UNIVERSITÄT GREIFSWALD
Wissen lockt. Seit 1456

Agglomeration, Human Capital, and Productivity in Vietnam

Inaugural – Dissertation

zur

Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

der

Rechts- und Staatswissenschaftlichen Fakultät

der

Universität Greifswald

vorgelegt von

Cuong Hoang Manh

Greifswald, 2023

Dekan: Prof. Dr. Volker Beckmann

1. Gutachter: Prof. Dr. Susanne Soretz
2. Gutachter: Prof. Dr. Daniel Schiller
3. Gutachter: PD Dr. Michal Myck

Tag der Disputation: 25. Januar 2023

Table of Contents

Acknowledgement

This dissertation was completed with much help, support, and encouragement from a number of people. First and foremost, I would like to express my profound gratitude to my first supervisor Prof. Susanne Soretz. Her academic insights, infinite encouragement, and great patience helped me stay energized and overcome numerous difficulties facing me throughout my PhD journey. I would also like to sincerely thank my second supervisor Prof. Daniel Schiller. His insightful comments and excellent suggestions, along with advice of Prof. Susanne Soretz, were essential to finalize studies in this dissertation. Generally, I feel lucky to have them as my supervisors. Their help, support, and advice are invaluable to me during my PhD.

Next, I would like to thank my parents, my parents-in-law, my wife, my beloved daughters, and my extended family, whose infinite love and unconditional support gave me strength to finish this thesis. I am also grateful to my uncle Dr. Nguyen Van Ngu who encouraged me to do a PhD in Germany. I also thank Assoc. Prof. Dr. Phan Tran Trung Dung and Dr. Phan Thanh Hung for inspiring me to do empirical studies when I lived and worked in Hanoi.

My sincere gratitude is also expressed to: Vietnam's Ministry of Education and Training, especially the International Cooperation Department, who made my PhD in Germany possible in the first place; to the University of Labor and Social Affair for allowing me to follow a full-time PhD program in Germany; to the General Statistics Office of Vietnam who has provided me valuable data for my researches; to the World University Service for their help to connect me with my supervisors. I would also like to gratefully thank Ms. Tran Thi Nga and Ms. Pham Thi Phuong for all their support with administrative procedures.

Finally, I would like to thank all my colleagues at the chair of AVWL, especially growth, structural change, and trade, the University of Greifswald. Their support made my life and work in Greifswald easier and more enjoyable. I would also like to thank all my friends for their support and encouragement during my stay in Germany, and Georgios Tsiachtsiras, Dr. Doan Thi Thanh Ha, and Dr. Nguyen Thi Xuan Thu for their helpful academic advice.

Chapter 1

Introduction

1.1. Introduction

The uneven spatial distribution of economic activities across regions within a country or a continent has been studied widely since Thünen (1826) created a basis for location theories with a monocentric model. It can be seen through the fact that approximately 74% of the total population of the European Union in 2014 lived in cities, whose regions made up only 17% of the land area of the Union at the time (Nabielek K. et al., 2016). This spatial pattern repeats itself in every developed country, while it is also the direction that developing countries have been moving forward, which is reflected by a statistic from the World Bank that the urbanization rate of the whole world has risen steadily from 33.6% in 1960 to 56% in 2020 and its growth rate in upcoming decades will be contributed mainly by developing countries because the vast majority of people in developed countries are already living in metropolitan areas. Likewise, production inputs such as labor, physical capital, and human capital tend to be spatially concentrated in a small number of regions in any country with high development of industrial and service sectors.

From the economic perspective, at the national level, Gross Domestic Product (GDP) per capita and its growth rate both show a clear positive relationship with urbanization rate and human capital level. At the regional level, the close link of regional prosperity with agglomeration and human capital is also often found in any country in the world. Even at the individual level, highest-paid workers and most profitable companies tend to find themselves located in regions with high level of urbanization or large share of high-educated workers. Such relations between income and space across different levels of geography are not random. As a rationale to this phenomenon, scholars in literature suggest that spatial concentration can generate positive externalities, and thus have a positive impact on productivity of not only regional economies but also individual firms and workers in those regions. Theoretically, spatial characteristics through which these externalities appear to be active is not only one. When the external benefits emerge owing to the large employment scale of a region, it is called urbanization economies (Rosenthal & Strange, 2004). When the externalities come from industrial concentration, they are referred to as localization economies or Marshallian externalities, acknowledging the pioneering contribution of Marshall (1890, Book 4, Chapter 10). When industrial variety is the

origin of these external gains, it is attributable to Jacobs externalities, following the book of Jacobs (1969). Put together, these regional properties commonly found in agglomerated regions, whose external effects are called the agglomeration externalities as a result. Finally, when spatial clusters of human capital lead to this type of benefits, it is named human capital externalities (HCE).

Empirically, there have been a number of studies providing evidence of these externalities. Most authors are unanimous that the agglomeration externalities exist, but still debate on which aspect of agglomeration is the most important source of these external effects. Furthermore, since firms are different in their characteristics, their potential gains from agglomeration may be so as well, but available evidence of this possibility remains few and far between. Next, there is also evidence of the positive external impact of HCE in literature and the number of studies in favor of the presence of this effect seems to dominate the ones opposing it. Nonetheless, not all papers apply an instrumental or experimental approach to overcome likely econometric problems facing the estimation of HCE. Furthermore, the existence of HCE is not confirmed by a number of influential studies, and the variable of urbanization is almost always missing in the estimation of HCE in literature, although highly educated workers tend to flock to cities (Carlino & Kerr, 2015). As a consequence, the evidence of HCE is also disputable. It is even harder to give inclusive conclusions on all the matters just mentioned above for the developing world where studies on these externalities with solid estimation strategies are still scarce (Combes & Gobillon, 2015). Meanwhile, it is worthwhile to conduct this sort of studies for developing countries, especially for emerging countries, because their spatial structure is more dynamic than and systematically different from developed countries, which may lead to different estimates of the agglomeration externalities between the two worlds.

This dissertation is implemented aiming to overcome the limitations of the literature as analyzed above. First, it attempts to verify the existence of the external effects of agglomeration and human capital in Vietnam – a typical developing and emerging country, employing advanced econometric methods along with micro-level data to obtain consistent estimators of the externalities. Second, it contributes evidence to the debate about the dominant force of agglomeration. Third, it finds out which characteristics help firms absorb more external benefits from the regional environment. Lastly, given solid proof of the slowdown in economic growth after its peak in any emerging countries and developed countries, efforts to boost up productivity are all valuable, and therefore up-to-date knowledge on the possibility to do that via regional agglomeration externalities helps. By providing new evidence of the externalities

and its heterogeneity, this dissertation may assist policy makers in adjusting regional structures and making economic and enterprise policies.

1.2. Conceptual background and empirical state of the art

1.2.1. Theory of agglomeration externalities

In searching for determinants and outcomes of the non-homogeneous distribution of economic activities in space, most regional and urban economists have considered agglomeration externalities, or agglomeration economies in other words, as the heart of their models after the pioneering work of Marshall (1890, Book 4, Chapter 10) (Chipman, 1970). This type of economies is defined by McDonald & McMillen (2010, p. 8) as the cost benefits from production and transaction activities that are located within boundaries of a region. Put differently, the spatial proximity helps reduce transportation cost of goods and services as well as the "transportation" cost of labor and knowledge. It is worth pointing out that the agglomeration economies can also emerge in a form of internal effects (Hoover, 1948; Isard, 1956; Brakman et al., 2009, p. 38 – 39). This happens when a firm with large-scale production activities concentrates all its establishments and facilities in a place to reduce its fixed costs per output, consequently gains increasing returns to scale[1]. In other words, the internal economies result from the spatial concentration of a single firm rather than close proximity of various firms in a region, which is this thesis' primary research of interest. Therefore, the terms "externalities" and "economies" are used interchangeably henceforth.

Duranton & Puga (2004) and Puga (2010) provide their extensive reviews and theoretical foundations for "causes" of agglomeration externalities, which is initiated by the analyses of Marshall (1890, p. 225). They point out that agglomeration gives rise to the economies through three main microfoundation channels. The first one is called sharing effects. In a region with higher geographical concentration of businesses, local demand and supply for intermediate goods grows accordingly, which improves production scale and number of local providers, and thus results in a diversity of local intermediate inputs with more affordable prices. Abdel-Rahman & Fujita (1990) formulate an urban model in which final goods are traded freely across regions in a completely competitive market, while sector-specific intermediate goods are transacted within borders of each region under condition of monopolistic competition. In equilibrium, returns to input factors are increasing within each sector-city pair, despite the

[1] A firm has an increasing return to scale technology if a x times rise in its production inputs leads to a more than x times rise in its outputs, given x being a positive real number.

returns to scale being constant in production of final goods, indicating productivity advantages of sharing intermediates in urban regions. Another sort of shared input that can be found in an agglomerated region is local infrastructure and public amenities such as airports, seaports, roads, hospitals, libraries, and parks. Given the high fixed costs of their construction and operation, each local firm and resident will bear a lower average cost from using these shared facilities in a region with a higher level of agglomeration. However, this source of benefit is not universal because when an area becomes overpopulated, external diseconomies may emerge from free riders and congestion. The final shared regional input documented in literature is the labor force, whose large size in cities plays a role of an employment shock absorber, which helps firm to reduce cost in expanding or diminishing their workforce in response to business cycles or local shocks.

The second channel is called matching effects operating through the labor force of agglomerated regions, which tends to be more skill-diverse and have a higher proportion of high-educated workers compared to sparse regions. Since each firm has its own demand for the skill composition of its workforce depending on its operating sector and technological level, locating in cities raises the firm's probability of finding workers that meet its needs in both quantity and quality. Similarly, when a region is more agglomerated, it more likely takes local firms less time and costs to discover local input suppliers that fulfill their production requirements. The third channel is learning effects. The spatial proximity in urban regions brings entrepreneurs and workers more chances to have face-to-face exchanges with their peers through which knowledge, skills, and experiences are transmitted (Marshall, 1890, p. 225) and new ideas are produced (Jacobs, 1969, p. 43 – 50), whose values add to the benefits of agglomeration. Considering urban regions as the cradle of innovation and invention owing to their diverse-knowledge environments is the core of Jacobs (1969, Chapter 2)'s reasoning for the economic importance of urbanization. Glaeser (1999) attempts to formalize the spillover theory of Marshall (1890, p. 225) by building a theoretical framework in which young and inexperienced workers move to cities to develop their skills through learning from their older and experienced peers.

However, agglomeration does not always generate positive externalities since it can bring negative effects as well when a region becomes more spatially concentrated than a certain level threshold (Capello, 2015, p. 20). In that case, local prices of immobile production inputs such as housing are getting too expensive, in addition to a more intense emergence of traffic congestion and urban pollution. These do harm to the performance of local firms and workers

and create forces pushing them away from settling down in an agglomerated area. These dispersion forces are also induced by transportation cost, though this type of cost gives rise to the concentration forces as well, as indicated by the definition of McDonald & McMillen (2010, p. 8) on agglomeration. Interaction between the two opposite forces underlies spatial equilibrium in regional and urban theories that attempt to explain spatial structure of regions and cities, whose models are rooted in location theories developed by German economists since the 19th century. Specifically, Thünen (1826) is the first to employ this framework with a spatial model in whose equilibrium all economic transactions are active in a single city center, and farmers with the highest productivity locate nearest the center and pay the highest land rents as well, while less productive famers operate father away and pay less rents. More than a century later, Alonso (1964), Mills (1967), and Muth (1969) produced benchmark models for later urban economists by putting the theory of Thünen (1826) into an urban setting to explain distribution of city sizes over space (Brakman et al., 2009, p. 36). Following the Alonso–Mills–Muth models, studies of Henderson (1974, 1977, 1988) characterize modern urban economics in a sense that city scale is driven by external economies and diseconomies, whose mechanisms are expressed above, and the transportation costs and non-urban regions are excluded from the analysis.

In a different direction, the model of Weber (1909) was among the first ones dealing with spatial distribution of industries rather than purely regions. Weber (1909) continues the framework of the concentration-dispersion interaction employed by Thünen (1826) but treats transportation costs as the direct dispersion force in an isodapane analysis to intuitively explain the location choice of manufacturing firms based on their cost-minimizing calculation and comparison with benefits from concentration. In a follow-up study, Christaller (1933) constructs a model with a honeycomb-shaped long-run spatial equilibrium that differentiates hierarchies of urban and non-urban regions with various functions, using geographical structure in Southern Germany as the evidence. Aiming to obtain a more general spatial equilibrium, Lösch (1940) rotates hexagonal networks of different goods around a common center by which industry-diverse and specialized cities are identified. The model of Lösch (1940) is proved using the spatial structure of the state of Iowa in the US. In the models of Christaller (1933) and Lösch (1940), goods and services that can reap higher benefits from internal economies of scale are produced in centers of larger hexagonal market areas. These location theories were then translated and synthesized by Isard (1956) as the base for regional economics, whose focus has been shifted towards explaining spatial disparities in regional growth and development

since then (Brakman et al., 2009, p. 46). Meanwhile, as Fujita et al. (1999, p. 33 – 34) conclude, subsequent regional studies based on the location central theory of Christaller (1933) and Lösch (1940) had been remaining ad hoc with loose theoretical foundations. To overcome the weaknesses of this theory but retain its quintessence of core-periphery regional structure, Krugman (1991) combines the traditional regional and urban models with new international trade theory in a multi-sector and multi-location framework. The dependence of agglomeration benefits on exogenous "iceberg-type" shipping costs of goods across regions is the basis for multiple equilibria, which can be stable or unstable, and show regional spatial structure of either core-periphery or equated depending on this type of cost. A further source of agglomeration economies embodied in Krugman (1991) is the so-called "home market effects" which describes profit-maximizing location choices of manufacturing firms in core regions to meet regional enormous demand for consumption with minimized transportation costs, while they serve peripheral markets from farther distances. The model of Krugman (1991) is considered as the foundation of new geographical economics (NGE), which has been evolving for the last decades from adding missing elements in the theory of Krugman (1991) to make it more complete.

Given the importance of agglomeration externalities in literature, the key interest of this dissertation is to find empirical evidence of these external effects. To do that, the thesis follows theories that stem from modern urban economics in which the dispersion force is caused by congestion or high price of immobile production factors rather than transportation costs. In other words, despite the innovative approach of NGE with the central role of transportation costs, it is beyond theoretical and empirical analyses in dissertation due to following reasons. First, the affordability of transportation costs nowadays has made this factor a less crucial role in determining agglomeration in comparison with a century or two ago (Glaeser, 2010, p. 7). This argument is supported theoretically by the tomahawk-shaped diagram 4.3a in Brakman et al. (2009, p. 143) which illustrates the spatial equilibria in Krugman (1991). Specifically, when transportation costs are low, agglomeration is predominant over spreading forces. Second, the NGE framework is relevant to regional analyses at larger geographical units (i.e., regions, countries), provided the foremost role of regional demand for the "home market" and cross-region interactions, while urban models are suitable for finer geographical units (i.e., districts, cities)[2] where agglomeration externalities and within-region transactions are more crucial

[2] Despite meaning probably larger geographical scales, region is used interchangeably with local, district, and city in theoretical analyses in this dissertation.

(Brakman et al., 2009, p. 288 – 289). This thesis conducts empirical analyses using characteristics of district-level regions, therefore the application of urban theories is more reasonable. Nevertheless, not only cities but also non-cities regions are included in this thesis's empirical calculation, consequently the resulting results also match the spirit of regional economics in general and central place theory in specific in a sense that all types of regions within a country are taken into account.

1.2.2. Theory of human capital externalities

Agglomeration is formulated commonly in literature as spatial concentration of a certain production input such as labor (e.g., Glaeser, 1992), physical capital (e.g., Ott and Soretz, 2010), and human capital (e.g., Rach, 1993; Acemoglu & Angrist, 2000; Moretti, 2004a; Gennaioli et al., 2013), with labor as the most widely used factor. When human capital is the case, its external effects are known as human capital externalities (HCE), which is another key research of interest in this thesis. Although the theory of HCE and mechanism behind it are broadly similar to agglomeration economies', the former arises from particular attributes of human capital, thus should be provided with a separate brief theoretical analysis. It should be noticed that human capital is used interchangeably with education or skill throughout this thesis, though it may mean health as well. According to Moretti (2004c), HCE can be viewed as the disparity between private and social benefits of human capital. Empirical evidence in literature shows that social return to education may be higher or lower than the private return, thus HCE can be either positive or negative. The positive influence is probably generated through the mechanism of the learning effects. More specifically, regions with higher level of human capital bring more chances of in-person meetings of workers with their high-skilled peers through which advanced skills and knowledge are shared, which improves productivity of local workers and firms as a result. This is the spirit of Lucas (1988) who is the pioneer in formulating human capital at both micro and macro levels in a single production function, and treats the latter term as positive externalities embodied in a technological term, leading to the increasing returns to scale.

The second source of positive HCE is the matching effects, which means that firms save more costs of searching for high-skilled workers if this local workforce is larger, holding other factors fixed. Alternatively, Moretti (2004c) explains productivity advantages created by a larger pool of highly educated workers using the model of Acemoglu (1996) in which physical capital and human capital are complementary to produce goods. If a group of local workers becomes more educated, local firms would raise their volume of physical capital accordingly,

expecting to hire those workers to maximize their profit. However, due to costly job recruitment, firms may end up employing workers whose education has not been improved yet, consequently equip these workers with higher amounts of invested physical capital, and thus improve their productivity compared to their peers in other regions. Interestingly, despite returns to scale of production function being constant in Acemoglu (1996), the complementarity between physical capital and human capital gives rise to HCE. This is also the outcome in the model of Acemoglu (1998) which underlines the complementarity of technological level embodied in physical capital and skills as a key factor to prove that an acceleration in share of highly educated workers in the US in the 1970s had an causal positive impact on productivity of this worker group in the 1990s.

Local human capital may also generate positive external effects owing to indirect channels, which can be seen from strong association of higher education with low crime and more votes for better political institutions. More specifically, higher schooling can lower local crime, which is a type of negative externalities, through raising opportunity costs of criminal activities due to resulting higher wage premium to legal jobs, and through making people more risk averse. Meanwhile, more-educated citizens tend to participate more actively in political outcomes owing to their higher turn-out rates, more demand for their authority, and richer knowledge about candidates in local and national elections, which leads to better political choices as a consequence. In other words, the presence of more highly-educated citizens comes with better security and more capable elected authority, thus indirectly enhancing productivity of local residents and businesses. Regarding the case of negative HCE, Moretti (2004c) mentions a possible cause related to education time. Specifically, an increase in mean schooling may imply that people with greater innate learning abilities tend to spend more time upgrading their education and academic degrees, thus joining the regional labor force later than their average peers. The late entry to the labor market of this group of productive workers can make social return to schooling smaller than the private return.

1.2.3. The evolution of empirical studies on agglomeration and human capital externalities

The key purpose of this dissertation is to verify the existence of agglomeration and human capital externalities, therefore it is worth giving a short overview of empirical directions that scholars have been following for the last several decades in testing and measuring the

externalities[3]. These effects are often formulated and estimated directly as a Hicks-neutral technology shifter (Hicks, 1963, p. 120) in a production function (e.g., Henderson, 1986; Lucas, 1988; Ciccone & Hall, 1996; Black & Henderson, 1999; Henderson, 2003; and Moretti, 2004b), or indirectly as a determinant of marginal product of labor in a wage regression (e.g., Rach, 1993; Moretti, 2004a; and Combes et al., 2010). These are the two approaches most employed in literature to estimate the regional externalities owing to wide usage of data on wage and production and their applicability to any cluster levels, from individuals to regions, even countries. Other indirect dependent factors also used to capture agglomeration and human capital economies are local employment, outputs, firm births, rents, and local growth in those values as surveyed in Rosenthal & Strange (2004) and Combes & Gobillon (2015). Regarding data, early empirical studies to test the spatial externalities have mainly conducted using aggregate data sets, such as Sveikauskas (1975), Nakamura (1985), Henderson (1986), Rach (1993), and Ciccone & Hall (1996), whose data was all recorded at a region level. Later, Glaeser & Maré (2001) and Henderson (2003) were the first studies using worker-level and plant-level data respectively to estimate scale externalities. Since then, the use of individual-level data in the strand of literature has become more preferable among regional and urban economists (e.g., Moretti, 2004a; Moretti, 2004b; Combes et al., 2008; Combes et al., 2010; Martin et al., 2011; Bratti & Leombruni, 2014; and Duranton, 2016), due to the more availability of this sort of data and the fact that its econometric application has large advantages in simultaneously tackling endogenous issues at both micro and region levels as well as separating the external effects from micro-level effects to obtain estimation results of the externalities with higher precision, which is beyond the capabilities of aggregate data.

Since available data sets in many countries in the world have been improving over time to cover finer levels of geography, more economic sectors, more years of sampling, and more information on characteristics of firms and their directors and workers. Similarly, empirical analyses have been expanding to deal with more possibilities of heterogeneity in the local external impacts. This research direction was first and earliest touched in debates on which type of agglomeration externalities is most important, known as "industrial scope" in Rosenthal & Strange (2004). Specifically, agglomeration is decomposed into localization and urbanization, following arguments of Marshall (1890, Book 4, Chapter 10) and Jacobs (1969, p. 43) respectively. Rosenthal & Strange (2004) takes an example of specialization in

[3] The term "externalities" used henceforth in this section may mean the external effects of either agglomeration, human capital, or both, given the close similarity in their theories and estimation strategies.

technology of Silicon Valley in the US as the typical evidence of economies resulting from high intensity of within-sector economic transactions as mentioned in Marshall (1890, p. 225). Jacobs (1969, p. 43 – 48), meanwhile, eyes to the development of New York, where strong growth of brassiere industry attributes to the innovations of tailors rather than the textile industry, to illustrate productivity spillover across various sectors in urban regions. This idea, according to Rosenthal & Strange (2004), indicates the relative measure of urbanization, while regional population or employment represents the absolute measure, whose productivity impact is known as urbanization economies. The argument of Jacobs (1969, p. 43) has sparked off an intense debate that, say, a baker becomes innovative owing to his close proximity to his local colleagues in bakeries or his neighbors and friends who do not work with cakes or cake making machines.

Theoretically, this is the dispute between regional economists over which one between location economies and urbanization economies, or between specialization and diversity, plays a dominant role in agglomeration economies. The external effects from specialization and diversity operate through the same channels as mentioned in section 1.1 with the primary difference being that the former comes into effect through intra-sector interactions, while the latter does it through between-sector economic activities. Until now, the dispute on the two connotations of agglomeration has not yet come to an end due to much heterogeneity among empirical tests. Digging into the sectoral aspect of the externalities but following another direction, a number of studies have made efforts to unfold the heterogeneous effects of local agglomeration and human capital between various economic sectors. For example, when comparing the manufacturing sector to its service counterpart, Blien et al. (2004) find that the former mainly benefits from localization, while urbanization is more crucial to the latter. Meanwhile, Melo et al. (2009) point out that the former gains more from agglomeration benefits compared to the latter. Estimating agglomeration economies for many detailed sectors, Brülhart & Mathys (2008) and Foster & Stehrer (2009) report the positive externalities for all except for the agricultural sector. In terms of technological level, high-tech sectors are considered to be more beneficial in a region with high level of agglomeration or education due to its ability to reap more benefits from the learning and matching effects in that region compared to low-tech sectors (Moretti, 2004b; Ehrl, 2013; and Liu, 2014).

The second type of heterogeneity is called "geographical scope" by Rosenthal & Strange (2004), which means that the intensity of externalities changes with various spatial distances to centers of agglomeration. In fact, papers of Rosenthal & Strange (2003), Fu (2007), and

Rosenthal & Strange (2008) show the attenuation of external effects of agglomeration and human capital over distance by comparing their estimated magnitudes between concentric spatial rings of various radii up to 100 miles. Desmet & Fafchamps (2005), however, find that strength of the externalities engages with the distances from centers of spatial rings following an inverted-U shaped relationship. Another way to assess the spatial scale of externalities is using regional variables measured at various geographical units, estimating their productivity impact separately, then making comparison between obtained estimators, as in Rosenthal & Strange (2001) and Martin et al. (2011). The rationale is that the strength of the externalities' mechanisms may vary across various geographical scales. For example, skills tend to be shared and acquired in a much finer spatial unit in comparison with the match of employees and employers in the labor market. Agglomeration economies as a sum of three mechanisms should show different influences between various geographical scales. Alternatively, a number of authors have chosen spatial lags models and make use of spatial econometrics to allow for the case that the externalities operate not only within a region but also spread to neighboring ones, or even to distant ones, with the influences being diminishing over distances. In other words, to measure the local economies accurately, productivity spillovers from close regions should be accounted for. However, according to Combes & Gobillon (2015), a major weakness of this method is that it is most efficient when combined with non-parametric assumption, but this is not the case of the local agglomeration and human capital estimation. Besides, this approach does anything but deal with key identification issues facing the estimation of the externalities.

The third case of heterogeneity is that the external impacts may vary across different temporal distances, whose topic was first discussed in detail by Glaeser & Maré (2001). So far, only the static effects of the externalities have been analyzed in this thesis, which is in fact taken most commonly in literature. Despite this, the impact on productivity may be either instantaneous or time-lagged. To identify the dynamic influences, previous empirical studies often choose growth in local employment, local income, or individual wage, especially over a long period of time such as several decades, as the explained factor, and the initial agglomeration or local education as explanatory variables. In the growth model of Lucas (1988), for instance, HCE acts as a key driver of long-run economic growth. Efforts to estimate the dynamic impact were made by, for example, Glaeser et al. (1992), Wheeler (2006), and Yankow (2006). To reconcile the static effects with the dynamic ones, Roca et al. (2017) even combine both types in a single econometric model in explaining regional income inequality.

Next, gains from local agglomeration and human capital may be heterogeneous across local regions of different sizes and countries of different stages of development. As discussed earlier, the negative externalities can exist, therefore agglomeration may have a bell-shaped relationship with productivity. In other words, smaller regions on the left-hand side of the bell could receive economies to scale, while bigger ones on the right-hand side suffer from the diseconomies. This non-linear pattern is not universal because it is found empirically significant in Au & Henderson (2006) and Graham (2007) but insignificant in Martin et al. (2011) and Duranton (2016). As regards countries' characteristics, Combes & Gobillon (2015) find that agglomeration economies tend to be stronger in developing countries compared to their developed counterparts, which seems to reflect a form of "catch-up effects". So far, there have not been any theories constructed to explain this phenomenon, bearing in mind that the consistent evidence for the developing world is still scarce (Combes & Gobillon, 2015).

The fifth sort of heterogeneity is that the externalities may vary across different groups of firm and individual. Beside own operating sector, characteristics of a firm that can influence its ability to gain benefits from the externalities of local agglomeration are, for example, age (e.g., McCann & Folta, 2011), size (e.g., Henderson, 2003; Martin et al., 2011; Combes et al., 2012), ownership type (e.g., Rigby & Brown, 2015). For the case of an individual, influential characteristics include skills, productivity level, education (e.g., Gould, 2007; Matano & Naticchioni, 2012; and Bratti & Leombruni, 2014), experience, age, gender (e.g., Phimister, 2005), immigration background, and race (e.g., Ananat et al., 2013).

Finally, the productivity impacts of the externalities can be found through their channels, as expressed in the first subsection. For instance, Audretsch & Feldman (1996) sought to measure the importance of each channel through observing concentration patterns of industries whose development is considered to depend heavily on the corresponding channel. Dumais et al. (1997), Rosenthal & Strange (2001), and Rigby & Essletzbichler (2002) follow this indirect approach to measure strength of the channels with industry-related indices, whose interpretations are doubted by Duranton & Puga (2004) and Rosenthal & Strange (2004) since they all are strongly correlated to the same variable: economic scale of a region. In another direction, Papageorgiou (2013) adopts a calibration strategy, which is still rarely used in regional studies (Combes & Gobillon, 2015), to estimate and confirm the contribution of the matching effects to the externalities. Overall, though microfoundation sources of agglomeration are vital to theoretical analyses on the externalities, empirical studies for these

sources remain immature and less informative because these are often unobservable and their influences are overlapped in practice (Rosenthal & Strange, 2004).

1.3. Structure of this dissertation

This thesis follows and supports the evolution of the strand of literature in various aspects. Firstly, given the benefit of micro-level data, this thesis also makes use of panel data on firm-level production inputs and outputs to estimate the external effects, which is formulated as technological shifters in a production function. As a result, a firm's productivity is employed as the primary explained variable in regression models. Secondly, this thesis deals with the industrial scope provided that evidence of the predominant externalities has not been inclusive thus far. Furthermore, the heterogeneity of the externalities in different sectors is also studied. Thirdly, addressing geographical and temporal scopes and estimating the contribution of each channel to economies are beyond this thesis's interest due to its limited scope and limitation of the data sets. Instead, knowledge on the decay of the externalities over distance is applied to discover a geographical unit that is the best to capture all three microfoundation sources. Fourthly, the thesis considers the presence of the diseconomies by testing a non-linear relationship between the external terms and productivity. Fifthly, this thesis pins interpretations of its results primarily for the developing world owing to the application of data from Vietnam. Finally, this thesis adds evidence of heterogeneous productivity gains from agglomeration.

The rest of this dissertation is structured in detail as follows. Chapter 2 solves the industrial scope and shows evidence of heterogeneous agglomeration effects across different characteristics of firms. To achieve this, a six-year panel data set collected from Vietnam is employed, and the estimation is based on a production function that its left-hand side is firm-level total factor productivity (TFP), while its right-hand side is local technology which contains the agglomeration terms. The regression process is split into two separate steps to tackle different potential econometric problems. In the first step, consistent values of TFP are obtained following a strategy that combines the control function approach with the instrumental variables (IV) technique to deal with the problem that production inputs change simultaneously with productivity. In the second step, TFP is regressed on agglomeration proxies and firm- and region-level controls, using multi-level fixed-effects (FE) to control for unobserved constant factors and industrial and local shocks. Obtained results point to urbanization as the most influential source among agglomeration forces. In addition, the agglomeration effects are found to be stronger for foreign-owned, small-sized, or young firms.

Next, chapter 3 aims to find the evidence of HCE applying an urban model in which entrepreneurs and workers migrate across regions at the cost of their human capital until a spatial equilibrium is reached. In equilibrium, they work in firms whose local technology contains the externalities of human capital and urbanization. To estimate, firm-level labor productivity is regressed on the external terms together with firm-level controls which include the firm and characteristics of its entrepreneur and workers, using a two-year panel data set from Vietnam. Identification concerns are dealt with using the two-stage least square (TSLS) method along with FE terms. The study also accounts for the heterogeneous gains from the externalities across industries of different technological levels. Again, the estimates show the predominant influence of urbanization, while the estimator of HCE switches from statistically significant to insignificant when urbanization is included in specification. Despite this, strong evidence of HCE is found in high-tech industries, while it does not for low-tech industries. In other words, the heterogeneous effects of local human capital among industries of different technological levels explains the ambiguous role of HCE in the whole economy. Finally, chapter 4 concludes, discusses, and suggests regional and education policies.

Chapter 2

The heterogeneous influences of agglomeration on total factor productivity in Vietnam

Although the heterogeneity in the productivity impacts of agglomeration for developed countries has been well documented so far, the study is still rarely carried out for emerging and developing countries. This chapter aims at verifying the existence of agglomeration externalities in Vietnam – a typical emerging country. Firstly, total factor productivity of each individual firm is measured consistently using the control function method in combination with the instrumental variable approach. These obtained values are then regressed on main variables of interest, which represent various facets of agglomeration, controlling for time-variant firm and regional characteristics. Next, making use of the multi-level fixed effects technique and detailed information provided by micro-level data attenuates concerns about self-selection and endogeneity in estimating the externalities. Overall, findings show the productivity-enhancing influences of employment density and industrial diversity but no significant evidence of productivity gains from regional specialization. Digging into various firms' characteristics, the most advantaged firms in highly agglomerated regions are demonstrated to be foreign-owned, small-sized, or young. Finally, several sensitivity checks show that the findings are robust across different productivity measures, industrial levels, samples, functional forms, and statistical techniques.

2.1. Introduction

The underlying theory describing the advantages of a geographically concentrated region, known as agglomeration externalities, has a long history of development in literature. It has been analyzed under three different forms related to industrial connections found in a region: intra-industry, inter-industry, and mixed-industry. The first form refers to Marshallian externalities, after Marshall (1890, Book 4, Chapter 10)'s original text. Marshall (1890, p. 225) points out that, unlike a dispersed industry, a clustered one can gain industry-specific economies of scale and boost regional growth because in such a condition, skill-specific labor force is larger, intermediate inputs market is more developed, costly machines are used more efficiently, and industrial knowledge is transferred more easily to stimulate innovation. This theory was later named Marshall-Arrow-Romer (MAR) externalities by Glaeser et al. (1992) who acknowledged the contributions of Arrow (1962) and Romer (1990) in formalizing

specific-sector spillovers of knowledge. The Marshallian external effects are also referred to as localization or specialization externalities (Rosenthal & Strange, 2004). The second form is based on the reasoning of Jacobs (1969, Chapter 2), hence the so-called Jacobian externalities. The key argument is that cities foster innovation owing to their advantageous environment for the birth of new ideas, new products, new types of work, and new industries. Finally, put the industrial structure of a region aside, the third considers employment size or density of a region to be the main source of external economies (Rosenthal & Strange, 2001; Rosenthal & Strange, 2004). They call these effects urbanization economies. As analyzed by Duranton & Puga (2004) and Puga (2010), these facets of agglomeration give rise to economies through three channels called the sharing effects, the learning effects, and the matching effects.

Up to now, a number of scholars have conducted empirical tests on these regional externalities. However, the discussion on which type of externalities dominates remains inclusive. Some authors such as Henderson (2003) and Martin et al. (2011) find evidence in favor of the Marshallian economies while others such as Glaeser et al. (1992), Ciccone & Hall (1996), and Combes et al. (2010) support the economies of urbanization and/ or industrial variety. It is even harder to draw an empirical conclusion for emerging and developing countries where such studies are still rare, especially the ones using micro-level data to overcome the inherent weakness of aggregate data on heterogeneous issues (Combes & Gobillon, 2015; Duranton, 2016). Key rationales behind this dispute according to Melo et al. (2009) are heterogeneities in industrial coverage and country origin of data employed for the estimation. In particular, there are significant differences in spatial pattern and its dynamic trend between highly developed countries and their low- or middle-income counterparts. For instance, developing countries have become more urbanized quickly over recent decades, but their urbanization rates remain low. In addition, the disparity between big and small cities is much larger for fast-growing developing countries where their development tends to be strongly associated with their integration to the global trade, leading to the economic and spatial dominance of a few megacities within each of those territories (Henderson et al., 2001; Duranton, 2008; Duranton, 2016; Razvadauskas, 2019). Razvadauskas (2019) points out that 26 out of in total 33 megacities in the world belong to developing countries in 2017 and more such cities will emerge in this country group in the next decade given the current population growth of other large cities. In summary, one should not assume that firms operating in countries of different development stages benefit similarly from agglomeration, and therefore it is well worth looking

for further evidence to this continuing debate on the dominant external economies, particularly in the emerging world.

Turning to the heterogeneous external gains corresponding to various characteristics of firms, a number of scholars have concentrated primarily on this possibility with suitable data at hand such as McCann & Folta (2011) and Rigby & Brown (2015) for the US and Canada respectively. Both studies verify that the strength and magnitude of the external effects vary significantly across firms of different characteristics. It is worth noting again that the inferences may show a different picture for emerging countries where this specific sort of study is still rarely implemented, and when it is, it focuses mainly on firm ownerships. The key reason for the contrast is that many attributes of these economies deviate considerably from those of advanced countries. In terms of ownership type, for example, foreign-owned enterprises (FOEs) operating in developing countries typically show their technological superiority compared to their native fellows (Tan & Meyer, 2011; Newman et al., 2015). This technology distance perhaps is much narrower in developed territories where many state-of-the-art innovations are introduced by their domestic-owned enterprises (DOEs). Turning to employment scale, the vast majority of firms in developing countries are small-sized and these firms, compared to developed countries, account for a higher share but are involved in smaller-scale production activities (Poschke, 2018). As regards age, long-lived firms are less present in emerging countries since economic booms for developed countries happened much longer ago. These systematic differences between firms of various types and countries of various development stages are suggestive of non-homogenous impacts of agglomeration. Therefore, providing evidence of this heterogeneity is another focus of this chapter.

In literature, Martin et al. (2011), Howard et al. (2014), and Gokan et al. (2019) conduct empirical tests which are most close to research questions solved in this chapter. The key contribution of the chapter are digging into aspects that are either neglected or incompletely analyzed in the three studies. First, the chapter deals with the question of how agglomeration induces unequal influences across many different firm characteristics, while Martin et al. (2011) examine only firm size, and Howard et al. (2014) and Gokan et al. (2019) mention only firm ownership. Second, the chapter shows efforts to yield a more consistent measure of TFP as the primary dependent variable, which assures the higher accuracy in identifying productivity impacts. Third, main hypotheses are formed on the basis of a wide range of contrasts and similarities between developed and developing countries in terms of agglomeration. Fourth, in the most recent study under the context of Vietnam, Gokan et al.

(2019) base their calculation entirely on cross-sectional data, therefore they capture the effects of the externalities accumulated over a long period of time. In other words, their estimates reflect a long-run impact of agglomeration. The key empirical calculation implemented in this study is different in a sense that it makes use of within-firm variation for 6 year – a short period of time, thus shows the short-run effects of agglomeration. Since the strength of the externalities may vary across time and different development states of countries, their long-run estimates are not necessarily the same as their short-run values.

