in Regions: Evidence from European Patent Data.' *European Economic Review*. 47: 687-710.

Campbell, J. and H. Hopenhayn. (2005). "Market Size Matters." *Journal of Industrial Economics*. 53:1-25.

Carlton, D. (1983). "The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables." *Review of Economics and Statistics*. 65: 440-449.

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

Cockburn, I. and R. Henderson. (1998). "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery." *Journal of Industrial Economics*. 46: 157-182.

Cohen, W. and D. Levinthal. (1989). "Innovation and Learning: The Two Faces of R&D." *Economic Journal*. 99: 569-596.

Coniglio, N. (2002). "Regional Integration and Migration: An Economic Geography Model with Heterogeneous Labour Force." *University of Glasgow Discussion Paper*.

Crozet, M. and P. Koenig. (2004). "EU Enlargement and the Internal Geography of Countries." *Journal of Comparative Economics*. 32: 265-279.

Crozet, M., Mayer, T. and J. Mucchielli. (2004). "How Do Firms Agglomerate? A Study of FDI in France." *Regional Science and Urban Economics*. 34: 27-54.

Davidson, R. and J. G. MacKinnon. (1981). "Several Tests for Model Specification in the Presence of Alternative Hypotheses." *Econometrica*. 49: 781-793.

Davidson, R. and J. G. MacKinnon. (2004). *Econometric Theory and Methods*. Oxford University Press. NY.

Dunne, T., Roberts, M. and L. Samuelson. (1989). "The Growth and Failure of U. S. Manufacturing Plants." *Quarterly Journal of Economics*. 104: 671-698.

Dupont, V. and P. Martin. (2006). "Subsidies to Poor Regions and Inequalities: Some Unpleasant Arithmetic." *Journal of Economic Geography*. 6: 223-240.

Eberts, R. and D. McMillen. (1999). "Agglomeration Economies and Urban Public infrastructure." in P. Cheshire and E. S. Mills, eds., *Handbook of Regional and Urban Economics*, Volume 3. Elsevier. Amsterdam.

Feenstra, R.C. and G.H. Hanson. (1997). "Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras." *Journal of International Economics*. 42: 371-393.

Feliciano, Z.M. (2001). "Workers and Trade Liberalization: The Impact of Trade Reforms in Mexico on Wages and Employment." *Industrial and Labor Relations Review*. 55: 95-115.

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

Galiani, S. and P. Sanguinetti. (2003). "The Impact of Trade Liberalization on Wage Inequality: Evidence from Argentina." *Journal of Development Economics*. 72: 497-513.

Glaeser, E. and D. Mare. (2001). "Cities and Skills." *Journal of Labor Economics*. 19: 316-342.

Goldberg, P.K. and N. Pavcnik. (2005). "Trade, Wages, and the Political Economy of Trade Protection: Evidence from the Colombian Trade Reforms." *Journal of International Economics*. 66: 75-105.

Greene, W. (2003). *Econometric Analysis 5th Edition*. Prentice Hall. Upper Saddle River, NJ.

Greene, W. (2006). "Censored Data and Truncated Distributions." in T. Mills and K. Patterson, eds., *Palgrave Handbook of Econometrics Volume 1; Theoretical Econometrics*. Palgrave Macmillan. London.

Guimaraes, P., Figueiredo, O. and D. Woodward. (2000). "Agglomeration and the Location of Foreign Direct Investment in Portugal." *Journal of Urban Economics*. 47: 115-135.

Hanson, G.H. (1997). "Increasing Returns, Trade and the Regional Structure of Wages." *Economic Journal*. 107: 113-133.

Hanson, G.H. (1998). "Regional Adjustment to Trade Liberalization." *Regional Science and Urban Economics*. 28: 419-444.

Hanson, G.H. (2001). "Scale Economies and the Geographic Concentration of Industry." *Journal of Economic Geography*. 1:255-276.

Head, K. and T. Mayer. (2004). "Market Potential and the Location of Japanese Investment in the European Union." *Review of Economics and Statistics*. 86: 959-972.

Head, K., Ries, J. and D. Swenson. (1999). "Attracting Foreign Manufacturing: Investment Promotion and Agglomeration." *Regional Science and Urban Economics*. 29: 197-218.

Helpman, E. (2006). "Trade, FDI, and the Organization of Firms." *Journal of Economic Literature*. 44: 589–630.

Henderson, J. V. (1986). "Efficiency of Resource Usage and City Size." *Journal of Urban Economics*. 19:47-70.

Henderson, J. V. (2003)."Marshall"s Scale Economies."*Journal of Urban Economics*. 53:1-28.

Henderson, J. V., Z. Shalizi, and A. J. Venables (2001). "Geography and Development." *Journal of Economic Geography*. 1: 81-105.

Holmes, T. and J. Stevens. (2002). "Geographic Concentration and Establishment Scale." *Review of Economics and Statistics*. 84: 682-690.

Hopenhayn, H. (1992). "Entry, Exit, and firm Dynamics in Long Run Equilibrium." *Econometrica*. 60: 1127-1150.

Jaffe, A., Trajtenberg, M. and R. Henderson. (1993). "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics*. 63: 577-598.

Jovanovic, B. (1982). "Selection and the Evolution of Industry." *Econometrica*. 50:

649-670.

