Journal of Asian Economics 86 (2023) 101620

Contents lists available at ScienceDirect

Journal of Asian Economics

journal homepage: www.elsevier.com/locate/asieco

Which firms benefit the most from agglomeration? New evidence from an emerging country with consistent measure of productivity

Manh Cuong Hoang [a,*], Daniel Schiller [b]

[a] University of Greifswald, Friedrich-Loeffler-Str. 70, D-17489 Greifswald, Germany
[b] Institute of Geography and Geology, University of Greifswald, Friedrich-Ludwig-Jahn-Str. 17a, D-17487 Greifswald, Germany

ARTICLE INFO

JEL Classification:
D24
O18
R10
R11
R12
R

Keywords:
Agglomeration
Urbanization
Specialization
Concentration
Productivity
Developing countries
Vietnam

ABSTRACT

Productivity influences of agglomeration for developed countries has been well documented so far, however, the studies are still rare for emerging and developing countries, especially ones focusing on firm heterogeneity. This article empirically investigates the effects of agglomeration on productivity using firm-level data from Vietnam – a typical emerging country. Firstly, the consistent productivity measure of each individual firm is yielded using the control function approach along with the instrumental variable procedure. Next, it is regressed on proxies of agglomeration, controlling firm and regional characteristics. Potential issues of self-selection and endogeneity are dealt with using the fixed effects technique and taking advantage of micro data. Findings show the productivity-enhancing influences of employment density and industrial diversity but no clear evidence on the productivity gains from specialization for a general firm. In addition, the most advantaged firms in highly agglomerated regions are proved to be foreign-owned, small-sized, or young. Finally, several sensitivity checks demonstrate that the estimated results are robust across various productivity measures, industrial levels, and samples.

1. Introduction

The underlying theory describing the advantages of geographically concentrated regions, known as agglomeration externalities, has a long history of analysis in literature. It evolves in three main directions: intra-industry, inter-industry, and external scale effects. The first approach is referred to Marshallian externalities, after Marshall (1920)'s pioneering analysis. This was later named Marshall-Arrow-Romer (MAR) externalities by Glaeser et al. (1992) who acknowledged the contributions of Arrow (1962) and Romer (1990) to the analysis of specific-sector spillovers of knowledge. The second approach is based on the reasoning of Jacobs (1969), hence the so-called Jacobian externalities. Finally, the third views city employment size or density as a main source of external economies, regardless of the industrial structure (Rosenthal & Strange, 2001; Rosenthal & Strange, 2004). Duranton and Puga (2004) underline these external effects with theoretical microeconomic foundations. Up to now, there have also been a number of empirical articles in literature that attempt to shed light on these spatial externalities. However, the issue of what spatial external effects play the most critical role in empirically-based productivity growth remains an inconclusive dispute. Some authors find evidence in favor of

Abbreviations: DOEs, Domestic-owned enterprises; FDI, Foreign direct investment; FOEs, Foreign-owned enterprises; GSO, General Statistical Office; POEs, Private-owned enterprises; SOEs, State-owned enterprises; TFP, Total factor productivity; VSIC, Vietnam Standard Industrial Code.

* Corresponding author.
E-mail address: cuong.hoang.m01@gmail.com (M.C. Hoang).

https://doi.org/10.1016/j.asieco.2023.101620
Received 18 December 2021; Received in revised form 17 October 2022; Accepted 14 April 2023
Available online 15 April 2023
1049-0078/

Marshallian economies (Henderson, 2003; Martin et al., 2011), while others support urbanization economies and/ or diversity benefits (Glaeser et al., 1992; Ciccone & Hall, 1996; Combes et al., 2010). The conclusion is even more ambiguous in 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). Furthermore, Melo et al. (2009) note that empirical evidence of the agglomeration effects in a region might be inconsistent with the findings in other regions due to regional heterogeneity. Meanwhile, typical agglomeration patterns in highly developed countries differ significantly from those in low- or middle-income countries. For instance, the urbanization growth rate of developing countries over recent decades has been higher, but its level has remained lower; the disparity between big and small cities is much larger for fast-growing developing countries, leading to the economic and spatial dominance of a few mega-cities within each of those territories (Henderson et al., 2001; Duranton, 2008; Duranton, 2016). All in all, one should not assume the same economic outcome from agglomeration in the countries of both groups, 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 external gains' heterogeneity, a number of scholars concentrate on this matter with suitable data at hand such as McCann and Folta (2011) and Rigby and Brown (2015) for the US and Canada. They all confirm that the agglomeration influences vary significantly with an individual firm's characteristics. Nevertheless, these conclusions again are not necessarily the same for the world of emerging countries where the difference in attributes of various economic subjects tend to deviate from that of advanced countries. In terms of ownership, for example, foreign-owned enterprises (FOEs) operating in developing countries typically have more advanced technology than native firms do (Tan & Meyer, 2011; Newman et al., 2015). This technology gap perhaps is much narrower in developed territories where many their domestic-owned enterprises (DOEs) are at the top of the world rankings with state-of-the-art innovations. As for employment scale, small firms in developing countries account for the largest share but are mainly involved in small-scale production activities (Poschke, 2018). Turning to age, there are less long-lasting firms in emerging countries since the economic boom happened to most of them only over recent decades, whereas the boom for developed countries happened much longer ago. Given the heterogeneity among firms of various types and countries at various development stages, the non-homogenous impact of agglomeration is the primary focus of this study.