This dissertation regards Vietnam as an ideal study case for the agglomeration analysis given a remark of Combes & Gobillon (2015) that empirical evidence for emerging and developing countries remains scarce. Meanwhile, Vietnam falls into this group of economies due to the fact that the country has been currently classified as a low-middle income country and has recorded a high average growth rate of about 6.4% over the last two decades (World Bank, 2019). Besides, similar to many other developing countries, the regional dynamics in Vietnam provide much variation for regional analyses. In particular, Vietnam has undergone a rapid urbanization phase after 1986 when the Doi Moi reform came into effects aiming to transform the country into a market-oriented and open economy. Specifically, urban population in Vietnam on average grew by 3.4% annually between 1986 and 2010 and 3% between 2011 and 2017, according to the World Bank (2011), OECD (2018), and the General Statistical Office (GSO). These numbers are higher than the figures of 2.5% and 0.88% found in Southeast Asian countries and advanced OECD members respectively. However, the urbanization rate of Vietnam reached merely 37.5% in 2017, which was much lower in comparison with the average level of about 81.5% in high-income countries. Still, 70% of Gross Domestic Product (GDP) of Vietnam was generated in cities in 2017, implying an enormous contribution of these urban regions in the economy, despite the low level of urbanization. This fact contrasts Vietnam, as an emerging country, to high-income countries, thus potentially reveals a divergence in the estimation results of the externalities. Another important reason to consider the context of Vietnam is the availability of micro-level data gathered by GSO, which contains information on employment, industry, and location of all firms in the country. Adding the time dimension to firm-level data opens up more possibilities to tackle econometric issues compared to the application of aggregate data mentioned in literature. Finally, the Vietnamese government expects to boost the national economic growth in coming decades through the development of cities. In particular, urbanization is encouraged strongly all over the country, whilst sparse production activities is discouraged. Therefore, understanding whether and how

heterogeneous this orientation enhances TFP of enterprises is important for designing urban policies that are capable of maximizing the positive impacts of the spatial externalities.

In this chapter, the external effects are estimated by regressing firm-level TFP on various variables of agglomeration and controls at firm and regional levels. Firm-level data was collected from the Annual Census of Enterprises between 2011 and 2016, provided by GSO in Vietnam. The explained variable, TFP, is calculated consistently for each manufacturing firm using Woodridge (2009)'s instrument approach on the basis of Levinsohn & Petrin (2003)'s framework. To prepare for the estimation, agglomeration proxies and other regional indices are computed based on district codes, 3-digit Vietnam Standard Industrial Code (VSIC) codes, and employment information available for all firms with business tax codes in Vietnam over the period 2011 – 2016. Main economic concerns, including time-variant unobserved missing variables, industrial and regional shocks, and potential self-selection are relieved by the application of multi-level fixed effects along with the short panel data. The results verify the existence of urbanization economies and diversity externalities in Vietnam. In addition, the heterogeneous effects of agglomeration are verified by adding product terms between each firm characteristics and connotations of agglomeration. These terms show that FOEs, small-sized, and young firms benefit the most from agglomeration.

The remainder of the chapter continues as follows. Section 2.2 provides literature background, section 2.3 explains the methodology of TFP estimation, and section 2.4 presents data used in the article, construction of main regional variables, and summary statistics. Next, section 2.5 discusses specification and related econometric issues, section 2.6 expresses estimated outputs, and section 2.7 reports findings from robustness checks. Finally, section 2.8 presents conclusions.

2.2. Literature review

Rosenthal & Strange (2004), Melo et al. (2009), and Combes & Gobillon (2015) provide intensive reviews of studies for developed countries with evidence of agglomeration. Combes & Gobillon (2015) show that various measures of agglomeration have been applied in the strand of literature. In particular, regional employment, population, or their density indices are often used to capture urbanization economies, while specialization is measured with either an absolute index such as industrial employment or with a relative index such as location quotient or employment share of the industry in the region. In terms of Jacobs externalities, it is often represented by a measure of industrial diversity which is constructed from employment shares

of local industries. Regardless of which measures are used to proxy agglomeration, Rosenthal & Strange (2004), Melo et al. (2009), and Combes & Gobillon (2015) find that authors in literature tend to reach conclusions favoring only one or two of the three above external effects. For more specific analyses, a number of influential studies are reviewed as follows.

In order to identify the source of knowledge spillovers in urban regions, Glaeser et al. (1992) make use of city-industry two-year data for the US and find that the employment scale of an industry expands more slowly in regions where the industry is more localized. Glaeser et al. (1992) come to a conclusion that a higher intensity of inter- rather than intra- industry interactions stimulates employment growth. Ciccone & Hall (1996) and Ciccone (2002) focus on the economies generated by county-level density in explaining the variance of labor productivity across states in the US and Europe respectively. They apply historical instruments to cross-sectional data and confirm the presence of urbanization economies on both sides of the Atlantic Ocean. In their studies, however, possible influences of Marshallian and Jacobs externalities are neglected. Henderson (2003) starts a trend in taking advantage of within-variation of observations using micro-level data to address the problem of unobserved heterogeneity in estimating local externalities. Henderson (2003) quantifies the impacts of agglomeration on productivity in the US for the period 1972 – 1992. Considering plant fixed-effects as the most reliable estimation technique among various applicable econometric methods, Henderson (2003) yields statistically significant evidence of within-industry externalities in only the high-tech sector, while the presence of urbanization economies are found in the machinery industry. Combes et al. (2010) explore French plant-level data, instrument for employment density with ecological and historical data, but treat localization as a control variable. Whether wages or productivity are defined as the explained factors, they find solid evidence of density economies. Martin et al. (2011) continue the use of French data but rely heavily on the first-difference generalized method of moments (FD-GMM) to tackle econometric concerns raised in the estimation of agglomeration. With the presence of both urbanization and localization proxies in their model as the main explanatory factors for plant-level productivity, they support only the positive externalities of localization. All in all, although there exists in literature evidence of the agglomeration externalities activated through both inter- and intra- industry relations, there has been no clear conclusion on which one is predominant over the other so far. Comparing results obtained in numerous previous articles, Combes & Gobillon (2015) comment that when a study verifies the existence of both

urbanization and localization economies, the latter tends to show a smaller role, while the evidence of Jacobs externalities is less robust across different studies.

Among the earliest scholars who tested theories of both Jacobs and Marshall for the developing world, Henderson et al. (2001) show the evidence for South Korea during its time as an emerging country. They apply the fixed effects method to a panel aggregate data set of 23 industries between 1983 and 1993 whose period provides a dynamic context for the regional analysis since there emerged a wave of production relocation from Seoul to other cities during this time. Regressing labor productivity on different external terms, they obtain evidence of industry scale effects, while the industrial variety is found active in only the high-tech economic sector. However, urbanization economies, which are captured by population, are not confirmed empirically. Using a different South Korean data set, Lee et al. (2005) exploit cross-sectional variations with productivity growth from 1981 to 1996 as the dependent variable and find the significant impact of only diversity when specialization is also included. Au & Henderson (2006) in their influential paper utilize city-level data from China in 1997 to estimate agglomeration influences, instrumenting for regional explanatory variables with historical data. They show that city-level economies of scale largely outweigh the diseconomies, and the external effects generated by industrial variety and urbanization are more strongly present than the Marshallian externalities are. Combes et al. (2013) follow Ciccone & Hall (1996) and Au & Henderson (2006) in applying historical instruments to a cross-sectional data set. In the study of Combes et al. (2013), data on wages is collected from household-level surveys conducted in Chinese cities in 2007, while agglomeration proxies are measured using city-level data. They demonstrate that the spatial density, with or without instrumented, explains significantly worker wages, while the role of industrial clusters is less clear. In extensive reviews of Duranton (2008) and Combes & Gobillon (2015), they notice that the evidence of spatial external economies in developing countries is mixed and the strength of the externalities is often found to be higher than those in developed territories. A rationale provided by Quigley (2009) is that duplicating successful ideas in emerging countries is highly profitable and there are more such opportunities in cities or specialized regions compared to the remaining regions.

Studies on agglomeration for Vietnam have so far focused mainly on its spatial patterns rather than its impacts on productivity. For example, Hamaguchi et al., 2012 examine the locational distribution of manufacturing industries in Vietnam between 2002 and 2007, while Nguyen et al. focus on how this regional spatial structure is affected by the emergence of multinational

firms. The exceptions are Howard et al. (2014), who use enterprise data gathered between 2002 and 2007 by GSO to estimate the productivity influences of manufacturing clusters, and Gokan et al. (2019), who employ the GSO data for the year 2012 to evaluate the impacts of agglomeration. A common feature of these two papers is that the concentration of firms rather than workers characterizes the agglomeration pattern of a local region or industry. However, according to Rosenthal & Strange (2001) and Martin et al. (2011), advantages of firm clusters mainly reflect the learning effects in which firms are the source of knowledge. Meanwhile, they prove that interactions between local employees play a dominant role in determining the agglomeration gains. In other words, in comparison with Howard et al. (2014) and Gokan et al. (2019), employment-based agglomeration indices used in this study should be better suited to the context of Vietnam because the main growth engine of the country is its cheap and abundant labor force rather than its innovative capability. This features a major difference between the study expressed in this chapter and the previous papers for Vietnam in literature.

Considering which one between inter-industry and intra-industry relations is more crucial to the influences of agglomeration, Henderson et al. (2001) argue that technological progress in developing territories is based heavily on adoption rather than innovation. Therefore, providing more chances to adopt new technology via both within- and cross-industry exchanges, the city environment should generate a more significant impact in comparison with specialized environments that can be seen even in sparse regions. In addition, based on a model adopted from the framework of Krugman (1991), Duranton (2008) discusses that, compared to developed countries, trade costs between regions are higher in developing countries due to barriers such as the less development of traffic infrastructure there. This may explain why the spatial structure of these countries tends to show the core-periphery pattern in a sense that the disparities between core cities and peripheral ones are much greater, and explain why the majority of megacities in the world are found in developing countries[4]. Furthermore, Duranton (2008) suggests that urban specialization is less attractive in the developing world due to the high transportation cost from city to city. Finally, the large magnitude of urbanization economies found in literature for developing countries is another possible indicator of its superiority over specialization economies.

When comparing Vietnam with a neighboring developing country – China, Malesky & London (2014) show that the two countries share many common features, especially persistent high

[4] Vietnam contributes a megacity, which is Ho Chi Minh City with the population of approximately 9 million in 2019.

growth rates, institutional system, and management policies of urban and industrialization. For instance, both apply a so-called Hukou system to restrict migration to big cities, especially to biggest ones to reduce pressure on urban infrastructure due to overcrowding, and thus limit the emergence of urbanization diseconomies in these regions. In addition, both promote industrialization and foreign investment through developing export processing zones, industrial parks, and special economic zones. In particular, according to GSO, there were 325 industrial parks in Vietnam in 2016 and their industrial structures are diverse. Some specialized in a few similar industries, while others served a variety of industries. Given these similarities and the evidence found in Au & Henderson (2006) and Combes et al. (2013) for China, the following hypothesis is proposed.

Hypothesis 1: Urbanization and industrial diversity have positive influences on productivity and are more influential than regional specialization in Vietnam.

The next aspect examined is whether or not firms with different ownership types benefit differently from agglomeration. In a rare test for a developed country to provide an answer to this question, Rigby & Brown (2015) find that FOEs do not differ significantly from DOEs in reaping productivity advantages from regional conditions. The rationale behind this result is probably the close similarity in many characteristics of the two firm groups in developing countries such as their technology, management efficiency, and the importance of regional labor force to their production process, which are related to the ability of firms to benefit from the spatial externalities. This contrasts with the developing world where DOEs tend to be inferior to FOEs in terms of those properties. In the context of Vietnam, understanding the role of each ownership type in the economy is crucial to suggest a plausible hypothesis. As for domestic firms, they can be either POEs or state-owned enterprises (SOEs). Before 1986, when the government passed the Doi Moi reform as an effort to bring the country out of its underdevelopment due to its decades-long central planning policy, SOEs were the main pillar of the national economy in Vietnam since private businesses were forbidden by law. Since 1986, the Vietnamese government has begun to encourage private firms, but still treated SOEs with protectionist and preferential policies. The strategy led to policy-driven rather than profit-maximization behavior of SOEs which made them less productive in comparison with private ownership. Admitting that mistake, the government has intensified privatizing or equitizing many SOEs since the early 2000s to transform them into private companies or listed companies while also remaining to be the majority shareholder. This process has altered the behavior of the remaining SOEs towards market orientation which is normally found in private companies.

Therefore, until the last decade, the difference in behavior of SOEs and POEs has become subtler.

Regarding foreign firms, the Vietnamese economy has been more and more open to the global economies since 1986 to attract foreign direct investment (FDI) aiming to boost economic growth, create jobs, and gain from technology diffusion pursuing the goal of industrialization and modernization. Until recently, the FDI sector has played a critical role in Vietnam's economy, especially in the manufacturing sector. For instance, FDI firms produced 55% of total manufacturing output and employed 50% of the manufacturing workforce in 2012 (Newman et al., 2015). One of the most appealing factors of Vietnam to foreign investors is an abundant, young, and cheap labor force, meaning that the labor market pooling may be more important to foreign-invested firms compared to their native fellows. Regional agglomeration also brings FDI firms a cost advantage in the global market compared to their home-country competitors who have limited access to such a labor force. This can be seen through a fact that 72% of the total value of exports from Vietnam were produced by FDI firms in 2016, however the value added by these production activities was generated mainly through manual work in factories producing manufacturing goods such as electronic devices, clothing, and footwear (Anh et al., 2019). Meanwhile, by concentrating on the domestic market, Vietnamese firms gain no larger advantage from the labor pool compared to their own domestic fellows as well as foreign firms who also locate their production site in the region. As a result, one could expect the performance of FOEs being superior to domestic firms when regions become more agglomerated. Furthermore, Lamin & Livanis (2013) report that newborn FOEs save more information cost in specialized regions in emerging countries due to earning knowledge and experience related to industry regulation and local institutions from incumbent FOEs and domestic local suppliers. Meanwhile, in the cross-sectional framework, the work of Gokan et al. (2019) is not able to verify that DOEs benefit from information spread either within or across industries. By using the term "agglomeration", this study henceforth means any between urbanization, industrial diversity, or specialization. The above reasons and evidence are the base to suggest

Hypothesis 2: Foreign-owned firms benefit the most from regional agglomeration.

Finally, firms of various sizes and ages might benefit differently from the externalities. Previous related studies in literature show that smaller size or younger age may bring firms a stroke of fortune in agglomerated areas. Organization theories suggest that when firms enlarge their employment scale, management tasks become more challenging. As a solution, they tend

to introduce more rigid regulations and foster culture of impersonal relationships to keep their system under control, whose side effects are the reduction in information flow within and from outside to their organization, in their flexibility to adapt quickly to regional changes, and in their ability to interact effectively with their local business partners (McCann & Folta, 2011; Millan et al. (2014); Knoben et al., 2016), which consequently lowers their ability to cultivate a productivity boost from agglomeration. Meanwhile, firms of modest scale are able to avoid such troubles owing to their size advantage. In addition, small firms are less risk averse compared to their larger business fellows. Similarly, immature firms can be blessed with better capacity to gain new knowledge and with higher flexibility to adopt novel routines in reaction to, say, changes in regional agglomeration (McCann & Folta, 2011).

From a regional science point of view, Rosenthal & Strange (2010) argue that smaller firms, and thus younger ones, are more dependent on regional factors to compensate for their weakness, therefore if agglomeration externalities exist in the region, they may be the most advantaged. For example, Tambunan (2011) reports that, in comparison with larger firms, small and medium-sized enterprises (SMEs) in Southeast Asia where Vietnam is located have poorer access to the formal capital market, while Schiffer & Weder (2001, p. 16) show that financial constraints are the largest obstacle to the development of small firms in almost any countries in the world regardless of their income levels. Due to this difficulty, when formal bank loans are not possible, small and young firms often stick to other channels to finance their business such as invoice finance, overdrafts, microloans, and informal loans (Nichter & Goldmark, 2009), whose lenders are obviously more present, diverse, and competitive in cities compared to rural regions, indicating an added benefit from locating in an urban area to this firm group. Furthermore, Nichter & Goldmark (2009) discuss that strong vertical and horizontal linkages built in urban areas could minimize the weaknesses of small-scale production facing small firms owing to joint production or the greater availability of local suppliers. In addition, the regional condition of large variety in supporting services such as consulting, employee training, or shipping can help enhance performance of small firms since they may not be able to afford to insource these services as large firms are capable of. By similar reasoning, young firms could benefit from a denser or more specialized region by learning from local experienced predecessors via the formation of larger social networks.

Studies that conduct empirical tests on the above arguments remain few and far between for developed countries with the proof of the heterogeneity in external gains is mixed for firm size and supportive for firm age, while the evidence is very rare for developing countries.

Henderson (2003) and Martin et al. (2011) find higher productivity gains to smaller firms when regions become more specialized in the US and France respectively. Studies of Andersson & Lööf (2011) with Swedish data and Rigby & Brown (2015) with Canadian data show no statistical difference in economic gains from urbanization and specialization between firms of various scales, while Combes et al. (2012) point out the dominance of large firms in gaining the productivity advantage from regional density in France. Turning to developing countries, to the author's best knowledge, Badr et al. (2019) is the only study that provides any evidence on the heterogeneity in firm size using cross-sectional Egyptian data. Relying solely on the OLS method, Badr et al. (2019) conclude that small firms benefit more than larger firms do from both intra- and inter- interactions. As for the age factor, McCann & Folta (2011) focus on the biotechnology industry in the US and report that immature firms derive higher productivity premium owing to industrial specialization compared to their mature fellows. Rigby & Brown (2015) reach the same conclusion for the manufacturing sector in Canda. So far, there has been no study reporting an estimated result on this effect for any developing country in the world, to the author's knowledge.

Regarding the relation between size and age, in a report supported by the World Bank to explain the importance of large firms in low and middle-income countries, Ciani et al. (2020) notice that these firms on average are about five years older than small firms. Meanwhile, Nichter & Goldmark (2009) aim to find out key factors contributing to the growth of small and medium-sized firms in developing territories. They show that market environment is such a growth booster, while it is not in the developed world, which can be explained by the fact that small businesses in high-income countries are fewer but larger than they are in the developing world (Poschke, 2018; Ciani et al. 2020). In addition, due to the earlier expansion of developed economies, firms there on average are older in comparison with developing economies. This information combined with the arguments and empirical evidence expressed earlier create the expectation of finding clearer proof for developing economies, particularly Vietnam where there is a lion's share of firms being immature or small-scale. It is worth pointing out that young age is not necessarily mean small scale, especially in low-income countries where nearly 50% of newborn firms are large-sized, while the number is about 33% for medium- and high-income countries (Ciani et al., 2020). Therefore, the deviation in the results when comparing between firms of various sizes and ages is also expected. On the whole, the following are suggested, ceteris paribus.

Hypothesis 3: Smaller-sized firms benefit more from regional agglomeration.

Hypothesis 4: Younger firms benefit more from regional agglomeration.

2.3. TFP estimation

This section focuses on methods used to estimate firm-level TFP, which is served as a dependent variable in the specification with agglomeration variables that whose details are set out later. To begin with, assuming that a firm i at time t produces goods according to a value-added-based production technology under the Cobb-Douglas form as

$$VA_{it} = A_{it}K_{it}^{\beta_k}L_{it}^{\beta_l} \tag{2.1}$$

where VA_{it} denotes firm value added, which equals gross revenue minus total intermediate input costs; K_{it} and L_{it} are capital and labor inputs respectively; and A_{it} refers to the firm's Hicks-neutral efficiency level during period t.

The logarithmic transformation of (2.1) is

$$va_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + v_{it} + \eta_{it} \tag{2.2}$$

where lower-case letters index natural logarithm of upper-case variables; $\ln(A_{it}) = \beta_0 + v_{it} + \eta_{it}$; parameters β_k and β_l represent capital and labor input elasticities respectively; β_0 is common technological knowledge to all sampled firms and technically it is the intercept parameter in the regression model; v_{it} is the key variable of interest – firm-level productivity, which is can be observed by the firm but cannot be by statisticians; And error term η_{it} is unobservable and assumed to be independent and identically distributed (i.i.d). To compare, v_{it} is a state variable, and thus has an impact on input choices of the firm[5], whereas η_{it} does not. This raises a concern about the contemporaneous positive correlation of v_t with both production inputs and outputs as v_t is unobservable, which is similar to the disturbance η_{it}. In practical terms, knowledge on their own productivity may guide firms to choose their input levels accordingly in order for profit-maximization. As a consequence, without controlling v_t, it enters into the disturbance η_{it}, therefore the application of the ordinary least squares (OLS) method to (2.2) could introduce an upward bias to estimated elasticities βs, then deflate the value of productivity as a result. A further problem is that the attrition of businesses over time might be not random but correlated to their productivity, causing the potential skewness of firm-level samples towards those with higher capital holding a similar level of productivity.

[5] v_t is assumed to be the only unobserved state variable for firms following Olley & Pakes (1996).

2.3.1. Olley & Pakes (1996)'s strategy

To deal with both problems expressed above, this chapter first adopts the model of firm behavior in Olley & Pakes (1996) (OP for short) featuring the dynamic arbitrary of v_t and its interdependence with production inputs. In this model, profit of a representative firm i at time t is assumed to be a function of its state variables as physical capital k_{it}, its productivity v_{it}, and state variables of all the firm's competitors which reflect the market structure. Such an active firm has to decide whether to exit from or to stay in the market at the beginning of every period. If exit is the case, it earns an amount of money Ψ from selling off its assets and disappears henceforth. Otherwise, the firm chooses labor inputs needed to combine with its state variables including its physical capital and its knowledge on its current productivity to produce goods and gain current profit. After that, the firm decides its investment level i_t so that physical capital evolves according to $k_{i,t+1} = k_{it} + i_{it}$. Meanwhile, v_{it}, as another state variable, is assumed to have the first-order Markov property, meaning that its value at $t + 1$ is conditional on the distribution of productivity in the market at time t.

Digging more deeply into the decision-making process of the firm to exit or continue in operation, it is reasonable to assume that motivation of the firm is to maximize its expected discounted value of its remaining profit after investment is made in the future. The firm chooses to stay in the market if this value is larger than Ψ, and thus this decision along with the choice of optimal investment are dependent on the firm's prediction of future state variables based on provided information at time t. In such a dynamic setting, the following Bellman equation[6] is proposed:

$$\begin{aligned} V_{it}(k_{it}, v_{it}) = \max\big[\Psi, sup_{I_{it}\geq 0}\, \Pi_{it}(k_{it}, v_{it}) - C(I_{it}) \\ + \rho E\{V_{i,t+1}(k_{i,t+1}, v_{i,t+1} | G_{it}\}\big] \end{aligned} \tag{2.3}$$

where $\Pi_{it}(k_{it}, v_{it})$ is the function of current profit with respect to state variables, $C(I_{it})$ represents the current expenses on investments, ρ is the discount factor, and G_{it} refers to information given at time t. The equation indicates that firm i would leave the market at time t if $\Psi > sup_{I_{it}\geq 0}\, \Pi_{it}(k_{it}, v_{it}) - C(I_{it}) + \rho E\{V_{i,t+1}(k_{i,t+1}, v_{i,t+1} | G_{it}\}$, or intuitively, if its liquidation value is larger than its expected returns in the future because, say, the current

[6] Compared with the original version, firm age as the third state variable in the model is dropped from this equation.

efficiency level v_{it} is not high enough. The solution to (2.3), under the Markov Perfect Nash equilibrium setting with functions $\underline{v}_{it}(.)$ and $I_{it}(.)$, is yielded with the exit indicator

$$\chi_{it} = \begin{Bmatrix} 1 \; if \; v_{it} \geq \underline{v}_{it}(k_{it}) \\ 0 \quad otherwise \end{Bmatrix} \tag{2.4}$$

and logarithm of investment demand function

$$i_{it} = I_{it}(v_{it}, k_{it}) \tag{2.5}$$

which implies that firm i adjusts their optimal investments i_{it} according to its state variables. Next, the monotonicity between i_{it} and v_{it} is assumed, which indicates that the firm makes further investments in year t when facing a positive productivity shock during that year. This allows function (2.5) to be inverted into

$$v_{it} = I_{it}^{-1}(v_{it}, k_{it}) = v(i_{it}, k_{it}) \tag{2.6}$$

for logarithm of investments $i_{it} > 0$. Equation (2.6) plays the role of a control function in this estimation because it makes v_{it} explicit with observables. Since v_{it} evolves according to a first-order Markov process, it could be presented as

$$v_{it} = E\left[v_{it} | v_{i,t-1}, \chi_{it} = 1\right] + \zeta_{it} \tag{2.7}$$

where $v_{i,t-1}$ is firm productivity at time $t - 1$ and ζ_{it} is the difference between the true value of productivity at time t and its expected value conditional on past information. In other words, ζ_{it} is the part of current productivity "beyond expectation", thus it represents innovation shock. Another assumption is that the capital input does not respond quickly to ζ_{it}, but l_{it} may do so. This allows for the possibility that the firm might hire more or lay off a certain number of workers according to its short-run variations in innovation. Putting the expression of v_{it} in (2.7) into (2.2) yields

$$va_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + E\left[v_{it} | v_{i,t-1}, \chi_{it} = 1\right] + \zeta_{it} + \eta_{it} \tag{2.8}$$

Next, plugging (2.6) into (2.2), and grouping the state factors and the intercept together with the expression

$$\phi(i_{it}, k_{it}) = \beta_0 + \beta_k k_{it} + v(i_{it}, k_{it}) \tag{2.9}$$

leads to

$$va_{it} = \beta_l l_{it} + \phi(i_{it}, k_{it}) + \eta_{it} \tag{2.10}$$

where

$$\mathrm{E}(\eta_{it} | l_{it}, k_{it}, i_{it}) = 0 \tag{2.11}$$

is assumed. Since the functional form of $\phi(.)$ is unknown, the application of parametric estimation becomes feasible by using a fourth-order polynomial in i_{it} and k_{it} and their cross products in place. Next, the consistent coefficient of β_l, denoted as $\hat{\beta}_l$, is calculated with no-intercept OLS estimation for (2.10) given that η_{it} is an i.i.d error term. This finishes the first stage without a result on capital elasticity. Notice that in this chapter, the hat symbol is added to a parameter to denote its estimated value, while it is added to a variable or a function to indicate its fitted value.

In the second stage, with value of $\hat{\beta}_l$ and corresponding fitted values of value added as $\hat{y}_{it}$ at hand, it becomes feasible to compute the fitted values of productivity using (2.9) for each specific possible value β_k^* of β_k as

$$\hat{v}_{it} = \hat{\phi}(i_{it}, k_{it}) - \beta_k^* k_{it} = \widehat{va}_{it} - \hat{\beta}_l l_{it} - \beta_k^* k_t \tag{2.12}$$

Next, the survival probability is estimated for the case of $\chi_{it} = 1$. Based on (2.4) with several transformations, that probability becomes

$$\begin{aligned} \mathrm{Pr}_{\mathrm{i,t-1}} &= \Pr\{X_{it} = 1 | \underline{v}_{it}, G_{t-1}\} \\ &= \Pr\{v_{it} \geq \underline{v}_{it} | \underline{v}_{it}, v_{t-1}\} \\ &= \mathrm{P_t}\{\underline{v}_{it}(i_{it}, k_{it}), v_{t-1}\} \end{aligned} \tag{2.13}$$

and both $\underline{v}_{it}$ and v_{t-1} can be expressed as a function of $i_{i,t-1}$ and $k_{i,t-1}$, thus

$$\mathrm{Pr}_{\mathrm{i,t-1}} = \omega\{i_{i,t-1}, k_{i,t-1}\} \tag{2.14}$$

where ω is a function with an unidentified form. Fitted value of the probability as $\widehat{Pr}_{i,t-1}$ then is yielded by regressing a probit model of χ_{it} on polynomials of $i_{i,t-1}$ and $k_{i,t-1}$. Next, the value $\hat{v}_{it}$ obtained from (2.12) is regressed on a fourth polynomial expansion in $\hat{v}_{i,t-1}$ and $\widehat{Pr}_{i,t-1}$ to approximate $E[v_t | v_{t-1}, \chi_{it} = 1]$, which is used along with obtained $\hat{\beta}_l$ and observed values of value added to calculate residuals for each given β_k^* and as

$$\widehat{\eta_{it} + \zeta_{it}} = va_{it} - \hat{\beta}_l l_{it} - \beta_k^* k_{it} - E\left[\nu_{it} | \widehat{\nu_{i,t-1}, Pr_{i,t-1}}\right] \tag{2.15}$$

After that, consistent $\hat{\beta}_k$ can be obtained by solving

$$min_{\beta_k^*} \left\{ \sum_i \sum_t \widehat{\eta_{it} + \zeta_{it}} \right\}^2 \tag{2.16}$$

In the final step, logarithm value of firm-level TFP in year t is computed as

$$lnTFP_{it} = va_{it} - \hat{\beta}_l l_{it} - \hat{\beta}_k k_{it} \tag{2.17}$$

2.3.2. Levinsohn & Petrin (2003)'s strategy

Similar to OP, Levinsohn & Petrin (2003) (LP for short) uses the control function to tackle the simultaneity issue of unobservable productivity. However, LP made a number of major changes in comparison with OP to improve its estimation efficiency, which makes LP the second method used in this chapter to estimate TFP. The first change is the use of a demand function of a certain intermediate input such as electricity consumption in LP to overcome weaknesses of the application of investment that its values may be merely zero for a high fraction of firms and may rise suddenly when a firm decides to make investment after a long time of monetary capital accumulation. Therefore, its inversion now is $\nu_{it} = m^{-1}(\nu_{it}, k_{it}) = \vartheta(m_{it}, k_{it})$ in place of (2.6) with m_{it} being the logarithm of intermediate inputs. However, LP notes that the choice between investment and intermediate inputs should be made based on the volume of null values of these potential control factors in data. The data set employed in this study indicates a considerable share of firms recorded with no annual investment, which may diminish the consistency of the estimation due to the resulting truncated probability distribution of investment. Meanwhile, power consumption, which is the essential intermediate input in the manufacturing sector, is reported with positive values in most of sampled firms. In addition, the non-monetary values and non-storable characteristics of electricity better reflect annual productivity shocks, and thus are superior to the features of material costs, whose values are used to measure TFP for Vietnam with the control function approach in Nguyen (2017). As a result, the demand function of power consumption is used in this study as the primary control function to estimate TFP. Furthermore, Levinsohn & Perin (2003) note that the monotonicity assumption should be checked to guarantee the suitability of the choice for the control function. Following the guidance of Levinsohn & Perin (2003), values of TFP are first obtained using

the method of LP, OP, or Woodridge (2009). Notice that the method of Woodridge (2009) will be expressed later. The logarithm of electricity usage or investments is then regressed on a fourth order polynomial in the logarithm of capital and logarithm of TFP. The resulting illustrative graphs with a fitting surface created after each regression shows the relationship between the control variable and the two state variables, whose details are found in Figures 2.1, 2.2, and 2.3. The three-dimensional surfaces show that holding capital constant, TFP computed by the method of LP or Woodridge (2009) using electricity consumption indicates a distinctive positive relationship, while the validity is very likely for the case of OP TFP. These results validate the application of the monotonicity condition, especially with the choice of the electricity usage for the control function.

Another change in LP is that the survival probability is no longer accounted for due to its ambiguous effects on final results in OP. Resulting from all of these changes, only $\hat{\beta}_l$ is calculated in the first stage since the function now also is non-linear in m_{it}. In the second stage, after the similar steps of transformation with $\mathrm{Pr}_{\mathrm{i,t-1}}$ being removed from and m_{it} being added to (2.15), residuals for each candidate value of β_k and β_m as β_k^* and β_m^* respectively are computed as

$$\widehat{\eta_{it} + \zeta_{it}} = y_{it} - \hat{\beta}_l l_{it} - \beta_k^* k_{it} - \beta_m^* m_{it} - E\left[\widehat{\nu_{it} | \nu_{i,t-1}}\right] \tag{2.18}$$

The consistent estimator $\hat{\beta}_k$ is then computed using the minimizing technique similar to OP.

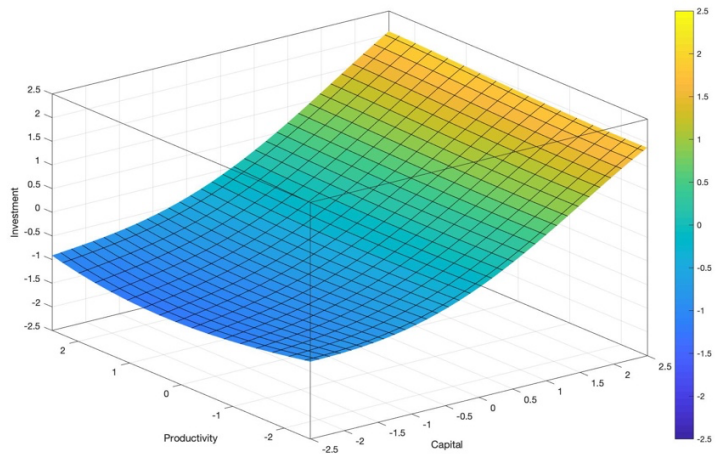

Figure 2.1. The visualization of the monotonicity condition, Olley & Pakes (1996)

Notes: The fitting surface in this graph is obtained by regressing the standardized logarithm of electricity consumption on a fourth order polynomial in standardized logarithms of firm-level capital and total factor productivity, which is measured following Olley & Pakes (1996).

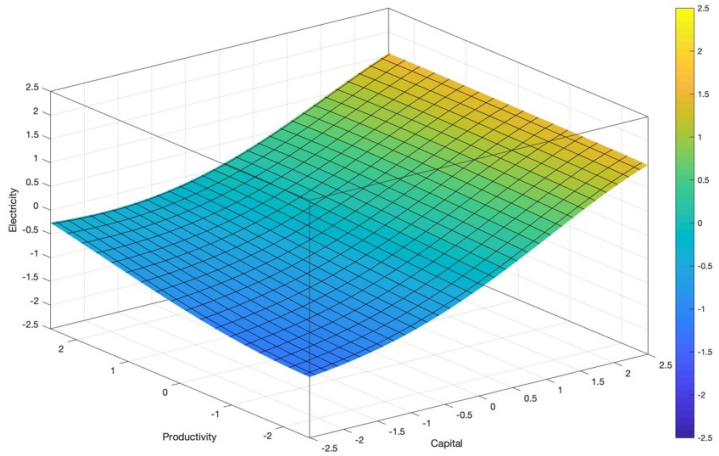

Figure 2.2. The visualization of the monotonicity condition, Levinsohn & Petrin (2003)

Notes: The fitting surface in this graph is obtained by regressing the standardized logarithm of electricity consumption on a fourth order polynomial in standardized logarithms of firm-level capital and total factor productivity, which is measured following Levinsohn & Petrin (2003).

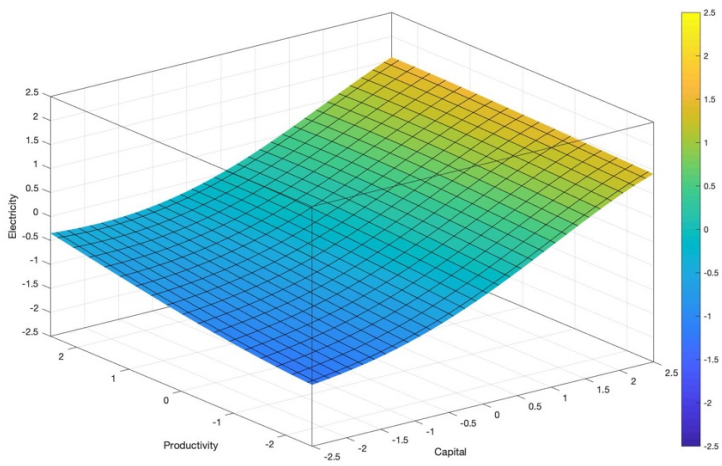

Figure 2.3. The visualization of the monotonicity condition, Woodridge (2009)

Notes: The fitting surface in this graph is obtained by regressing the standardized logarithm of electricity consumption on a fourth order polynomial in standardized logarithms of firm-level capital and total factor productivity, which is measured following Woodridge (2009).

2.3.3. Wooldridge (2009)'s strategy

Wooldridge (2009) inherits the control function approach used in OP and LP but proposes an instrumental tactic under a form of the generalized method of moments (GMM) to further improve efficiency of the two methods. First, if the control proxy chosen is energy, equation (2.11) with m_{it} used in place of i_{it} is expanded to become

$$E\left(\eta_{it} | l_{it}, k_{it}, m_{it}, l_{i,t-1}, k_{i,t-1}, m_{i,t-1}, \dots, l_{i1}, k_{i1}, m_{i1}\right) = 0, \ t = 1,2, \dots, T \qquad (2.19)$$

which allows the disturbance η_{it} to be serially dependent. Next, equation (2.6) is rewritten in the LP setting as

$$E\left(v_{it}|k_{it}, l_{i,t-1}, k_{i,t-1}, m_{i,t-1}, \dots, l_{i1}, k_{i1}, m_{i1}\right) = E\left(v_{it}\middle|v_{i,t-1}\right) \tag{2.20}$$

Since $v_{i,t-1} = \vartheta(m_{i,t-1}, k_{i,t-1})$, plugging into (2.20) yields

$$E\left(v_{it}|k_{it}, l_{i,t-1}, k_{i,t-1}, m_{i,t-1}, \dots, l_{i1}, k_{i1}, m_{i1}\right) = f\left[\vartheta(m_{i,t-1}, k_{i,t-1})\right] \tag{2.21}$$

where, identical to $\vartheta(.)$,$f(.)$ also has no predetermined functional form. Using (2.7) without the probability part and (2.21) yields

$$v_{it} = f\left[\vartheta(m_{i,t-1}, k_{i,t-1})\right] + \zeta_{it} \tag{2.22}$$

Since it is possible to rewrite (2.9) and (2.10) as

$$va_{it} = \beta_l l_{it} + \beta_k k_{it} + \vartheta(m_{it}, k_{it}) + \eta_{it}, \qquad t = 1,2, \dots, T \tag{2.23}$$

Plugging (2.22) into (2.23) obtains

$$va_{it} = \beta_l l_{it} + \beta_k k_{it} + f\left[\vartheta(m_{i,t-1}, k_{i,t-1})\right] + \mu_{it}, \qquad t = 2, \dots, T \tag{2.24}$$

where $\mu_{it} = \eta_{it} + \zeta_{it}$. The two equations (2.23) and (2.24) are important to estimate β_l and β_k. The composite error μ_{it} in (2.24) helps transform the zero condition in (2.19) into

$$E\left(\mu_{it}|k_{it}, l_{i,t-1}, k_{i,t-1}, m_{i,t-1}, \dots, l_{i1}, k_{i1}, m_{i1}\right) = 0, t = 2, \dots, T \tag{2.25}$$

According to Ackerberg et al. (2006), estimating β_l using (2.24) as OP and LP do in their first stage of regression raises a concern that the obtained OLS estimates may be contaminated due to the likelihood that the firm does not only choose l_{it} given v_{it} but also decide m_{it} simultaneously. In other words, l_{it} should be treated similarly to m_{it} as a non-parametric function of state variables. If this is the case, (2.23) has an identification issue, and thus the estimator of β_l using either the OP or LP procedure is still biased upwards. To solve this potential issue, this study strengthens the LP model with the IV estimation proposed by Wooldridge (2009) who makes use of the GMM framework and generates a single-step process instead of the two-stage process of OP and LP. The method is denoted shortly as LP-W to emphasize the use of Wooldridge's method based on the core ideas of LP. In addition, Figure 2.3 strongly supports the application of the monotonicity condition with LP-W. Henceforth, this study considers LP-W to be the most reliable and indeed the standard approach to tackle simultaneity bias. Despite this, TFP is also calculated with the OLS, LP, and OP approaches to check on how findings on agglomeration economies vary across different productivity measures.