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

Krugman, P. and R. Livas. (1996). “Trade Policy and the Third World Metropolis.” *Journal of Development Economics*. 49: 137-150.

Lall, S., Shalizi, Z. and U. Deichmann. (2004). “Agglomeration Economies and Productivity in Indian Industry.” *Journal of Development Economics*. 73: 643-673.

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

Liu, L. (1991). “Entry-Exit and Productivity Changes: An Empirical Analysis of Efficiency Frontiers.” Ph.D. Thesis, University of Michigan.

Lucas, R.E. (1988). “On the Mechanics of Economic Development.” *Journal of Monetary Economics*. 22: 3-42.

Malmberg, A., Malmberg, B. and P. Lundequist. (2000). “Agglomeration and Firm Performance: Economies of Scale, Localisation, and Urbanisation among Swedish Export Firms.” *Environmental and Planning A*. 32: 305-321.

Mansori, K. (2003). “The Geographic Effects of Trade Liberalization with Increasing Returns in Transportation.” *Journal of Regional Science*. 43: 249-268.

McFadden, D. (1984). “Econometric Analysis of Qualitative Response Models.” in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics, Volume 2*. Elsevier. Amsterdam.

Melitz, M. (2003). “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity.” *Eonometrica*. 71: 1695-1725.

Ministry of Land, Infrastructure and Transportation. (2007). “Survey on the Current Situation of Rural Communities,” in Japanese. http://www.mlit.go.jp/.

Moretti, E. (2004a). “Workers’ Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions.” *American Economic Review*. 94: 656-690.

Moretti, E. (2004b). “Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Section Data.” *Journal of Econometrics*. 121: 175-212.

Moretti, E. (2004c). “Human-Capital Externalities in Cities.” in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economics, Volume 4*. Elsevier. Amsterdam.

Nakamura, R. (1985). “Agglomeration Economies in Urban Manufacturing Industries: A Case of Japanese Cities.” *Journal of Urban Economics*. 17: 108-124.

OECD. (2003). *OECD Science, Technology, and Industry Scoreboard*. OECD. Paris.

Olley, S. and A. Pakes. (1996). “The Dynamics of Productivity in the Telecommunications Equipment Industry.” *Econometrica*. 64: 1263-1298.

Ottaviano, G., Tabuchi, T. and J. F. Thisse. (2002). “Agglomeration and Trade Revisited.” *International Economic Review*. 43: 409-435.

Overman, H. and L. A. Winters. (2006). “Trade and Economic Geography: The Impact of EEC Accession on the UK.” *CEPR Discussion Paper*. no. 5574.

Paluzie, E. (2001). “Trade Policy and Regional Inequalities.” *Papers in Regional Science*. 80: 67-85.

Pavcnik, N. (2002). “Trade Liberalization, Exit, and Productivity Improvements:

Evidence from Chilean Plants." *Review of Economic Studies*. 69: 245-76.

Pavcnik, N. (2003). "What Explains Skill Upgrading in Less Developed Countries?" *Journal of Development Economics*. 71: 311-328.

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

Redding, S. and A.J. Venables. (2004). "Economic Geography and International Inequality." *Journal of International Economics*. 62: 53-82.

Rice, P., Venables, A. J. and E. Patacchini (2006). "Spatial Determinants of Productivity: Analysis for the Regions of Great Britain." *Regional Science and Urban Economics*. 36: 727-752.

Roberts, M. and J. Tybout. (1997). "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs." *American Economic Review*. 87: 545-564.

Romer, P. (1990). "Endogenous Technological Change." *Journal of Political Economy*. 98: S71-102

Rosenthal, S. and W. Strange. (2004). "Evidence on the Nature and Sources of Agglomeration Economies." in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economics*, volume 4. Elsevier. Amsterdam.

Syverson, C. (2004). "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy*. 112: 1181-1222.

Tybout, J. (2000). "Manufacturing Firms in Developing Countries: How Well They Do and Why?" *Journal of Economic Literature*. 38: 11-44

UN-HABITAT. (2003). “Slums of the World: The Face of Urban Poverty in the New Millennium?” http://www.unhabitat.org/.

Venables, A.J. (2005). “Spatial Disparities in Developing Countries: Cities, Regions and International Trade.” *Journal of Economic Geography*. 5: 3-21.

Yale Global Forum. (2007). http://yaleglobal.yale.edu/globalization/.

Yeaple, S. (2005). “A Simple Model of Firm Heterogeneity, International Trade and Wages.” *Journal of International Economics*. 65: 1-20.

# APPENDICES

**Appendix A. Expansion of Abbreviations**

In this appendix, we expand on several abbreviations used in the chapter 2 and provide more detailed explanations how those results are derived.