Studies most closely related to this research are those of Martin et al. (2011) who examine agglomeration impacts for France and Howard et al. (2014) and Gokan et al. (2019) who conduct the research for Vietnam. This article contributes to the strand of literature by focusing on several aspects that are either neglected or incomplete in the three studies. First, it deals with the question of how agglomeration induces unequal influences across various firm characteristics, whose evidence remains scarce for the developed and developing world. Second, by utilizing a more consistent measure of productivity as the primary dependent variable, a higher accuracy of the estimation of productivity impacts is assured. Third, contrasts and similarities in agglomeration influences between developed and developing countries are analyzed to reach main hypotheses. Fourth, in the context of Vietnam, Gokan et al. (2019) exploit between-firm variation based on cross-sectional data and consequently identify the existence of spatial externalities accumulated over a long period of time. As a result, their estimates should be interpreted as the long-run effects. This study is different since it takes advantage of within-firm variation for a short period of time, thus captures the short-run effects of agglomeration.

Vietnam is regarded as an ideal study case for the agglomeration analysis considering a Combes and Gobillon (2015)'s note about the insufficiency of empirical evidence for developing countries. The status of an emerging country stems from the fact that it has been currently classified as a low-middle income country and has had a high average growth rate of about 6.4% over the last two decades (World Bank, 2019). Along with many other developing countries, Vietnam has undergone a rapid urbanization phase in recent decades, particularly after the 1986 Doi Moi reform. Since then, according to the World Bank (2011), OECD (2018) and the General Statistical Office (GSO) in Vietnam, the urban population on average grew by 3.4 % annually during the period 1986–2010 and 3 % during the period 2011–2017, which is higher than the average of Southeast Asian countries (2.5 %) as well as advanced OECD member countries (0.88 %). By 2017, however, the urbanization rate reached merely 37.5 %, which was still much lower than the level of about 81.5 % in high-income countries, and lower than that of Southeast Asia. However, urban areas played a crucial role in Vietnam's economy by contributing 70 % of the country's Gross Domestic Product (GDP) in 2017. These differences reflect the contrast between emerging countries in general, Vietnam in particular and high-income countries, thus potentially revealing a disparity in the estimated results of agglomeration externalities. Another reason to choose Vietnam is the availability of a wealth of micro information on employment, industry, and location offered by GSO, which covers all firms registered in the country. This helps to construct more accurate regional variables for different geographical units and levels of industry, and thus overcome the shortcomings of aggregate data mentioned in literature. Finally, one of the Vietnamese government's key policies on urban and economic growth for the period 2015–2025 is the encouragement to urbanize more regions all over the country, whilst simultaneously discouraging sparse production activities. Therefore, understanding the potential economic gains for firms due to the agglomeration process and the heterogeneity of its influences is meaningful for Vietnam's urban policy makers when planning economic development strategies.

In this paper, firms' productivity is regressed on agglomeration proxies and control variables at firm and regional levels. Firm-level data was collected from the annual censuses on enterprises between 2011 and 2016, provided by GSO in Vietnam. The primary measure of productivity is total factor productivity (TFP) calculated consistently for each firm that is active in the manufacturing sector using Woodridge (2009)'s instrument approach on the basis of Levinsohn and Petrin (2003)'s framework. To prepare for the estimation, main regional variables, particularly proxies for agglomeration forces, are computed based on location codes, industrial codes, and employment information available for all firms with business tax codes in Vietnam between 2011 and 2016. When calculating regional variables, the preferred geographical unit is a district, and the industrial level is described by a 3-digit Vietnam Standard Industrial Code (VSIC) level. The main regression makes use of panel data and fixed effects to control for time-invariant unobserved factors, industrial and regional shocks, and potential self-selection bias, using observations from firms with at least 20 employees and without changing their main industry and location during sampling periods. The results provide a strong indication of the predominant role of urbanization economies and industrial diversity. In addition, interaction terms show that FOEs, small-sized, and young firms benefit the most from agglomeration.