To continue, the unspecified functions in (2.23) and (2.24) can be approximated using a fourth polynomial expansion in variables inside the parentheses plus their interaction terms as in LP. This leads to the following assumptions

$$\vartheta(k_{it}, m_{it}) = \gamma_0 + \varphi(k_{it}, m_{it})\gamma \tag{2.26}$$

and

$$f\left[\vartheta\left(m_{i,t-1}, k_{i,t-1}\right)\right] = \alpha_0 + \alpha_1[\varphi(k_{it}, m_{it})\gamma] + \alpha_2[\varphi(k_{it}, m_{it})\gamma]^2 + \\ \dots + \alpha_n[\varphi(k_{it}, m_{it})\gamma]^n \tag{2.27}$$

because function $f(.)$ can be expressed with the nth-degree polynomials in $\vartheta(.)$. This study follows the standard procedure with $n = 1$ and $\alpha_1 = 1$. Plugging (2.26) and (2.27) into (2.23) and (2.24) gains

$$y_{it} = \tau + \beta_l l_{it} + \beta_k k_{it} + \varphi(m_{it}, k_{it})\gamma_1 + \eta_{it}, t = 1,2, \dots, T \tag{2.28}$$

and

$$y_{it} = \sigma + \beta_l l_{it} + \beta_k k_{it} + \varphi\left(m_{i,t-1}, k_{i,t-1}\right)\gamma_1 + \mu_{it}, t = 2, \dots, T \tag{2.29}$$

respectively.

Now, this system of equations can be used to estimate consistent parameters with the IV approach as expressed in Wooldridge (2010, Chapter 11). The crucial factor is orthogonality conditions (2.19) and (2.25) giving the instrument vector

$$\mathrm{Z}_{it} = \begin{pmatrix} 1, l_{it}, k_{it}, \varphi(m_{it}, k_{it}) \\ 1, l_{i,t-1}, k_{it}, \varphi\left(m_{i,t-1}, k_{i,t-1}\right) \end{pmatrix} \tag{2.30}$$

where the above part is used to identify (2.28) and the under part is for (2.29). Finally, for each $t > 1$, the GMM estimators of input parameters can be yielded from the equation system of the residual matrix

$$r_{it} = \begin{pmatrix} y_{it} - \tau - \beta_l l_{it} - \beta_k k_{it} - \varphi(m_{it}, k_{it})\gamma_1 \\ y_{it} - \sigma - \beta_l l_{it} - \beta_k k_{it} - \varphi\left(m_{i,t-1}, k_{i,t-1}\right)\gamma_1 \end{pmatrix} \tag{2.31}$$

and

$$E[\mathrm{Z}_{it}' r_{it}] = 0, t = 2,3, \dots, T. \tag{2.32}$$

After obtaining the estimators of production inputs, the logarithm values of TFP are computed as in (2.17).

2.4. Data and agglomeration variables

2.4.1. Data

There are two stages of estimation in this study. The regression output of the first stage is firm-level productivity, whose values are used as the explained variable in the second stage to identify the agglomeration effects. Despite the difference, both stages employ mainly data from a single source, which is the Annual Census of Enterprises conducted by GSO in Vietnam for the period 2011-2016. The primary purposes of these national censuses are to assess the development of business establishments and their workforce throughout the country and to calculate important economic measures such as Gross Regional Domestic Product (GRDP). As a result, reports produced by these censuses are pivotal to the Vietnamese government in making and adjusting economic and enterprise policies. Given its importance, the enterprise census covers annual basic information of all firms with business tax codes in Vietnam, collected from local tax offices and authorities where firms are under their administration. The comprehensive coverage of all firms brings a big advantage to this data since the construction of multi-level regional variables becomes feasible with the calculation based on firm-level information[7]. The available information consists of the firm's ownership type, industry code, operating status, location, revenues, profits, number of workers, and tax code. Firms are identified with their unique tax code. The code system used to classify various industries during the 2011 – 2016 period is the most recent adjusted 5-digit VSIC in 2007.

For more detailed firm-level information, GSO had to carry out censuses by means of face-to-face interviews with directors or chief accountants of firms. Another census method was sending questionnaires with detailed instructions to sampled firms and requiring them to complete the forms in time. Due to the heavy load of these tasks, merely a sizeable fraction of firms were sampled annually to investigate. Exceptions were made in 2011 and 2016 when all firms were surveyed. In the remaining years (2012 – 2015), firms were chosen to sample based on their industry, ownership, region, and size[8]. More concretely, GSO sampled all firms being

[7] Since a firm might have multiple establishments located in different regions, the application of data at an establishment level may be more assuring compared to a firm level for the accuracy of region-level measures. However, this should not be a concern based on features found in data of the year 2011 when establishment-level information is available to extract. Specifically, multi-establishment firms with a difference in district-level locations between their headquarters and any their establishments make up only approximately 2.6% of all firms operating in Vietnam during the year. Furthermore, the workforce in the headquarters of firms commonly is the largest among their establishments, thus regional variables calculated using GSO firm-level data are negligibly different from their values using GSO establishment-level data.

[8] In 2012, for instance, GSO surveyed 168,854 firms in total, while data on location, industry, workforce, and revenues is available for a total of 358,558 firms.

state-owned or foreign-invested or belonging to ten of the least populous provinces in Vietnam during this period. Domestic private-owned enterprises (POEs) were all surveyed if they had at least 20 employees[9], otherwise they were sampled randomly. These surveys provide further firm-level information including value of assets, investments, energy use, labor costs, interest costs, profit, taxes, and other financial obligations to the authority. A limitation of these censuses is that several useful variables such as education level of workers, and income from international trade are available in only a few years. Furthermore, information on working hours and materials was beyond the investigation.

In the stage of estimating firm-level productivity, detailed information on inputs, outputs, and investment of each firm is needed to estimate the function (2.2). Production inputs are made up of capital, labor, and a certain control variable, which is electricity consumption if the method of LP or LP-W is applied, and investments if OP is the case. The labor input of a firm as l_{it} in function (2.2) is measured as total number of employees and employers of the firm recorded at the end of the year. Regarding capital stock k_{it}, its measurement is more complex due to its involvement with depreciation and devaluation of fixed assets over time, whose rates vary across firms, industries, regions, and years. Scholars who measure productivity in literature tend to calculate the depreciated values of physical capital using the so-called perpetual inventory method (PIM), which requires long panel data and details on specific types of capital inputs, and original value and corresponding depreciation method for each type. With such data at hand, the application of PIM is feasible. However, in the case of the GSO data, such full information is obtainable for merely a few years, and 6 years of data is rather short[10]. Furthermore, the choice of strongly unbalanced panel data in this study results in difficulties in tracking annual fluctuation of fixed assets to meet requirements for PIM. Therefore, the original value of total fixed assets recorded for each firm at the end of the year[11] is employed in this study as a proxy for capital input, whose values are deflated using the deflator of Development Investment Capital of Society provided by GSO.

[9] This firm-size threshold was not constant across different regions and years. The number tend to be higher for more populous province-level regions and years without a comprehensive census. For example, in 2012, GSO randomly sampled 20% of domestic POEs with fewer than 20 employees in most provinces. Nevertheless, the corresponding figure for the two biggest province-level regions (Hanoi and Ho Chi Minh) was 20% of firms with 20 to 50 employees and 10% of those with under 20 employees.

[10] To construct a longer panel data set, the data is also available to collect for years before 2011 and after 2016. Nevertheless, some information essential to this study was missing during those periods. Specifically, data on energy consumption was unavailable before 2011, while interest cost was beyond the censuses after 2016.

[11] This choice of timing fits the characteristic of capital accumulation employed in the model of Levinsohn & Perin (2003).

As for the control functions, this study uses annual electricity consumption for production to represent intermediate input m_{it} when estimating with the LP or LP-W technique[12], and annual investment i_{it} when the OP method is the case. Instead of other types of energy and intermediate inputs, the electricity consumption is chosen because its physical value is not affected by the price change, and its non-storable form prevents its reported value from being distorted by fluctuation in annual inventory which would be a problem if material or fuel is taken into consideration. Turning to value added va_{it} in equation (2.2), it is computed commonly by subtracting material costs from gross revenue. However, data on this cost element is unavailable, therefore va_{it} is calculated based on cost and profit components which constitute value added, following Ha & Kiyota (2014), who make use of firm-level data from GSO for another strand of literature. More specifically, nominal value added of a firm in a year is calculated as the summation of change in accumulated depreciation between the beginning and end of that year, total employment costs, net operating profit, indirect taxes, and interest costs[13]. Its real values are then obtained using a manufacturing value-added GDP deflator published by GSO. Since different industries should show different elasticities of capital and labor, the TFP estimation is conducted separately for each of 21 2-digit VSIC manufacturing industries[14] applying the Stata command *prodest* created by Rovigatti & Mollisi (2018).

The final aspect needed to pay attention to is data-cleaning procedures before the regression is implemented. To create regional variables out of firm-level data, observations are dropped if any of their codes of tax, location, industry is found missing, or if the value of workforce is negative, whose cases make up a tiny share of total observations. Next, firms with less than 20 workers are removed from all regression samples to mitigate possible effects of random sampling[15]. In addition, the industry code of multi-industry firms is assigned to their main industry. Finally, firms who changed their 3-digit VSIC or district location between 2011 and 2016 are also eliminated as a solution to potential data input errors and a treatment to mitigate the possible problem of endogeneity in location choice which is discussed in the next section.

[12] As far as the authors are aware, this is the first research that attempts to estimate TFP in Vietnam making use of the firm's power consumption in the control function in the framework of LP or LP-W.
[13] It is worth pointing out that the calculation in this study is more accurate than Ha & Kiyota (2014) since information on interest costs is unavailable during sampling years of their data set.
[14] Industry 12 (tobacco) and industry 19 (coke and refined petroleum products) are dropped from the productivity calculation because they have too few observations in order to yield statistically significant estimates..
[15] Although this measure helps, it greatly reduces the sample size.

2.4.2. Agglomeration proxies

This subsection expresses indices used to proxy for agglomeration, which are constructed based on information on labor, location, and industry of all firms with registered tax codes in Vietnam, using the data set mentioned above. Two measures are employed to capture urbanization economies. The first one is the logarithm of employment density, used to represent the absolute sense of urbanization, and the second one is industrial diversity, used to catch the relative sense of urbanization, or Jacob externalities in other words. Making use of the firm-level data set from GSO, the density is calculated as

$$DEN_t^r = \ln \frac{\sum_{i \in r} e_{it}^r}{a_r^t} \tag{2.33}$$

where e_{it}^r is the workforce of firm i located in region r and a_{rt} denotes its land area at time t. Data on regional land area a_{rt} is gathered from provincial-level statistical yearbooks published annually in Vietnam. An advantage of this density-based index over other absolute employment measures such as regional employment is that it is less impacted by zoning idiosyncrasies, thus better displaying the true scale effects (Combes et al., 2010). For example, if region 1 is twice larger in employment but four times smaller in land area than region 2, it means that region 2 is twice denser than region 1. In this case, the spatial proximity, which drives the agglomeration mechanisms, is reflected more accurately with the employment density in comparison with the regional employment.

The index of diversity is computed following Martin et al. (2011) as

$$DIV_t^{sr} = -\ln \left[\sum_{s' \in S|\{s\}} \left(\frac{e_t^{s'r}}{e_t^r - e_t^{sr}} \right)^2 \right] \tag{2.34}$$

where e_t^{sr} refers to number of workers in the industry-region sr; $e_t^{s'r}$ is the number of employees working in each remaining industry other than s within region r; e_t^r denotes the total number of workers in region r at time t. The diversity level reaches the minimum value of zero if only two industries are active in the region, while it rises with an increase in the number of industries and the equal degree of their regional shares of employment. The DIV measure, in fact, is the logarithm of the inverse of the Herfindahl-Hirschman Index (HHI) constructed on employment shares of all active industries s within region r, as

$$HHI_t^r = \sum_s \left(\frac{e_t^{sr}}{e_t^r}\right)^2 \tag{2.35}$$

The sum of squared shares of all industries in terms of employment will surge when there are fewer industries in the region, or when some industries achieve employment growth at the expenses of others. It implies that HHI represents the industrial concentration, or the lack of diversity in a region in other words. Thus, its inverted value is considered to illustrate the local industrial diversity. In comparison with the logarithm of the inverse of HHI, the employment share of the own industry is excluded from DIV. As noted by Combes & Gobillon (2015), this exclusion is necessary to enable a better interpretation given the presence of a specialization variable in the model.

The final agglomeration proxy is specialization. This study uses the logarithm of sum of location quotient (LQ) plus one to capture the MAR externalities, calculated as

$$SPE_t^{sr} = \ln\left(1 + \frac{\frac{e_t^{sr}}{e_t^r}}{\frac{e_t^s}{e_t}}\right) \tag{2.36}$$

where e_t^{sr} and e_t^r are known above, e_t^s is defined as the number of workers in industry s at the national level, and e_t refers to the total employment of the whole business enterprise sector in year t. Location quotient is the complex fraction in parentheses, which reflects the relative concentration of industry s in region r by comparing the employment share of the industry in the region as fraction e_t^{sr}/e_t^r to its share in the whole country as fraction e_t^s/e_t. The value of LQ is close to zero when too few people are working in the industry in the region compared to the presence of this industry at the national level. LQ falling into between 0 and 1 indicates a specialization level below national average, while it means the opposite when LQ is larger than 1. LQ equal to 1 implies that the regional share of employees in a specific industry is equivalent to its national share, and therefore the relative specialization strength of that industry is the same as the national average. LQ is equal to 5 for an industry, for example, meaning that the industry in question is 5 times more concentrated in terms of employment in the local region than on the national average. SPE is the logarithm-transformed version of sum of LQ plus one, thus has the minimum value of zero. Compared to LQ, this logarithm form is more convenient to interpret the estimation results because the higher the value of this index shows, the higher the degree of regional specialization it implies.

2.4.3. Geographical units and summary statistics

Since regional variables are aggregated using the micro-level data based on the choice of geographical boundaries, the identification of their unit type is essential. To begin with, there are three administrative subdivisions available in the data set, including province, district, and commune, with the last one being the finest level of geography. In 2016, there were officially 63 province-level, 793 district-level, and 11,162 commune-level regions in Vietnam. Intuitively, on average, a province is equivalent to a circle of space with a radius of approximately 40.9 km, while those numbers for a district and a commune are about 12.2 and 3.1 km respectively, based on the author's calculation. It is tempting to calculate regional variables and then estimate agglomeration externalities separately for each of the three spatial units to see how the estimates of the externalities change with scale units. However, there are good reasons not to do so but to conduct the estimation only with the most appropriate unit.

The first reason is given from the theoretical perspective. On the one hand, the map drawn using a very fine unit of geography may provide a misleading picture of agglomeration level because spatial clusters are broken down into many separate areas with borders (Head & Mayer, 2004). On the other hand, external benefits through information spillover are mainly found at finer levels of geography, say a commune or district, the gains from sharing intermediate inputs and high-quality infrastructure tend to be found at larger geographical levels, say district or province, meanwhile the labor market pooling generates the positive externalities across all geographical scopes (Rosenthal & Strange, 2001; Rosenthal & Strange, 2004; and Combes & Gobillon, 2015). Therefore, the district level reconciles those arguments, and thus should be considered as the most suitable spatial unit to capture the externalities.

Other reasons are provided from the technical perspective. Due to the administrative merger or division of a number of communes and districts during the years 2011-2016, data from all sources is geo-coded to the code system of the GSO's 2016 census, using commune identities as the common base. However, when the administrative changes took place at a sub-commune level, the geo-coding could not work properly. For example, imagine a case that several wards and their population belonging to commune A in 2011 were set to be under administration of commune B from 2012. It means that even if no worker had left from or come to settle in commune A between 2012 and 2016, the commune B in 2011 was different from it in 2016 in terms of both land area and labor. This commune-level distortion could not be corrected by geocoding without information at a finer level of geography, but it functions well for districts and provinces owing to the commune-level information. Next, it is not possible to calculate

commune-level employment density due to the unavailability of data on land area as fine as at the commune level between 2011 and 2016. Finally, to estimate agglomeration economies, an interaction term of region and year is added to the specification to control for regional shocks, which requires information at a broader geographical level than the level used for the primary analysis. Since there is no larger administrative unit than a province in Vietnam, doing a regional analysis at the province level is not empirically appropriate. For all reasons mentioned above, this study considers districts integrated with 3-digit VSIC codes as the spatial and industrial unit in the baseline analysis, even though the regression is also conducted at a 2-digit VSIC scale as a sensitivity check.

After the cleaning process, the final sample used to estimate the agglomeration externalities is an unbalanced data set spanning six years from 2011 to 2016 with 26,987 individual firms and a total of 80,638 observations. The statistical description, correlation matrices of variables, and the distribution of key firm-level and region-level variables for the estimation of TFP and agglomeration are shown in tables from 2.1 to 2.7 and Figure 2.4. It is worth pointing out some features of data seen from these tables and figures. Table 2.1 indicates that, compared to other years, firms observed in 2011 or 2016 are more present in the regression sample owing to the comprehensive coverage of the GSO census in these two years. Table 2.2 details that there are only 55,176 out of 80,638 observations with available information on investment. These missing values include firms that either did not report their annual investments or invested less than 1 million VND per year, thus their logarithmic values are negative or incalculable. As discussed earlier, this high share of missing values makes the application of LP and LP-W superior to OP. Table 2.3 implies that all variables entering into the main specification of agglomeration estimation vary across years because the standard deviation (SD) for within variation is all positive, thus these variables are all suitable for the application of the fixed-effects estimation method. In addition, the year-to-year changes of the employment density seem not as dramatic as specialization and diversity because the disparity between between-group SD and within-group SD is much wider for the density. Table 2.5 shows a negative correlation between the regional specialization with other agglomeration variables but a quite strong positive correlation between the density and the industrial diversity. Table 2.4 indicates that the majority of firms in Vietnam are POEs. FOEs make up more than 27% of the total number of observations, which is considerable compared to a mere fraction of about 2.8% of SOEs in the final sample. Graphs in Figure 2.4 show that the kernel distributions of variables essential to the agglomeration regressions are positively skewed towards smaller firms,

younger firms, less dense regions, and less specialized industries, while the distribution of industrial diversity over regions is more symmetric.

Table 2.1. Temporal composition of the final sample

Year	Observations	Percent	Cum. Percent
2011	14,484	18.46	18.46
2012	12,837	15.92	34.38
2013	11,817	14.65	49.03
2014	13,261	16.45	65.48
2015	10,567	13.10	78.58
2016	17,272	21.42	100.00
Total	80,638	100.00	

Table 2.2. Summary of statistics for productivity regressions

Variable	Obs	Mean	Std. dev.	Min	Max
ln (Value Added)	80,638	8.857	1.767	0.015	19.08
ln (Capital)	80,638	4.614	1.254	2.996	11.353
ln (Labor)	80,638	9.171	2.130	1.212	18.133
ln (Electricity)	80,638	12.044	2.373	2.303	24.693
ln (Investment)	55,176	7.484	2.148	0.088	17.508

Notes: The unit of value added, capital stock, and investment is million VND, the unit of electricity usage is Wh.

Table 2.3. Summary of statistics for agglomeration regressions

Variable		Obs	Mean	Std. dev.	Min	Max
ln (Productivity)	overall	80,638	2.961	1.085	-5.466	10.694
	between			1.018	-5.441	8.947
	within			0.451	-3.651	9.172
ln (Employment density)	overall	80,638	5.839	2.139	-2.583	10.986
	between			2.219	-2.123	10.986
	within			0.136	3.604	7.490
Specialization	overall	80,638	1.419	1.019	0.003	6.928
	between			1.009	0.004	6.928
	within			0.152	-0.543	3.685
Diversity	overall	80,638	2.365	0.647	0.031	3.644
	between			0.649	0.037	3.643
	within			0.144	0.559	4.326
Competition	overall	80,638	1.691	1.095	0	8.415
	between			1.085	0	6.826
	within			0.214	-0.927	5.897
Regional Share of FOEs	overall	80,638	0.069	0.103	0	0.7
	between			0.086	0	0.7
	within			0.044	-0.142	0.555
ln (Regional Average Wages)	overall	80,638	4.069	0.353	1.852	6.473
	between			0.328	2.522	6.290
	within			0.189	2.501	5.635
Firm Age	overall	80,638	8.961	8.043	0	70
	between			7.387	0	67.5

	within			1.468	5.961	11.961
Foreign Ownership	overall	80,638	0.262	0.435	0	1
	between			0.396	0	1
	within			0.048	-0.571	1.096
State Ownership	overall	80,638	0.179	0.099	0	1
	between			0.079	0	1
	within			0.039	-0.752	0.826

Notes: The unit of average wages is million VND a year. Regional variables are measured at a district and 3-digit VSIC level.

Abbreviation: FOEs, foreign-owned enterprises.

Table 2.4. Ownership structure of firms in the final sample

Firm type	State-owned	Domestic private-owned	Foreign-owned	Total
Number of observations	2,286	56,383	21,969	80,638
Share	2.835 %	69.921 %	27.244 %	100 %

Table 2.5. Correlation matrix of variables used in the baseline agglomeration estimation

	1	2	3	4	5	6	7	8	9	10
1. ln (Productivity)	1.0000									
2. ln (Employment density)	0.1985	1.0000								
3. Specialization	-0.0127	-0.4495	1.0000							
4. Diversity	0.1105	0.4426	-0.1803	1.0000						
5. Competition	-0.1161	0.2579	0.1973	0.1083	1.0000					
6. Regional Share of FOEs	0.1216	0.1664	-0.0406	-0.1172	0.0962	1.0000				
7. ln (Regional Average Wages)	0.2425	0.5174	-0.2699	0.2907	0.0698	0.3000	1.0000			
8. Firm Age	0.0626	0.1563	-0.0338	0.1213	-0.0286	-0.0860	0.1099	1.0000		
9. Foreign Ownership	0.1958	0.1403	0.0095	-0.0158	-0.0301	0.4192	0.2167	-0.0730	1.0000	
10. State Ownership	0.0257	0.0383	-0.0075	0.0610	-0.0870	-0.0635	0.0226	0.2666	-0.1089	1.0000

Notes: Regional variables are measured at a district and 3-digit VSIC level.

Abbreviation: FOEs, foreign-owned enterprises.

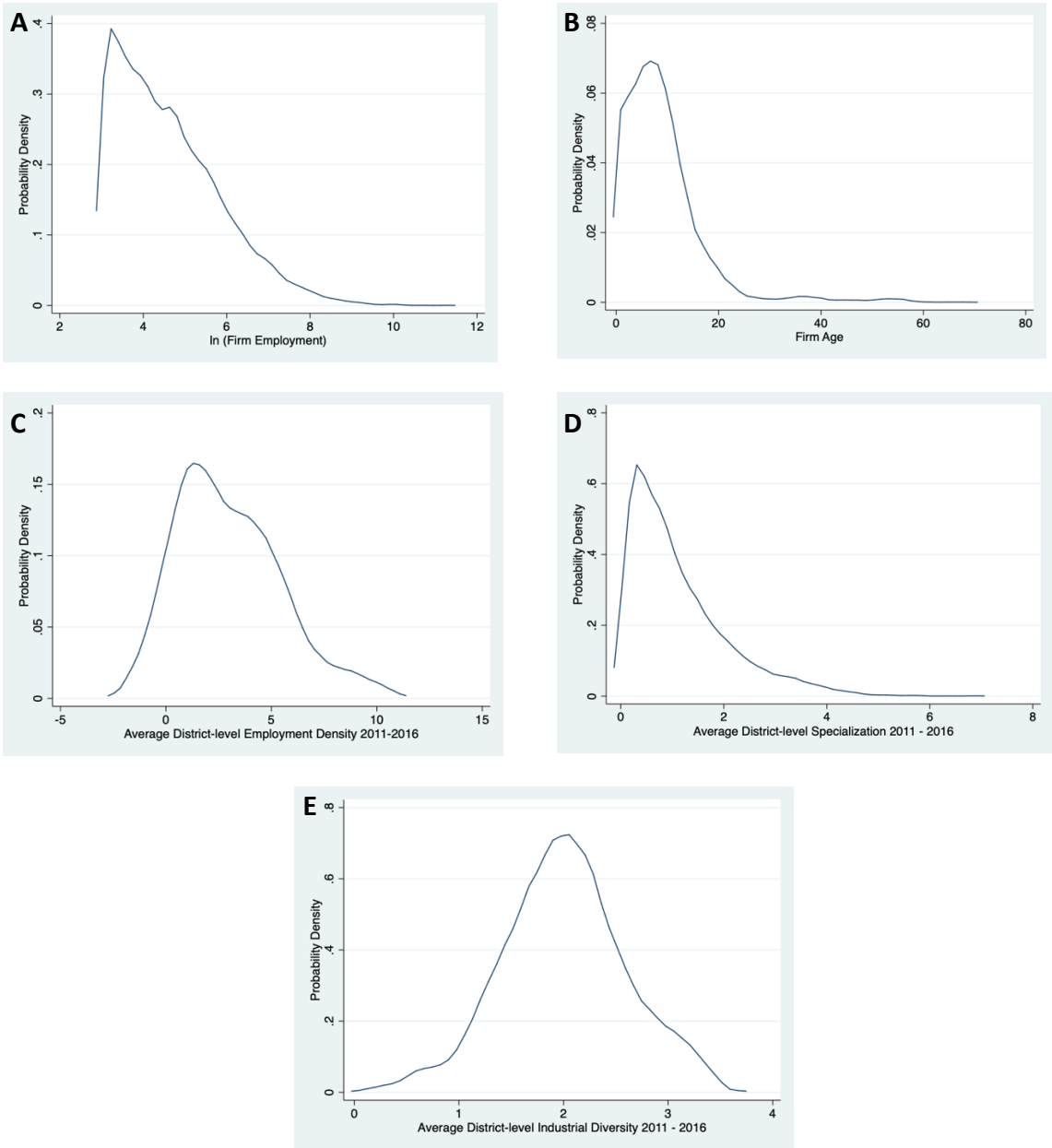

Figure 2.4. Kernel density distributions of key firm-level and region-level continuous variables

Notes: In graphs A and B, variables are firm-level. Region-level variables are averaged over a region during the period 2011-2016 in graphs C and E, and over an industry in a region during the time in graph D. In graphs C and E, each observation represents a certain district. In graph D, each observation represents a certain industry in a certain district.

As can be seen from the industrial structure of the final sample expressed in Table 2.6, the most influential industries are the ones producing food products, wearing apparel, non-metallic mineral products, and fabricated metal products (except machinery and equipment), which are only 4 out of 21 industries but make up about 46% of observations in total. According to the statistical classification of economic activities constructed by the European Community (NACE), these four industries are classified as either low-technology or medium-low-

technology, which implies that innovation or the pool of highly skilled specialized workers may be found less important for the Vietnamese manufacturing sector compared to other economies with the dominance of high-technology industries. This is confirmed again in Table 2.7, which shows the strong presence of low-tech and medium-low-tech industries in the estimation sample. In terms of spatial structure, the map of Vietnam in Figure 2.5 illustrates to what extent the disparity in annual average employment density is large between district-level regions during the 2011 – 2016 period. It can be seen that districts with density above 1,000 employees per square km are primarily located in or around the two biggest province-level regions of Vietnam - Hanoi and Ho Chi Minh City. The remaining densest regions are mainly located along the coastline from north to south.

Table 2.6. Industrial structure of the final sample – two-digit VSIC

2-digit VSIC	Manufacturing Industries at two-digit VSIC	No of Obs	No of Firms	Avg Obs per Firm
10	Manufacture of food products	10,989	3,525	3.1
11	Manufacture of beverages	898	284	3.2
13	Manufacture of textiles	3,701	1,139	3.2
14	Manufacture of wearing apparel	9,809	3,281	3.0
15	Manufacture of leather and related products	3,216	1,010	3.2
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	4,317	1,704	2.5
17	Manufacture of paper and paper products	3,649	1,130	3.2
18	Printing and reproduction of recorded media	1,811	682	2.7
20	Manufacture of chemicals and chemical products	3,232	1,035	3.1
21	Manufacture of pharmaceuticals, medicinal chemical, and botanical products	856	225	3.8
22	Manufacture of rubber and plastic products	5,843	1,920	3.0
23	Manufacture of other non-metallic mineral products	8,993	3,050	2.9
24	Manufacture of basic metals	1,921	668	2.9
25	Manufacture of fabricated metal products, except machinery and equipment	7,202	2,774	2.6
26	Manufacture of computer, electronic and optical products	1,883	673	2.8
27	Manufacture of electrical equipment	1,948	582	3.3
28	Manufacture of machinery and equipment n.e.c	1,365	444	3.1
29	Manufacture of motor vehicles; trailers and semi-trailers	1,027	277	3.7
30	Manufacture of other transport equipment	1,295	424	3.1
31	Manufacture of furniture	4,918	1,633	3.0
32	Other industries	1,765	527	3.3

Table 2.7. Industrial structure of the final sample – technological level

Technological Level	2-digit and 3-digit Manufacturing Industries	No of Obs	No of Firms
High-technology	21, 26, 303	2,747	900
Medium-high-technology	20, 252, 27, 28, 29, 302, 304, 309, 310, 325	13,416	4,226
Medium-low-technology	182, 22, 23, 24, 251, 259, 301	24,549	8,647
Low-technology	10, 11, 13, 14, 15, 16, 17, 181, 31, 32 excluding 325	43,046	14,169

Note: The high-tech classification of manufacturing is defined based on the Statistical classification of economic activities in the European Community (NACE)

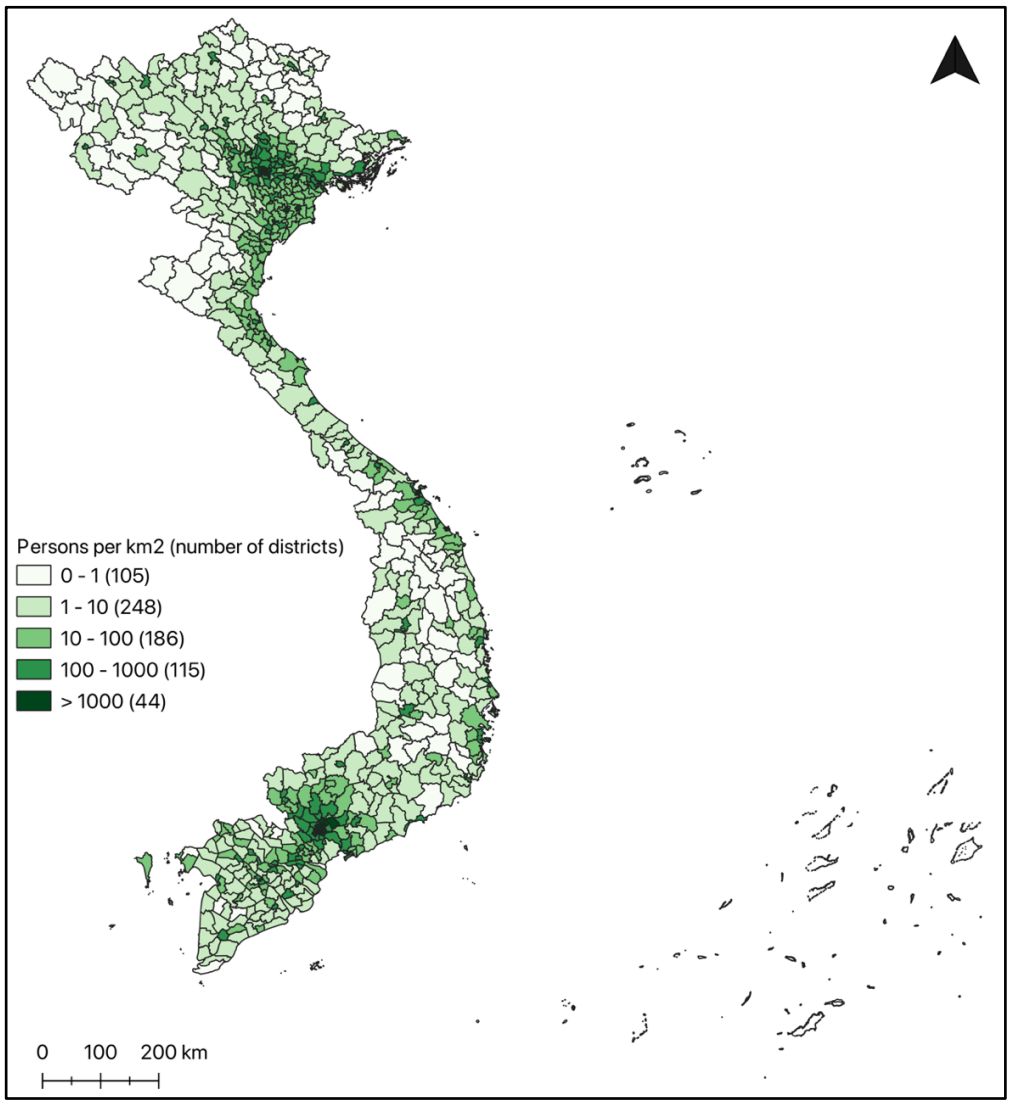

Figure 2.5. Map of Vietnam and annual district-level average employment density

Notes: The map is drawn using employment data from the censuses of enterprises conducted by the General Statistical Office from 2011 to 2016, land-area data is extracted from provincial-level statistical yearbooks, and spatial data is published by the Ministry of Natural Resources and Environment in Vietnam.

2.5. Empirical strategy

2.5.1. Baseline specification

To construct a specification to estimate the agglomeration effects, assume that firms in region r are homogenous. Following Combes & Gobillon (2015), a representative firm i in the region in sector s at time t produces goods according to a technology that is Leontief in intermediates inputs, denoted as M_{irt}, but is Cobb-Douglas in physical capital, labor, and local technology, denoted as K_{irst}, L_{irst}, and A_{rst} respectively. Therefore, the firm's production function is written as

$$Y_{irst} = \min\left(\frac{M_{irst}}{a}, A_{rst}K_{irst}^{\beta_k} L_{irst}^{\beta_l}\right), \qquad 0 < a, \beta_k, \beta_l < 1 \tag{2.37}$$

where a is the fixed share of intermediate inputs in output, and Y_{it} represents firm output. Importantly, A_{rt} is assumed to be a function of agglomeration forces as

$$A_{rst} = (DEN)_{rt}^{\beta_1}(SPE)_{srt}^{\beta_2}(DIV)_{srt}^{\beta_3} \tag{2.38}$$

where DEN, SPE, and DIV are the connotations of agglomeration calculated according to expressions (2.33), (2.36), and (2.34) respectively.

Given that M_{irt} is constantly proportional to Y_{it} when the firm scales up or down production, profit maximization of (2.37) gives

$$\ln\left(\frac{p_{irst}Y_{irst} - b_{irst}M_{irst}}{p_{irst} - ab_{irst}}\right) - \beta_k \ln(K_{irst}) - \beta_l \ln(L_{irst}) = \ln(A_{rst}) \tag{2.39}$$

where p_{irst} refers to the average revenue per unit of goods produced; b_{irst} is the price of intermediate inputs. The expression $\frac{p_{irst}Y_{irst}-b_{irst}M_{irst}}{p_{irst}-ab_{irst}}$ represents deflated value-added of the firm. The left-hand side of (2.39) is in fact the logarithm of value-added-based TFP, whose computation method is expressed in section 3 of this chapter. To estimate agglomeration economies, (2.39) is transformed further into the regression model

$$\begin{aligned} lnTFP_{it}^{sr} = \alpha + \beta_1 DEN_t^r + \beta_2 SPE_t^{sr} + \beta_3 DIV_t^{sr} + \beta_4 X_{1_{it}}^{sr} + \beta_5 X_{2_t}^{sr} + X_{3_t} \\ + f_i + \epsilon_{it} \end{aligned} \tag{2.40}$$

where $lnTFP_{it}^{sr}$ is the natural logarithm of TFP of firm i in region r in industry s at time t;

DEN, SPE, and DIV are the employment density, specialization, and diversity indices respectively as mentioned above; vectors $X_{1_{it}}^{sr}$ and $X_{2_t}^{sr}$ contain time-variant firm-level and regional controls respectively; vector X_{3_t} includes time, time-industry, and time-region fixed effects; f_i is firm fixed effects; and ϵ_{it} is an usual error term assumed to be uncorrelated with explanatory variables.

The firm-level controls included in the baseline specification are composed of foreign ownership and state ownership. This application is inspired by the findings in Waldkirch (2014) that domestic firms tend to grow faster owing to their restructuring after receiving investments from foreign investors. In addition, Douma et al. (2006), Huang & Shiu (2009), and Ongore (2011) provide evidence for India, Taiwan, and Kenya respectively that firms improve their performance when their percentages of state ownership are lower or when their percentages of foreign ownership are higher. It is worth emphasizing that the foreign ownership is a continuous and time-variant index referring to the fraction of a firm's property owned by foreigners, ranging from 0 to 1. The same definition applies to state ownership. Meanwhile, the ownership type, which is used later in the econometric models with interaction terms, reflects the dominant shareholders who own more than 50% of the firm. As a result, the ownership type is a categorical variable with three different values: state-owned, privately domestic-owned, and foreign-owned.

Turning to controls at the region level, they comprise industrial competition strength, regional fraction of FOEs, and the logarithm of regional average wages. The first one is employed following Glaeser et al. (1992) who find the significant evidence of the positive impact of competition on regional employment growth. Its underlying mechanism is suggested by Porter (1990, Chapter 2) that fierce competitiveness gives firms incentives to innovate and become more efficient in order to survive and grow. In a follow-up study, Porter (2000) describes competition as a channel through which the regional agglomeration, in particular specialization, generates the externalities. The rationale for this relation is that regional competition emerges and becomes stronger when regional specialization is formed by more firms. Porter (2000) argues that locating next to others in the same business environment gives firms even more incentives to find innovative ways of boosting their productivity to outperform not only their local competitors but also their neighbors who work in other industries. To capture this effect, the regional competitive intensity is measured based on the structure of the employment market as

$$COM_t^{sr} = -ln\left[\sum_i \left(\frac{e_{it}^{sr}}{e_t^{sr}}\right)^2\right] \tag{2.41}$$

When there is only a firm operating in the market, competition is absent, correspondingly COM is equal to zero as the minimum value. The degree of competition rises with the emergence of more firms in the same industry s in region r and the expansion of their workforce, which makes the employment structure more balanced, resulting in a rise in COM. In contrast, when the employment structure in industry s in region r becomes less and less balanced due to the stronger development of only a modest fraction of firms, COM is decreasing, implying a less competitive market.