*Derivation of Utility Differences (10a) and (10b)*

By substituting equations (5a), (5b), (7), (9a) and (9b) into (6), the difference in the utility level of the entrepreneurs (equation 10a) and industrialists (equation 10b) are derived as a function of their share in the North, transport costs, the parameters. Parameters, $A_1$, $A_2$, $C_1$, $C_2$, $C_3$ and $C_4$, in equations (10a) and (10b) take the following form.

$$A_2 = \frac{(2\beta-\gamma-\delta)\left((6\beta-5\gamma)\beta+(5\beta-3\gamma)\delta+\gamma^2+(2\beta-\gamma+\delta)\delta L\right)}{\left((2\beta-\gamma)^2-\delta^2\right)^2}$$

$$A_1 = A_2 + \frac{(L+2)(\gamma-\delta)}{2(\beta-\gamma)(2\beta-\gamma-\delta)}$$

$$C_1 = \frac{2\left((2\beta-\gamma-\delta)^2(3\beta-\gamma+2\delta)\alpha-2\left(\beta(\gamma-2\beta)+\delta^2\right)\delta\theta_M\right)}{\left((2\beta-\gamma)^2-\delta^2\right)^2 A_1}$$

$$C_2 = \Big[2\Big((2\beta-\gamma-\delta)^2(3\beta-\gamma+2\delta)\alpha-\Big((2\beta-\gamma)^2\beta-(2\beta-\gamma)^2\delta$$
$$-(3\beta-2\gamma)\delta^2+\delta^3\Big)\theta_M\Big)\Big]\Big/\left((2\beta-\gamma)^2-\delta^2\right)^2 A_2$$

$$C_3 = C_2 - \frac{2(\beta-\gamma)\theta_M}{A_2(2\beta-\gamma-\delta)^2}$$

$$C_4 = C_1 - \frac{2(3\beta - \gamma - 2\delta)\theta_M}{A_1(2\beta - \gamma - \delta)^2}$$

Note that $A_1 \geq A_2 > 0$, $C_1 > C_4$ and $C_2 > C_3$ given the parametric assumptions.

*Derivation of the Maximum Transport Cost That Assures Trade* ($t_{\max}$)

The maximum transport cost that assures trade between the North and the South $t_{\max}$ is defined by $t_{\max} = \min\left\{t : p_s^{ro} - t - x_s \geq 0,\ \ r,o = N,S \ \ s = H,M,\ \ r \neq o\right\}$.

Setting the difference between price and marginal costs including transport costs $p_s^{ro} - t - x_s$ to zero, we obtain

$$t_{\max} = \frac{2(\beta - \gamma)\left((2\beta - \gamma - \delta)\alpha - (2\beta - \gamma)\theta_M\right)}{4\beta(\beta - \gamma) + \gamma^2 - \delta^2}.$$

Although the denominator of $t_{\max}$ is positive, the numerator can take either sign depending on the productivity difference. Thus, we must set the upper bound, $\theta_{M\max}$, to assure that maximum transport costs are positive:

$$\theta_{M\max} = \frac{\alpha(2\beta - \gamma - \delta)}{2\beta - \gamma}.$$

It can be shown that $C_4 > 0$ when the productivity difference is less than $\theta_{M\max}$. Finally, as shown in the Appendix B, $t_{\max} > C_1$ is necessary so that both types of firms initially disperse between regions and this condition is satisfied when total mass of labor is greater than $L_{\min}$, which is derived as:

$$L_{\min} = \frac{1}{(2\beta-\gamma+\delta)(2\beta\gamma-\gamma^2-\delta^2)((2\beta-\gamma-\delta)\alpha-(2\beta-\gamma)\theta_M)}\Big[2\Big\{(2\beta-\gamma-\delta)\alpha$$
$$\left(6\beta^3-\gamma^3+2\gamma^2\delta-2\gamma\delta^2-\delta^3-(9\gamma-7\delta)\beta^2+(5\gamma^2-8\gamma\delta+\delta^2)\beta\right)$$
$$+2(2\beta-\gamma+\delta)((2\beta-\gamma)\beta-\delta^2)\delta\theta_M$$
$$+(2\beta-\gamma)\left(6\beta^3+(3\gamma-\delta)\delta^2-(7\gamma-\delta)\beta^2+2(\gamma^2-2\delta^2)\beta\right)\theta_M\Big\}\Big].$$

*Derivation of Utility Differences (14a) and (14b)*

Following the same steps as in the derivation of (10a) and (10b), the difference in the utility level of the entrepreneurs (equation 14a) and industrialists (equation 14b) are derived as a function of their share in the North and transport costs. The parameters in the equations, $q_i(t_F, t)$, $i = 1, \ldots, 4$, are defined as follows:

$$q_1(t_F,t) = \frac{-1}{4(\beta-\gamma)(2\beta+\gamma K_F)^2}\Big[t\Big\{-8(\beta-\gamma)(3\beta+\gamma+2\gamma K_F)\alpha$$
$$+\left(12\beta^2+4(3K_F+L)\beta\gamma+(K_F(3K_F+2L-2)-4)\gamma^2\right)t$$
$$-2(4\beta+2\gamma+3\gamma K_F)(K_F(t_F+\theta_F)+\theta_M)\gamma\Big\}\Big]$$

$$q_2(t_F,t) = q_1(t_F,t) - \frac{t\theta_M}{2(\beta-\gamma)}$$

$$q_3(t_F,t) = q_1(t_F,t) - \frac{t\theta_M}{\beta-\gamma}$$

$$q_4(t_F,t) = q_1(t_F,t) - \frac{3t\theta_M}{2(\beta-\gamma)}$$

For $t > 0$, $q_1(t_F,t) > q_2(t_F,t) > q_3(t_F,t) > q_4(t_F,t)$.