Despite the reasoning of Porter (2000) about the strong relationship between COM and SPE, local competition has a potential impact on productivity through incentives of firms rather than three key mechanisms through which the agglomeration produces the external benefits as analyzed earlier. Therefore, this study treats COM only as a regional control to explicitly take into account its possible productivity influences embodied in agglomeration effects. Next, based on the persuasive evidence found in Takii (2005) and Newman et al. (2015), the second regional control used in the model is the share of regional firms owned by foreigners, used to control for technological spillovers from advanced foreign firms to domestic firms as well as to other local FOEs. The third control is the logarithm of regional average wages added to proxy for regional stock of human capital, which is shown to play an important role in local growth in Acs & Armington (2004).

2.5.2. Econometric issues of agglomeration estimation

The regression of firm-level productivity on regional agglomeration under the existence of several factors may produce estimators which deviate significantly from their true values, as discussed at length in Henderson (2003) and Combes & Gobillon (2015). The first possible factor is the existence of unobserved variables that have some explanatory power to productivity but are correlated to regional variables of interest. At the firm level, some characteristics beyond the data set such as the excellence of director board, productive working culture, the high share of skilled workers, or the application of state-of-the-art technology may help to induce TFP of a firm since whose advantages can produce higher value added compared to its business fellows even though they all manufacture with the same combination of physical capital and labor input. If these firm-level attributes are exogenous to the explanatory variables

on the right-hand side of (2.40), especially agglomeration proxies, the absence does not cause any trouble. However, in fact, highly capable managers tend to flock in cities, specific-skilled workers are more commonly found in specialized regions, and the development of supporting technology is spurred more greatly in industrial-diverse regions. As a result, such firm-level unobserved factors may bias upwards the estimation results of the externalities.

At the region level, there are many possible factors potentially endogenous to the external terms. For example, compared to other regions, metropolitan or specialized regions often have better infrastructure such as broader roads, easier access to airports and seaports, and the availability of industry-specific infrastructure, which greatly facilitate the production and business activities of local firms. Another case is that capable urban authorities might know the bright side of spatial proximity, thus they spur agglomeration simultaneously with making effective economic policies to improve productivity of firms in their regions. Furthermore, from the temporal perspective, any urban or industrial shocks that have a meaningful impact on performance of local firms also are able to push estimators of agglomeration away from their true estimate. Without controlling for these unobserved regional conditions, it is hard to separate their effects from the agglomeration externalities, thus their estimators may be biased as a consequence. The next econometric problem is the endogenous location choice or reverse causality in other words, which may remain even if the missing factors mentioned above are measurable and included in the specification. The reason is that productive firms could choose to locate in agglomerated regions to benefit from the externalities, to make use of well-developed regional infrastructure, or to gain from business-friendly policies there. This unobserved sorting behavior could bias the estimation results of agglomeration in the same ways as omitted variables probably cause. The final possible problem is the attrition effects. It arises from the competitive environment of agglomerated regions where productive firms survive, while their unproductive fellows leave the market after their bankruptcy. If this is the case, the evidence of the productivity-improving effects in cities may actually be the outcome of a selection process rather than, say, of technological spillover effects.

A solution to the econometric problems triggered by region-level factors as mentioned above is to instrument for proxies of agglomeration with historical variables, pioneered by Ciccone & Hall (1996) who make use of data on railway and population a century before, and followed by Combes et al. (2010) and Duranton (2016). However, this strategy suffers from several major limitations: data for such long-lagged variables is rarely available, especially in developing countries; when it is available, it is only suitable to instrument for the current

urbanization, leaving industrial-related variables such as specialization untouched[16]; since instrumental variables (IVs) of this kind are time-invariant, to make them usable for panel data, practitioners must sacrifice its time dimension by doing regression with all time-averaged variables, or making use of data of only one certain year. Another strategy in literature without such limitations is to first difference data to deal with omitted variables at both firm and region levels before the differenced agglomeration measures are instrumented with agglomeration variables lagged two or three years to tackle the remaining econometric problems. This is the core idea of the FD-GMM estimation, applied by Henderson (2003) and Martin et al. (2011), which is superior to the use of historical information in the way that the possible endogeneity of every explanatory variable is taken care of using its own past information in the data set. Nevertheless, given the unbalanced nature of micro-level data used in the strand of literature, the application of this technique requires observations to be present in the data set for at least 3 to 4 consecutive years. In other words, mainly survivors are accounted for in the regression sample using the FD-GMM method. This tends to sharply reduce sample size, and thus make the sample more balanced artificially. If the reasons behind firms' survivability are linked to both regional characteristics and years-lagged instruments, the estimated results are not free from biases (Levinsohn & Petrin, 2003). Combes & Gobillon (2015) argue that it is hard to guarantee the exogeneity of short-lagged instruments when both the instruments and the endogenous regional variables are extracted from the same data, therefore they recommend against the use of FD-GMM in agglomeration estimation.

To tackle the potential problems of endogeneity in estimating the externalities but avoid the weaknesses of the methods mentioned above, this study combines the fixed-effects technique applied in Henderson (2003) and Combes & Gobillon (2015) with several additional statistical measures. Specifically, to deal with omitted variables, the firm fixed effects term f_i is inserted into the specification (2.40). By removing observations of firms changing their location or industry during sampling periods as well[17], the term f_i does not only capture all time-invariant firm-level influential characteristics but also constant regional and industrial factors. The elimination of relocating and industry-changing firms is also useful in the sense that the regression does not have to face the potential endogeneity of their location and industry choices, therefore alleviating the sorting problem. Next, vectors ${X_1}_{it}^{sr}$ and ${X_2}_{t}^{sr}$ which contains

[16] The rationale is that, unlike regional characteristics, industrial structure in the distant past is often not helpful to explain the structure at present.

[17] This implementation drops about 11.8% of total observations of the sample at the stage of cleaning, whose influence on the sample size is still very modest compared to the act of keeping only survivors in the data set.

time varying controls are placed in (2.40) to mitigate the problems caused by unobserved time-variant variables. The combination between the FE terms and firm-level controls also help remedy the problem of the sorting behavior (Combes & Gobillon, 2015) and partially the attrition effects. Finally, to control for temporal shocks at the province level and the 2-digit industry level, the FE terms of year, year-industry, and year-province embodied in the vector X_{3_t} are also inserted into (2.40). In essence, these FE terms capture the annual trend of TFP of all firms across different districts within each province and across different 3-digit industries within each 2-digit industry, thus take into account the effects of, for instance, a massive improvement in province-level infrastructure or a technological shock to the whole industry on local firms' productivity.

Besides the above arguments, this study is in favor of the FE technique due to following additional reasons: when applying the Hausman test to the baseline estimation, random effects estimates are rejected in favor of the FE method; Henderson (2003) concludes that the application of firm FE along with other FE terms to deal with potential problems of endogeneity is superior to and more influential than the use of the FD-GMM technique; The regression is also implemented with the FD-GMM method as a robustness check, however the obtained results are highly changeable across different ways to construct the specification and show all unrealistically high magnitudes of the externalities, which may stem from the slow changes of regional variables over time in combination with the strongly unbalanced but rather short data set, therefore the FD-GMM results are beyond the report of this study; Combes et al. (2012) put considerable efforts towards taking explicitly the issue of attrition into account and they receive almost-unchanged estimation results; Gokan et al. (2019), following the test strategy of Combes et al. (2012), find no significant distortion from the selection effects in Vietnam; Finally, as a sensitivity check, the baseline results of this study are compared to the ones from the adjusted sample with two biggest province-equivalent cities in Vietnam (Hanoi and Ho Chi Minh) excluded. If the potential bias caused by the selection effects and reserve causality is serious, the primary estimates would change dramatically in terms of their magnitudes or even signs. However, estimated results in section 7 confirm the robustness of estimators yielded from regression with such a truncated sample. On the whole, the combination of multi-level fixed effects, controls, and the removal of moving firms is considered to be sufficient to tackle the potential endogenous issues in estimation of agglomeration in this study.

2.6. Primary Results

2.6.1. Production function results

Table 2.8 details coefficients of production input elasticities for 21 separate 2-digit industries using various strategies as expressed earlier. Considering the estimation results using the baseline method of LP-W, all elasticities are positive and significant and vary from industry to industry. Regarding capital input, its lowest estimated elasticities are 0.04 and 0.075 belonging to the business lines producing clothing (coded as 14) and leather products (coded as 15) respectively – labor intensive industries, while two highest values are 0.476 for the manufacture of rubber and plastic products (coded as 22) and 0.573 for the production of beverages products (coded as 11). Turning to the role of labor input across various industries, output is least sensitive to changes in labor quantity in the industry of electrical equipment (coded as 27) with the labor elasticity of 0.503. In contrast, output responses most strongly to a change in workforce in the industry of clothing with the coefficient of 1.026, which is unsurprising because of the high demand for manual workforce over machinery in this industry. Moreover, this is the only industry with the labor elasticity above 1 along with the capital estimator of 0.040 – the lowest elasticity as mentioned above. In the remaining industries, estimated elasticities of the two inputs fall into the middle range of the two above extremes. To wrap up, the remarkable heterogeneity of estimated coefficients highlights the importance of performing separate regressions for various industries, as conducted in this study, rather than grouping all observations into a single sample for the production function estimation.

Table 2.8. Results from production function estimation

2-digit VSIC	OLS		OP		LP		LP-W	
	$\hat{\beta}_l$	$\hat{\beta}_k$	$\hat{\beta}_l$	$\hat{\beta}_k$	$\hat{\beta}_l$	$\hat{\beta}_k$	$\hat{\beta}_l$	$\hat{\beta}_k$
10	0.800***	0.396***	0.672***	0.264***	0.706***	0.264***	0.696***	0.282***
	(0.010)	(0.006)	(0.017)	(0.086)	(0.015)	(0.054)	(0.011)	(0.022)
11	0.807***	0.630***	0.673***	0.706***	0.608***	0.519***	0.492***	0.573***
	(0.053)	(0.025)	(0.105)	(0.101)	(0.096)	(0.125)	(0.049)	(0.081)
13	0.781***	0.330***	0.701***	0.302***	0.722***	0.264***	0.750***	0.241***
	(0.015)	(0.007)	(0.030)	(0.068)	(0.026)	(0.055)	(0.016)	(0.038)
14	0.977***	0.122***	0.962***	0.084***	0.920***	0.052***	1.026***	0.040**
	(0.009)	(0.005)	(0.015)	(0.024)	(0.014)	(0.015)	(0.009)	(0.016)
15	0.857***	0.164***	0.859***	0.146***	0.800***	0.068**	0.826***	0.074***
	(0.012)	(0.008)	(0.025)	(0.046)	(0.022)	(0.030)	(0.012)	(0.024)
16	0.845***	0.303***	0.771***	0.364***	0.781***	0.201***	0.801***	0.225***
	(0.017)	(0.008)	(0.031)	(0.042)	(0.026)	(0.060)	(0.020)	(0.028)
17	0.910***	0.303***	0.833***	0.145	0.833***	0.165**	0.858***	0.172***
	(0.018)	(0.010)	(0.032)	(0.105)	(0.027)	(0.065)	(0.019)	(0.042)

18	0.834***	0.313***	0.732***	0.230**	0.688***	0.257***	0.688***	0.269***
	(0.022)	(0.010)	(0.048)	(0.108)	(0.040)	(0.075)	(0.026)	(0.047)
20	0.733***	0.470***	0.652***	0.307***	0.688***	0.287***	0.630***	0.329***
	(0.021)	(0.011)	(0.040)	(0.090)	(0.036)	(0.073)	(0.021)	(0.047)
21	0.895***	0.421***	0.779***	0.298**	0.791***	0.301***	0.635***	0.336***
	(0.038)	(0.025)	(0.083)	(0.122)	(0.077)	(0.086)	(0.039)	(0.072)
22	0.775***	0.346***	0.660***	0.381***	0.660***	0.453***	0.592***	0.476***
	(0.013)	(0.008)	(0.027)	(0.054)	(0.025)	(0.047)	(0.014)	(0.035)
23	0.892***	0.334***	0.742***	0.323***	0.776***	0.302***	0.766***	0.337***
	(0.012)	(0.006)	(0.020)	(0.051)	(0.019)	(0.050)	(0.014)	(0.023)
24	0.851***	0.429***	0.696***	0.542***	0.760***	0.216**	0.716***	0.211***
	(0.035)	(0.017)	(0.058)	(0.104)	(0.054)	(0.105)	(0.045)	(0.053)
25	0.821***	0.344***	0.660***	0.404***	0.702***	0.437***	0.673***	0.301***
	(0.013)	(0.006)	(0.026)	(0.078)	(0.024)	(0.099)	(0.015)	(0.031)
26	0.793***	0.270***	0.717***	0.320***	0.700***	0.294***	0.669***	0.326***
	(0.021)	(0.013)	(0.038)	(0.077)	(0.034)	(0.056)	(0.020)	(0.050)
27	0.683***	0.388***	0.613***	0.114	0.613***	0.191**	0.553***	0.221***
	(0.023)	(0.015)	(0.039)	(0.078)	(0.038)	(0.093)	(0.022)	(0.056)
28	0.847***	0.313***	0.863***	0.279***	0.809***	0.179	0.772***	0.215***
	(0.030)	(0.016)	(0.056)	(0.073)	(0.060)	(0.113)	(0.035)	(0.074)
29	0.796***	0.399***	0.608***	0.513***	0.718***	0.485***	0.654***	0.380***
	(0.037)	(0.026)	(0.075)	(0.120)	(0.069)	(0.117)	(0.036)	(0.096)
30	1.069***	0.242***	0.920***	0.231	0.904***	0.125	0.972***	0.157*
	(0.031)	(0.018)	(0.064)	(0.141)	(0.060)	(0.138)	(0.043)	(0.085)
31	0.884***	0.200***	0.789***	0.183***	0.811***	0.196***	0.907***	0.216***
	(0.011)	(0.007)	(0.025)	(0.040)	(0.019)	(0.016)	(0.015)	(0.026)
32	0.795***	0.273***	0.724***	0.184	0.724***	0.136*	0.706***	0.155***
	(0.019)	(0.012)	(0.034)	(0.119)	(0.029)	(0.081)	(0.019)	(0.052)

Notes: Standard errors in parentheses of OLS, bootstrapped standard errors in parentheses of OP, LP, and LP-W

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 2.9. Correlation matrix of TFP values yielded using different methods

	1	2	3	4
1. Logarithm of OLS TFP	1.0000			
2. Logarithm of OP TFP	0.6699	1.0000		
3. Logarithm of LP TFP	0.6812	0.8046	1.0000	
4. Logarithm of LP-W TFP	0.7379	0.7669	0.8937	1.0000

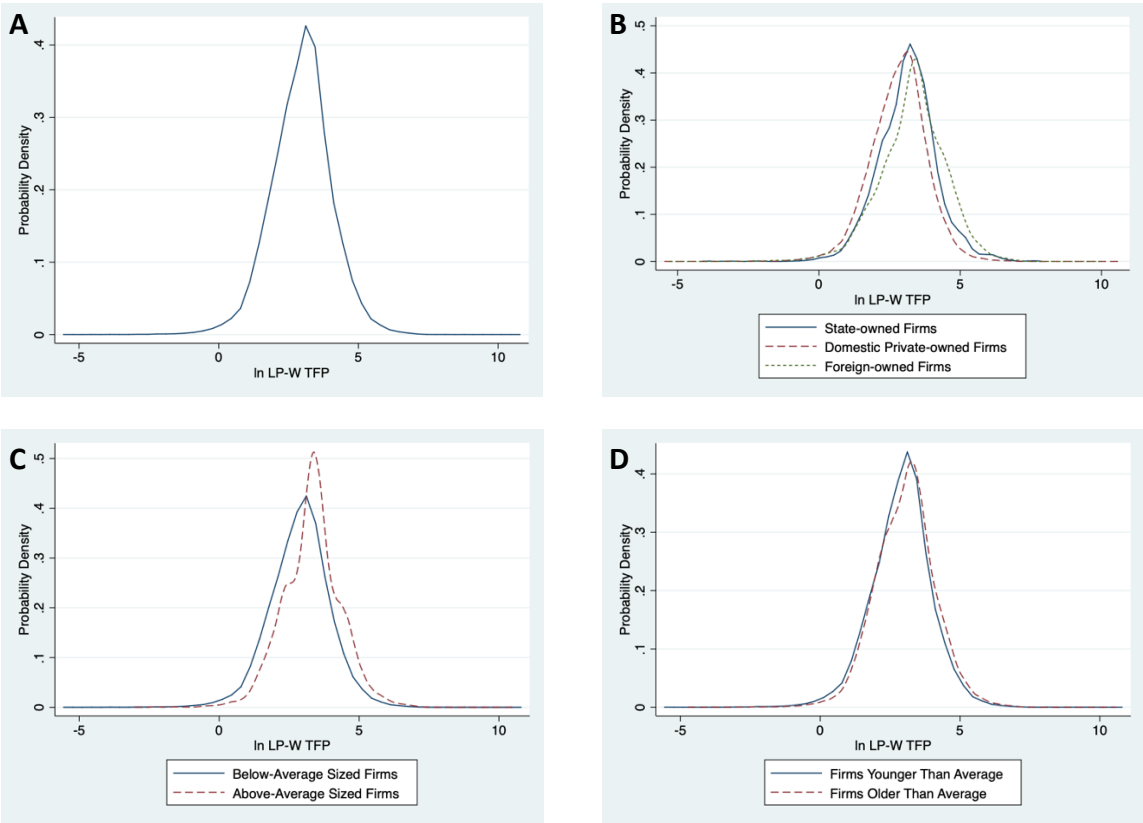

Figure 2.6. Kernel density distributions of LP-W TFP by different firm-level characteristics

Notes: In graph C, the average firm size is 304.2053, calculated using the full final sample; In graph D, the average age is 8.961, calculated similarly.

Next, it is worth looking at Table 2.8 to compare the outcomes yielded using LP-W to the remaining methods. When compared to all three control function approaches, OLS input estimators show a systematic disparity in their magnitude. As discussed earlier, due to the positive correlation between productivity and production inputs, uncontrolled productivity embodied in the error term could bias input parameters of the production function upwards, and ultimately underestimate real firm performance. This prediction is confirmed in this study as can be seen from Table 2.8 because on average the OLS estimators of capital and labor are about 10.8% and 6.9% higher respectively compared with their LP-W versions. Similarly, the OLS elasticities tend to be higher than their OP and LP counterparts. Obviously, it was anticipated that using OLS productivity would underestimate the magnitude of the agglomeration externalities as a consequence. In terms of LP versus LP-W, except for some outliers, the latter tends to produce smaller $\hat{\beta}_l$ but larger $\hat{\beta}_k$, which is a reasonable outcome given that the LP-W method is applied to correct the potential upward bias of $\hat{\beta}_l$ facing LP. Meanwhile, the OP estimators can be lower or higher than their LP and LP-W corresponding coefficients depending on specific industries. This sort of compensation leads to the trivial

disparity of productivity values using each of the trio LP, OP and LP-W, and explains why their correlation coefficients are very high, especially between LP and LP-W, as detailed in Table 2.9. Nonetheless, keeping in mind that the LP-W technique is still considered to produce the most consistent estimates of productivity, and thus a consistent agglomeration estimation. Finally, graph A in Figure 2.6 shows the probability distribution of LP-W TFP of observations in the full sample, while graphs B, C, and D illustrate the distribution for observations from firms of different characteristics which are key to the analysis on the heterogeneous effects of agglomeration conducted later. As shown by these graphs and the t-test on the statistical difference between various distributions, firms that are foreign-owned, large-sized, or old tend to be more productive than their remaining fellows.

2.6.2. Baseline estimates

This subsection presents and discusses baseline results on the agglomeration estimation, detailed in Table 2.10. Columns (1) to (4) show univariate and multivariate estimates of agglomeration proxies in the specification without any control variables, however their interpretation is suppressed until the full specification in column (8) is mentioned. Basically, the significance levels of all three connotations of agglomeration remain unchanged moving from the univariate regressions to the multivariate one, while their magnitudes increase considerably, particularly for DEN and DIV – the ones with statistically significant results. This increase stems from a statistical fact that although density is positively correlated to diversity across districts and years as shown earlier in Table 2.4, their annual changes go in an opposite direction, which can be seen indeed through the correlation coefficient of -0.2228 between demeaned values of DEN and DIV, bearing in mind that the FE technique exploits only these temporal changes rather than cross-firm and cross-region differences. In other words, when districts become denser, their industrial structure tends to be less diverse in Vietnam between 2011 and 2016. Therefore, estimating DEN in the fixed-effects context without the inclusion of DEN will introduce downward bias to the estimator of DIV, and vice versa for the case of DEN. This suggests the importance of entering various proxies of agglomeration into the specification for more accurate estimation.

Table 2.10. Agglomeration effects on firm productivity, District/3-digit VSIC

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables of interest	DEN	0.084*** (0.025)	-	-	0.136*** (0.028)	0.135*** (0.028)	0.129*** (0.028)	0.166*** (0.029)	0.129*** (0.028)
	SPE	-	0.014 (0.018)	-	0.018 (0.019)	0.018 (0.019)	0.014 (0.019)	0.003 (0.021)	0.015 (0.019)
	DIV	-	-	0.047*** (0.017)	0.083*** (0.020)	0.083*** (0.020)	0.082*** (0.020)	0.073*** (0.020)	0.077*** (0.020)
Firm controls	Foreign ownership	-	-	-	-	0.110** (0.054)	-	0.083 (0.054)	0.107** (0.054)
	State ownership	-	-	-	-	0.026 (0.054)	-	0.037 (0.056)	0.025 (0.054)
Regional controls	Competition	-	-	-	-	-	0.043*** (0.011)	0.053*** (0.012)	0.043*** (0.011)
	Fraction of FOEs	-	-	-	-	-	0.244 (0.167)	0.223*** (0.044)	0.242 (0.167)
	Regional average wages	-	-	-	-	-	0.196*** (0.028)	0.197*** (0.027)	0.196*** (0.028)
	Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	FE of time-industry and time-region	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
	Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	No. of obs	80,638	80,638	80,638	80,638	80,638	80,638	80,638	80,638
	No. of firms	26,987	26,987	26,987	26,987	26,987	26,987	26,987	26,987
	F-Test	11.165***	0.550	7.493***	10.931***	7.383***	15.580***	13.898***	11.925***
	R^2	0.071	0. 070	0. 070	0. 071	0. 071	0. 073	0. 048	0. 073

Notes: Cluster-robust standard errors in parentheses calculated at a province and 2-digit VSIC level.

Abbreviation and denotation: FE, fixed effects; ln LP-W TFP, the logarithm of total factor productivity computed using the method of Woodridge (2009); DEN, employment density; LQ, location quotient; DIV, diversity; COM, competition; FOEs, foreign-owned enterprises.

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Next, control variables added in succession in columns (5) and (6) of Table 2.10 to see how their inclusion influences the estimated externalities. It is well recognized that estimated coefficients for agglomeration proxies remain stable with the addition of various controls. In particular, two parameters of urbanization are always positive and significant whereas that of specialization remains statistically insignificant. Regarding magnitude, allowing for control factors leads to a minor decrease in the strength of computed agglomeration impacts. It is worth mentioning that, to avoid the risk of producing artificially inflated standard errors (SEs) when micro-level explained variables are regressed on regional regressors, SEs are clustered at a province and 2-digit VSIC level throughout all regressions on agglomeration in this study, following the guide of Moulton (1990) and Cameron et al. (2015).

As regards columns (7) and (8) of Table 2.10, they are alike in the sense that all controls are included, but the FE terms of time-industry and time-region are dropped from (7) in order to emphasize the importance of controlling unobserved regional and industrial shocks which tend to happen in agglomerated areas and are likely positive in fast-growing economies. Moving from (7) to (8) with the complete set of FE, the magnitude of DEN is declined by approximately 22.3%, while it is almost unchanged for DIV. The fall in the estimated coefficient of DEN is in line with the anticipation analyzed above, and thus underlines the existence of upward bias when regional shocks are not taken into consideration. All in all, the results in column (8) confirm the presence of urbanization economies in Vietnam. In general, a double in employment density leads to a gain of approximately 12.9% in the productivity of local firms, holding other factors fixed. This conclusion is consistent with the findings of other authors in literature, such as Rosenthal & Strange (2004) and Combes et al. (2010), despite differences in specification and the ways to construct regional variables. When compared with US and European counterparts where productivity elasticities to urbanization vary between 3 and 8% (Rosenthal & Strange, 2004), the density elasticity is a bit larger for Vietnam. This difference, again, supports the observation of Combes & Gobillon (2015) about the urbanization benefits in developed countries versus developing countries. In terms of industrial variety – the relative aspect of urbanization, the estimated results verify its positive influence on productivity with a statistically significant coefficient of 0.077, which echoes the ideas of Jacobs (1969, p. 43) and the findings of Glaeser et al. (1992).

As for the specialization externalities, although its estimator is positive as expected, it is statistically insignificant, implying an ambiguous role of industrial concentration on TFP of local firms. This outcome might be driven by the dominance of low-technology industries and manual work in the Vietnamese manufacturing sector, whose growth relies less on a large local specialized labor force. The underdevelopment of supporting industries in Vietnam as noted by Tuan & Yoshi (2010) is an additional reason. Intuitively, locating in a specialized region becomes less meaningful for a firm if it eventually has to import the majority of its intermediate inputs from, say, China and South Korea for its massive production. To sum up, the estimation results demonstrate the existence of agglomeration economies in Vietnam and suggest inter-industry rather than intra-industry relationships as the main driver of the external benefits, and thus approves *Hypothesis 1.* This is in line with the findings of Au & Henderson (2006) and Combes et al. (2013) on the strong influence of urbanization economies in China – another emerging country. The last point to analyze in this subsection is estimators of controls. The

estimates show that a firm performs better with the more intense presence of foreign investors, while its TFP is not enhanced significantly with the higher control level of the state, as expected. In terms of regional controls, their coefficients display all positive signs and significant impacts as predicted, except for the fraction of FOEs in the region. These results support the findings of Glaeser et al. (1992) on the presence of Porter externalities and the findings of Acs & Armington (2004) on positive influence of local human capital, but do not confirm the spillover effects generated by regional FOEs.

2.6.3. Interaction between agglomeration effects and firm characteristics

As the positive influences of agglomeration have been proved, the question now is how firms of different characteristics gain differently from these externalities. This is dealt with technically by regressing the specification (2.40) again with the addition of interaction terms of a firm-level characteristic of interest with agglomeration variables. The estimated results are displayed in Table 2.11. Considering the case where agglomeration economies are heterogeneous across firms of various ownerships, a categorical variable of ownership is inserted into (2.40) along with its product terms with DEN, SPE, and DIV using the effects on domestic private-owned firms as the benchmark. Among the estimates of these terms in column (1), only the interactions of foreign-owned firm types with DEN and SPE are statistically significant. The positive coefficients imply that FOEs receive more external benefits than SOEs and POEs do when their production region becomes denser or their industry is more localized. The insignificant interactions of state-owned status with agglomeration proxies indicate that SOEs do not gain from agglomeration more than POEs, which suggests that the comparison between DOEs and FOEs may be more meaningful. Therefore, in column (2) of Table 2.11, the variable of ownership status is transformed into a binary variable which equals 1 for domestic-owned status as the base group and 2 for foreign-owned. The resulting estimated parameters are almost similar to what they are in column (1), implying the advantage of FOEs over DOEs in reaping benefits from a more agglomerated region. Specifically, FOEs are predicted to be 18.6 percentage points more productive relative to DOEs when local employment density doubles. Plotted using the results displayed in column (2), graph A in Figure 2.7 presents these disparities intuitively with a more inclined line for FOEs in the relationship between density and productivity. Interestingly, the graph also shows that although productivity of DOEs lags behind FOEs in high-density regions, the former wins over the latter in low-density regions. Notice that the main effects of SPE remains insignificant despite its significant product term with the foreign-owned status, therefore its graph is not plotted

because the slopes of resulting marginal lines are likely unreliable. In brief, these estimation results are in favor of *Hypothesis 2* and thus echo the conclusion of Howard et al. (2014) and Gokan et al. (2019) about the advantage of FOEs in Vietnam.

Table 2.11. Agglomeration effects with interaction terms of agglomeration and firm characteristics, District/3-digit VSIC

		(1)	(2)	(3)	(4)
Variables of interest	DEN	0.075*** (0.028)	0.075*** (0.028)	0.128*** (0.028)	0.153*** (0.029)
	SPE	-0.024 (0.021)	-0.022 (0.021)	0.010 (0.019)	0.053** (0.022)
	DIV	0.079*** (0.022)	0.079*** (0.022)	0.086*** (0.020)	0.064*** (0.024)
Interactions with state-owned ownership	DEN * Ownership type = State-owned	0.009 (0.020)	-	-	-
	SPE * Ownership type = State-owned	0.045 (0.046)	-	-	-
	DIV * Ownership type = State-owned	-0.006 (0.051)	-	-	-
Interactions with foreign-owned ownership	DEN * Ownership type = Foreign-owned	0.186*** (0.024)	0.186*** (0.024)	-	-
	SPE * Ownership type = Foreign-owned	0.119*** (0.033)	0.117*** (0.033)	-	-
	DIV * Ownership type = Foreign-owned	-0.021 (0.036)	-0.021 (0.036)	-	-
Interactions with firm size	DEN * Firm size	-	-	-0.000 (0.000)	-
	SPE * Firm size	-	-	0.000 (0.000)	-
	DIV * Firm size	-	-	-0.00003** (0.00001)	-
Interactions with firm age	DEN * Firm age	-	-	-	-0.008*** (0.001)
	SPE * Firm age	-	-	-	-0.005*** (0.002)
	DIV * Firm age	-	-	-	-0.0001 (0.002)
Firm-level variables	Ownership type = State-owned	-0.0624 (0.193)	-	-	-
	Ownership type = Foreign-owned	-1.048*** (0.191)	-1.046*** (0.192)	-	-
	Firm size	-	-	0.0001* (0.0001)	-
	Firm age	-	-	-	0.056*** (0.014)
	FE of time, time-industry, and time-region	Yes	Yes	Yes	Yes
	Firm FE	Yes	Yes	Yes	Yes
	No. of obs	80,638	80,638	80,638	80,638
	No. of firms	26,987	26,987	26,987	26,987
	F-Test	11.87***	16.36***	8.20***	10.52***
	R^2	0.075	0.075	0.074	0.075

Notes: Cluster-robust standard errors in parentheses calculated at a province and 2-digit VSIC level. Each column of this table replicates column (8) of Table 2.10, adding the full interaction terms of agglomeration variables and a certain firm characteristic, including: a categorical variable of ownership type with domestic private-owned type as the base group in column (1); a binary variable of ownership type with domestic-owned type as the base group in column (2); and firm size and firm age in columns (3) and (4) respectively. The control variables *state ownership* and *foreign ownership* in columns (1) and (2) are removed from the regressions due to their strong correlation with interaction terms. Control variables unrelated to the interactions are not shown for brevity.

Abbreviation and denotation: FE, fixed effects; DEN, employment density; SPE, specialization; DIV, diversity.

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

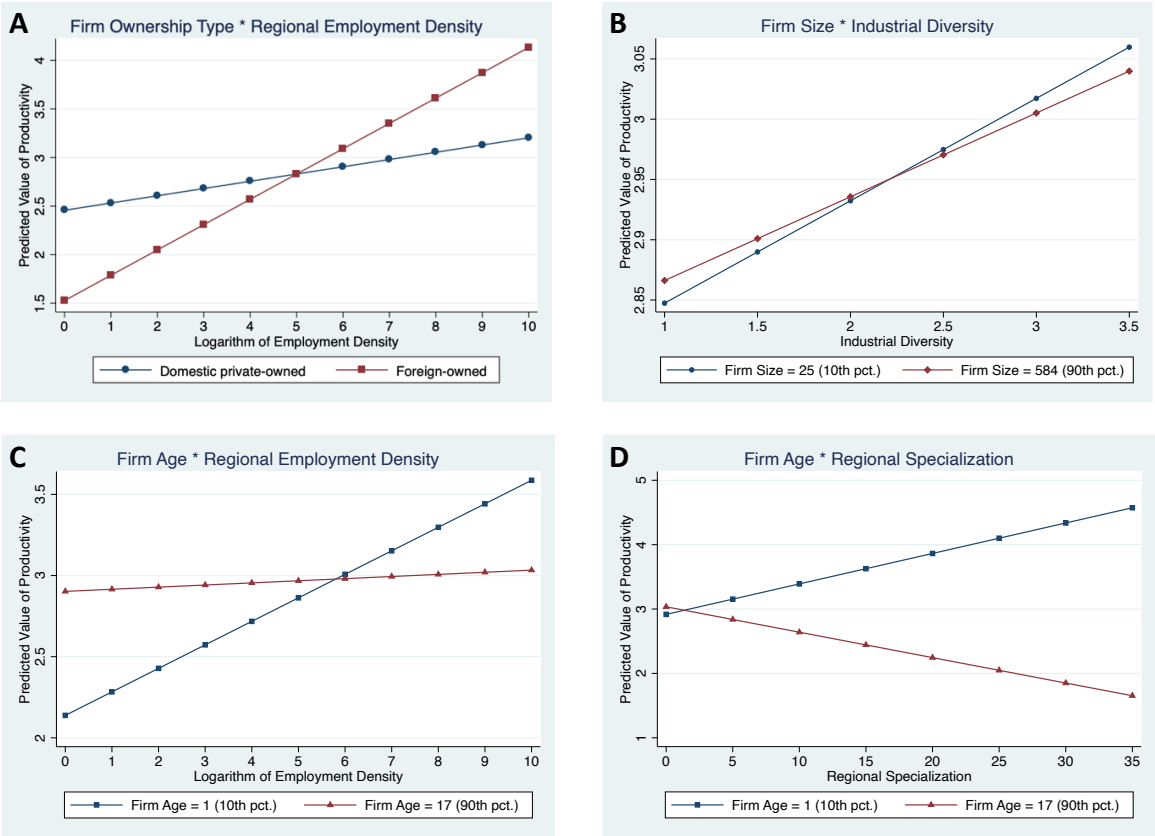

Figure 2.7. Graphs of several interactions between agglomeration and firm characteristics

Notes: To generate each figure, the predicted value of the logarithm of LP-W total factor productivity on the y-axis is yielded using estimated coefficients and the mean value of the remaining explanatory variables. This predicted value is then plotted separately for two different subsamples against one of agglomeration variables on the x-axis. Graphs A and B are produced following the regression results shown in column (2) of Table 2.11, while graphs C and D are based on columns (3) and (4) respectively.

Abbreviation; pct., percentile.

Turning to the possibility that the external benefits change across different firm sizes. Column (3) of Table 2.11 shows such evidence with the introduction of the firm employment scale and its multiplicative terms with DEN, SPE, and DIV to the specification. Of three terms, only the one of employment size with industrial variety is statistically significant, whose coefficient is -0.00003 at the 95% level of confidence, meaning that a firm benefits more from Jacob

externalities by a coefficient of 0.003 compared to itself or another firm whose workforce is 100 workers larger. In a nutshell, smaller-sized firms gain a merely marginal benefit when the industrial composition in their region becomes more varied, given that the main effect of DIV is much stronger with the estimator of 0.086. This minor difference is also seen in graph B of Figure 2.7 which shows productivity gains from diversity between firms of the 10th versus 90th percentiles of the size distribution. The upward-sloping line with a slightly lower gradient for larger-scale firms reflects the negative estimate as well as the modest magnitude of the product term. This result is consistent with the findings of Rosenthal & Strange (2010) and Martin et al. (2011), and therefore proves *Hypothesis 3*, despite that the productivity differential is small and only industrial variety is meaningful to the comparison.

Finally, the results in column (4) of Table 2.11 verify the heterogeneity in agglomeration impacts over firms of different ages with negative and statistically significant coefficients at the 99% confidence level of the terms of ages interacted with DEN and SPE. Surprisingly, the main effects of SPE now are also significant, which validates the statistical inferences for both these connotations of agglomeration in their relationship with firm ages. Specifically, adding one year old to age of a firm reduces its external benefits by 0.8% and 0.5% when local employment density and the value of employment location quotient plus one double respectively. Graph C of Figure 2.7 confirms the marginal cost of aging paid by the urbanization economies with the upward line that is steeper for younger firms who fall into the 10th percentile of the age distribution versus older firms at the 90th percentile. It is noteworthy from the graph C that, despite gaining higher marginal external benefits, immature firms remain overwhelmed by performance of their local mature fellows in less dense regions – specifically the ones with less than about 365 workers[18] per square kilometer. Similarly, graph D of Figure 2.7 illustrates the productivity gain of younger firms from specialized regions with an upward-sloping marginal line. Somewhat surprisingly, the slope turns downward for older firms, meaning that they suffer from diseconomies as their industries become more and more specialized in the region. These almost symmetrically opposite impacts on firms of different age subgroups cancel out each other in the full sample, which may explain why regional specialization appears to have no significant effect on a generic firm found in the baseline estimation. In other words, there is evidence of the Marshall externalities but only for young firms. All in all, the results support *Hypothesis 4*, and are in line with the findings of McCann & Folta (2011) and Rigby & Brown (2015). Interestingly, smaller firms are found to benefit

[18] This number is the x-axis coordinate of the intersection point between two lines in graph C of Figure 2.7.

more from a diverse regional condition, while younger firms enjoy that premium in a denser or more specialized environment. This divergence is expected since the sample overlap between small firms and young ones is just partial.

2.7. Robustness check

To assess how the estimates of agglomeration are sensitive to various choices of variables, functional forms, regression samples, and statistical techniques, a number of checks are implemented by adjusting the baseline specification and regression sample to such choices. Columns (1) – (4) of Table 2.12 show that the productivity contribution of urbanization is highly robust to the alternatives of productivity measures, especially between the measures dealing with potential simultaneity biases. It is worth pointing out that the application of LP-W TFP produces estimates of the agglomeration externalities that are almost identical to the results using LP or OP but are significantly higher than the estimators generated with OLS TFP. Even the coefficient of SPE in the sample with OLS TFP is statistically significant and displays the negative sign. This divergence is not unexpected given that the estimation of TFP by means of OLS tends to generate inflated input elasticities, thus underestimating productivity records. A take-away point from these findings is that a switch from OLS to the control function approach has a notable influence on the estimation of the externalities, while the difference generated by the latter methods is negligible.

The following robustness check is displayed in column (7) of Table 2.12, whose results are obtained using industry-related variables measured at the 2-digit level instead of the 3-digit level. Moving from the column (7) to the column (4), the magnitude of three main regional variables of interest changes slightly but their significance levels remain constant, featuring the stability of estimates to different choices of industrial units. The next aspect to examine is whether the existence of multi-establishment firms leads to potential measurement errors in estimating the productivity impacts of agglomeration. Specifically, a multi-plant firm may be influenced by changes in regions where its other plants are located rather than only by the region where it is home to its main factory, which makes plant-level dependent variables more ideal in comparison with the firm-level ones. However, plant-level TFP is infeasible to obtain using the firm-level data from Vietnam as well as many other countries in the world because the data on capital assets is observable merely at the firm level[19] despite some establishment-

[19] This is the common difficulty and limitation in estimating productivity using micro-data since capital information is often reported at firm-level but not establishment-level.

level information being available. Furthermore, given that headquarter plants commonly play a key role in firm production, and multi-plant firms make up a very small fraction of the total number of manufacturing firms in Vietnam[20], its presence in the sample should not cause any big issue. Nonetheless, to guarantee such a case, it is worth comparing the estimation results with and without multi-plant firms, following Martin et al. (2011). As can be seen from column (5) of Table 2.12, the removal of multi-plant firms from the regression sample generates a very subtle difference in the estimates in comparison with the baseline results in the column (4), which eases the concern about the cross-region operation of these firms.

Table 2.12. Agglomeration effects with a variety of TFP measures, units of geography-industry, and the exclusion of the two biggest cities in Vietnam

	(1) OLS	(2) LP	(3) OP	(4) LP-W	(5) LP-W	(6) LP-W	(7) LP-W
Sample	Full				Single-plant firms	Drop two biggest cities	Full
Unit	District & 3-digit VSIC						District & 2-digit VSIC
DEN	0.089*** (0.029)	0.133*** (0.028)	0.129*** (0.028)	0.129*** (0.028)	0.128*** (0.030)	0.144*** (0.029)	0.119*** (0.027)
SPE	-0.041** (0.020)	0.015 (0.019)	0.015 (0.019)	0.015 (0.019)	0.020 (0.020)	0.014 (0.021)	0.031 (0.021)
DIV	0.077*** (0.020)	0.080*** (0.019)	0.077*** (0.020)	0.077*** (0.020)	0.078*** (0.023)	0.084*** (0.022)	0.063*** (0.022)
Foreign ownership	0.102* (0.054)	0.107** (0.054)	0.102* (0.054)	0.107** (0.054)	0.117** (0.057)	0.081 (0.060)	0.106** (0.054)
State ownership	-0.001 (0.055)	0.022 (0.054)	0.020 (0.055)	0.025 (0.054)	0.022 (0.075)	0.046 (0.060)	0.024 (0.055)
Competition	0.055*** (0.011)	0.044*** (0.011)	0.042*** (0.011)	0.043*** (0.011)	0.043*** (0.012)	0.049*** (0.013)	0.040*** (0.011)
Fraction of FOEs	0.269 (0.164)	0.236 (0.165)	0.240 (0.166)	0.242 (0.167)	0.390 (0.247)	0.203 (0.167)	0.268 (0.167)
Regional average wages	0.196*** (0.029)	0.195*** (0.029)	0.194*** (0.028)	0.196*** (0.028)	0.173*** (0.031)	0.216*** (0.034)	0.194*** (0.028)
FE of time, time-industry, and time-region	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs	80,638	80,638	80,638	80,638	68,983	62,691	80,638
No. of firms	26,987	26,987	26,987	26,987	25,738	19,856	26,987
F-Test	12.193***	12.056***	11.793***	11.925***	11.138***	9.057***	10.601***
R^2	0.024	0.073	0.070	0.073	0.077	0.076	0.073

Notes: Cluster-robust standard errors in parentheses are calculated at a province and 2-digit VSIC level. Columns in this table replicate column (8) in Table 2.10 with the following changes: the TFP indices are estimated with the ordinary least square in column (1), and with the methods of Levinsohn & Petrin (2003) and Olley & Pakes (1996) in columns (2) and (3) respectively; in column (5), only observations from single-plant firms are employed; in

[20] In 2016, approximately 4% of manufacturing firms were multi-plant in Vietnam.

column (6), observations from the two biggest cities in Vietnam are removed from the regression sample; in column (7), all regional variables are measured at a district and/ or 2-digit VSIC level; control variables are not shown for brevity.

Abbreviation and denotation: FE, fixed effects; DEN, employment density; SPE, specialization; DIV, diversity.

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 2.13. Agglomeration effects with different forms of agglomeration variables

	(1) DEN^2	(2) SPE^2	(3) DIV^2	(4) $DEN \times SPE$
DEN	0.128** (0.058)	0.129** (0.028)	0.131** (0.029)	0.109*** (0.031)
SPE	0.015 (0.019)	0.039 (0.036)	0.014 (0.019)	-0.033 (0.038)
DIV	0.077*** (0.020)	0.077*** (0.020)	0.108 (0.074)	0.076*** (0.020)
DEN^2	0.000 (0.006)	-	-	-
SPE^2	-	0.039 (0.036)	-	-
DIV^2	-	-	-0.007 (0.015)	-
$DEN \times SPE$	-	-	-	0.010 (0.007)
FE of time, time-industry, and time-region	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
No. of obs	80,638	80,638	80,638	80,638
No. of firms	26,987	26,987	26,987	26,987
R^2	0.073	0.073	0.073	0.073

Notes: The regressions in this table replicate column (8) in Table 2.10 adding the squared term of DEN as DEN^2 in column (1), the squared term of SPE as SPE^2 in column (2), the squared term of DIV as DIV^2 in column (3), and the interaction term of DEN and SPE in column (4).

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

The next inspection is to see whether the drop of observations of the two biggest province-level regions in Vietnam – Hanoi and Ho Chi Minh City from the full sample causes any extreme downward shift in estimated coefficients. If it is the case, it means that the potential sorting behavior of productive firms towards the most urbanized regions remains a problem despite the application of several remedies. However, the results shown in column (6) of Table 2.12 indicate very minor shifts, which are even upward rather than downward as predicted, implying that the reverse causality should not be a concern in this study. Next, since the estimate of specialization externalities is statistically insignificant in the baseline regression, it is worth testing the possibility of non-linear external terms as well, which allows for case that the diseconomies emerge when a region becomes too agglomerated. To argue, it is possible that density and specialization have an interactive impact on TFP because, say, the

specialization externalities might only be active in a highly dense region. Furthermore, the density diseconomies might exist beyond a certain threshold of DEN. Columns (1) – (4) of Table 2.13 indicate that these possibilities are not empirically supported because squared terms of agglomeration proxies and the interaction term between DEN and SPE are all insignificant.

Table 2.14. Agglomeration effects with standard errors clustered at various levels

Column	(1)	(2)	(3)	(4)
Cluster level of standard errors	Unclustered	Province	2-digit VSIC	Province/ 2-digit VSIC
DEN	0.129*** (0.023)	0.129*** (0.032)	0.129*** (0.033)	0.129*** (0.028)
SPE	0.015 (0.013)	0.015 (0.020)	0.015 (0.025)	0.015 (0.019)
DIV	0.077*** (0.016)	0.077*** (0.020)	0.077*** (0.016)	0.077*** (0.020)
Foreign ownership	0.107*** (0.040)	0.107* (0.055)	0.107 (0.065)	0.107** (0.054)
State ownership	0.025 (0.050)	0.025 (0.052)	0.025 (0.056)	0.025 (0.054)
Competition	0.043*** (0.009)	0.043*** (0.013)	0.043*** (0.013)	0.043*** (0.011)
Fraction of FOEs	0.242 (0.153)	0.242*** (0.090)	0.242* (0.125)	0.242 (0.167)
Regional average wages	0.196*** (0.021)	0.196*** (0.032)	0.196*** (0.027)	0.196*** (0.028)
FE of time, time-industry, and time-region	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
No. of obs	80,638	80,638	80,638	80,638
No. of firms	26,987	26,987	26,987	26,987
R^2	0. 073	0. 073	0. 073	0. 073

Notes: Columns in this table replicate column (8) in Table 2.10 with the following changes: standard errors in parentheses are unclustered in column (1), while their cluster-robust values are calculated at a province level, a 2-digit VSIC industry level, and a province and 2-digit industry level in columns (2), (3), and (4) respectively.

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

The final discussion is the problem of statistical inference. It involves both the intra-sector and intra-region correlations between regional variables that are regressed with a micro-level dependent variable. These correlations can make computed standard errors biased downwards, as emphasized theoretically and empirically by Moulton (1990), and recently by Cameron et al. (2015). More specifically, regional explanatory variables of the model are perfectly or strongly correlated within the same region or the same industry, which can artificially inflate the t-statistics of βs in the specification (2.40), and consequently affect the precision of statistical interference. One possible solution used commonly in literature is to obtain clustered

standard errors by defining each industry, region, or pair of industry and region as a cluster to take the within correlation into account. However, this is not a perfect solution due to the fact that regional factors may remain strongly correlated to each other across defined clusters. To comprehensively solve the problem, Cameron et al. (2015) suggest a multi-way clustering method to address potential autocorrelation across all the three directions of industry, region, and industry-region. Unfortunately, the multi-way clustering method does not work well with multi-level fixed effects – the main econometric methodology used in this article. Furthermore, the inclusion of time-region and time-industry fixed effects in the specification of this study also help deal with this problem of strong correlation. Hence, the solution of clustering robust standard errors at the province – 2-digit industry in all regressions is viewed as enough to solve the issue. To perform a robustness check, comparison of the sensitivity of cluster choices is implemented, whose results set out in Table 2.14 show a very slight change in standard errors across different cluster levels, especially for agglomeration estimators. Therefore, the potential problem of strong within correlations should not be a worry in this study.

2.8. Conclusions

This study contributes to the strand of literature by focusing on verifying the existence of the agglomeration externalities in a developing country – Vietnam. Furthermore, specific research questions answered in this chapter are which one between inter- and intra-industry relationships is dominant in agglomeration economies and which firms reap most external benefits. To introduce plausible hypotheses, this study discusses empirical proof about the contrasting background between developing and developed countries in terms of geographical concentration, dominant industries, and firm characteristics. As the baseline explained variable, firm-level TFP in this study is computed exploiting the power of control functions and instruments under the GMM context to deal with simultaneity problems raised in the production function estimation. Meanwhile, potential econometric problems in estimating the agglomeration effects are tackled, making use of multi-level fixed effects and micro data. The results reveal that the externalities are active mainly through density and industrial variety and winning firms in agglomerated regions tend to be foreign-owned, small, or young. These findings are robust to different TFP measures, functional forms, and to the subtraction of multi-plant firms or largest cities in Vietnam.

Chapter 3

The impact of human capital externalities on firm performance in Vietnam: integrating entrepreneurs and workers

Previous studies in agglomeration literature tend to include solely urban economies or human capital externalities. Meanwhile, entrepreneurship literature tends to focus primarily on entrepreneurs' human capital but ignore the influences of workers' and regional agglomeration on firms. This study analyzes all those region- and micro-level factors as determinants of firm productivity under the spatial equilibrium of a theoretical model in which entrepreneurs and workers migrate across regions based on their human capital. To gain sound empirical evidence, the unique establishment-level database from Vietnam and historical instruments for the external effects are employed. Estimated results confirm the positive externalities of employment scale, which is robust across various samples, regional measures, and regression methods, while the evidence of human capital spillover is only found in high-tech economic sectors. The ambiguous role of regional higher education in the economy may be driven by the predominance of low-tech industries and less extensive-knowledge business services in Vietnam.

3.1. Introduction

The impact of education on productivity has been widely documented in literature. The related analyses in the context of a country are conducted following two different directions of the impact: region-level and micro-level. The former commonly known as human capital externalities (HCE) is modeled theoretically and verified empirically in the pioneering studies of Lucas (1988), Rauch (1993), and Moretti (2004a). These externalities along with urbanization economies (UE) are among the primary determinants of agglomeration forces. A common channel through which both external effects may boost the productivity of local firms is that the spillover of knowledge is more intensive in regions with higher levels of education or larger population (Combes & Gobillon, 2015). In fact, more populated regions tend to have a bigger share of residents with high intellectual ability (Bacolod et al., 2010) and workers with higher education (Eeckhout et al., 2014). However, previous regional studies very often avoid introducing HCE and UE simultaneously in a single specification, despite the two variables

being strongly correlated (Carlino & Kerr, 2015; Combes & Gobillon, 2015). In such a framework, the estimate of regional human capital without an identification strategy should be interpreted with caution because it also captures the influences of non-human capital factors. Even if data is available to the application of the instrumental variables (IV) approach, the regression still faces the risk of using invalid instruments caused by the exclusion of UE from the specification.

From the micro-level perspective, human capital might affect firms' performance through entrepreneurs and workers, which act as two separate inputs in a production function in Lucas (1978) and Gennaioli et al. (2013). So far, regional models that explicitly take entrepreneurial factors into account are rarely formulated (Glaeser et al., 2010). From the empirical point of view, Moretti (2004) and Duranton (2016a) note that education, especially higher education, is the popular measure of human capital in literature, which is also the case in this study. La Porta & Shleifer (2008) and Syverson (2011) show that it is important to account for both human capital inputs in regression when explaining firm-level productivity, especially in developing countries. Entrepreneurship studies of Van der Sluis et al. (2008), Doms et al., (2010), and Unger et al. (2011) provide reasons and evidence that human capital measured through formal education, especially higher education, enhances productivity of entrepreneurs by giving them a better ability to sense profit-making opportunities, to accomplish entrepreneurial tasks, and to accumulate new entrepreneurial knowledge over time. Nevertheless, there are complaints in literature that many authors focus on entrepreneurs but overlook regional characteristics, owing to the limited scope of surveys on firm owners. Furthermore, regional studies on firm-level outcome tend to include regional externalities in the specification, but neglect entrepreneurial factors (Doms et al., 2010; Glaeser et al., 2010).

Gennaioli et al. (2013), as far as the author knows, is the unique paper that accounts for all the aspects mentioned above theoretically and empirically within a regional framework. In their regressions with firm-level productivity, they use cross-sectional data, which covers the large part of the globe, and use sub-national geographical units for their regional analyses. Therefore, their statistical inferences should be considered as the cross-country evidence rather than the cross-region evidence for a country. In addition, they base their empirical calculations on the method of ordinary least squares (OLS), thus neglect the problems of omitted variables and reverse causality facing the estimation of the external effects. This study follows the theoretical framework of Gennaioli et al. (2013) with a number of different directions. First and most importantly, the identification issues in estimating UE and HCE are handled using an IV

approach with historical instruments along with the methods of first-difference (FD) and fixed-effects (FE) as robustness checks. Second, further firm-level variables, in particular demographic characteristics of entrepreneurs and workers, are accounted for in the specification. Third, this study uses data sets that cover a large number of firms operating within a country with a fine unit of geography referred to as the region, which is more suitable for the regional context and for capturing human capital spillover compared to the application of large geographical units with cross-country and cross-continent regional data (Fu, 2007; Rosenthal & Strange, 2008). Fourth and finally, this study pays attention to the heterogeneity of the external effects. All these research directions constitute the first contribution of this study, given that its primary purpose is to verify the existence of HCE from higher education.

So far, there have been a number of studies on the external effects of human capital for developed countries, especially the US. Although these externalities are verified by many scholars, the confirmation is far from universal. It may vary depending on the country, the measure of regional human capital, the choice of analysis's geographical unit, the inclusion of other factors mentioned above, and the estimation strategy employed. Furthermore, studies on HCE for developing countries are still rare, especially the ones with a decent identification strategy. Knowing whether or not the external effect of human capital is causal is important for policy makers, particularly in developing countries where their national development budget is modest compared to their developed counterparts. If, for example, a larger regional share of university-educated employees is demonstrated to boost productivity of local firms and workers, lawmakers may consider allocating a higher share of the national budget to higher education. This, in fact, has been a debate for a long time in Vietnam over the low level of government spending on this level of education. Specifically, according to the World Bank, that number was only 0.33% for Vietnam in 2016, which was much lower in comparison with other developing countries in Asia such as Thailand, Malaysia, and China with 0.64%, 1.13%, and 0.87% respectively. The disparity is even larger when compared to developed countries. Those reasons motivate the implementation of this study, which concentrates on the influences of regional human capital on firm productivity for Vietnam – a typical developing country. To the author's best knowledge, there are no prior studies on HCE for Vietnam, and Duranton (2016b) is the only other paper that attempts to deal with the identification problem for both HCE and UE using data from a developing country. These constitute the second contribution of this study to the narrow strand of literature.

Vietnam is considered to be an interesting case to analyze the external effects of human capital due to several further reasons. Firstly, similar to most other developing countries, the average education level of workers in Vietnam is low in comparison with their developed counterparts. In particular, the publicly accessible Barro-Lee data set shows that the share of population aged 15 to 64 with completed tertiary education in Vietnam in 2015 was only 4.09%, while the numbers were 23.64%, 27.54%, and 24.24% for the UK, South Korea, and Australia respectively – three developed countries. Likewise, the economic sectors that have the high demand for blue-collar workers tend to dominate developing economies, while they are in more balance with the other sectors in developed economies. Meanwhile, the less strong presence of university-educated labor force and high-tech economic sectors could lead to the less important roles of HCE in the developing world in general and in Vietnam in specific. Therefore, the fact that there have been numerous studies on this matter for developed countries do not assure that the similar evidence will be found in their developing counterparts. Secondly, as a fast-growing country with dramatic changes in socio-economic conditions since the 1990s, Vietnam appears to be an ideal case to apply the TSLS method with historical instruments for recent regional levels of human capital in order to gain the causal inference. Finally, unique firm-level data sets provided by GSO cover all firms and their affiliated establishments in Vietnam with the rich information on directors and workers. This allows the possibility of using both regional and micro-level factors within a single specification and dealing with the heterogeneity in productivity influences on firms from different business sectors.

Using OLS with location and sector FE for the data set in 2016 to fit the spatial model of Gennaioli et al. (2013), this study finds the significant influences of human capital at both the region level and the firm level on firm-level productivity. However, the estimate of HCE becomes statistically insignificant when the sampling year is 2011, or when the regression method of FD or firm FE is applied to control for local and sector economic shocks and all unobserved constant factors. To measure the causal impact of HCE, historical instruments are employed including the past share of university-educated workers and student population, while instruments for UE are the past population. These instruments are considered to be exogenous sources since they were recorded before many society-changing events happened in Vietnam, while regional variables of interest were measured after these events. Making use of these instruments with the method of two stage least squares (TSLS) yields, again, the insignificant estimator of HCE. On the contrary, the estimated coefficient of UE stays strongly significant across all years of sampling, sample sizes, and different regression techniques. To

test the possibility of non-linear relationships or omitted interactions, the specification is also added with non-linear terms of HCE or UE, or the interaction term of the two external effects. The regression results show that those concerns have no empirical support. Since the evidence of HCE is found in only the case of the 2016 sample with OLS, it is called weak. Surprisingly, when the TSLS regression is carried out separately for high-tech and low-tech sectors, the HCE estimator obtained is positive and significant for the former but negative and insignificant for the latter. Thus, the ambiguous evidence of HCE may be attributable to the dominance of the low-tech economic sectors in Vietnam.

This chapter is structured as follows. Section 3.2 reviews the literature and suggests hypotheses. Section 3.3 expresses a theoretical model that is the base for the empirical application. Section 3.4 specifies variables used in the regression and discusses econometric issues raised in the estimation of HCE, while section 3.5 describes the main features of data sets and discusses the validity of instruments. Section 3.6 reports estimated results with or without an IV approach and assesses their robustness. Finally, section 3.7 concludes.

3.2. Literature review

After the discussions of Marshall (1890, Book 4, Chapter 10) and Jacob (1969, Chapter 2), a number of theories have been developing for many decades to model HCE. In a growth context, Lucas (1988) is considered as the pioneer in treating average level of human capital as an external effect in an aggregate production function with Hicks-neutral technology. In an urban context, Rauch (1993) considers local HCE as a public good. In order for the empirical application, Rauch (1993) follows the model on public good of Roback (1982) in which the free mobility of firms and consumers leads to a spatial equilibrium where wages and land rents are identified. In an urban growth context in Black & Henderson (1999), UE has an interdependent relationship with HCE and the two agglomeration terms jointly determine the urban economy in an endogenous growth model. In a regional context, Gennaioli et al. (2013) introduce an innovative model based on the combination of Lucas (1978) and Lucas (1988) with urban models. Their model takes into account all HCE, UE, and human capital of entrepreneurs and workers in a single framework in explaining firm productivity. The utility-maximizing behaviors of entrepreneurs and workers, given their partial mobility across regions and the strength of the external effects, lead to a unique spatial equilibrium.

Regarding the empirical aspect, the existence of local HCE has not received affirmative statements from economists so far. Considered as a pioneer in the application of micro data to

the strand of literature, Rauch (1993) applies the Mincerian approach (Mincer, 1974) with data censused for US cities, relying on the OLS method, and controlling for individual human capital and regional amenities. Rauch (1993) shows that the average schooling level has a positive influence on wages of local workers. To overcome the limitation of Rauch (1993) due to the absence of an identification strategy, Acemoglu & Angrist (2000) exploit the exogenous variation of compulsory schooling laws and child labor laws at the state level in the past in dealing with the potential endogeneity of average schooling when it is regressed with individual wages in the US. Moving from OLS to the IV technique, the evidence of HCE changes from being strong to being weak, along with the dramatic decrease in the magnitude of estimated coefficients. With the availability of panel data at hand, Moretti (2004a) focuses on controlling unobserved characteristics of individuals and regions and employs the share of college-educated workers in the labor force rather than schooling level as in previous studies. Moretti (2004a) instruments for regional education with age structure and education subsidy programs in the US in the past and finds evidence that supports the existence of spillovers from higher education. In a follow-up study, Moretti (2004b) chooses plant-level productivity instead of wages as the explained variable. Moretti (2004b) makes use of plant-level data and bases the empirical strategy entirely on the fixed-effects (FE) technique to deal with econometric concerns. Moretti (2004b) confirms that production plants become more productive when human capital is higher in the region, holding other factors fixed. Furthermore, Moretti (2004b) finds that high-tech plants absorb most spillovers from HCE.

Ciccone and Peri (2006) proceed with the use of US data using the shift-share approach to control for the imperfect substitute between high and low-educated workers, which is neglected in prior studies on wage impact that follows the Mincerian approach. Another novelty of their study is that controlling individual education is not needed. To deal with identification issues, they employ the historical demographic structure of city population following Moretti (2004a). They show that the impact of average schooling is statistically insignificant and give their different estimates in comparison with previous studies credit for their explicit controls on the imperfect substitution between workers of various skills. Rosenthal & Strange (2008) focus on the absolute number of highly educated workers rather than the relative values to represent the external effects of human capital. Utilizing instruments from geologic characteristics, they find strong evidence of human capital spillover in the US. They also report that the strength of HCE diminishes quickly over distance, implying a local impact. Doms et al. (2010) take a different direction with the use of the dependent variable as performance of young firms or income of

self-employees. They exploit a novel entrepreneurship data set in the US and estimate by means of location-industry FE to deal with region-level omitted variables. After controlling for education of entrepreneurs, they find an ambiguous income impact of regional education, this suggests the necessity of allowing for human capital of entrepreneurs when estimating HCE. Milan et al. (2014) follow the entrepreneurship approach of Doms et al. (2010) but employ panel data from European countries. In particular, they condition out a more comprehensive set of demographic characteristics of entrepreneurs in explaining their income. Relying on the method of random effects (RE), they report that entrepreneurs benefit from regional education. Bratti & Leombruni (2014) make use of the information on the change in graduates from Italian colleges and historical demographic structure to instrument for the change in regional education in recent times. They verify the existence of HCE only for high-skilled workers. To wrap up, the empirical proof of HCE for developed countries, especially the US, is numerous and tends to suggest that local HCE are active and positive. Despite this, the conclusion is far from inconclusive because of the weak or insignificant results from a number of influential studies.

Turning to some empirical evidence for the developing world, Liu (2007) adopts the standard estimation procedure in Moretti (2004a) in an attempt to estimate the external effects of education across Chinese provinces. Liu (2007) exploits the household-level data surveyed for urban regions in 1988 and 1995, measures regional education as average schooling, and controls for a number of individual and regional characteristics. The two instruments applied in Liu (2007) are the Compulsory Education Law which came into effect in 1986 and the share of the highly educated population in 1990. Liu (2007) reports the strong evidence of HCE in China. Muravyev (2008) considers the average education level right before the fall of the Soviet Union as exogenous variations to the influence of regional human capital on wages in Russian cities in the post-Soviet time. Muravyev (2008) concludes that the more the share of a highly educated city population is, the higher wages citizens of the city earn. Liu (2014) proceeds with another HCE study for China, but moves on to the regression of production function with firm-level panel data in the 1990s. Liu (2014) applies multi-level FE and TSLS as alternative regression methods, using the explained variable as firm-level total factor productivity (TFP) and the spatial distribution of library books as an instrument for schooling. The results of Liu (2014) confirm the existence of local HCE regardless of whether FE or TSLS is employed. However, Liu (2014) doubts the choice of library books as a valid instrument since such a sort of city amenities might be endogenous due to its possible correlation to both firm-level TFP

and other city characteristics. It is worth noting that, similar to Moretti (2004b), Liu (2014) reports the dominance of high-tech firms in reaping benefits from regional education.

A common feature of the above studies' specification, regardless of whether they are conducted for developed or developing countries, is that regional scale of population or employment is not included. In fact, Combes & Gobillon (2015) and Duranton (2016b) note that this is a common feature of most studies on HCE in literature. As discussed in chapter 2, despite the mixed evidence in literature, UE might exist and cast the direct influence on wages of workers and productivity of firms. Meanwhile, total local population or employment tends to be used as the denominator in the measure of local average human capital, leading to the strong and positive correlation between the UE and HCE. Moreover, there is evidence that human capital and urbanization have a two-direction relationship in regional context (Duranton & Puga, 2014). It means that estimating HCE without taking UE properly into account may result in upward-biased results. This said, one can argue that the application of TSLS helps relax these worries. However, the instruments chosen in literature are not criticism proof. Many instruments for local human capital in previous studies may be correlated with local population, thus become invalid if this scale variable is excluded from specification. For example, the implementation of the compulsory schooling laws and child labor laws might be overseen more carefully in urban areas compared to rural areas. Next, the age structure could be systematically different between rural and urban regions, with the share of old people expected to be higher in rural areas. Even the application of local human capital in the past may be invalidated if one fails to control local employment due to the mutual relationship of the two regional variables as mentioned above.

To avoid these potential econometric problems, a few studies have attempted so far to allow for UE in regional studies on HCE. The first to mention is Fu (2007), whose specification includes not only local employment density as a proxy of UE, but also other types of agglomeration externalities. Fu (2007) reports the positive and significant for all terms of external effects, including HCE, at the very fine geographic unit – block level for Boston, the US. In the more recent paper, Gennaioli et al. (2013) verify the main predictions of their regional theory by using data that covers massive land areas of the globe. In a single regression, they find the significant and positive estimates for both UE and HCE at the sub-national geographical level. However, both Fu (2007) and Gennaioli et al. (2013) have the identification problem since they use cross-sectional data sets but rest their whole statistical inference on merely OLS with location FE. With micro-level panel data at hand, Combes et al. (2008) and

Liu (2014) estimate HCE by means of multi-level FE for France and China respectively. Combes et al. (2008) go even further by taking explicit account of the internal migration of workers across French cities and make an instrument for UE. Both Combes et al. (2008) and Liu (2014) find strong evidence of the two external effects but Combes et al. (2008) emphasize the dominant role of UE over the other agglomeration externalities in explaining the distribution of productivity across space. Duranton (2016b), to the author's knowledge, is the only paper that has an identification strategy for both UE and HCE in a single estimation. Using individual wages as the dependent variable, soil characteristics and number of higher education facilities to instrument for HCE, and historical population to instrument for UE in Colombia, Duranton (2016b) finds that HCE is positive and strongly significant without UE included, but turns to insignificant with the inclusion. Moreover, the evidence of UE is found strongly robust across various specifications. This verifies the concerns as mentioned above that the absence of UE in the specification could result in the upward-biased estimate of HCE. Thus, along with the conclusions of Combes et al. (2008a) and the findings of chapter 2 on the dominance of UE in Vietnam, the following hypothesis is proposed, given that this study's focus is the estimate of HCE.

Hypothesis 1: Urbanization economies are dominant over human capital externalities in Vietnam.

Finally, based on findings of Moretti (2004b) and Liu (2014), the following is suggested.

Hypothesis 2: Firms in higher-tech and more knowledge-intensive economic sectors benefit more from human capital externalities in comparison with the remaining sectors.

3.3. Theoretical framework

The primary objective of this study is to verify the presence of local HCE through testing the hypotheses stated in the previous section. To achieve this with firm-level productivity as the explained variable, it is ideal to allow for not only region-level effects (HCE and UE) but also firm-level production factors, especially firm-level human capital, in a single model. This section presents the conceptual micro foundation for such an empirical application following Gennaioli et al. (2013).

3.3.1. Setting

In the context of a country, "highly productive" regions denoted as G have productivity $\tilde{A}_G$ and make up a share of ϑ in the number of regions that is normalized to 1. Meanwhile, each "low productive" region denoted as U is endowed with productivity $\tilde{A}_U < \tilde{A}_G$, and make up the

remaining share of $1-\vartheta$ of the number of regions in total. An agent j's utility function is assumed to be

$$u(c,a) = c^{1-\rho_j} a^{\rho_j}, \quad 0 < \rho < 1 \tag{3.1}$$

where ρ refers to the share of the agent's budget devoted to accommodation, and c and a represent consumption and accommodation services respectively. The number of agents is normalized to 2, with a half being "rentiers" without human capital and another half being "laborers" with h units of human capital for each. Each rentier has 1 unit of accommodation to rent to laborers, M units of land, and K units of physical capital to rent to firms. The endowment of initial regional human capital of the region $i = G, U$ is exogenous following Pareto distribution with the range $[\underline{h}, +\infty)$ where $\underline{h} > 1$ and $\underline{H_i}$ measures the mean value. Each laborer uses their human capital to work as either an entrepreneur or a worker, depending on which occupation maximizes his or her income. As an entrepreneur with h human capital working in firm j, he or she hires production factors and produces consumption goods in region i according to the Cobb-Douglas production function

$$y_{i,j} = A_i h^{1-\alpha-\beta-\gamma} H_{i,j}^{\alpha} K_{i,j}^{\beta} M_{i,j}^{\gamma}, \quad \alpha + \beta + \gamma < 1 \tag{3.2}$$

where $H_{i,j}$ is total human capital of workers, $K_{i,j}$ is physical capital, $M_{i,j}$ is land areas, and A_i is regional total factor productivity. Genaioli et al. (2013) follow Lucas (1988) to model A_i as

$$A_i = \tilde{A}_i (E_i(h)^{\omega} L_i)^{\xi}, \quad \omega \geq 1,\ \xi > 0 \tag{3.3}$$

where $\tilde{A}_i$ represents productivity-related regional characteristics such as geographical advantages or institution and $E_i(h)$ denotes the average human capital of L_i laborers working in region i; and parameter ω represents the productivity influence of the quality of human capital, while ξ refers to the strength of external effects. Thus, the terms $E_i(h)^{\omega\xi}$ and L_i^{ξ} capture HCE and UE respectively since the transformed version of $(E_i(h)^{\omega} L_i)^{\xi}$ is $E_i(h)^{\omega\xi} L_i^{\xi}$. Productivity A_i is treated as given until the spatial equilibrium is reached so that the location choice of laborers is considered in advance.

Rentiers in region i earn $\lambda_i M$ by providing land and $\zeta_i K_i$ by providing physical capital to firms' production with λ_i and ζ_i as rental rates. In addition, rentiers earn ν_i by renting accommodation to entrepreneurs and workers. While physical capital is fully mobile, laborers are partially mobile, and the land area M and a normalized number of 1 rented accommodation in each region are immobile. If choosing to become a worker, a laborer endowed with h human capital

earns wage $w_i h$ with w_i being the wage rate. Finally, an entrepreneur gains profit $\pi_i(h)$ from his or her firm's production of consumption goods, whose price is normalized to 1 throughout this model, given that price-related values are all deflated in the empirical application.

3.3.2. Occupation choice in equilibrium

To maximize his or her profit from firm j in region i, an entrepreneur solves

$$\max A_i h_{i,j}^{1-\alpha-\beta-\gamma} H_{i,j}^{\alpha} K_{i,j}^{\beta} M_{i,j}^{\gamma} - w_i H_{i,j} - \zeta_i K_{i,j} - \lambda_i M_{i,j} \tag{3.4}$$

Assuming that firm j employs an entrepreneur with h_j, meaning a share of h_j/H_i^E of all entrepreneurs' human capital available in region i, and employs other production factors with the same proportion. The rationale behind this assumption is that firm size is proportional to entrepreneur's human capital, as in Lucas (1978). As a result,

$$H_{i,j} = \frac{h_j}{H_i^E} \cdot H_i^W, \qquad K_{i,j} = \frac{h_j}{H_i^E} \cdot K_i, \qquad M_{i,j} = \frac{h_j}{H_i^E} \cdot M \tag{3.5}$$

Plugging (3.5) into (3.2) and adding up all firm-level equations in each region yields the aggregate regional production function

$$Y_i = A_i (H_i^E)^{1-\alpha-\beta-\gamma} (H_i^W)^{\alpha} K_i^{\beta} M^{\gamma} \tag{3.6}$$

where Y_i denotes total outputs in region i. It is worth noting that subscript i of land area M is already removed because regional land area is normalized to 1, thus is the same everywhere. The first-order conditions are

$$w_i = \frac{\partial Y_i}{\partial H_i^W} = \alpha . A_i (H_i^E)^{1-\alpha-\beta-\gamma} (H_i^W)^{\alpha-1} K_i^{\beta} M_i^{\gamma} \tag{3.7}$$

$$\pi_i = \frac{\partial Y_i}{\partial H_i^E} = (1-\alpha-\beta-\gamma) . A_i (H_i^E)^{-\alpha-\beta-\gamma} (H_i^H)^{\alpha} K_i^{\beta} M_i^{\gamma} \tag{3.8}$$

$$\zeta = \frac{\partial Y_i}{\partial K_i} = \beta . A_i (H_i^E)^{1-\alpha-\beta-\gamma} (H_i^H)^{\alpha} K_i^{\beta-1} M_i^{\gamma} \tag{3.9}$$

meaning the marginal product of workers' human capital, entrepreneurial human capital, and physical capital respectively. Thereby, working in firm j in region i, an entrepreneur with h_j unit of human capital earns profit $\pi_i(h)$ equal to $h_j \cdot \partial Y_i / \partial H_i^E$. Meanwhile, a worker with such an amount of human capital earns a wage level equal to $h_j \cdot \partial Y_i / \partial H_i^W$. A laborer j with human

capital h_j chooses an occupation of an entrepreneur if and only if $h_j \cdot \partial Y_i / \partial H_i^E > h_j \cdot \partial Y_i / \partial H_i^W$, and an occupation of a worker if $h_j \cdot \frac{\partial Y_i}{\partial H_i^E} < h_j \cdot \partial Y_i / \partial H_i^W$. This implies that the entrepreneur's human capital is often considered to be higher than workers' in this model, which is proved to be reasonable in practice using the data set for this study. In equilibrium, laborers have no incentive to switch between the two occupations. This indifference means

$$\frac{\partial Y_i}{\partial H_i^E} = \frac{\partial Y_i}{\partial H_i^W} \tag{3.10}$$

Next, transforming the right-hand sides of (3.7), (3.8), and (3.9), then using (3.6) yields

$$\frac{\partial Y_i}{\partial H_i^W} = \frac{\alpha Y_i}{H_i^W} \tag{3.11}$$

$$\frac{\partial Y_i}{\partial H_i^E} = \frac{(1-\alpha-\beta-\gamma)Y_i}{H_i^E} \tag{3.12}$$

$$\frac{\partial Y_i}{\partial K_i} = \frac{\delta Y_i}{K_i} \tag{3.13}$$

Plugging (3.11) and (3.12) into (3.10) obtains

$$\frac{\alpha Y_i}{H_i^W} = \frac{(1-\alpha-\beta-\gamma)Y_i}{H_i^E} \tag{3.14}$$

Total human capital H_i in region i is supplied by entrepreneurs and workers in the region, meaning that $H_i = H_i^E + H_i^W$, which is combined with (3.10) to yield

$$H_i^E = \left(\frac{1-\alpha-\beta-\gamma}{1-\beta-\gamma}\right) \cdot H_i, \qquad H_i^W = \left(\frac{\alpha}{1-\beta-\gamma}\right) \cdot H_i \tag{3.15}$$

These equations reflect the occupational distribution of laborers in region i. Next, due to the full mobility of physical capital, its marginal output ζ is constant across regions regardless of whether they are highly or low productive. As a result, using equation (3.13) leads to

$$\frac{K_G}{K_U} = \left(\frac{A_G}{A_U}\right)^{\frac{1}{1-\beta}} \left(\frac{H_G}{H_U}\right)^{\frac{1-\beta-\gamma}{1-\beta}} \tag{3.16}$$

By using equation (3.7), which represents worker's wages per unit of his or her human capital, the wage disparity between regions G and U is

$$\frac{w_G}{w_U} = \frac{A_P}{A_U}\left(\frac{K_G}{K_U}\right)^{\beta}\left(\frac{H_G}{H_U}\right)^{-\beta-\gamma} \tag{3.17}$$

Plugging (3.16) into (3.17) yields

$$\frac{w_G}{w_U} = \left(\frac{A_G}{A_U}\right)^{\frac{1}{1-\beta}}\left(\frac{H_G}{H_U}\right)^{\frac{-\gamma}{1-\beta}} \tag{3.18}$$

Substituting the terms of regional productivity in (3.18) with its expression in (3.3) obtains

$$\frac{w_G}{w_U} = \left(\frac{\tilde{A}_G}{\tilde{A}_U}\right)^{\frac{1}{1-\beta}} \cdot \left(\frac{E(h_G)^{\psi}L_G}{E(h_U)^{\psi}L_U}\right)^{\frac{\xi}{1-\beta}} \cdot \left(\frac{H_U}{H_G}\right)^{\frac{\gamma}{1-\beta}} \tag{3.19}$$

3.3.3. Migration decision

Turning back to utility function (3.1) to make explicit migration behaviors of laborers. In equilibrium, equation (3.10) implies that both entrepreneurs and workers earn the same amount of income per unit of their human capital, equivalent to the wage rates w_i. Thus, laborers in region i earn an aggregate income of $w_i.H_i$. Plugging into (3.1) obtains

$$u_{w,i}(c,a) = (w_i h)^{1-\rho} \cdot \left(\frac{w_i h}{\theta_i}\right)^{\rho} = \frac{w_i h}{\theta_i^{\rho}} \tag{3.20}$$

where θ_i refers to the rental rates. Next, in maximizing their utility, laborers spend a share of ρ of their income on their accommodation, and the remaining share of $1-\rho$ on their consumption. Therefore, the demand for rental accommodation in region i is $\rho \cdot w_i \cdot H_i/\theta_i$,. Because of the assumed unitary supply of accommodation, equating the supply and demand leads to $\rho \cdot w_i \cdot H_i/\theta_i = 1$. It means that

$$\theta_i = \rho \cdot w_i \cdot H_i \tag{3.21}$$

Plugging into (3.20) yields

$$u_{w,i}(c,a) = \frac{w_i^{1-\rho}}{\rho^\rho} \cdot \frac{h}{H_i^\rho} \tag{3.22}$$

Regarding migration decision, assuming that at $t = 0$, wages and interest rates are higher in regions G. Thus, initially, labor and capital tend to move to these regions. In addition, assuming that a migrant has to pay migration costs as an amount of $\phi > 0$ from his or her human capital. Hence, a laborer with human capital h_j decides to migrate if and only if $u_{j,G} \geq u_{j,U}$. This means, by using (3.22),

$$\frac{w_G^{1-\rho}}{\rho^\rho} \cdot \frac{h-\phi}{H_G^\rho} \geq \frac{w_U^{1-\rho}}{\rho^\rho} \cdot \frac{h}{H_U^\rho} \tag{3.23}$$

and consequently,

$$\frac{w_G^{1-\rho}(h-\phi)}{H_G^\rho} \geq \frac{w_U^{1-\rho}h}{H_U^\rho} \tag{3.24}$$

It implies that, in this model, laborers with higher levels of human capital tend to migrate first since they are more likely to afford the migration costs $\phi > 0$ and meet the condition of inequation (3.24). However, the more the laborers sort into highly productive regions G, the more the total human capital H_G is, and thus the stronger the demand for accommodation becomes. This leads to higher rental rates η_G as expressed in (3.21), which, as a result, decreases the utility of laborers in regions G as expressed in (3.20). In other words, more and more laborers migrate from low productive regions to highly productive regions make the left-hand side's value of inequation (3.24) rising, whereas the values of its right-hand side are falling. In equilibrium, there exists a threshold of human capital denoted as h_m that inequation (3.24) is equalized. In such a scenario, all laborers initially live in regions U with human capital $h_j \geq h_m$ have moved to regions G, while laborers with lower human capital have stayed. As a result, in spatial equilibrium,

$$\frac{w_G^{1-\rho}(h_m-\phi)}{H_G^\rho} = \frac{w_U^{1-\rho}h_m}{H_U^\rho}$$

$$\Leftrightarrow \frac{w_G}{w_U} = \left(\frac{H_G}{H_U}\right)^\rho \cdot \frac{h_m}{h_m-\phi}$$

$$\Leftrightarrow \frac{h_m - \phi}{h_m} = \left(\frac{w_U}{w_G}\right)^{1-\rho} \cdot \left(\frac{H_G}{H_U}\right)^{\rho}$$

$$\Leftrightarrow h_m \left[1 - \left(\frac{w_U}{w_G}\right)^{1-\rho} \cdot \left(\frac{H_G}{H_U}\right)^{\rho}\right] = \phi \tag{3.25}$$

Using equations (3.19) and (3.25) yields

$$h_m \cdot \left[1 - \left(\frac{\tilde{A}_U}{\tilde{A}_G}\right)^{\frac{1-\rho}{1-\beta}} \cdot \left(\frac{E(h_U)}{E(h_G)}\right)^{\frac{\omega\xi(1-\rho)}{1-\beta}} \cdot \left(\frac{L_U}{L_G}\right)^{\frac{\xi(1-\rho)}{1-\beta}} \cdot \left(\frac{H_G}{H_U}\right)^{\frac{\rho(1-\beta)+\gamma(1-\rho)}{1-\beta}}\right] = \phi \tag{3.26}$$

Since $E(h_G) = H_G/L_G$ and $E(h_U) = H_U/L_U$, plugging into (3.26) with some further transformations obtains

$$\frac{\phi}{h_m} = 1 - \left(\frac{\tilde{A}_U}{\tilde{A}_G}\right)^{\frac{1-\rho}{1-\beta}} \cdot \left(\frac{L_G}{L_U}\right)^{\xi(\omega-1)\frac{(1-\rho)}{1-\beta}} \cdot \left(\frac{H_G}{H_U}\right)^{\frac{(\gamma-\omega\xi)(1-\rho)+\rho(1-\beta)}{1-\beta}} \tag{3.27}$$

which shows the equilibrium condition with the external effects allowed for.