*Derivation of the Maximum Transport Cost that Assures Trade in the Open Economy*

The parameters $t_{2\max}$ and $t_{F\max}$ are the maximum transport costs which assure trade between the North and the South and between the two countries, respectively. These are obtained in the same way as $t_{\max}$ is derived. Specifically, by setting the difference between price and marginal costs including transports costs to zero, we obtain

$$t_{2\max} = \frac{2(\beta-\gamma)\alpha + (\theta_F - \theta_M)\gamma K_F - (2\beta-\gamma)\theta_M}{2\beta + \gamma K_F}$$

$$t_{F\max} = \min\left\{ \frac{2(\beta-\gamma)\alpha + (\theta_F - \theta_M)\gamma K_F - (2\beta-\gamma)\theta_M}{2(\beta-\gamma) + \gamma K_F}, \frac{2(\beta-\gamma)\alpha - 2\beta\theta_F + \gamma\theta_M}{2\beta} \right\}$$

Although the numerator of $t_{2\max}$ and $t_{F\max}$ can take either sign depending on the productivity difference, they are always positive as long as $\theta_M$ and $\theta_F$ are in the space "*pd*":

$$pd = \{(\theta_M, \theta_F) \in R^2 : 2(\beta-\gamma)\alpha + (\theta_F - \theta_M)\gamma K_F - (2\beta-\gamma)\theta_M > 0, \\ 2(\beta-\gamma)\alpha - 2\beta\theta_F + \gamma\theta_M > 0, \theta_M > 0, \theta_F \geq 0\}$$

It can be also shown that $q_4(t_F, t) = 0$ has a positive solution with respect to domestic transport costs when the productivity differences are in the space "*pd*". Finally, as noted in section 2.4, $q_1(t_F, t)$ must be less than zero at $t = t_{2\max}$ so that both types of firms initially disperse between regions. This condition is satisfied when total mass of labor is greater than $L_{2\min}$:

$$L_{2\min}=\frac{1}{2\gamma\left(2\beta+\gamma K_F\right)\left(2\left(\beta-\gamma\right)\alpha+\left(\theta_F-\theta_M\right)\gamma K_F-\left(2\beta-\gamma\right)\theta_M\right)}\Big[2\left(\beta-\gamma\right)\Big(12\beta^2$$

$$+\left(\left(5K_F+6\right)K_F+4\right)\gamma^2+8\left(1+2K_F\right)\beta\gamma\Big)\alpha+\Big\{2\left(2\beta+\gamma K_F\right)\left(4\beta+2\gamma+3\gamma K_F\right)t_{F\max}$$

$$+\left(4\beta^2+8\left(K_F+1\right)\beta\gamma+\left(3\left(K_F+2\right)K_F+4\right)\gamma^2\right)\theta_F\Big\}\gamma K_F$$

$$+\left(24\beta^3+2\left(9K_F+2\right)\beta\gamma^2K_F+\left(4+\left(3K_F^2+K_F+2\right)K_F\right)\gamma^3+4\left(9K_F+1\right)\beta^2\gamma\right)\theta_M\Big].$$

**Appendix B. Proofs**

In this appendix, we provide a proof for each of the major results and propositions in the chapter 2.

*Result 1 in Section 2.3*

First, we prove that $C_3 > C_4$ and $C_2 > C_4$. As shown in Appendix A, the sign of the difference $C_3 - C_4$ depends on a cubic function of $\delta$ and the coefficient on the third-degree term is positive. At $\delta = \gamma$, the difference is positive and the slope is negative. At $\delta = 0$, the difference is positive and the slope is negative. Therefore, for a nonnegative $\delta$, we have $C_3 > C_4$. From $C_2 > C_3$, we have $C_2 > C_4$.

Second, we prove that $C_1 > C_3$. The sign of $C_1 - C_3$ depends on a quadrant function of $\delta$ and the coefficient of $\delta^4$ is positive. Thus, $C_1 - C_3$ has at most three local extreme or inflection points. At $\delta = \gamma$, the difference is positive. Since the slope is positive and the second derivative is negative, this point locates to the left of the middle critical point of the three inflection points. At $\delta = 0$, it can be shown that $C_3 > t_{\max}$ which implies $C_3 > C_1$ because $t_{\max} > C_1$ from Appendix A, i.e. the difference is negative. Therefore, there exists a $\delta^{z1}$ between zero and $\gamma$ such that $C_1 = C_3$ at $\delta = \delta^{z1}$ and for any $\delta \in \left(\delta^{z1}, \gamma\right]$, $C_1 > C_3$.

Finally, we prove that $C_1 > C_2$. The sign of $C_1 - C_2$ depends on a cubic function of $\delta$ and the coefficient of $\delta^3$ is negative. At $\delta = \gamma$, the difference and the slope is positive. At $\delta = \delta^{z1}$, from the above, we know that $C_1 = C_3$. Since $C_2 > C_3$, $C_2$ must be greater than $C_1$ at $\delta = \delta^{z1}$. Since at $\delta = 0$, the slope remains positive, there exists a $\delta^{z2}\left(> \delta^{z1}\right)$ such that for any $\delta \in \left(\delta^{z2}, \gamma\right]$, we have $C_1 > C_2$.

Result 1 in section 2.3 follows if we define $\delta^*$ as $\delta^* = \delta^{z2}$.