3.3.4. Spatial equilibrium

Assuming that total human capital in the country is constant and exogenous, denoted as $\underline{H}$. In addition, assuming that initial total human capital of regions G and regions U are $\underline{H_G}$ and $\underline{H_U}$ respectively, which are also exogenous. Since ϑ is the share of regions being highly productive, and $1 - \vartheta$ is the remaining share, this leads to

$$\underline{H} = \vartheta \underline{H_G} + (1 - \vartheta)\underline{H_U} = \vartheta H_G + (1 - \vartheta) H_U \tag{3.28}$$

It implies that the relationship between the change in human capital in regions G and regions U denoted as ΔH_G and ΔH_u respectively is

$$\Delta H_G = \frac{1 - \vartheta}{\vartheta} \Delta H_U \tag{3.29}$$

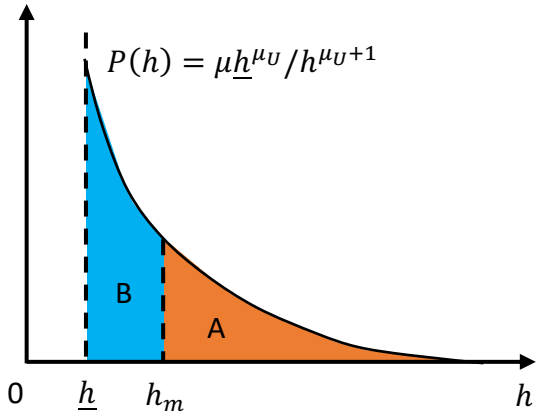

Figure 3.1. Illustration of the Pareto distribution of human capital in a low productive region at $t = 0$

Laborers' initial total human capital in each region U is assumed to follow a Pareto distribution with the probability density function (PDF) as $P(h) = \mu \underline{h}^{\mu_U}/h^{\mu_U+1}$ where parameter $\mu_U > 1$. Figure 3.1 is drawn to illustrate the graph of this PDF, keeping in mind that h_m might be any value within the range of $[\underline{h}, +\infty)$. This distribution means that before the migration starts, total human capital in each region U is $\int_{\underline{h}}^{+\infty} h \cdot \left(\mu \underline{h}^{\mu_U}/h^{\mu_U+1}\right) \cdot dh$. This is equivalent to the regional mean value of h times the sum of the areas of A and B, therefore this sum also represents the labor force in each region U. In equilibrium, all laborers in regions U with human capital $h \geq h_m$ already moved to regions G. This implies that $\int_{h_m}^{+\infty} h \cdot \left(\mu \underline{h}^{\mu_U}/h^{\mu_U+1}\right) \cdot dh$ units of human capital, or actually the part of labor force in area B in Figure 3.1, have moved out of each region U. Given that, initially, total human capital in each region U is $\underline{H}_U$, the amount of human capital has left this region when the equilibrium is reached is

$$\Delta H_U = \frac{\int_{h_m}^{+\infty} h \cdot \frac{\mu_U \underline{h}^{\mu_U}}{h^{\mu_U+1}} \cdot dh}{\int_{\underline{h}}^{+\infty} h \cdot \frac{\mu_U \underline{h}^{\mu_U}}{h^{\mu_U+1}} \cdot dh} \cdot \underline{H}_U$$

$$= \frac{\frac{\mu_U}{-\mu_U+1} \cdot \underline{h}^{\mu_U} \cdot h^{-\mu_U+1}\Big|_{h_m}^{+\infty}}{\frac{\mu_U}{-\mu_U+1} \cdot \underline{h}^{\mu_U} \cdot h^{-\mu_U+1}\Big|_{\underline{h}}^{+\infty}} \cdot \underline{H}_U$$

$$= \left(\frac{h_m}{\underline{h}}\right)^{-\mu_U+1} \cdot \underline{H}_U \qquad (3.30)$$

Plugging into (3.29) yields

$$H_G = \underline{H}_G + \Delta H_G = \underline{H}_G + \underline{H}_U \frac{1-\vartheta}{\vartheta}\left(\frac{\underline{h}}{h_m}\right)^{\mu_U - 1}$$

$$\Leftrightarrow h_m = \underline{h}.\left[\frac{H_G - \underline{H}_G}{\underline{H}_U} \cdot \frac{\vartheta}{1-\vartheta}\right]^{\frac{1}{1-\mu_U}} \tag{3.31}$$

Next, as mentioned above, area A in Figure 3.1 indicates the part of labor force of each region U that has migrated to regions G in equilibrium, this means

$$L_G = L_{G0} + \frac{1-\vartheta}{\vartheta} \cdot \int_{h_m}^{+\infty} \frac{\mu_U \underline{h}^{\mu_U}}{h^{\mu_U+1}} \cdot dh \cdot L_{U0}$$

and

$$L_U = L_{U0} - \int_{h_m}^{+\infty} \frac{\mu_U \underline{h}^{\mu_U}}{h^{\mu_U+1}} \cdot dh \cdot L_{U0}$$

where L_{U0} measures the labor force in each region U before the migration starts. Consequently, this leads to

$$L_G = L_{G0} + \frac{1-\vartheta}{\vartheta} \cdot \underline{h}^{\mu_U} \cdot h_m^{-\mu_U} \cdot L_{U0} \tag{3.32}$$

and

$$L_U = L_{U0} - \underline{h}^{\mu_U} \cdot h_m^{-\mu_U} \cdot L_{U0} \tag{3.33}$$

As a result, the employment difference between regions G and U is

$$\frac{L_G}{L_U} = \frac{L_{G0} + \frac{1-\vartheta}{\vartheta} \cdot \underline{h}^{\mu_U} \cdot h_m^{-\mu_U} \cdot L_{U0}}{L_{U0} - \underline{h}^{\mu_U} \cdot h_m^{-\mu_U} \cdot L_{U0}} \tag{3.34}$$

Another assumption is that, initially, labor force is evenly distributed between regions G and U, implying that $L_{P0} = L_{U0}$. Plugging this and (3.31) into (3.34) yields

$$\frac{L_G}{L_U} = \frac{1 + \frac{1-\vartheta}{\vartheta} \cdot \left(\frac{H_G - \underline{H}_G}{\underline{H}_U} \cdot \frac{\vartheta}{1-\vartheta}\right)^{\frac{\mu_U}{\mu_U - 1}}}{1 - \left(\frac{H_G - \underline{H}_G}{\underline{H}_U} \cdot \frac{\vartheta}{1-\vartheta}\right)^{\frac{\mu_U}{\mu_U - 1}}} \tag{3.35}$$

Using (3.27), (3.31), and (3.35) obtains the following equation

$$\frac{\phi}{\underline{h}} \left(\frac{\vartheta}{1-\vartheta}\right)^{\frac{1}{\mu_U - 1}} \left(\frac{H_G - \underline{H}_G}{\underline{H}_U}\right)^{\frac{1}{\mu_U - 1}} = 1 - \left(\frac{\tilde{A}_U}{\tilde{A}_G}\right)^{\frac{1-\rho}{1-\beta}} \cdot$$

$$\left(\frac{1 + \frac{1-\vartheta}{\vartheta} \cdot \left(\frac{H_G - \underline{H}_G}{\underline{H}_U} \cdot \frac{\vartheta}{1-\vartheta}\right)^{\frac{\mu_U}{\mu_U - 1}}}{1 - \left(\frac{H_G - \underline{H}_G}{\underline{H}_U} \cdot \frac{\vartheta}{1-\vartheta}\right)^{\frac{\mu_U}{\mu_U - 1}}}\right)^{\frac{\xi(\omega-1)(1-\rho)}{1-\beta}} \cdot \left(\frac{(1-\vartheta)H_G}{\underline{H} - \vartheta H_G}\right)^{\frac{(\gamma-\omega\xi)(1-\rho)+\rho(1-\beta)}{1-\beta}}$$

(3.36)

At $t = 0$, $H_G = \underline{H}_G$, the left-hand side of (3.36) is equal to zero. Assuming that $(\gamma - \omega\xi)(1-\rho) + \rho(1-\beta) > 0$, which is reasonable since it requires that the returns to land areas β is larger than the parameter showing the importance of human capital quality ψ times the strength of external effects ξ. In that case, once migration starts, the left side of (3.36) is increasing in H_P, while it is opposite for the right-hand side. This leads to a unique and stable spatial equilibrium given that the full agglomeration in regions G is not the case. The reason is that the right-hand side of (3.36) reaches negative infinity when H_G approaches the level of $\underline{H}$. Hence, provided the unique equilibrium, this model gains the following equation for firm-level output, which is the combination of (3.2) and (3.3), as

$$y_{i,j} = \tilde{A}_i [E_i(h)^{\omega} L_i]^{\xi} \, h_{i,j}^{1-\alpha-\beta-\gamma} H_{i,j}^{\alpha} K_{i,j}^{\beta} M_{i,j}^{\gamma} \tag{3.37}$$

This equation allows a direct empirical application by implying in an equilibrium setting that a firm's performance is determined by both region-level and firm-level factors, which in particular include all HCE, UE, and firm-level human capital.

3.4. Econometric issues

3.4.1. Specification

To estimate (3.37), further transformations are needed. First, dividing both sides of this equation by firm's total employment $l_{i,j}$ and transforming terms of total human capital into its mean values times its labor force yields

$$\frac{y_{i,j}}{l_{i,j}} = \tilde{A}_i[E_i(h)^{\omega}L_i]^{\xi} \cdot \left[E\left(h_{i,j}^{E}\right)\right]^{1-\alpha-\beta-\gamma} \cdot \left(\frac{l_{i,j}^{E}}{l_{i,j}}\right)^{1-\alpha-\beta-\gamma} \cdot \left[E\left(h_{i,j}^{W}\right)\right]^{\alpha} \cdot \left(\frac{l_{i,j}^{W}}{l_{i,j}}\right)^{\alpha} \cdot \left(\frac{K_{i,j}}{l_{i,j}}\right)^{\beta} \cdot \left(\frac{M_{i,j}}{l_{i,j}}\right)^{\gamma} \tag{3.38}$$

Assuming that $E(h)_i \cong e^{\overline{\theta_i}B_i}$ where B_i is the share of regional employees with a university degree and $\overline{\theta_i}$ is average returns to university education. Similarly, at the firm level, $E\left(h_{i,j}^{E}\right) \cong e^{\overline{\theta_{i,j}^{E}}B_{i,j}^{E}}$ and $E\left(h_{i,j}^{W}\right) \cong e^{\overline{\theta_{i,j}^{W}}B_{i,j}^{W}}$ are assumed where $B_{i,j}^{E}$ refers to a binary variable that equals to 1 if an entrepreneur owns a university degree, and 0 otherwise; $B_{i,j}^{W}$ is the share of firm employees with some university education; and $\overline{\theta_{i,j}^{E}}$ and $\overline{\theta_{i,j}^{W}}$ are average private returns to higher education of entrepreneurs and workers respectively. These approximations and the logarithm transformation change (3.38) into

$$\begin{aligned}\ln\left(\frac{y_{i,j}}{l_{i,j}}\right) = {} & \ln\left(\tilde{A}_i\right) + \omega\xi\overline{\theta_i}B_i + \xi\ln(L_i) + (1-\alpha-\beta-\gamma)\overline{\theta_{i,j}^{E}}B_{i,j}^{E} \\ & + (1-\alpha-\beta-\gamma)\ln\left(\frac{l_{i,j}^{E}}{l_{i,j}}\right) + \alpha\overline{\theta_{i,j}^{W}}B_{i,j}^{W} + \alpha\ln\left(\frac{l_{i,j}^{W}}{l_{i,j}}\right) \\ & + \beta\ln\left(k_{i,j}\right) + \gamma\ln\left(m_{i,j}\right)\end{aligned} \tag{3.39}$$

where $k_{i,j}$ and $m_{i,j}$ represent physical capital per employee and land per employee respectively, keeping in mind that an employee means either an entrepreneur or a worker. This equation implies that firm-level labor productivity is determined by factors including productive-related regional characteristics $\tilde{A}_i$; regional share of university-educated employees B_i, or actually HCE; logarithm of total employment in the region $\ln(L_i)$, or actually UE; Entrepreneur and workers' education as $B_{i,j}^{E}$ and $B_{i,j}^{W}$ respectively; physical capital per employee ($k_{i,j}$); land per

employee ($m_{i,j}$); and firm-level share of managerial and non-managerial working positions as $l_{i,j}^{E}/l_{i,j}$ and $l_{i,j}^{W}/l_{i,j}$ respectively.

Following (3.39), the regression model of this study is written as follows

$$\ln\left(\frac{y_{i,j}}{l_{i,j}}\right) = \delta_0 B_i + \delta_1 \ln(L_i) + \delta_2 B_{i,j}^{E} + \delta_3 B_{i,j}^{W} + \delta_4 \ln(l_{i,j}) + \delta_5 \ln(e_{i,j}) + \delta_6 \ln(k_{i,j}) + X_{i,j} + \psi + \chi + \epsilon_{i,j} \quad (3.40)$$

where $\ln(k_{i,j})$ is considered as a substitute for $\ln(m_{i,j})$ since the firm-level land area is not available in the data; $\ln(e_{i,j})$ is log of electricity consumption per employees, taken to approximate $\ln(k_{i,j})$; $X_{i,j}$ is a vector of firm-level factors including entrepreneur, workers, and firm's characteristics; ψ and χ are the fixed effects terms of sector and province respectively, provided that regional variables are measured at a finer geographic level - district; $\epsilon_{i,j}$ is an error term, which might contain other regional factors that have productivity influences on firm, including some factors in $\tilde{A}_i$; and $\ln(l_{i,j})$ is log of the firm's workforce, used to approximate terms $l_{i,j}^{E}/l_{i,j}$ and $l_{i,j}^{W}/l_{i,j}$ in (3.39). Although these changes do not affect the application of (3.39) in estimating HCE and UE, it becomes difficult to compare the estimates of (3.40) to the original parameters in (3.39). Thus, the estimation parameters are all simplified into δs in (3.40) in which δ_0 and δ_1 are the primary interest of this study to verify the hypotheses.

3.4.2. Econometric concerns

To gain the consistent estimates of δ_0 and δ_1, econometric issues emerging in the regression for (3.40) should be taken care of. The main concerns consist of omitted variables and the endogenous location choice of firms and workers, relatively similar to the main identification concerns facing the estimation of agglomeration externalities expressed in chapter 2 of this dissertation. To avoid such repeated content for UE, this subsection focuses on specific potential issues in estimating HCE.

The first problem is caused by firm-level omitted variables that are correlated to both regional education and firm-level productivity. Such uncontrolled firm characteristics may contaminate the estimate of δ_0. This problem is less concerning in this study owing to the inclusion of $B_{i,j}^{E}$, $B_{i,j}^{W}$, and ψ next to the main firm-level production factors on the right-hand side of (3.40), since these terms tend to be strongly correlated to B_i (Doms et al., 2010; & Glaeser et al., 2010).

Obviously, it should be easier to find a university-educated entrepreneur or worker in a region with a higher share of university graduates. The correlation matrix in Table 3.4 in the next section confirms the strong correlation between B_i, $B_{i,j}^E$, and $B_{i,j}^W$ with the baseline data used in this study. As a result, without controlling $B_{i,j}^E$ and $B_{i,j}^W$, variable B_i would capture the effects of micro-level human capital as well. This is the reason why HCE is statistically significant without $B_{i,j}^E$ being included in the specification in Doms et al. (2010), but its significance disappears with the inclusion. In addition, firms' labor productivity is more correlated to the average level of labor productivity and human capital within their economic sectors in the region in comparison with the whole local region (Hendricks, 2011), thus it is important to control for sector-specific characteristics with the term ψ of sector FE. To reduce further risks of omitting other influential factors, a rich set of firm-level factors as vector $X_{i,j}$ is added in (3.40). In particular, $X_{i,j}$ contains demographic characteristics of entrepreneurs, including age and its squared term, gender, and ethnic group, whose significance is proved in the intensive entrepreneurship review of Parker (2018). The remaining firm-level controls are average age of workers and its squared term, age of firm, state ownership, and foreign ownership. All in all, with these solutions, unobserved firm-level factors should not be a concern in this study, given that its focus is the consistent estimate of HCE.

The second problem is omitted variables at the region level, which occurs when unobserved and idiosyncratic regional features in the error terms are correlated to both regional education and firm's productivity. A clear source of this issue is regional characteristics embodied in $\tilde{A}_i$ that local institutions and natural conditions are the two typical examples. These factors are not taken explicitly into account in the specification (3.40), while they serve as predictors of firm-level labor productivity in (3.39). The problem complicating the estimation of HCE is that residents in a region with more highly educated workers tend to demand and vote for a more effective local authority. As a result, it is difficult to disentangle the productivity impact of this local authority's economic policies from the spillover of regional human capital. The same scenario applies to regions with natural advantages. It means that, without controlling for $\tilde{A}_i$, the estimate of HCE could be upward-biased. To deal with this problem, the province FE term χ is included in (3.40) to control for provincial characteristics that are constant across districts – the primary geographic level of regional analyses in this study. Fortunately, idiosyncratic economic and education policies are mainly delivered by authorities at the provincial level rather than at the district level in Vietnam. Put differently, the institutional environment may

be heterogeneous across provinces but tends to be homogeneous across districts. As a result, the province FE term should be enough to control for the districts' institutional setting. Turning to the natural conditions such as favorable climate and access to sea, this FE term helps reduce partly the potential bias because province-level characteristics are often strongly correlated to district-level characteristics. However, admittedly, the inclusion of χ might be not enough in the case that the first-nature conditions are district-specific.

Another consequential unobservable regional factor is idiosyncratic shocks, which simultaneously influence firm-level productivity and local human capital. Economic local shocks such as the foundation or the bankruptcy of a large corporation can generate such an impact. For example, in the case of a bankruptcy, this event leads to a sudden rise in local unemployment rate, which is a negative demand shock to local businesses. At the same time, highly educated workers unemployed due to this shock could migrate to other regions in seeking a new occupation, provided that they are often considered to be more mobile compared to their less educated peers. This consequently leads to a decrease in the share of the highly educated labor force. As a result, the positive correlation between productivity of local firms and the university share may be significant but the causal relationship does not exist. By similar logic, other demand-side or supply-side local shocks, which might be positive or negative, are possible causes of the spurious estimate of HCE if they are not controlled properly.

The final potential econometric concern is reverse causality. This emerges when the influence comes from regional productivity to regional human capital rather than the opposite direction. Theoretically, the equilibrium setting of the production function (3.37), as the base for regressions in this study, embodies this problem since laborers endowed with highest levels of human capital migrate from low productive regions to highly productive counterparts until the spatial equilibrium is reached. In practical terms, local areas with a large number of highly productive firms may have more generous local budgets to build their modern amenities. The presence of both these productive businesses and local amenities help attract highly educated workers to come to work and settle down, and thus leads to the rise in the share of these workers in the region. This can be seen from the report of Abel & Deitz (2011) that a significant share of the highly educated workers in the local labor force of US regions is supplied by immigration rather than local universities. In another situation, high-tech and knowledge-intensive businesses, which tend to have high labor productivity, may make their location choices based on the share of university-educated workers in the local region. As a consequence, the OLS estimate of HCE is upward-biased under the presence of this location selection.

A better solution to deal with the caveats discussed above, in comparison with the only measure of including ψ and χ in (3.40), is the application of panel data with multi-level FE as expressed in chapter 2 of this thesis. However, the panel data used in this study is available for only two years (2011 and 2016), which reduces the effectiveness of multi-level FE. Specifically, when firm FE is employed to this sort of data, only observations that appear at least twice in the sample are usable. This action leads to an artificially balanced panel because it covers only survivors over 5 years of sampling. For the case of Vietnam, the data set for this study shows that only 28.45% of firms above 10 employees remained active throughout the period 2011–2016. If the survival ability of these firms over such a relatively long period is significantly correlated with regional education, the biased estimate of HCE is unavoidable. This can happen, for instance, when highly educated customers make more strict demands on quality of goods, which eliminates low productive firms from the local markets as a consequence. This relationship is supported empirically by Millan et al. (2014) for the European Union, though Doms et al. (2010) find it insignificant for the US. The problem is milder in the case of the sample with 6 consecutive years because the survival pattern is more diverse in unbalanced panel data. Another limitation of the two-year panel data is that the absorption of provincial and industrial shocks through time-province and time-industry FE terms becomes less efficient due to the absence of the information for the period 2012–2015.

Above all, whether or not the data is available for 6 years, the application of multi-level FE still suffers from a number of weaknesses in estimating the external effects. First and foremost is its incapability of controlling district-specific shocks. Second, it could not deal with local shocks or location choices made before 2011. Third, there could be a lot of other hidden channels through which district-level time-varying characteristics are correlated to both local university graduates and firm-level productivity. If this problem exists, it is beyond the capabilities of the FE method. Meanwhile the IV approach has been known for a long time about its capability of tackling the problem of endogeneity by exploiting exogenous variations (Wooldridge, 2015, Chapter 15). To overcome these limitations and deal with identification concerns mentioned above, this study relies primarily on the TSLS method with historical instruments in the spirit of Moretti (2004a) and Duranton (2016b) for the statistical inferences. The information on the instruments are expressed in detail in the next section. It is worth noting that the technique of multi-level FE is still applied, but serves as a robustness check.

3.5. Data and instrument

3.5.1. Data, variables, and regional context

To estimate UE and HCE with a firm-level outcome, this study employs the unique firm-level data collected from the Comprehensive Censuses of Enterprises in Vietnam for the years 2011 and 2016[21], which covers 343,214 and 517,203 firms respectively. Basically, the data contains the information that is similar to the data sets expressed in the previous chapter, but adds the contents on the education and demographic characteristics of directors and workers. It is worth mentioning that this study considers a firm director as an entrepreneur working in that firm, and uses the two terms interchangeably afterwards. This viewpoint follows a definition in Parker (2018, p. 7) that an entrepreneur is someone who takes risks to earn profit from doing business. Variables are calculated based on these data sets as follows. First, due to the focus on the spillover from higher education, the education level $B_{i,j}^{E}$ for each director is a binary variable, which gains the value of 1 if he or she owns a university degree and 0 otherwise. Meanwhile, the variable $B_{i,j}^{W}$ is referred to as the share of firm workers with university education. Other controls for characteristics of entrepreneurs include log of age, squared log of age, a dummy variable of gender, and a categorical variable of ethnicity to identify whether he or she is Kinh ethnic – the dominant ethnic group in Vietnam, minority ethnic, or foreign. The remaining controls for workers are log of average age and its squared term. All these demographic features are controlled following the entrepreneurship theories and evidence expressed in Doms et al. (2010), Millan et al. (2014), and Parker (2018).

Turning to firm-level non-human capital factors, the production factors $k_{i,j}$ and $e_{i,j}$ are measured with per-worker values of fixed assets and annual electricity consumption respectively. Values of fixed assets are deflated using the deflator of development investment capital of society, similar to the previous study in this dissertation. Other characteristics of firms included in the specification are age, state ownership, and foreign ownership. Regarding the explained variable, this study adheres to the theoretical model by using log of revenue per employee as the variable $y_{i,j}/l_{i,j}$, although value-added per employee is also used as a robustness check. The values of these output variables are both deflated using province-level deflators, collected from the Province-level Statistical Yearbooks published annually by Vietnamese provincial authorities. The clean-up procedure for the regression is similar to the

[21] So far, GSO has carried out this sort of census with detailed information on education level of directors and workers only for these two years, therefore the data of any other years is unusable to estimate HCE.

previous study in this dissertation, except for three important differences. First, this study includes observations from both industrial and service sectors. This helps generate statistical inferences with the interpretation for the economy in general rather than for a single manufacturing field as in chapter 2. Second, the baseline regression sample removes observations from firms with 10 employees or fewer to avoid noise commonly found in micro-sized firms' reports, especially with the presence of the service sector in the data. The problem of random sampling should not be a worry in this study because the enterprise censuses cover the firms of all sizes in 2011 and 2016. After cleaning up, the cross-sectional samples of 2011 and 2016 remain 73,326 and 96,817 observations respectively. Despite this, the other regression samples which cover only firms above 5 or 20 employees are also used as a sensitivity check. The third and also final main difference is that the removal of firms that change their district-level location between 2011 and 2016 is not applied to the cross-sectional data of each year. However, this removal procedure is adopted when the panel data is employed for the regression to tackle the potential biases caused by movers.

The next aspect is the construction of regional variables. Since a number of firms have multiple affiliated establishments, the establishment-level information from the censuses for 2011 and 2016 is exploited to obtain more accurate regional measures. The available information for each establishment include the sector and location codes, the tax identification to match with its parent firm, number of employees, and education attainment structure of the establishment's directors and workers[22]. It is worth highlighting that the regression is still conducted at the firm level because the information on other production inputs and outputs is missing at establishment level. To adhere to the theoretical model expressed in section 3.3, the variables $\ln(L_i)$ and B_i are defined primarily as district-level total employment and the district-level share of university-educated employees respectively. Their formulas are

$$\ln(L_i) = \ln\left(\sum_{j \in i} l_{i,j}\right)$$

and

$$B_i = \frac{\sum_{j \in i} l_{i,j}^B}{L_i}$$

[22] Unfortunately, the reports on the average education of establishment's managers and workers is not available in 2016. To fill that gap, this establishment-level education information is assumed to be similar to its firm-level information.

where $l_{i,j}$ and $l_{i,j}^{B}$ is the workforce and the number of workers with university education respectively of establishment j in district i. The alternative measure of $\ln(L_i)$, which is used as a robustness check, is logarithm of district-level population, extracted from the Province-level Statistical Yearbooks. Finally, there are four instruments used in the TSLS method in total to tackle the potential endogeneity of B_i and $\ln(L_i)$. This historical information is gathered from the Vietnam Population and Housing Census in 1999. The detailed list of these instruments and their usages is shown in Table 3.1.

Tables 3.2 and 3.3 provide the summary statistics for main variables and instruments used in this study for the years 2011 and 2016 respectively. The sample of 2016 is preferable as baseline because its 17 years after the time of instruments (1999) is more assuring for the exogeneity condition compared to the 12-year gap of the sample of 2011. The rationales behind this argument are expressed in the next subsection. Table 3.4 shows the correlation matrix of main variables for the regression, while Table 3.5 shows the sectoral structure of the regression samples. It can be seen that observations in the main samples are quite balanced between the industry and the service sector. It is worth pointing out several interesting demographic statistics using Table 3.2 and Table 3.3. In 2011, about 61% of firm directors owned at least a university degree, while the number for workers was merely 15,8%, implying a higher level of human capital that is much higher for entrepreneurs on average. This fact supports the model's assumption on the human capital advantage of entrepreneurs over workers. The numbers both increased by about 6 percentage points over the period 2011–2016, meaning the increase in the supply of highly educated employees over the course of 5 years. Next, 4 out of 5 directors were men, which was almost constant over the two years. The persistent trend of the higher percentage of men involving entrepreneurial activities is also reported in Kerr et al. (2017) and Parker (2018, p. 300) for most countries in the world, especially in developing countries. Finally, only about 10% of all entrepreneurs were either foreign or minority ethnic, and entrepreneurs were approximately 12 years older than workers on average.

Table 3.1. Summary of instrumental variables

Instrumental variables	Denotation	For endogenous variables
District-level share of university-educated population aged 50 or higher in 1999	IV_1	B_i
District-level share of student population aged from 20 to 24 in 1999	IV_2	

Logarithm of district-level population aged 20 or higher in 1999	IV_3	$\ln(L_i)$
Logarithm of district-level population aged 60 or higher in 1999	IV_4	

Table 3.2. Summary statistics of main variables in the sample of 2011

Variable	Obs	Mean	Std. dev.	Min	Max
Ln(Revenue per employee)	73,326	5.567	1.508	0.012	13.012
Ln(Value added per employee)	69,523	4.100	1.056	0.001	12.947
Ln(Capital per employee)	73,326	4.177	1.606	0.0001	13.659
Ln(Electricity per employee)	73,326	6.394	1.753	0.046	17.305
Ln(Number of employees)	73,326	3.572	1.123	2.303	11.346
Entrepreneur's university education	73,326	0.611	0.488	0	1
Workers' university education	73,326	0.158	0.220	0	1
Male entrepreneur	73,326	0.803	0.398	0	1
Foreign entrepreneur	73,326	0.023	0.149	0	1
Minority ethnic entrepreneur	73,326	0.071	0.257	0	1
Entrepreneur's age	73,326	44.109	9.705	18	88
Squared entrepreneur's age	73,326	2,039.765	888.114	324	7,744
Worker's average age	73,326	32.019	5.763	23.8	65
Squared worker's age	73,326	1058.456	398.009	566.44	4,225
Firm age	73,326	6.304	6.269	0	66
State ownership	73,326	0.017	0.096	0	1
Foreign ownership	73,326	0.072	0.253	0	1
Ln(District employment)	73,326	10.612	1.433	4.304	12.690
Ln(District population)	73,326	12.208	0.631	9.278	13.642
District university education	73,326	0.158	0.107	0	0.601

Notes: The unit of revenue, value added, and capital stock is million VND per year, the unit of electricity usage is Wh.

Table 3.3. Summary statistics of main variables and instruments in baseline sample of 2016

Variable	Obs	Mean	Std. dev.	Min	Max
Ln(Revenue per employee)	96,817	5.888	1.469	0.004	14.138
Ln(Value added per employee)	92,984	4.415	1.052	0.010	14.672
Ln(Capital per employee)	96,817	4.438	1.621	0.002	12.518
Ln(Electricity per employee)	96,817	5.848	2.057	0.009	19.311
Ln(Number of employees)	96,817	3.523	1.094	2.303	10.489
Entrepreneur's university education	96,817	0.678	0.467	0	1
Workers' university education	96,817	0.212	0.251	0	1

Male entrepreneur	96,817	0.794	0.404	0	1
Foreign entrepreneur	96,817	0.019	0.138	0	1
Minority ethnic entrepreneur	96,817	0.079	0.270	0	1
Entrepreneur's age	96,817	45.253	9.909	17	89
Squared entrepreneur's age	96,817	2146.042	936.603	289	7,921
Worker's average age	96,817	33.978	5.747	23	65
Squared worker's age	96,817	1,187.563	411.169	529	4,225
Firm age	96,817	8.026	6.672	0	71
State ownership	96,817	0.008	0.074	0	1
Foreign ownership	96,817	0.071	0.254	0	1
Ln(District-level employment)	96,817	10.821	1.406	4.585	12.894
Ln(District-level population)	96,817	12.275	0.632	8.700	13.808
District-level university education	96,817	0.206	0.124	0.013	0.580
$\boldsymbol{IV_1}$	96,817	0.062	0.073	0	0.287
$\boldsymbol{IV_2}$	96,817	0.185	0.158	0.004	0.701
$\boldsymbol{IV_3}$	96,817	11.416	0.620	6.668	12.670
$\boldsymbol{IV_4}$	96,817	9.386	0.642	4.615	10.418

Notes: The unit of revenue, value added, and capital stock is million VND per year, the unit of electricity usage is Wh.

Table 3.4. Correlation matrix of main variables in the baseline sample in 2016

	1	2	3	4	5	6	7	8	9	10
1. Ln(Revenue per employee)	1.000									
2. Ln(Value added per employee)	0.652	1.000								
3. Ln(Capital per employee)	0.376	0.399	1.000							
4. Ln(Electricity per employee)	0.106	0.058	0.092	1.000						
5. Ln(Number of workers)	-0.042	0.092	0.028	-0.286	1.000					
6. Entrepreneur's education	0.122	0.208	0.015	-0.043	0.171	1.000				
7. Workers' education	0.225	0.325	0.004	0.038	-0.124	0.322	1.000			
8. Ln(District employment)	0.182	0.202	-0.066	0.071	0.079	0.324	0.278	1.000		
9. Ln(District population)	0.100	0.083	-0.085	0.069	0.022	0.179	0.108	0.717	1.000	
10. District university education	0.154	0.197	-0.097	0.001	-0.074	0.318	0.471	0.479	0.183	1.000

Table 3.5. Structure of economics sectors in the final cross-sectional samples

	Sectors	No. of obs of 2011	No. of obs of 2016
Industry	Manufacturing	23,788	27,720
	Construction	13,358	17,628
	Mining and quarrying; Electricity, gas and water supply	4,023	5,448
Services	Market Services (Trade; Transportation; Accommodation and food; and Business and administrative services)	31,011	44,198
	Non-market services (Public administration; Community, Social and other services and activities)	1,144	1,824
	Total	73,324	96,818

Regarding patterns of urbanization and regional higher education in Vietnam, Table 3.6 presents how uneven the spatial distribution is by listing the most and least populous province-equivalent regions. As can be seen, the seven most populous provinces make up in total approximately a third of the whole population, while the number is merely about 3.8% for the seven least populous provinces, given that there are 63 province-equivalent regions in Vietnam. In addition, the spatial concentration varies across regions. Among the seven most populous regions, Ho Chi Minh and Binhduong are strongly urbanized with the urbanization rate of above 75%, which is contrast to Thanh Hoa and Nghe An, where the rates are only 17.1% and 15.1% respectively. In terms of economic aspect, Ho Chi Minh City, Hanoi, Hai Phong, Dong Nai, and Binh Duong jointly contribute to nearly 50% of Vietnam's GDP in 2016. One may expect that these uneven and diverse spatial patterns of regional population and economy provide much variations across regions, thus are useful for the regional analysis.