*Proof of Proposition 1*

From result 1, we know that for any $\delta \in \left(\delta^*, \gamma\right]$, $C_1 > C_2$, $C_3 > C_4$, $C_1 > C_3$ and $C_2 > C_4$ hold. For $t \in \left(C_1, t_{\max}\right]$, $C_i - t < 0,\ \ i = 1,\ldots,4$ and $\partial\left(V_l^N - V_l^S\right)/\partial\lambda_l < 0$. Thus, a symmetric equilibrium is the stable (initial) equilibrium. For $t \in \left(C_2, C_1\right)$, $C_1 - t > 0$ and $C_i - t < 0,\ \ i = 2,3,4$, we have $\partial\left(V_H^N - V_H^S\right)/\partial\lambda_H > 0$. In the latter case, symmetric equilibrium is no longer stable and high-productivity firms move to the North. When $\lambda_H > 1/2$, $V_M^N - V_M^S < 0$ and $\partial\left(V_M^N - V_M^S\right)/\partial\lambda_M < 0$ hold, causing low-productivity firms to relocate to the South until $V_M^N - V_M^S = 0$ or $\lambda_M = 0$. When $\lambda_H > 1/2$ and $\lambda_M < 1/2$, $V_H^N - V_H^S > 0$ always holds since $C_2 - t < 0$. Therefore, $\lambda_H = 1$. Finally, when $\lambda_H = 1$, $\lambda_M = 0$ is never attained because $A_2\left(C_3 - t\right) > A_1\left(C_4 - t\right)$ always holds. Consequently, $\lambda_H = 1$ and $\lambda_M \in \left(0, 1/2\right)$ is the equilibrium.

Solving $V_M^N - V_M^S = 0$ for $\lambda_M$ and differentiating it with respect to *t* shows that $d\lambda_M/dt<0$, which implies that as transport costs decrease, the share of low-productivity firms in the North increases when *t* falls. For $t\in(C_3,C_2)$, when $\lambda_H=1$, $V_H^N-V_H^S>0$ holds since $A_1\geq A_2>0$. Thus, entrepreneurs have no incentive to relocate from the North. When $\lambda_H=1$, $V_M^N-V_M^S=0$ and $\partial\left(V_M^N-V_M^S\right)/\partial\lambda_M<0$ hold for $\lambda_M\in(0,1/2)$ and $V_M^N-V_M^S<0$ holds for $\lambda_M\in[1/2,1]$ because $C_i-t<0,\ i=3,4$. Therefore, $\lambda_H=1$ and $\lambda_M\in(0,1/2)$ is the equilibrium. Again $d\lambda_M/dt<0$. After *t* becomes less than $C_3$, $V_M^N-V_M^S=0$ and $\partial\left(V_M^N-V_M^S\right)/\partial\lambda_M<0$ hold for $\lambda_M\in[1/2,1]$ and $V_M^N-V_M^S>0$ holds for $\lambda_M\in(0,1/2)$. Hence $\lambda_H=1$ and $\lambda_M\in[1/2,1]$ is the equilibrium and $d\lambda_M/dt<0$.

From Result 1 of section 2.3, we know that for $\delta<\delta^{z1}$, $C_2>C_1$ and $C_3>C_1$ hold. Again, for $t\in(C_1,t_{\max}]$, there exists a symmetric equilibrium Then, for $t<C_1$, sector *H* will relocate to the North because $\partial\left(V_H^N-V_H^S\right)/\partial\lambda_H>0$. Since $C_3-t>0$, this relocation makes $V_M^N-V_M^S>0$ and low-productivity firms also relocates to the North. Hence, $\lambda_M>1/2$. When $\lambda_M>1/2$, $V_H^N-V_H^S>0$ always holds. As a result, $\lambda_H=1$ and $\lambda_M\in[1/2,1]$ is the stable spatial equilibrium. Also, $d\lambda_M/dt<0$ holds.

*Proof of Proposition 2*

Since $t(t_F)$ is a linear function of $t_F$ with a positive slope, for any $t_c \in (t(0), t(t_{F\max}))$, there exists a $t_{FC} \in (0, t_{F\max})$ such that $t_c = t(t_{FC})$. As explained in section 2.4, when $t_F$ decreases, $q_3(t_F, t) + q_4(t_F, t)$ shifts downward when depicted against *t* (figure 2.2). $q_3(t_F, t) + q_4(t_F, t) > 0$ holds for $t_F \in (t_{FC}, t_{F\max}]$ given $t = t_c$. It can also be shown that $q_1(t_F, t) + q_2(t_F, t) > 0$ because $q_1(t_F, t) > q_2(t_F, t) > q_3(t_F, t) > q_4(t_F, t)$. Therefore, $\lambda_H = 1$ and $\lambda_M = 1$ is a stable equilibrium because $V_H^N - V_H^S > 0$ and $V_M^N - V_M^S > 0$ hold.

After the international transport costs reach $t_{FC}$, further decline will cause $q_3(t_F, t) + q_4(t_F, t)$ to be negative, but $q_1(t_F, t) + q_2(t_F, t)$ remains positive. Therefore, $V_H^N - V_H^S > 0$, $V_M^N - V_M^S < 0$ and $\partial\left(V_M^N - V_M^S\right) / \partial\lambda_M < 0$ hold when $\lambda_H = 1$ and $\lambda_M = 1$. This implies that low-productivity firms are the first to relocate to the South. By solving $V_M^N - V_M^S = 0$ for $\lambda_M$ and substituting this into (14a), we have $V_H^N - V_H^S = (\lambda_H - 1/2)\left[q_1(t_F, t)q_4(t_F, t) - q_2(t_F, t)q_3(t_F, t)\right] / q_4(t_F, t)$ and it can be shown that $\left[q_1(t_F, t)q_4(t_F, t) - q_2(t_F, t)q_3(t_F, t)\right] / q_4(t_F, t) > 0$ for $t_F \in (0, t_{FC})$. Thus, as long as $\lambda_M \in (0,1)$, i.e. $\lambda_M$ is determined by $V_M^N - V_M^S = 0$, $\lambda_H = 1$ is always an equilibrium. Note that, with $q_3(t_F, t) > q_4(t_F, t)$, when $\lambda_H = 1$, low-productivity firms do not fully agglomerate in the South, i.e.

$V_M^N - V_M^S < 0$ does not occur.

Finally, by setting $\lambda_H = 1$ and solving $V_M^N - V_M^S = 0$ for $\lambda_M$ and differentiating it with respect to $t_F$, we have $d\lambda_M / dt_F = \left[ q_3(t_F, t) q_4'(t_F, t) - q_4(t_F, t) q_3'(t_F, t) \right] / 2q_4(t_F, t)^2 > 0$. Thus, as international transport costs decrease, the share of low-productivity firms in the North decreases as long as it is determined by $V_M^N - V_M^S = 0$.

*Proof of Proposition 3*

As shown in the proof of proposition 2, at $t_F = t_{FC}$, low-productivity firms begin relocating to the North. By setting $\lambda_H = 1$ and solving $V_M^N - V_M^S = 0$ for $\lambda_M$, we obtain the share of low-productivity firms in the North. We can show that $dt_{FC} / d\theta_M > 0$, $dt_{FC} / d\theta_F < 0$ and $d\lambda_M / d\theta_F > 0$.

**Appendix C. Regional Plant Distribution and Parameter Estimates**

Regional plant distribution by industry is presented in table C.1. Parameter estimates of equations (19) and (20) in the chapter 3 are presented in tables C.2 and C.3, respectively.

**Table C.1: Plant Location by Industry and Region in Chile 1998-2003 Average**

| | ISIC | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Region | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
| 1 | 2.6 | 1.3 | 1.1 | 3.5 | 3.3 | 4.3 | 4.8 | 2.8 |
| 2 | 2.0 | 0.3 | 0.6 | 2.0 | 4.9 | 5.5 | 14.0 | 3.0 |
| 3 | 1.3 | 0.1 | 0.3 | 0.5 | 1.0 | 2.3 | 8.9 | 1.3 |
| 4 | 3.8 | 0.5 | 1.0 | 1.4 | 1.1 | 3.5 | 2.1 | 1.0 |
| 5 | 12.1 | 5.7 | 4.2 | 4.6 | 6.2 | 6.6 | 8.0 | 4.3 |
| 6 | 5.1 | 0.1 | 2.9 | 1.2 | 0.9 | 2.0 | 3.3 | 1.6 |
| 7 | 6.9 | 1.4 | 10.9 | 3.1 | 0.9 | 3.3 | 0.5 | 1.6 |
| 8 | 14.6 | 6.3 | 27.7 | 7.3 | 4.8 | 11.8 | 7.0 | 11.2 |
| 9 | 2.7 | 1.2 | 8.9 | 1.3 | 0.7 | 5.6 | 1.0 | 2.8 |
| 10 | 8.0 | 0.9 | 6.9 | 2.0 | 2.0 | 5.0 | 0.3 | 2.4 |
| 11 | 0.7 | 0.1 | 0.9 | 0.6 | 0.2 | 0.2 | 0.0 | 0.0 |
| 12 | 1.6 | 0.3 | 1.9 | 0.8 | 0.6 | 0.9 | 0.0 | 0.4 |
| 13 | 38.7 | 81.6 | 32.9 | 71.5 | 73.4 | 49.1 | 50.2 | 67.6 |

Source: INE, Chilean Manufacturing Census, Various Years.
Note: ISIC Code 31 Food Processing, 32 Textiles, 33 Wood, 34 Paper, 35 Chemicals, 36 Glass, 37 Basic Metals, 38 Machinery