As regards regional university education, this data is illustrated in the map of mainland Vietnam in Figure 3.3. It is worth noting that the boxes on the map point to the capitals or economic centers of the seven most populous provinces in Vietnam. There are two patterns emerging from the map. First, the presence of highly-educated workers tends to be more intense in central districts and cities than in suburban areas. This is reasonable because the manufacturing factories, which attract many blue-collar workers, tend to be in suburban areas, while there are more white-collar workers found in the city centers. The strong correlation between regional employment and university education at the district level is also seen in Figure 3.2 with an upward regression line. Second, there are several regions that have a high share of university-educated employees but a low scale of employment. In the map, they are among largest-sized districts with the darkest color. In Figure 3.2, they are observations with high values of B_i but lie below the regression line. These regions are characterized with the lion share of the labor

Table 3.6. Most and least populated province-equivalent regions in Vietnam, 2016

Population rank	Province-level regions	Population (million)	Share of urban population (%)
1	Ho Chi Minh City	8.479	81.2
2	Hanoi	7.591	53.6
3	Thanh Hoa	3.571	17.1
4	Nghe An	3.203	15.1
5	Dong Nai	2.951	35.0
6	Binh Duong	1.996	76.5
7	Hai Phong	1.985	46.7
57	Ninh Thuan	0.601	36.2
58	Dak Nong	0.594	15.2
59	Dien Bien	0.568	15.1
60	Cao Bang	0.524	23.2
61	Kon Tum	0.508	35.6
62	Lai Chau	0.436	17.4
63	Bac Kan	0.308	18.8
	The whole country	**91.223**	**34.5**

Source: Province-level Statistical Yearbooks in 2016

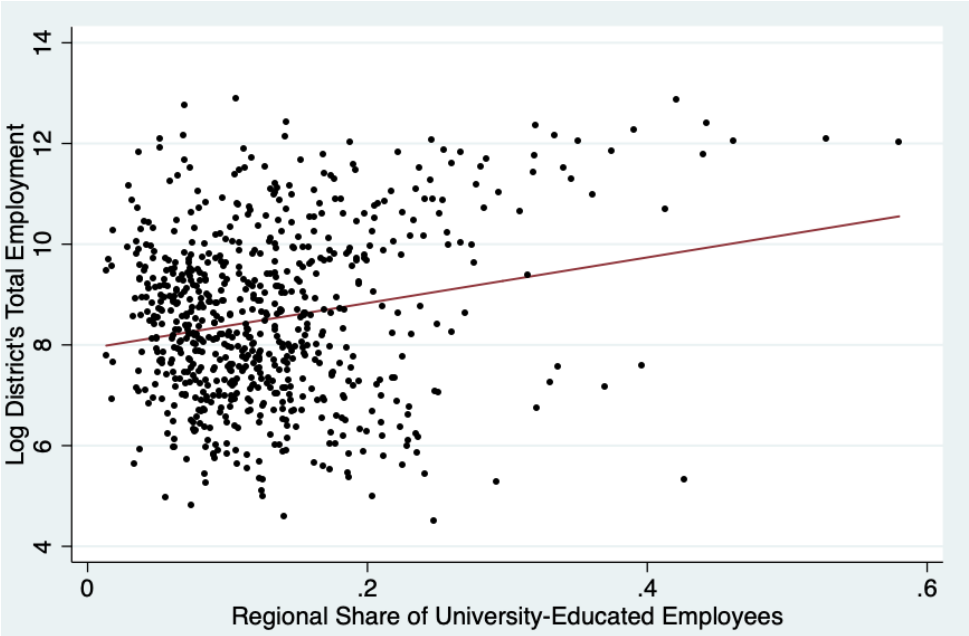

Figure 3.2. Urbanization and regional human capital, 2016

Notes: The vertical axis represents the log values of district-level total employment, while the horizontal axis indicates the district-level share of workers with university education. These variables are computed using data from the Comprehensive Census of Enterprises conducted by GSO for the year 2016. There are 707 districts.

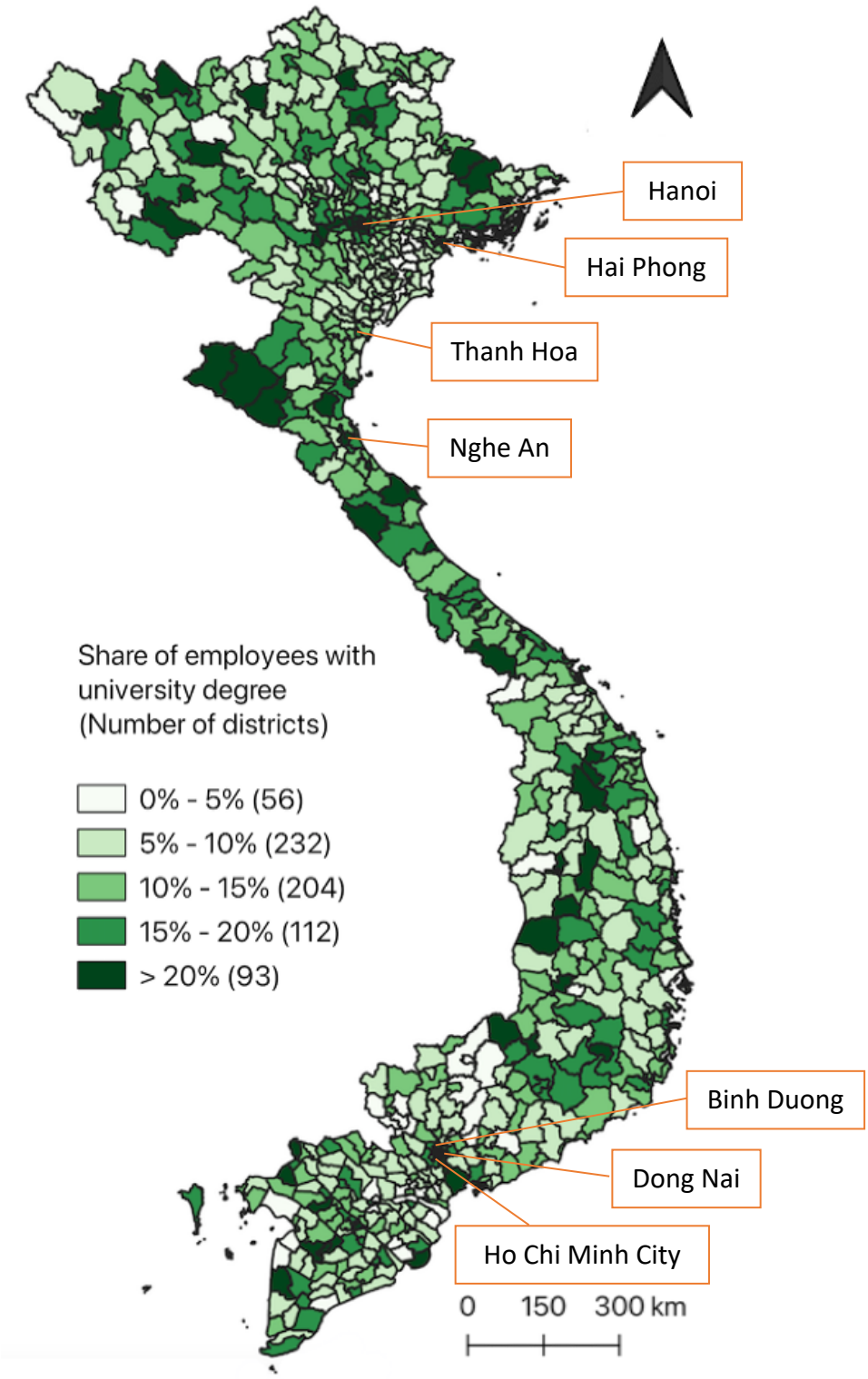

Figure 3.3. Map of mainland Vietnam with share of university-educated employees in district-level regions in 2016

Notes: The map is drawn using data from the comprehensive census of enterprises conducted by the General Statistical Office in 2016 and spatial data published by the Ministry of Natural Resources and Environment in Vietnam. The boxes on the map point to the capitals or economic centers of the seven most populous provinces in Vietnam.

force working as farmers who are excluded in the censuses of enterprises. Therefore, compared to other regions, the measure of university education for these outliers tends to depend on a smaller number of local firms, which are mainly micro-sized firms with several workers. This might suggest the issues of occupational selection, which is found more commonly in agriculture-dominant low-income regions (van der Sluis et al., 2005). Specifically, the most educated children of local farmers are more likely to choose to become entrepreneurs or employees for local firms rather than farmers or self-employees if they decide to stay in the local region, because they can earn higher returns to university education in the region with the lack of higher education. Therefore, one can argue that in regions with a high share of farmers in the labor force, using the education index based on enterprise data is not suitable to capture the external effects of regional higher education. However, the counter-argument is that technical and business knowledge spreads mainly across local enterprises and their workers, thus the spillover of regional human capital can be captured without the inclusion of the agricultural sector. Finally, if these outliers are caused by measurement errors to some extent, this should not be a big concern owing to the use of regional instruments in this study.

3.5.2. The validity of instruments

To address the endogeneity concerns raised in the regression of firm productivity on the external effects and to obtain causal inferences, this study employs the TSLS technique with four historical instruments listed in Table 3.1. There are two requirements that these instruments must meet to be valid in the TSLS regression. In particular, the instruments must be the predictors of the endogenous variables conditional out other variables in the specification (the relevant condition) but uncorrelated to unobservable factors in the error terms (the exogeneity condition). In other words, in the case of B_i and $\ln(L_i)$, the valid instruments are the ones that influence productivity only through their correlations to the spatial distribution of human capital and employment. There are two stages of regression. In the first stage, the endogenous variables B_i and $\ln(L_i)$ are regressed separately on exogenous instruments along with all firm-level variables and the province and sector FE terms. Intuitively, this stage helps "remove" the endogenous variations from B_i and $\ln(L_i)$ to yield their fitted values denoted as $\hat{B}_i$ and $\ln(\hat{L}_i)$. These values are thus considered as the "exogenous" versions of B_i and $\ln(L_i)$. In the second stage, the specification (3.40) is estimated only with OLS using $\hat{B}_i$ and $\ln(\hat{L}_i)$ to substitute for B_i and $\ln(L_i)$ respectively. All in all, the validity of the instruments is crucial to obtain the consistent estimates of δ_0 and δ_1.

The validity is IV_1 is the first issue that should be discussed. Its usage is inspired from a popular instrument for education level of workers in wage regression in labor economics literature, namely education of his or her parents. The application of this instrument is based on the evidence that children of high-educated parents tend to obtain a higher level of education afterwards in comparison with less-educated parents. Therefore, the requirement of significant correlation to the endogenous variable (education level of workers) is met. Besides, putting the role of a worker's education aside, it is hard to point out any other universal channels through which the education level of his or her parents might influence his or her productivity or the productivity of firms where the worker works, conditional on other worker's characteristics. This assures the exogeneity condition for a valid instrument. As a generalization to the regional level, the education level of previous population generations should be a suitable instrument for the current generation's education of regional employment in explaining firm-level productivity, controlling other influential regional factors. Another feature that makes IV_1 a suitable instrument is the inclusion of only the population aged 50 or higher in 1999. Given that the official retirement ages in Vietnam in 2016 for male and female are 60 and 55 respectively, most workers observed in the instrument were already out of the labor market in 2016 when they were all at least 67 years old. This logic suggests that IV_1 is very likely the exogenous source to B_i.

The next instrument is IV_2, which is employed following closely to Duranton (2016b) who uses the number of local universities, or the supply of local university courses in other words, as an exogenous variation to B_i. The rationale behind this choice is that residents living in a district with one or more universities are more likely to receive formal education from these HE facilities due to the cost advantages compared to ones living in other districts. Similarly, IV_2 can predict B_i through the distribution of university students across districts in the past. The concrete reasons are as follows. Typically, 18-year-old people in Vietnam in 1999 after finishing high school can start their next study life following three different options. The first is to enroll in a vocational school, which requires 2 years or less to finish. The second is to enroll in a junior college, which normally requires 3 years to finish, and the third is a university with 4 years of studying needed. This means that if a person aged 20–24 was still studying, he or she was very likely a student of a certain junior college or a university. Besides, after graduation from a junior college, one can register for an "upgrade" course in a university and spend at least 2 more years to obtain a degree in that university. Furthermore, between 2000 and 2010, many junior colleges in Vietnam were "upgraded" to universities when they had met

criteria set by the Ministry of Education. Finally, students tend to live as close as possible to their universities. Its consequence is that the district where a university is located tends to host the largest number of students, while the surrounding districts host less, and far away districts host the least. All in all, IV_2 is a good predictor of the future distribution of highly educated workers across space, and thus superior to the number of local higher education institutes as a valid instrument.

The common and most important feature of IV_1 and IV_2 is the distant-past nature of their values. Their observations were made in 1999 before the introduction of drastic education reform programs in later years. In particular, the Vietnamese government passed the Resolution No. 14/2005/NQ-CP of November 02, 2005 aiming to fundamentally and comprehensively improve the higher education system. The resolution made many novel policies including encouraging the development of privately-owned and foreign-owned universities, which reduced the near monopoly of state-owned university education facilities at the time; increasing the enrollment scale of universities in order to raise the ratio of students to population; and expanding the transition programs from junior colleges to universities. To strengthen further these education policies, the Vietnamese parliament passed the Law on Higher Education in 2012. This law made the resolution in 2005 into law and added several innovative policies such as decomposing higher education programs into academic-oriented and application-oriented, boosting technology transfer between universities and firms, and encouraging business corporations to build their own universities to meet their labor demand. Due to these national-scale reforms, according to the Ministry of Education, the number of universities in Vietnam has grown quickly from 70 in 1999 to 235 in 2016. This more-than-triple increase in the number of universities over 17 years has substantially altered the spatial distribution of highly educated workers across districts. To summarize, the education policy changes between 2005 and 2012 aimed at considerably improving both the quality and quantity facets of university education. This helped alter the state-controlled and central planning higher education system to become more market-oriented. In the spirit of Muravyev (2008), these sorts of society-changing policies can be considered as "natural experiments". IV_1 and IV_2 are observed before the "experiment", thus their fit to the exogeneity condition is secured.

Regarding the solution to counter the potential endogeneity of $\ln(L_i)$, this study employs IV_3 and IV_4 following the instrument choices of Combes et al. (2010) and Duranton (2016b). These authors instrument for regional employment in France and Colombia respectively with its information on population recorded from 50 to 150 years before the year of research samples.

The choice is based on the central argument that population and employment structures are persistent over time, hence they remain correlated to each other even after a long time. Moreover, the distant gap of time reduces the likelihood that those long-lagged instruments are significantly correlated with any current local characteristics that affect firm productivity. In the case of Vietnam, this sort of population information before the year 2000 is only available through population censuses conducted in 1979, 1989, and 1999. Among these years, only the population census in 1999 provides information as detailed as the commune level, which is strictly needed to match to the district-level data in 2016. The reason, as expressed in chapter 2, is that there are the mergers or divisions of many communes and districts between 2000 and 2016. Therefore, commune-level historical data is needed to match correctly to district-level information of the year 2011 or 2016. However, a drawback to this choice is that the chronological distance between 1999 and 2016, the latter country of the two years, is only 17 years. This might be too short to ensure the exogeneity condition in comparison with the gap of at least 50 years in Combes et al. (2010) and Duranton (2016b).

Despite this, one can argue that this choice is still safe since Vietnam has witnessed fundamentally dynamic changes in socio-economic conditions since the *Doimoi* reform in 1986. Observations in 1999 are made before the happening of "experimental" events that influence enormously the employment distribution across districts through local population growth and internal migration. Those events include the approval of the Enterprise Law in 2000, the Competition Law in 2005, and the Foreign Investment Law in 2011. In addition, the event that Vietnam became an official member of the World Trade Organization (WTO) in 2006 is also influential. Since then, the growing influx of Foreign Direct Investment (FDI) to Vietnam has dramatically changed districts and provinces where FDI firms are located. These regions have attracted a huge volume of domestic immigrants to work in large factories or for supporting services, thus altering the spatial distribution of employment across space. To further distance from the information on employment in 2011 and 2016, this study chooses only a segment of population in 1999 as the instruments. With IV_3, the group of population under 20, which represented future labor force, was removed from the total population in 1999. Meanwhile, IV_4 includes only the population over 60, who were almost out of the labor market in 2011 and 2016. All in all, IV_3 and IV_4 are considered to be valid instruments for $\ln(L_i)$. Besides, the sample of 2016 is preferable as the baseline sample because its further distance to 1999 is "safer" to the exogeneity condition compared to the sample of 2011.

Turning to the statistics and tests for the instruments, Table 3.7 reports the pairwise correlations

between B_i, $\ln(L_i)$, and the four instruments using the baseline sample. The first-stage estimation results presented in Table 3.8 confirm the strong prediction of all instruments with significant estimators and high R-squared values. The risk of weak instruments should also not be the case because the values of F-statistics are well above the value threshold of 10 as mentioned in Stock & Yogo (2005). Finally, the implementation of overidentification tests is feasible owing to the availability of more than one instrument for each endogenous variable. The zero hypothesis of this test is that instruments are uncorrelated to the error terms. Therefore, it is expected that this the P-value of this test is well above 0.1 to not reject the hypothesis. In practice, the results reported in the next sections always report the high p-value for the overidentification test when the TSLS method with multiple instruments and the full specification is applied. To wrap up, this study considers instruments listed in Table 3.1 to be valid for district-level university education and regional employment when estimated with a firm's labor productivity.

Table 3.7. Correlation matrix of main regional variables and instruments, 2016

	1	2	3	4	5	6
1. District-level university education as B_i	1.000					
2. Ln(District-level employment) as $\ln(L_i)$	0.481	1.000				
3. IV_1	0.812	0.507	1.000			
4. IV_2	0.766	0.565	0.874	1.000		
5. IV_3	0.352	0.571	0.316	0.341	1.000	
6. IV_4	0.299	0.448	0.273	0.253	0.951	1.000

Table 3.8. First-stage estimation, 2016

Endogenous variable	B_i		$\ln(L_i)$	
$IV1$	1.021*** (0.152)	-	-	-
$IV2$	-	0.360*** (0.041)	-	-
$IV3$	-	-	0.966*** (0.158)	-
$IV4$	-	-	-	0.690*** (0.178)
Province/Sector FE	Yes	Yes	Yes	Yes
F-test	240.324	220.431	25.078	19.760
R^2	0.844	0.779	0.698	0.662

Notes: Cluster-robust standard errors in parentheses calculated at the province level. The explained variables are the endogenous variables. All regressions include the full set of firm-level factors expressed in the specification (3.40). Only the estimators of instruments are reported for brevity.

*** Significant at 1%

3.6. Results

3.6.1. Results without an instrumental approach

This subsection starts with estimation results without the application of an IV approach, reported in Table 3.9. The regressions are conducted across columns with various regression methods and samples. Columns (1) and (2) show the OLS results for the cross-sectional samples of 2011 and 2016 respectively, mirroring exactly the specification in (3.40) and treating B_i and $\ln(L_i)$ as exogenous variables. Columns from (3) to (5) display the estimated results using the panel data. The method used for column (3) remains the OLS, with the use of year-province and year-industry FE in place of the normal FE terms of province and sector. This helps control local and sectoral shocks in 2016 with the year 2011 as the base year, owing to the emergence of the time dimension. In columns (4) and (5), the techniques of FD and FE are applied respectively along with the removal of observations that change their district-level location to reduce the problem of location endogeneity. In essence, the FD method treats 2011 and 2016 as times t and $t+1$, transforms (3.40) into a first-differenced equation with the subtraction of the t values from the $t+1$ values on both sides of the equation, then estimates the new equation with OLS. By doing this, all unobserved constant factors at the firm-level and district-level are eliminated to alleviate the issue of omitted variables. Meanwhile, the Hausman test supports the application of the multi-level FE technique against the RE method in column (5), with the technical details like the previous chapter. Basically, FD and FE deal with the same issue of unobserved time-invariant variables, thus they should produce the same results. The slight difference in the estimates between columns (4) and (5) caused by the difference in the impacts of FE terms in the two samples. In general, independent variables on the right-hand side of the specification (3.40) explain well the variations in firm-level labor productivity with all methods, which goes up to 50.1% in column (1) with OLS and the sample of 2011.

As for the main regional factors of interest, the estimates of both δ_0 and δ_1 are both positive across various samples and methods, as expected from the theoretical model. However, while $\hat{\delta}_1$ is always strongly significant, $\hat{\delta}_0$ tends to be insignificant. The exception is the baseline sample of 2016 in column (2) where $\hat{\delta}_1$ is significant, despite a significance level of only 10%. Moving from columns (1) – (3) to (4) – (5), the magnitude of $\hat{\delta}_1$ increases from around 0.04 to around 0.1. This difference could be interpreted as the short-run versus long-run effects of urbanization, as discussed in the previous chapter, implying that UE is stronger in the short-

Table 3.9. Estimation results without an identification strategy

Column		(1)	(2)	(3)	(4)	(5)
Method		OLS	OLS	Pooled OLS	FD	FE
Sample		2011	2016	2011 & 2016		
Main firm-level variables	Capital $k_{i,j}$	0.276*** (0.007)	0.292*** (0.009)	0.291*** (0.007)	0.258*** (0.016)	0.254*** (0.014)
	Electricity $e_{i,j}$	0.140*** (0.005)	0.040 (0.029)	0.069*** (0.022)	0.031*** (0.011)	0.026** (0.009)
	Labor $l_{i,j}$	0.053*** (0.013)	0.051** (0.020)	0.057*** (0.014)	-0.156*** (0.033)	-0.170*** (0.019)
	Entrepreneur's education $B_{i,j}^{E}$	0.017 (0.013)	0.025** (0.012)	0.024** (0.015)	0.017 (0.016)	0.012 (0.017)
	Workers' education $B_{i,j}^{w}$	0.838*** (0.073)	0.983*** (0.053)	0.941*** (0.065)	0.287*** (0.030)	0.355*** (0.038)
Regional variables	District university education B_i	0.139 (0.173)	0.325* (0.181)	0.249 (0.175)	0.136 (0.422)	0.225 (0.314)
	District employment $\ln(L_i)$	0.043*** (0.008)	0.033*** (0.008)	0.039*** (0.007)	0.102*** (0.031)	0.095*** (0.034)
Control entrepreneur's characteristics	Male	-0.069*** 0.015)	-0.048*** (0.010)	-0.056*** (0.011)	-	-
	Foreign	0.017 (0.051)	0.029 (0.039)	0.032 (0.043)	-	-
	Ethnic minority	-0.013 (0.041)	0.252** (0.115)	0.168* (0.095)	-	-
	Age	0.028*** (0.004)	0.034*** (0.004)	0.031*** (0.003)	0.048*** (0.004)	0.044*** (0.004)
	Squared Age	-0.0003*** (0.00004)	-0.0004*** (0.00004)	-0.0003*** (0.00003)	-0.001*** (0.00004)	-0.0005*** (0.00004)
Control workers' characteristics	Age	0.075*** (0.011)	0.092*** (0.012)	0.085*** (0.013)	0.035*** (0.009)	0.043*** (0.008)
	Age Squared	-0.001*** (0.0002)	-0.001*** (0.0002)	-0.001*** (0.0001)	-0.001*** (0.0001)	-0.001*** (0.0001)
Control other firm characteristics	Firm Age	0.009*** (0.001)	0.007*** (0.001)	0.008*** (0.001)	-	-
	State ownership	0.219*** (0.048)	0.037 (0.049)	0.156*** (0.039)	0.294*** (0.057)	0.255*** 0.055
	Foreign ownership	0.045 (0.051)	-0.260 (0.164)	-0.153 (0.108)	0.027 (0.038)	0.045 0.032
	Time-Province FE	No	No	Yes	No	Yes
	Time-Sector FE	No	No	Yes	No	Yes
	Province FE	Yes	Yes	No	Yes	No
	3-digit Sector FE	Yes	Yes	No	Yes	No
	Firm FE	No	No	No	No	Yes
	No. of obs	73,326	96,817	170,126	43,581	163,414
	R^2	0.501	0.485	0.495	0.175	0.257

Notes: Cluster-robust standard errors in parentheses calculated at the province level. The explained variable is firm labor productivity in all columns. A full set of controls is included in columns (1) – (3), while several controls are removed in columns (4) – (5).

Abbreviation: FE, fixed effects; OLS, ordinary least square; FD, first-difference; RE, random effects.

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

-run. In conclusion, the OLS, FD, and FE results support the existence of UE, while the evidence of HCE is weak, keeping in mind the causal impacts of these external effects is uncertain without an IV or experimental approach. Turning to main firm-level variables, signs of their estimators are all positive as expected, except for $l_{i,j}$, which is negative in the regression with FD and FE. This is unsurprising given that $l_{i,j}$ appears also as the denominators in the computations of other per-worker production factors. Since $\ln(x_{i,j}/l_{i,j}) = \ln(x_{i,j}) - \ln(l_{i,j})$, the estimator of $\ln(l_{i,j})$ in the specification reflects the net effects, thus can be negative. In terms of firm-level human capital-related factors, while the university education of workers shows a significant impact on labor productivity in all columns, the evidence for entrepreneurs is only found significant with OLS in the sample of 2016 or the pooled sample of the two years. Despite this, the magnitude of 0.025 in column (2) is modest, meaning that a firm with a university-educated entrepreneur has 2.5% higher labor productivity than another firm without such an entrepreneur, holding other factors fixed.

Next, it is worth highlighting the estimated results of other firm-level controls[23]. Interestingly, female entrepreneurs perform better than their male peers despite the strong presence of gender inequality in Vietnam. These findings are at odds with the evidence on the superior performance of male entrepreneurs expressed in Parker (2018, p. 309 & 322) who also points out that the entrepreneurship of women in developing countries is constrained by their lower human capital, childcare responsibility, and more restrictive access to external finance. Turning to ethnic characteristics, foreign and minority-ethnic entrepreneurs are found not to perform significantly better than their Kinh-ethnic peers. The exception comes from the regression for 2016 in column (2). It shows that an entrepreneur belonging to an ethnic minority group gives a 25.2% advantage to labor productivity of his or her own firm in comparison with other firms whose directors are Kinh-ethnic. In terms of age-related terms, their estimators are all statistically significant across columns. Their results from the baseline sample in column (2) indicate that the relationship between firm's productivity and ages of entrepreneur and workers is inverted-U shaped, whose graphs peak at the ages of 42.5 and 46 respectively. These

[23] Since the time-invariant or less-time-variant nature of several controls is not suitable for the methods of FD and FE, which is based on the within-variation, they are dropped from the regressions in column (4) and (5).

quadratic patterns are consistent with the findings of Doms et al. (2010). Furthermore, when a firm is a year older than another firm, labor productivity of the former is predicted to be 0.7% higher than the latter, as implied by the estimator of firm age in column (2). Finally, the estimated coefficients of state ownership tend to be positive and significant across columns, while foreign ownership is always insignificant. The productivity advantage caused by the more intense presence of state in ownership of a firm is surprising because it is expected that state power comes with bureaucratic and low-productive business behaviors. However, it might point to the outcome of the Vietnamese government's long-implemented policies towards firms with state ownership. In their efforts to reform the state-owned sector, they have retained mainly the most profitable state-owned firms, and fostered to privatize or dissolve the remaining lower-productive firms since the 2000s. For this reason, state ownership is positively correlated to labor productivity, but the relationship may not be causal. Dealing with this problem is beyond the scope of this study, given that state ownership is not the variable of interest, and whether this variable is included in the regression, the estimates of other variables remain almost unchanged.

3.6.2. Results with an instrumental approach

This subsection turns to the TSLS results using mainly the cross-sectional baseline sample of 2016, which is presented in Table 3.10. Column (1) mirrors the results of column (2) of Table 3.9 as a benchmark to make comparison with the instrumental estimates. Intuitively, the OLS estimators are placed next to the IV results to assess the influences of "endogeneity removers". Columns (2) – (5) report the estimated results with B_i instrumented with IV_1 and IV_2 and/or with $\ln(L_i)$ instrumented with IV_3 and IV_4 as explained in the previous section. It is worth pointing out that the estimates of firm-level factors are very stable over different columns, therefore the interpretation focuses on merely the estimated coefficients of B_i and $\ln(L_i)$. Firstly, column (2) reports the results with both B_i and $\ln(L_i)$ and the full set of controls included. Again, the evidence of UE is verified with an estimated coefficient of 0.035 at the 5% significance level, meaning that a double in district-level total employment leads to a 3.5% increase in labor productivity of firms operating in the district. Most importantly, this relationship is now proved to be causal, owing to the application of instruments. The estimate of B_i, however, is not significantly different from zero. Moving from OLS in column (1) to TSLS in column (2), where the endogeneity issue of the external effects is taken into account with instruments, the estimator of $\ln(L_i)$ rises by a minor margin, from 0.033 to 0.35. The comparison test shows that this change is statistically insignificant. This is consistent with the

findings of Combes et al. (2010) and Duranton (2016b) about the slight and insignificant changes in the estimate of UE when the historical instruments are employed. This may suggest the mild problem of endogeneity facing the estimation of $\ln(L_i)$. In contrast, the estimated coefficient of HCE changes quickly from 0.325 at the 10% significance level to an insignificant coefficient of only 0.031. It implies that most variations in regional university education in relation to firms' labor productivity is not exogenous, and the instruments of B_i are at work.

Table 3.10. Baseline results with an identification strategy, 2016

	Column	(1)	(2)	(3)	(4)	(5)
	Method	OLS	TSLS	No controls	Only UE	Only HCE
Main firm-level variables	Capital $k_{i,j}$	0.292*** (0.009)	0.291*** (0.009)	0.300*** (0.007)	0.291*** (0.009)	0.291*** (0.009)
	Electricity $e_{i,j}$	0.040 (0.029)	0.040 (0.029)	0.041 (0.031)	0.040 (0.029)	0.040 (0.029)
	Labor $l_{i,j}$	0.051** (0.020)	0.052*** (0.010)	0.068*** (0.019)	0.052*** (0.019)	0.053*** (0.020)
	Entrepreneur's education $B_{i,j}^{E}$	0.025** (0.012)	0.029** (0.012)	0.025* (0.013)	0.031** (0.012)	0.037*** (0.013)
	Workers' education $B_{i,j}^{w}$	0.983*** (0.053)	1.005*** (0.054)	1.020*** (0.049)	1.008*** (0.064)	1.002*** (0.053)
Regional variables	District university education B_i	0.325* (0.181)	0.031 (0.239)	0.026 (0.264)	-	0.291* (0.147)
	District employment $\ln(L_i)$	0.033*** (0.008)	0.035** (0.015)	0.042** (0.017)	0.034*** (0.011)	-
	Province FE	Yes	Yes	Yes	Yes	Yes
	3-digit Sector FE	Yes	Yes	Yes	Yes	Yes
	Controls	Yes	Yes	No	Yes	Yes
	1st stage stat for B_i	-	64.74	66.81	-	173.74
	1st stage stat for L_i	-	25.61	25.35	40.09	-
	Over-id p-value	-	0.760	0.864	0.678	0.076
	No. of obs	96,817	96,809	96,809	96,809	96,809
	R^2	0.484	0.198	0.191	0.198	0.197

Notes: Cluster-robust standard errors in parentheses calculated at the province level. The explained variable is firm labor productivity in all columns. Column (1) mirrors column (2) of Table 3.9. A full set of controls is included in all columns. The estimates of controls are hidden for brevity. The regressions for columns (2) – (5) are conducted using the Stata command *ivreghdfe*, which removes 8 singleton observations from the baseline sample.

Abbreviation: FE, fixed effects.

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Secondly, column (3) displays the estimated results without any firm-level controls. The small change in the estimates of UE and HCE indicates the modest impacts of controls on the model. Next, to assess the influence of the addition of $\ln(L_i)$ on the estimation of B_i and vice versa,

the variable B_i is removed from the specification in column (4), while the variable $\ln(L_i)$ is absent in column (5). Interestingly, when comparing column (4) to column (2), the estimator of $\ln(L_i)$ remains almost unchanged in the absence of B_i. This implies that the validity of IV_3 and IV_4 and the estimation of $\ln(L_i)$ are not affected by the removal or inclusion of B_i. On the contrary, the coefficient of B_i in column (5) turns to significant with the magnitude of 0.291. However, this result is suspicious because its overidentification test fails, keeping in mind that the test passes in all columns (1) – (4). This could strongly imply that the high correlation of IV_1 and IV_2 with $\ln(L_i)$ has violated their validity in the specification that $\ln(L_i)$ is excluded. Thus, it confirms the concerns on the validity of instruments for HCE in the absence of UE raised in section 2. The situation that the HCE estimator turns to insignificant when UE is included in the specification is also found in Duranton (2016b). All in all, the baseline results suggest the strong evidence of UE, but the weak evidence of HCE, thus support *Hypothesis 1.*

3.6.3. Robustness checks

There are several robustness checks carried out in this study to assess the sensitivity of externalities' estimators. First, columns (2) – (5) in Table 3.11 presents the TSLS results using the baseline sample of 2016 along with the smaller sets of instruments, while all instruments are employed in column (1) as a benchmark. As can be seen, the significant estimate of UE and the insignificant estimate of HCE are robust to various combinations of instruments, with slight changes in the estimated magnitudes of UE. The exception is seen in column (5) when only IV_1 and IV_4 are employed, the estimated coefficient of $\ln(L_i)$ becomes statistically insignificant. At the same time, the first-stage F-statistics of $\ln(L_i)$'s estimation decreases to only 9.41, below the threshold of 10, implying the issue of a weakness instrument. Therefore, the combination of more instruments is preferable in this study.

Next, Table 3.12 reports the TSLS results using various samples and specifications. Again, column (2) of Table 3.12 duplicates column (2) of Table 3.10 to show the benchmark IV estimates of UE and HCE. Moving to column (1), which employs the sample of 2011 instead of 2016, the significance levels of estimated results remain unchanged. Although the sign of B_i's estimator switches, it is not significantly different from zero. The similar scenarios emerge in columns (3) and (4) where the coverage of the sample is enlarged or contracted with different firm-size thresholds. In column (5), $\ln(L_i)$ is measured with district-level population instead, giving the significant coefficient of 0.42, which is close to the benchmark result. Meanwhile, the magnitude of B_i's estimator rises considerably, but it remains insignificant. Surprisingly,

when the method applied in column (5) is OLS[24] rather than TSLS, the estimate of $\ln(L_i)$ is statistically insignificant. This might relate to a capability of the IV approach to overcome the issue of measurement errors. It is worth highlighting that the TSLS estimators of population-based and employment-based measures are very close, leading to two implications. First, there might be some measurement errors in district-level population, which prevent it from reflecting true urbanization. Second, it shows that the instruments are at work. As regards columns (6) – (7), which test the possibility of non-linear influences of the external effects. The insignificant results of all the external terms indicate that that functional form does not fit. Turning to columns (8), whose results aim at verifying the possibility that there exists the interaction between B_i and $\ln(L_i)$, which may explain the statistical insignificance of B_i. The insignificant estimates of that interaction term and the main term of B_i disprove such a possibility.

Table 3.11. TSLS results with various combinations of instruments, 2016

Column	(1)	(2)	(3)	(4)	(5)
Number of instruments	4	3	3	2	2
District university education B_i	0.031 (0.239)	0.051 (0.256)	-0.115 0.179)	-0.052 (0.238)	-0.124 (0.354)
District employment $\ln(L_i)$	0.035** (0.015)	0.032* (0.017)	0.048** (0.020)	0.042** (0.021)	0.050 (0.039)
IV_1	Yes	Yes	Yes	Yes	Yes
IV_2	Yes	No	Yes	No	No
IV_3	Yes	Yes	Yes	Yes	No
IV_4	Yes	Yes	No	No	Yes
1st stage stat for B_i	65.13	50.11	127.03	60.13	51.25
1st stage stat for $\ln(L_i)$	25.71	32.17	14.78	14.24	9.41
Over-id p-value	0.786	0.654	0.563	-	-
No. of obs	96,809	96,809	96,809	96,809	96,809
R^2	0.198	0.198	0.197	0.197	0.197

Notes: Cluster-robust standard errors in parentheses calculated at the province level. The explained variable is firm labor productivity in all columns. Column (1) mirrors column (2) of Table 3.10. A full set of firm-level variables is included in all columns, their estimates are hidden for brevity.

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

[24] This result is kept out of the table for brevity.

Table 3.12. TSLS results with various samples and specifications, 2011, 2016

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year of sampling	2011	2016						
Sample	>10	>10	>5	>20	Popu	Non-lin 1	Non-lin 2	Interact
B_i	-0.090 (0.232)	0.031 (0.239)	-0.004 (0.212)	-0.003 (0.305)	0.289 (0.183)	1.278 (0.864)	0.173 (0.315)	3.325 (4.337)
$\ln(L_i)$	0.041*** (0.016)	0.035** (0.015)	0.037*** (0.013)	0.042** (0.017)	0.042** (0.021)	0.013 (0.022)	0.253 (0.215)	0.070** (0.030)
$B_i \# \ln(L_i)$	-	-	-	-	-	-	-	-0.266 (0.333)
B_i^2	-	-	-	-	-	-1.644 (1.016)	-	-
$\ln(L_i)^2$	-	-	-	-	-	-	-0.011 (0.011)	-
IV_1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV_2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
IV_3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV_4	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV_1^2 and IV_2^2	-	-	-	-	-	Yes	-	-
IV_3^2 and IV_4^2	-	-	-	-	-	-	Yes	-
$IV_1 \# IV_3$	-	-	-	-	-	-	-	Yes
$IV_1 \# IV_4$	-	-	-	-	-	-	-	Yes
1st stage stat for B_i	39.84	25.71	27.19	26.35	132.50	1,405.6	22.08	64.20
1st stage stat for $\ln(L_i)$	28.29	65.13	61.37	63.55	64.74	14,565.4	43.99	161.20
1st stage stat for B_i^2	-	-	-	-	-	15,865.9	-	-
1st stage stat for $\ln(L_i)^2$	-	-	-	-	-	-	21.67	-
1st stage stat for $B_i \# \ln(L_i)$	-	-	-	-	-	-	-	182.38
Over-id p-value	0.950	0.786	0.515	0.221	0.204	0.303	0.798	0.717
No. of obs	73,313	96,809	147,747	58,202	96,809	96,809	96,809	96,809
R^2	0.222	0.198	0.160	0.241	0.198	0.198	0.198	0.198

Notes: Cluster-robust standard errors in parentheses calculated at the province level. The explained variable is firm labor productivity in all columns. Column (2) mirrors column (2) of Table 3.10. A full set of firm-level variables is included in all columns, their estimates are hidden for brevity. The sample covers only firms above 10 employees in columns (1) – (2) and (5) – (7), while the size thresholds for columns (3) and (4) are 5 and 20 respectively. Column (5) uses district-level population as L_i. Columns (6) and (7) add the squared term of B_i and L_i respectively. Column (8) adds the interaction term of B_i and L_i.