**Table C.2: Parameter Estimates of the Self-Selection Model: All Plants**

| | (1) | (2) | (3) | (4) | | (5) | | (6) | |
|---|---|---|---|---|---|---|---|---|---|
| Parameter | *α* | *α* | *α* | *α* | *γ* | *α* | *γ* | *α* | *γ* |
| Percentile | 10 | 50 | 90 | 10 | | 50 | | 90 | |
| Intercept | -16.875$^{***}$ | -17.207$^{***}$ | -35.539$^{**}$ | 2.087 | | 12.123$^{**}$ | | -24.118 | |
| | (4.856) | (5.092) | (14.811) | (4.424) | | (5.790) | | (18.241) | |
| ln*LOC* | 1.910$^{***}$ | 2.400$^{***}$ | 5.531$^{***}$ | 2.193$^{***}$ | -2.619$^{***}$ | 0.451 | -3.065$^{***}$ | 6.400$^{***}$ | -1.090 |
| | (0.339) | (0.261) | (0.755) | (0.587) | (0.641) | (0.459) | (1.137) | (0.934) | (0.733) |
| ln*HHI* | -0.442 | -1.302$^{**}$ | -3.554$^{**}$ | -1.228$^{*}$ | 0.810 | -0.615 | 0.607 | -2.319 | -4.265$^{**}$ |
| | (0.579) | (0.567) | (1.617) | (0.650) | (0.772) | (0.606) | (1.855) | (1.710) | (1.769) |
| ln*SCL* | 1.319$^{***}$ | 1.194$^{***}$ | 3.455$^{***}$ | 0.312 | -3.034$^{***}$ | -0.208 | -1.340 | 3.746$^{***}$ | -4.012$^{**}$ |
| | (0.324) | (0.337) | (0.976) | (0.248) | (0.751) | (0.317) | (1.848) | (1.016) | (1.972) |
| ln*AG* | 1.021$^{***}$ | 1.681$^{***}$ | 0.851 | 0.396$^{*}$ | 3.095$^{***}$ | 0.572$^{*}$ | -1.970 | -0.234 | 3.729* |
| | (0.338) | (0.352) | (1.012) | (0.220) | (0.818) | (0.299) | (2.832) | (1.017) | (2.070) |
| ln*HC* | 2.112 | 5.644$^{***}$ | 3.556 | -2.430 | 1.863 | 2.738$^{*}$ | -12.322$^{*}$ | 8.406 | -0.407 |
| | (1.638) | (1.660) | (4.825) | (1.502) | (1.361) | (1.609) | (6.776) | (5.284) | (2.509) |
| ln*DIST* | 1.035$^{***}$ | 1.445$^{***}$ | 3.652$^{***}$ | 0.040 | 1.606$^{***}$ | 0.270 | 2.986$^{**}$ | 2.973$^{***}$ | 1.366 |
| | (0.292) | (0.303) | (0.878) | (0.244) | (0.326) | (0.303) | (1.391) | (0.952) | (0.858) |
| *σ* | 1.422 | 1.495 | 4.274 | 0.608 | | 1.042 | | 3.345 | |
| Log Likelihood | -287.040 | -297.379 | -463.688 | -257.087 | | -256.614 | | -457.055 | |
| OBS. | 173 | 173 | 173 | 173 | | 173 | | 173 | |

Note: *, ** and *** indicate statistical significance at the 10%, 5% and 1%, respectively. Value in parenthesis is the standard error. Dependent variable is the measure of regional productivity distribution and the row labeled 'Percentile' indicates which measure is used.

**Table C.3: Parameter Estimates of the Self-Selection Model: Exporting Plants**

| | (1) | (2) | (3) | (4) | | (5) | | (6) | |
|---|---|---|---|---|---|---|---|---|---|
| Parameter | *α* | *α* | *α* | *α* | *γ* | *α* | *γ* | *α* | *γ* |
| Percentile | 10 | 50 | 90 | 10 | | 50 | | 90 | |
| Intercept | -36.410** | -61.264*** | -88.290 | -67.855*** | | -81.024*** | | -208.434*** | |
| | (15.619) | (15.912) | (90.850) | (13.774) | | (13.151) | | (48.840) | |
| ln*LOC* | 5.252*** | 8.123*** | 27.346*** | 3.208** | -0.743 | 9.127*** | -1.276** | 8.845*** | -1.186 |
| | (1.218) | (1.350) | (7.862) | (1.561) | (0.528) | (2.324) | (0.628) | (3.100) | (1.191) |
| ln*HHI* | -2.637 | -1.271 | -27.438** | -5.213** | -1.677 | -4.468 | -3.413*** | 3.933 | -17.822*** |
| | (1.934) | (1.990) | (11.835) | (2.348) | (1.098) | (2.825) | (1.363) | (6.317) | (4.537) |
| ln*SCL* | 3.963*** | 7.751*** | 5.368 | 3.450*** | -1.566*** | 8.489*** | -1.850** | 10.988*** | -11.826*** |
| | (1.027) | (1.130) | (6.201) | (0.712) | (0.385) | (0.733) | (0.854) | (2.841) | (3.942) |
| ln*AG* | -0.416 | -2.116 | 4.476 | 3.081*** | 2.231*** | -2.257** | 2.251** | 9.782*** | 10.076** |
| | (1.299) | (1.345) | (7.591) | (0.790) | (0.523) | (1.092) | (0.973) | (3.991) | (4.481) |
| ln*HC* | 3.404 | 12.232** | 10.663 | 3.378 | 2.129* | -5.499 | -1.674 | 6.094 | -23.094*** |
| | (5.424) | (5.949) | (32.436) | (5.167) | (1.304) | (5.589) | (1.363) | (14.443) | (7.846) |
| ln*DIST* | 1.813** | 4.240*** | 2.817 | 2.062*** | 0.905*** | 4.866*** | 0.269 | 9.221*** | 2.618*** |
| | (0.908) | (0.971) | (5.485) | (0.692) | (0.202) | (0.750) | (0.239) | (2.286) | (0.773) |
| ln*EXP* | 4.855*** | 4.861*** | 15.625** | 8.990*** | 0.060 | 5.901*** | 6.667*** | 10.346*** | 10.230** |
| | (1.905) | (0.914) | (6.604) | (2.315) | (1.119) | (1.601) | (1.919) | (1.896) | (4.455) |
| *σ* | 3.295 | 3.664 | 19.713 | 0.097 | | 1.154 | | 7.089 | |
| Log Likelihood | -269.634 | -281.270 | -446.486 | -250.969 | | -259.996 | | -377.632 | |
| OBS. | 160 | 160 | 160 | 160 | | 160 | | 160 | |