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 3.13. Results with value-added per employee

	Column	(1)	(2)	(3)	(4)
	Method	TSLS	TSLS	FD	FD
	Explained variable	Rev	VA	Rev	VA
Firm-level variables of interest	Capital $k_{i,j}$	0.291*** (0.009)	0.242*** (0.008)	0.258*** (0.016)	0.236*** (0.008)
	Electricity $e_{i,j}$	0.040 (0.029)	0.019* (0.011)	0.031*** (0.011)	0.018*** (0.003)
	Labor $l_{i,j}$	0.052*** (0.010)	0.085*** (0.007)	-0.156*** (0.033)	-0.079*** (0.018)
	Entrepreneur's education $B_{i,j}^{E}$	0.029** (0.012)	0.065*** (0.008)	0.017 (0.016)	0.008 (0.011)
	Workers' education $B_{i,j}^{w}$	1.005*** (0.054)	0.888*** (0.083)	0.287*** (0.030)	0.241*** (0.027)
Regional variables	District university education B_i	0.031 (0.239)	0.698*** (0.165)	0.136 (0.422)	-0.021 (0.293)
	District employment $\ln(L_i)$	0.035** (0.015)	-0.031 (0.023)	0.102*** (0.031)	0.073** (0.028)
	Province FE	Yes	Yes	Yes	Yes
	Industry FE	Yes	Yes	Yes	Yes
	Controls	Yes	Yes	Yes	Yes
	1st stage stat for B_i	25.61	25.52	-	-
	1st stage stat for L_i	64.74	65.56	-	-
	Over-id p-value	0.760	0.227	-	-
	No. of obs	96,809	92,974	43,581	39,969
	R^2	0.198	0.250	0.175	0.155

Notes: Cluster-robust standard errors in parentheses calculated at the province level. Column (1) mirrors column (2) of Table 3.10, column (3) mirrors column (4) of Table 3.9. A full set of controls is included in all columns, their estimates are hidden for brevity.

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Finally, columns (2) and (4) of Table 3.13 show the estimated results with the dependent variable as value-added per employee, whose value-added measure is calculated similarly to its measure in chapter 2. To serve as benchmarks, the estimates using the revenue-based explained variables are expressed in column (1) and (3). Interestingly, the estimated results of the externalities terms are starkly different between columns (1) and (2), despite firm-level estimators being stable. Specifically, when the value-added measure is the case, the TSLS estimate of B_i becomes strongly significant, but the estimator of $\ln(L_i)$ turns insignificant. Surprisingly enough, the strong difference disappears when the FD method is applied instead in columns (3) and (4). Since the evidence of the external effects obtained from column (4) is consistent with all previous results using the revenue-based explained variable, the surprising estimates in column (2) might be attributable to the combination of the value-added-based variable and cross-sectional data. The reason is that the value-added measure may contain firm-

specific measurement errors, especially in the service sector where the measurement of the breakdown of the costs may vary from firm to firm. If these firm-level errors are significantly correlated to regional measures, using regional instruments could not tackle this problem. The FD method, meanwhile, can handle this issue well by treating firm-specific measurement errors as a time-invariant firm-level factor, because accounting methods of firms tend to remain stable over time. Consequently, the results presented in Table 3.13 are considered as the evidence of robustness rather than the unusual evidence. This also gives a relief to the use of value-added in production estimation in chapter 2, where the main regression technique used to estimate agglomeration effects is firm FE. Despite this, column (2) sounds a warning of the application of value-added with cross-sectional data. A more careful investigation on this problem is beyond the scope of this study. To summarize, the baseline estimates of urbanization economies and human capital externalities are proven to be robust across various methods, samples, and measurement choices in this study.

3.6.4. Heterogeneous effects of regional university education

Another way to interpret the insignificant impact of HCE is that it shares the mechanism of knowledge spillover with UE. As a result, the estimates obtained in this study imply that when UE and HCE jointly predict labor productivity, the quantity of knowledge source (labor force) might be dominant over its quality (education level). This is reasonable for Vietnam – a developing country, whose economy grows through the cheap and abundant labor force rather than innovation in the first decades of this millennium. This argument is strengthened with the same findings of Duranton (2016b), who does wage regression for Colombia – another developing country, despite the different specification and variable construction. Duranton (2016b) attributes the ambiguous role of HCE to the absence of the informal sector in the data, citing the evidence of higher returns to regional university education to workers in this sector. However, this argument is untestable for Vietnam due to the lack of data in this sector in 2011 and 2016. Rather, this study is inclined towards the predominance of low-tech and less knowledge-intensive economic sectors in Vietnam as the primary reason. One could argue that, say, a local economy packed with many textile and footwear factories would benefit more from a larger labor force than from a higher local share of university-educated workers. In contrast, the region with higher intensity of financial and insurance sectors are expected to benefit more from human capital spillover. *Hypothesis 2*, in fact, already hypothesized this difference.

Table 3.14. TSLS results with high-tech versus low-tech firms, 2011, 2016

Column	(1)	(2)	(3)	(4)	(5)	(6)
Year of sampling	2011			2016		
Sectors	High-tech	Low-tech	Difference	High-tech	Low-tech	Difference
District university education B_i	0.610** (0.267)	-0.247 (0.247)	Yes***	0.543** (0.213)	-0.096 (0.244)	Yes***
District employment $\ln(L_i)$	0.044 (0.027)	0.047*** (0.017)	No	0.044* (0.025)	0.036** (0.015)	No
Over-id p-value	0.680	0.127	-	0.542	0.671	-
No. of obs	9,704	63,609	-	12,513	84,296	-
R^2	0.243	0.221	-	0.212	0.197	-

Notes: Cluster-robust standard errors in parentheses calculated at the province level. The explained variable is firm labor productivity in all columns. Columns (1), (2), (4), and (5) mirror the regression in column (2) of Table 3.10 with different samples. The estimates of firm-level variables are hidden for brevity.

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

To test this hypothesis, the main sample of each year is decomposed into two subsamples. The first one contains only observations from high-tech, medium-high-tech, and knowledge-intensive two-digit VSIC economic sectors, abbreviated as high-tech. The second one covers the remaining sectors, which are medium-low-tech, low-tech, and less knowledge-intensive, abbreviated as low-tech. These sectors are classified according to the official classification on high-tech industry and knowledge-intensive services of Eurostat – the statistical office of the European Union. Each subsample is then regressed with the TSLS technique and a full set of firm-level variables, like column (2) of Table 3.10. The results are reported in Table 3.14. As expected, the estimate of B_i for high-tech firms is positive and statistically significant in both 2011 and 2016, whereas it is negative and insignificant for low-tech firms. This is in line with the findings of Moretti (2004b) and Liu (2014), though they neither control for UE in their model nor adopt an IV approach. Considering the results from column (4) to interpret, it means that a 1% increase in district-level share of university-educated workers makes labor productivity of high-tech firms in the district rise by 0.543%, which, interestingly, is within the range between 0.5% and 0.7% reported in Moretti (2004b) for manufacturing plants in the US. To test the statistical difference between estimators in columns (1) versus (2), and (4) versus (5), the regression with the full sample covering all sectors is rerun, adding a dummy variable of high-tech sectors and its interaction terms with all main variables and controls. The significance levels of these interactions are shown in columns (3) and (6). As can be seen, the difference in the productivity influences of district-level university education on the high-tech versus low-tech firms is statistically significant. As a result, *Hypothesis 2* is proved.

Regarding the existence of UE, their positive and significant estimate is stable across technology levels, except for the coefficient for high-tech firms in the sample of 2011, which is significant at only the 11.4% significance level. Despite this, the difference in urbanization estimators between high-tech and low-tech firms is not statistically significant. To conclude in the regional context of Vietnam, the quantity of knowledge spreaders captured by UE is crucial universally, while the knowledge content of spreaders captured by HCE is mainly important to the high-tech sectors. Furthermore, the modest importance of these sectors in the Vietnamese economy may be the main driver of the ambiguous evidence of regional human capital's spillovers. Given that the predominance of low-tech sectors is a pattern across all developing countries, the spatial model of Gennaioli et al. (2013) might be more suitable to predict the external effects of human capital in developed countries.

3.7. Conclusions

This study aims to estimate the external effects of human capital along with urbanization for Vietnam since this evidence remains scarce for developing countries. For the empirical tests, the specification is constructed out of the Gennaioli et al. (2013)'s regional model in which entrepreneurs and workers choose their occupation and location based on their human capital to optimize their utility. These behaviors along with production factor choices of entrepreneurs to maximize benefits of their firms lead to a unique spatial equilibrium, which is the setting for statistical interpretations in this study. The endogeneity issues facing the estimation of HCE and UE are tackled with the exploitation of exogenous sources from the past and the method of TSLS. The key results show the ambiguous productivity impact of regional university education, while the external effects of urbanization are found to be significantly positive. These findings may be typical for developing countries, where the contribution of high-tech economic sectors to the economies remains small compared to developed countries. Therefore, this study makes a call to further evidence for other developing countries, provided that the inclusion of both UE and HCE in a specification and the application of the IV approach are necessary.

Chapter 4

Conclusions and policy suggestions

4.1. Summary

This dissertation is carried out to verify the existence of the externalities in Vietnam – an emerging and developing country which has experienced substantial changes in socio-economic conditions over the last decades. The key rationale behind the implementation of this research is to overcome the scarcity of studies in the spatial externalities in literature for developing countries. To achieve such an objective, chapter 2 focuses on the agglomeration externalities and their heterogeneity in firm characteristics. Considering the importance of the manufacturing sector in Vietnam and requirements for the high accuracy of intermediate input measures in the estimation of firm-level TFP, only manufacturing firms are employed in the regression sample of this chapter. Since TFP, as the only explained variable, must be calculated instead of purely collected from the data set, its calculation methods should be under scrutiny. To avoid introducing upward biases to estimated input elasticities caused by the simultaneity problem facing the production function estimation with OLS, the control function techniques are applied following OP, LP, and Woodridge (2009). However, the method of Woodridge (2009) is considered as the most reliable one owing to the addition of the IV-GMM strategy to solve some weaknesses found in OP and LP. With estimated TFP measures at hand, the key regression model is constructed on a production function in which its technology shifter contains the external terms. Profit-maximizing behavior of firms allows the production function to be transformed into the baseline specification with the logarithm of LP-W TFP on the left-hand side and agglomeration proxies and controls, which are calculated using the firm-level data set from GSO and provincial level statistical yearbooks, on the right-hand side. Econometric problems commonly found when estimating this sort of specification are caused by missing variables, self-selection, and shocks at regional or industrial levels. To tackle these problems, fixed effects terms of firm, year, year-region, and year-industry are applied using a six-year panel data set from Vietnam, along with the removal of movers from the final estimation sample. By exploiting within-firm, within-region, and within-industry variations over 6 years, this study is considered to capture the short-run rather than long-run effects of agglomeration.

The baseline results confirm the presence of urbanization economies and Jacobs externalities in the short run, while the impact of regional specialization is ambiguous. Therefore, in line with literature, this study confirms the existence of the agglomeration externalities, but only urban externalities are found matter in Vietnam. Hence, the estimates are in favor of the external effects of urbanization in the debate on the dominant source of the externalities. Despite this, the author considers the evidence to be region-specific since it may be strongly associated with features of developing and emerging countries, especially those economies with the dominance of low-tech industries and the underdevelopment of the intermediate input market in the manufacturing sector. In the next step, variables of firm-level characteristics are inserted in addition to their product terms with proxies of agglomeration into the baseline regression to verify the heterogeneous effects of the externalities. Yielded estimators and resulting marginal graphs show that urbanization economies benefit mainly foreign-owned firms and young ones, while the industrial diversity helps small firms. Furthermore, it is interesting that the specialization externalities are found active for young firms, while it does harm to old ones. All in all, these findings demonstrate that the external effects are not the same for every firm, but vary across firms of different ownerships, sizes, and ages.

In chapter 3, the external effect of human capital is the primary research of interest. Given that local human capital is strongly correlated to local urbanization as well as micro-level human capital of entrepreneurs and workers, separating the productivity influences of each factor from the others is essential to gain sound empirical evidence. To achieve that purpose, this study makes use of a model in which these factors along with production inputs jointly explain firm-level labor productivity under a spatial equilibrium condition when laborers become indifferent to the choice between an entrepreneurial job and a worker one, and highest-educated entrepreneurs and workers no longer have incentives to use their human capital to migrate to productive regions. Despite covering many firm-level and region-level determinants of firm-level productivity and employing of a firm-level two-year panel data set which covers almost all economic sectors, the specification made straight out of this model is not enough to assure an unbiased estimate of HCE, given the biasing forces of potential identification issues. As a solution, this study uses data collected from population censuses in the distant past to instrument for both regional variables of urbanization and human capital. Beside the fact that statistical tests indicate these instruments to be very likely valid, they are also considered to be exogenous due to the fast-changing environment of the Vietnamese economy and society over the last several decades.

With regional variables constructed on the establishment-level information, the baseline estimates do not confirm the presence of HCE in Vietnam, despite the coefficient of UE remains positive and significant. Furthermore, it shows that the inclusion of UE and other firm-level terms of human capital is essential since the estimator of HCE, even when instrumented, turns to significant in the specification without these controls. Especially, the exclusion of UE invalidates the use HCE since exogeneity tests become failed, implying the certain existence of endogeneity when HCE is estimated without the inclusion of UE, which echoes concerns of Carlino & Kerr (2015) about the common absence of UE in the regression of HCE in literature. Moreover, the insignificant evidence of HCE is confirmed in almost all robustness checks across various regression techniques, measures of UE, combinations of instruments, and functional forms. However, when the regression sample is split into sub-samples according to technological levels, the new estimates show the significant presence of HCE only in high-or-medium-high-tech and knowledge-intensive industries, whose economic sector plays a less important role in the Vietnamese economy. This finding suggests that human capital may spill over more strongly in a country where these highly productive sectors make up larger share in GDP, thus offers a rationale for the ambiguous impact of HCE in Vietnam – the economy developed based on medium-or-low-tech and less-knowledge-intensive industries.

4.2. Discussions

This subsection expresses some limitations of studies conducted in this dissertation and point out directions for future research. First, several aspects of the TFP estimation in chapter 2 should be discussed. Despite electricity consumption being superior to investment in making unobserved productivity explicitly through the monotonicity assumption, its validity might become less assured when firms suddenly switch to energy saving technology. In that case, a decrease in electricity consumption may not necessarily be an outcome of a decrease in productivity but the opposite – an improvement, which invalidates the application of the monotonicity condition. However, the clear monotonicity shown in Figure 2.3 for LP-W TFP implies that the impact of this possible firm-level technology shock is negligible. The reason is that firms tend to install only a few electricity saving machines and equipment to replace obsolete ones each year rather than replace them all at once, thus could not cause a massive reduction in electricity consumption holding other input factors fixed. Additionally, if the shocks happen at the industrial scale, they are less problematic since the production function estimation is carried separately for each industry. Nevertheless, when the firm-level data on

energy saving technology is available, this factor should be inserted as a determinant into the demand function of energy to guarantee the greater accuracy of the TFP estimation.

In addition, and more crucially, when information on materials is available, one can compute revenue-based TFP and rerun the baseline regression of agglomeration with this index as an important robustness check because there is evidence from Gandhi et al. (2017) that the revenue-based measure is often divergent from its value-added-based version. Specifically, compared to revenue-based productivity, the value-added-based one tends to show larger variance and higher summation of input elasticities, whose difference levels depend on the cost fraction of intermediate inputs in production of the industry under study. As a result, using value-added-based TFP often leads to larger heterogeneity across various industries. Nonetheless, this should not cause any big differences between the estimators using the two measures in chapter 2 since industrial heterogeneity is well controlled with fixed-effects terms. Gandhi et al. (2017) also note that choosing the revenue-based approach instead of the value-added one changes the distribution of TFP more substantially compared to the application of an advanced econometric method to deal with the potential endogeneity facing the estimation of productivity. Therefore, future studies with available data can estimate agglomeration with both TFP measures and then make a comparison to guarantee the consistency of statistical inferences in the externalities.

Next, it is worth pointing out that the explanatory power of the agglomeration model in chapter 2 is limited as the value of R-squared is merely 0.076, despite this low value being found in influential papers in literature as well. This is not surprising since only within variations are considered in regression with fixed-effects, while between variations are often much greater, thus the estimation with OLS tends to produce larger R-squared. Nevertheless, afterward studies with richer data sets at the firm level or even the plant level can exploit further time-variant micro-level factors that better explain within-firm variations of TFP to raise the prediction ability of the model. According to the review on determinants of firm productivity by Syverson (2011), such potential factors may be details about worker quality, adaptation of new technology, research and development investment, and level of innovation in products.

Another aspect that is worthwhile to discuss is the ambiguous evidence of HCE in the low-or-medium-low-tech and less-knowledge-intensive industries shown in Table 3.14 of chapter 3. According to Ciccone & Peri (2006) and Combes & Gobillon (2015)'s remarks about the estimates of HCE in literature, this result may purely reflect the imperfect substitutability between high- and low-educated workers. Particularly, when the regional fraction of

university-educated workers increases, non-university-educated workers become scarcer in the region. Due to the imperfect substitution between the two groups of workers, the lower supply of the low-educated workers helps raise their marginal productivity, and thus their wages, which pushes up wage costs of firms as a consequence, especially the ones depending on this labor force the most – low-tech and less-knowledge intensive firms. As a result, the improvement in regional human capital can generate not only the external benefits but also the extra cost to low-tech firms, which might explain why its HCE estimator, which reflects the net effect, is statistically insignificant and even negative. By similar logic, the strong evidence of HCE for high-tech firms may stem from the positive impact of the substitution effects on industries with higher demand for highly educated workers. Despite this, one can argue that the share of university-educated workers in the labor force was still low in Vietnam in 2011 and 2016 compared to other emerging countries as well as developed ones. Therefore, it is reasonable to say that a high share of regional human capital could not make low-tech firms incur higher labor costs in Vietnam at the time. In addition, there is new evidence that the substitutability between workers of different education groups is in fact high rather than low as being commonly referred to in literature (Bils et al., 2022). Therefore, the possibility of this substitution effect should not cast doubts on the unbiasedness of the results on the heterogeneity of HCE across firms of different technology levels in chapter 3, especially for the case of Vietnam. However, future studies can put more efforts on accounting for this effect explicitly to see whether it leads to a significant change in the estimates of HCE.

4.3. Policy suggestions

The findings expressed in this dissertation may give local firms and policy makers useful guidance on making policies. Firstly, the strong and robust evidence of urbanization economies found in both chapters 2 and 3 supports economic development policies of the Vietnamese government for the period 2021 – 2030. Specifically, in the 13th National Party Congress of the Communist Party Vietnam taking place in 2021, the Vietnamese government formulated regional development strategies based on urban centers and set a goal to raise the national urbanization rate to 50% by the end of this decade. However, positive and significant estimates of the Jacobs externalities found in this study imply that regional policies to raise the level of industrial variety should be also considered. Meanwhile, the insignificant role of regional specialization in the externalities sends a message to the government and local authorities that localized industries are operating in a less efficient way than what it is supposed to be. A

possible long-run solution is to promote more strongly the development of the domestic intermediate input market in Vietnam.

Secondly, the evidence of the heterogeneous effects of agglomeration across different firm-level characteristics sounds a warning to domestic, large, and old firms about their limitations in reaping benefits from an agglomerated environment. A possible measure, which is not beyond their control, is to construct more efficient organizing principles that is at least equivalent to their development level in order to receive more information spillover from the region. Finally, the estimated results of HCE in chapter 3 support economic objectives set by the Vietnamese government that high-tech industries and knowledge-intensive service sectors have a heavier weight in the economy. For example, the digitalized sectors are planned to make up a GDP share of 20% by 2025, and 30% by 2030, while the percentage was only 10% in 2021. The presence of human capital spillover to high-tech firms found in this dissertation suggests that these economic policies should be implemented simultaneously with the plans to expand the scale of university-educated labor force.

References

Abdel-Rahman, H., & Fujita, M. (1990). Product variety, Marshallian externalities, and city sizes. *Journal of regional science*, 30(2), 165-183.

Abel, J. R., & Deitz, R. (2011). Do colleges and universities increase their region's human capital?. *Journal of Economic Geography*, 12(3), 667-691.

Acemoglu, D. (1996). "A micro foundation for social increasing returns in human capital accumulation". *Quarterly Journal of Economics*, 779–804.

Acemoglu, D. (1998). Why do new technologies complement skills? Directed technical change and wage inequality. *The quarterly journal of economics, 113*(4), 1055-1089.

Acemoglu, D., & Angrist, J. (2000). How large are human-capital externalities? Evidence from compulsory schooling laws. *NBER macroeconomics annual*, 15, 9-59.

Ackerberg, D.A., Caves, K., & Frazer, G., 2006. Structural identification of production functions, mimeo, UCLA Department of Economics.

Acs, Z., & Armington, C. (2004). Employment growth and entrepreneurial activity in cities. *Regional Studies*, 38(8), 911-927.

Ananat, E., Fu, S., & Ross, S. L. (2013). Race-specific agglomeration economies: *Social distance and the black-white wage gap* (No. w18933). National Bureau of Economic Research.

Andersson, M., & Lööf, H. (2011). Agglomeration and productivity: evidence from firm-level data. *The annals of regional science,* 46(3), 601-620.

Anh, B. K., Thai, N. Q., & Trinh, B. (2019). Foreign Direct Investment (FDI) in Vietnam Economy. *Theoretical Economics Letters*, 9(04), 986.

Arrow, K. J. (1971). The economic implications of learning by doing. In *Readings in the Theory of Growth* (pp. 131-149). Palgrave Macmillan, London.

Au, C. C., & Henderson, J. V. (2006). Are Chinese cities too small?. *The Review of Economic Studies*

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American economic review*, 86(3), 630-640.

Bacolod, M., Blum, B. S., & Strange, W. C. (2010). Elements of skill: traits, intelligences, education, and agglomeration. *Journal of Regional Science*, 50(1), 245-280.

Badr, K., Rizk, R., & Zaki, C. (2019). Firm productivity and agglomeration economies: evidence from Egyptian data. *Applied Economics*, 51(51), 5528-5544. ISO 690

Bils, M., Kaymak, B., & Wu, K. J. (2022). *Labor Substitutability among Schooling Groups* (No. w29895). National Bureau of Economic Research.

Black, D., & Henderson, V. (1999). A theory of urban growth. *Journal of political economy*, 107(2), 252-284.

Blien, U., & Suedekum, J. (2004). Local economic structure and industry development in Germany, 1993-2001. *Available at SSRN 602403*.

Brakman, S., Garretsen, H., & Van Marrewijk, C. (2009). *The new introduction to geographical economics*. Cambridge University Press.

Bratti, M., & Leombruni, R. (2014). Local human capital externalities and wages at the firm level: Evidence from Italian manufacturing. *Economics of Education Review*, 41, 161-175.

Brülhart, M., & Mathys, N. A. (2008). Sectoral agglomeration economies in a panel of European regions. *Regional Science and Urban Economics*, 38(4), 348-362.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317-372.

Capello, R. (2015). *Regional economics*. Routledge.

Carlino, G., & Kerr, W. R. (2015). Agglomeration and innovation. *Handbook of regional and urban economics*, 5, 349-404.

Ciani, A., Hyland, M. C., Karalashvili, N., Keller, J. L., Ragoussis, A., & Tran, T. T. (2020). *Making It Big: Why Developing Countries Need More Large Firms*. The World Bank.

Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review*, 46(2), 213-227.

Ciccone, A., & Hall, R. E. (1996). Productivity and the Density of Economic Activity. *The American Economic Review*, 86(1), 54-70.

Ciccone, A., & Peri, G. (2006). Identifying human-capital externalities: Theory with applications. *The Review of Economic Studies*, 73(2), 381-412.

Chipman, J. S. (1970). External economies of scale and competitive equilibrium. *The Quarterly Journal of Economics*, 84(3), 347-385.

Combes, P. P., Démurger, S., & Li, S. (2013). Urbanisation and migration externalities in China.

Combes, P. P., Duranton, G., & Gobillon, L. (2008). Spatial wage disparities: Sorting matters!. *Journal of urban economics*, 63(2), 723-742.

Combes, P. P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6), 2543-2594.

Combes, P. P., Duranton, G., Gobillon, L., & Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration economics* (pp. 15-66). University of Chicago Press.

Combes, P. P., & Gobillon, L. (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics* (Vol. 5, pp. 247-348). Elsevier.

Desmet, K., & Fafchamps, M. (2005). Changes in the spatial concentration of employment across US counties: a sectoral analysis 1972–2000. *Journal of economic geography*, 5(3), 261-284.

Douma, S., George, R., & Kabir, R. (2006). Foreign and domestic ownership, business groups, and firm performance: Evidence from a large emerging market. *Strategic Management Journal*, 27(7), 637-657.

Dumais, G., Ellison, G., & Glaeser, E. L. (2002). Geographic concentration as a dynamic process. *Review of economics and Statistics*, 84(2), 193-204.

Duranton, G. (2008). From cities to productivity and growth in developing countries. *Canadian Journal of Economics/Revue canadienne d'économique*, 41(3), 689-736.

Duranton, G. (2016a). Determinants of city growth in Colombia. *Papers in Regional Science*, 95(1), 101-131.

Duranton, G. (2016b). Agglomeration effects in Colombia. *Journal of Regional Science*, 56(2), 210-238.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2063-2117). Elsevier.

Duranton, G., & Puga, D. (2014). The growth of cities. In *Handbook of economic growth* (Vol. 2, pp. 781-853). Elsevier.

Doms, M., Lewis, E., Robb, A., 2009. Local labor market endowments, new business characteristics, and performance. *Journal of Urban Economics* 67 (1), 61–77.

Eeckhout, J., Pinheiro, R., & Schmidheiny, K. (2014). Spatial sorting. *Journal of Political Economy*, 122(3), 554-620.

Ehrl, P. (2013). Agglomeration economies with consistent productivity estimates. *Regional Science and Urban Economics*, 43(5), 751-763.

Foster, N., & Stehrer, R. (2009). Sectoral productivity, density and agglomeration in the Wider Europe. *Spatial Economic Analysis*, 4(4), 427-446.

Fu, S. (2007). Smart café cities: Testing human capital externalities in the Boston metropolitan area. *Journal of Urban Economics*, 61(1), 86-111.

Fujita, M., Krugman, P. R., & Venables, A. (1999). *The spatial economy: Cities, regions, and international trade*. MIT press.

Gandhi, A., Navarro, S., & Rivers, D. (2017). How heterogeneous is productivity? A comparison of gross output and value added. *Journal of Political Economy, forthcoming*.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2013). Human capital and regional development. *The Quarterly journal of economics*, 128(1), 105-164.

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

Glaeser, E. L. (Ed.). (2010). *Agglomeration economics*. University of Chicago Press.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of political economy*, 100(6), 1126-1152.

Glaeser, E. L., Rosenthal, S. S., & Strange, W. C. (2010). Urban economics and entrepreneurship. *Journal of urban economics*, 67(1), 1-14.

Gokan, T., Kuroiwa, I., & Nakajima, K. (2019). Agglomeration economies in Vietnam: A firm-level analysis. *Journal of Asian Economics*, 62, 52-64.

Gould, E. D. (2007). Cities, workers, and wages: A structural analysis of the urban wage premium. *The Review of Economic Studies*, 74(2), 477-506.

Graham, D. J. (2007). Variable returns to agglomeration and the effect of road traffic congestion. *Journal of Urban Economics*, 62(1), 103-120.

Ha, D. T. T., & Kiyota, K. (2014). Firm-Level Evidence on Productivity Differentials and Turnover in Vietnamese Manufacturing. *The Japanese Economic Review*, 65(2), 193-217.

Head, K., & Mayer, T. (2004). The empirics of agglomeration and trade. In *Handbook of regional and urban economics* (Vol. 4, pp. 2609-2669). Elsevier.

Henderson, J.V. (1986), "Efficiency of Resource Usage and City Size," *Journal of Urban Economics*, 18, 47-70.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of urban economics*, 53(1), 1-28.

Henderson, V., Lee, T., & Lee, Y. J. (2001). Scale externalities in Korea. *Journal of Urban Economics*, 49(3), 479-504.

Hendricks, L. (2011). The skill composition of US cities. *International Economic Review*, 52(1), 1-32.

Hicks, J. (1963). *The theory of wages*. Springer.

Hamaguchi, N., Ishizuka, F., & Sakata, S. (2012). Liberalization, integration, and industrial location in Vietnam. In *Economic Integration and the Location of Industrie*s (pp. 43-87). Palgrave Macmillan, London.

Howard, E., Newman, C., Rand, J., & Tarp, F. (2014). *Productivity-enhancing manufacturing clusters: Evidence from Vietnam* (No. 2014/071). WIDER Working Paper.

Huang, R. D., & Shiu, C. Y. (2009). Local effects of foreign ownership in an emerging financial market: Evidence from qualified foreign institutional investors in Taiwan. *Financial Management*, 38(3), 567-602.

Jacobs J. (1969). *The Economy of Cities*. Vintage, New York, NY.

Kerr, S. P., Kerr, W. R., & Xu, T. (2018). Personality traits of entrepreneurs: A review of recent literature. *Foundations and Trends® in Entrepreneurship*, 14(3), 279-356.

Knoben, J., Arikan, A., van Oort, F., & Raspe, O. (2016). Agglomeration and firm performance: One firm's medicine is another firm's poison. *Environment and Planning A*, 48(1), 132-153.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, 99(3), 483-499.

La Porta, Rafael and Andrei Shleifer (2008). "The Unofficial Economy and Economic Development." *Brookings Papers on Economic Activity*, 2008: 275-352.

Lamin, A., & Livanis, G. (2013). Agglomeration, catch-up and the liability of foreignness in emerging economies. *Journal of International Business Studies*, 44(6), 579-606.

Lee, B. S., Sosin, K., & Hong, S. H. (2005). Sectoral manufacturing productivity growth in Korean regions. *Urban Studies,* 42(7), 1201-1219.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2), 317-341.

Liu, Z. (2007). The external returns to education: Evidence from Chinese cities. *Journal of Urban Economics*, 61(3), 542-564.

Liu, Z. (2014). Human capital externalities in cities: evidence from Chinese manufacturing firms. *Journal of economic geography*, 14(3), 621-649.

Lucas Jr, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, 508-523.

Lucas Jr, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, 22(1), 3-42.

Malesky, E., & London, J. (2014). The political economy of development in China and Vietnam. *Annual Review of Political Science*, 17, 395-419.

Marshall, A. (1890). Principles of economics Macmillan. *London (8th ed. Published in 1920)*.

Martin, P., Mayer, T., & Mayneris, F. (2011). Spatial concentration and plant-level productivity in France. *Journal of Urban Economics*, 69(2), 182-195.

Matano, A., & Naticchioni, P. (2012). Wage distribution and the spatial sorting of workers. *Journal of Economic Geography*, 12(2), 379-408.

McCann, B. T., & Folta, T. B. (2011). Performance differentials within geographical clusters. *Journal of Business Venturing*, 26(1), 104-123.

McDonald, J. F., & McMillen, D. P. (2010). *Urban economics and real estate: theory and policy*. John Wiley & Sons.

Melo, P. C., Graham, D. J., & Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional science and urban Economics*, 39(3), 332-342.

Mincer, J. (1974) *Schooling, Experience, and Earnings*. New York: Columbia University Press, for the National Bureau of Economic Research

Millan, J. M., Congregado, E., Roman, C., Van Praag, M., & Van Stel, A. (2014). The value of an educated population for an individual's entrepreneurship success. *Journal of business venturing*, 29(5), 612-632.

Mills, E. S. (1967), An aggregate model of resource allocation in a metropolitan area, *American Economic Review*, 57: 197–210.

Moretti, E. (2004a). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of econometrics*, 121(1-2), 175-212.

Moretti, E. (2004b). Workers' education, spillovers, and productivity: evidence from plant-level production functions. *American Economic Review*, 94(3), 656-690.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The review of Economics and Statistics*, 334-338.

Muravyev, A. (2008). Human capital externalities Evidence from the transition economy of Russia 1. *Economics of Transition*, 16(3), 415-443.

Muth, R. (1969), *Cities and Housing*, Chicago: University of Chicago Press.

Nabielek K. et al. (2016), Cities in Europe. PBL Netherlands Environmental Assessment Agency, The Hague.

Nakarnura, R. (1985), "Agglomeration Economies in Manufacturing," *Journal of Urban Economics*, 17, 108-124.

Newman, C., Rand, J., Talbot, T., & Tarp, F. (2015). Technology transfers, foreign investment and productivity spillovers. *European Economic Review*, 76, 168-187.

Nichter, S., & Goldmark, L. (2009). Small firm growth in developing countries. *World development*, 37(9), 1453-1464.

Nguyen, T. X. T., & Diez, J. R. (2017). Multinational enterprises and industrial spatial concentration patterns in the Red River Delta and Southeast Vietnam. *The Annals of Regional Science*, 59(1), 101-138.

Nguyen, H. Q. (2017). Business reforms and total factor productivity in Vietnamese manufacturing. *Journal of Asian Economics*, 51, 33-42.

OECD (2018), *OECD Urban Policy Reviews: Viet Nam,* OECD Publishing, Paris.

Olley, G., Pakes, A. (1996). The dynamics of productivity in the telecommunication equipment industry. *Econometrica 64*, 1263–1297.

Ongore, V. O. (2011). The relationship between ownership structure and firm performance: An empirical analysis of listed companies in Kenya. *African Journal of Business Management*, 5(6), 2120-2128.

Ott, I., & Soretz, S. (2010). Productive public input, integration and agglomeration. *Regional Science and Urban Economics*, 40(6), 538-549.

Papageorgiou, T. (2013). Worker sorting and agglomeration economies. *Penn State University, Penn State University Park*.

Parker, S. C. (2018). *The economics of entrepreneurship*. Cambridge University Press. Kindle Edition.

Phimister, E. (2005). Urban effects on participation and wages: Are there gender differences?. *Journal of Urban Economics*, 58(3), 513-536.

Porter, M. E. (1990). The competitive advantage of nations. *Competitive Intelligence Review*, 1(1), 14-14.

Porter, M. E. (2000). Location, competition, and economic development: Local clusters in a global economy. *Economic development quarterly, 14*(1), 15-34.

Poschke, M. (2018). The firm size distribution across countries and skill-biased change in entrepreneurial technology. *American Economic Journal: Macroeconomics*, 10(3), 1-41.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of regional science, 50*(1), 203-219.

Quigley, J. M. (2009). Urbanization, agglomeration, and economic development. *Urbanization and growth*, 115.

Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: evidence from the cities. *Journal of urban economics*, 34(3), 380-400.

Razvadauskas, F. V. (2019). Megacities: Developing country domination. *Euromonitor International. Available online: https://go. euromonitor. com/strategy-briefing-cities-2018-megacities. html (accessed on 1 March 2021).*

Rigby, D. L., & Essletzbichler, J. (2002). Agglomeration economies and productivity differences in US cities. *Journal of Economic Geography*, 2(4), 407-432.

Rigby, D. L., & Brown, W. M. (2015). Who benefits from agglomeration?. *Regional Studies*, 49(1), 28-43.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of political Economy*, 90(6), 1257-1278.

Roca, J. D. L., & Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, 84(1), 106-142.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, 98(5, Part 2), S71-S102.

Rosenthal, S. S., & Strange, W. C. (2001). The determinants of agglomeration. *Journal of urban economics*, 50(2), 191-229.

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2119-2171). Elsevier

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2), 373-389.

Rosenthal, S. S., & Strange, W. C. (2010). Small establishments/big effects: Agglomeration, industrial organization and entrepreneurship. In *Agglomeration economics* (pp. 277-302). University of Chicago Press.

Rovigatti, G., & Mollisi, V. (2018). Theory and practice of total-factor productivity estimation: The control function approach using Stata. *The Stata Journal*, 18(3), 618-662.

Schiffer, M., & Weder di Mauro, B. (2001). *Firm size and the business environment* (No. BOOK). The World Bank; International Finance Corporation.

Stock, J. H., & Yogo, M. (2002). Testing for weak instruments in linear IV regression.

Sveikauskas, L. (1975), "Productivity of Cities," *Quarterly Journal of Economics*, 89, 393-413.

Syverson, C. (2011). What determines productivity?. *Journal of Economic literature*, 49(2), 326-65.

Takii, S. (2005). Productivity spillovers and characteristics of foreign multinational plants in Indonesian manufacturing 1990–1995. *Journal of Development Economics*, 76(2), 521-542.

Tambunan, T. T. H. (2011). Development of small and medium enterprises in a developing country: The Indonesian case. *Journal of Enterprising Communities: People and Places in the Global Economy*.

Tan, D., & Meyer, K. E. (2011). Country-of-origin and industry FDI agglomeration of foreign investors in an emerging economy. *Journal of International Business Studies*, 42(4), 504-520.

Tuan, N. P., & Yoshi, T. (2010). Organisational capabilities, competitive advantage and performance in supporting industries in Vietnam. *Asian Academy of Management Journal*, 15(1).

Unger, J. M., Rauch, A., Frese, M., & Rosenbusch, N. (2011). Human capital and entrepreneurial success: A meta-analytical review. *Journal of business venturing*, 26(3), 341-358.

Van der Sluis, J., Van Praag, M., & Vijverberg, W. (2005). Entrepreneurship selection and performance: A meta-analysis of the impact of education in developing economies. *The World Bank Economic Review*, 19(2), 225-261.

Von Thünen, J. H. (1826), Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalo¨konomie, Hamburg: Perthes.

Waldkirch, A. (2014). Foreign firms and productivity in developing countries. *Department of Economics, Colby College. Available at [Accessed 2 July 2018].*

Weber, A. (1909), Über den Standort der Industrien, Tübingen: JCB Mohr.

World Bank. 2011. *Vietnam urbanization review: technical assistance report (English)*. Washington, DC: World Bank.

World Bank, 2019, accessed on 19 August 2019, https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=VN

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics letters*, 104(3), 112-114.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.