Note: *, ** and *** indicate statistical significance at the 10%, 5% and 1%, respectively. Value in parenthesis is the standard error. Dependent variable is the measure of regional productivity distribution and the row labeled 'Percentile' indicates which measure is used.

**Appendix D. Alternative Specification of the Plant-Level Production Function**

In the appendix, we test the absorptive capacity hypothesis against an alternative common specification. In most human-capital spillover studies, it is assumed that any plant, regardless of the skill intensity, can benefit from spillovers (e.g. Moretti, 2004a). Thus, the production function to be estimated under the alternative hypothesis can be written as:

(D.1) $$\ln Y_{irt} = \theta_0 + \theta_1 \ln L^s_{irt} + \theta_2 \ln L^u_{irt} + \theta_3 \ln M_{irt} + \theta_4 \ln E_{irt} + \theta_5 \ln K_{irt} + \theta_6 \ln G_{irt} + \theta_7 S_{irt} + \omega_{irt} + \varepsilon_{irt}.$$

Note that plant's skill intensity, $\eta_{irt-1}$, does not scale $S_{irt}$ in equation (D.1), i.e., the regional skill intensity can enhance every plant's productivity. Given plants' predetermined capital stocks, plants first realize their productivity level and then, choose the level of conventional inputs. Equation (6) in the chapter 4 is rewritten as follows:

(D.2) $$\omega_{irt} = \omega\left(\ln M_{irt}, \ln K_{irt}\right).$$

By substituting (D.2) into (D.1):

(D.3) $$\ln Y_{irt} = \theta_1 \ln L^s_{irt} + \theta_2 \ln L^u_{irt} + \theta_4 \ln E_{irt} + \theta_6 \ln G_{irt} + \theta_7 S_{irt} + \varphi\left(\ln M_{irt}, \ln K_{irt}\right) + \varepsilon_{irt},$$

where

(D.4) $$\varphi\left(\ln M_{irt}, \ln K_{irt}\right) = \theta_0 + \theta_3 \ln M_{irt} + \theta_5 \ln K_{irt} + \omega\left(\ln M_{irt}, \ln K_{irt}\right).$$

As before, we can consistently estimate equation (D.3) using OLS by approximating $\varphi(\cdot)$ with a third-order polynomial. Then, we carry out the J test (Davidson and MacKinnon, 1981) between equations (7) and (D.3). The results of the J test are shown in the table D.1 along with the coefficient estimate of $\theta_7$. We first consider the null hypothesis that equation (D.3) is preferred to equation (7), which is rejected for every industry. However, the null hypothesis that equation (7) is preferred to equation (D.3) is not rejected for seven industries. Although J test strongly supports equation (D.3), it should be noted that J test tends to over-reject when the number of explanatory variables which appear in the alternative hypothesis but do not appear in the null hypothesis is large (Davidson and MacKinnon, 2004). Ten variables in the alternative do not appear in the null in this case due to the different specifications of plants' productivity, $\omega_{irt}$ (See equations 6 and D.2).

Next, the parameter measuring human-capital spillovers is significant for only three industries under the specification (D.1) while the parameter is significant for five industries under the absorptive capacity hypothesis reported in table 4.2. In the table D.1, parameters for chemical and machinery industries are significant at one percent. These two industries are often regarded as high-tech (e.g. OECD, 2003), and hence, knowledge should be an important production factor for every plant in these industries. Hence, every plant appears to benefit from spillovers. On the contrary, our results in table 4.2 show that only high skill-intensive plants can benefit from spillovers even in industries considered low-tech, e.g. food processing and paper industries. The specification in equation (D.3) is a poor fit relative to that in equation (7) because it ignores plant heterogeneity.

**Table D.1: Alternative Specification and Test of Nonnested Models**

| | ISIC | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
| $S$ | 0.166 | 0.233 | 0.346 | 0.547 | 1.059*** | 0.133 | 0.623* | 0.334*** |
| | (0.107) | (0.238) | (0.185) | (0.366) | (0.195) | (0.193) | (0.382) | (0.127) |
| J test | | | | | | | | |
| $H_0 : S \quad H_1 : \eta S$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $H_0 : \eta S \quad H_1 : S$ | 0.155 | 0.518 | 0.253 | 0.927 | 0.000 | 0.801 | 0.407 | 0.143 |
| OBS. | 4504 | 1991 | 1318 | 923 | 1860 | 583 | 253 | 2489 |

Note: * and *** indicate statistical significance at the 10% and 1%, respectively. Value in parenthesis is the bootstrapped standard error based on 200 iterations. Values for J test correspond to p-value. Dependent variable is the log of output.