The remainder of the paper continues as follows. Section 2 provides literature background, Section 3 explains the methodology of TFP estimation, and Section 4 presents data used in the article and summary statistics. Next, Section 5 covers the construction of main regional variables, discusses specification and related issues, Section 6 expresses estimated outputs, and Section 7 reports findings from robustness checks. Finally, Section 8 presents conclusions and discussion.

2. Literature review

In regard to the reasons behind the economic gains found in urban environments, Jacobs (1969) argues that larger cities benefit from having a greater variety of industries allowing for the circulation of cross-industry ideas, information, which stimulates innovation. Rosenthal and Strange (2004) in their extensive survey suggest that cities might gain benefits from spatial concentration of economic activities, regardless of their industrial structure. They call these effects urbanization economies. Alternatively, supporters for MAR externalities argue that the geographical proximity of activities within an industry-region benefits both the industry and the region (Glaeser et al., 1992). This advantage is also referred to as localization or specialization externalities (Rosenthal & Strange, 2004). As analyzed by Duranton and Puga (2004) and Puga (2010), agglomeration gives rise to economies through three channels called sharing effects (the geographical concentration of firms helps their local suppliers to earn the economies of scale, thus firms are advantageous from sharing their cost-optimal suppliers and high-quality infrastructures), learning effects (there are more opportunities to learn know-how from others for employees and to learn new technologies and management techniques for employers in a dense urban area), and matching effects (pairs of workers and employers, suppliers and customers, or financiers and entrepreneurs match better in a populated region). It is worth pointing out that these channels can be established through both within- and cross-industry interactions. In terms of the temporal dimension, Rosenthal and Strange (2004) and Combes and Gobillon (2015) note that a match between an employer and his or her new employees will generate an immediate impact on productivity, while knowledge spillovers might bring the external gains with a time lag. As a result, the externalities, which influence productivity through either or all of three channels mentioned above, can be static, dynamic, or both. However, given that more attention has been paid to the static effects of agglomeration in the existing literature, as can be seen from extensive reviews of Rosenthal and Strange (2004) and Combes and Gobillon (2015), it is also the focus of this study.

As far as empirical evidence found in developed countries is concerned, Rosenthal and Strange (2004), Melo et al. (2009), and Combes and Gobillon (2015) provide thorough reviews of related studies. They show that each paper in literature tends to reach conclusions favoring only one or two of the three above external effects. The influential paper of Glaeser et al. (1992) makes use of regional data for the US and finds that a higher intensity of inter- rather than intra- industry interactions stimulates employment growth. Ciccone and Hall (1996) and Ciccone (2002) exploit cross-sectional regional data using instrument variables as historical values and show the statistically positive impact of employment density on labor productivity in the US and Europe respectively. Henderson (2003) starts a trend in taking advantage of within-variation of observations to address the problem of unobserved heterogeneity in determining local effects. He quantifies the influences of localization and industrial diversity on productivity in high-tech and machinery sectors in the US for the period 1972–1992. He yields statistically significant evidence of within-industry externalities in solely the high-tech sector. Combes et al. (2010) explore French longitudinal data and instrument spatial density with geographical and historical data. Whether wages or productivity are defined as explained factors, they find solid evidence of scale economies. Martin et al. (2011) continue the use of French panel data but employ the first-difference generalized method of moments (FD-GMM) to deal with potential endogenous issues. They enter both urbanization and localization measures into the specification to explain the productivity of production plants but support only positive externalities of the latter. All in all, although literature gives supportive evidence of the existence of agglomeration economies from both inter- and intra- industry relations, there is no clear conclusion whether one is predominant over the other. Combes and Gobillon (2015) make a remark that if urbanization and localization economies are both found in a study, the latter tends to show a smaller role, and the impact of industrial diversity is found less robust over various studies.

Among the earliest scholars that tested both Jacobs's and Marshall's theories for a developing country, Lee et al. (2005) show the evidence for South Korea during its time as an emerging country. They exploit cross-sectional variations with productivity growth being the dependent variable. In the model with both specialization and diversity included, the latter is the only significant and positive effect. Au and Henderson (2006) in their influential paper use data of 285 Chinese cities in 1997 to estimate net agglomeration effects, utilizing historical information as instruments for regional explanatory variables. They find that economies at the scale of a city largely outweigh the diseconomies, and the external effects of industrial variety and urbanization predominate over Marshallian externalities. Combes et al. (2013) adopt the instrument variables strategy of Ciccone and Hall (1996) and Au and Henderson (2006) to a cross-sectional data set collected at household level from China. Their results demonstrate the predominance of spatial density over industrial clusters in improving workers' wages. In their extensive reviews, Duranton (2008) and Combes and Gobillon (2015) notice that the evidence of spatial external economies in developing countries is mixed and the elasticities of firm performance or worker wages with respect to local scale are found to be higher than those in developed territories. A possible reason proposed by Quigley (2009) is that duplicating successful ideas in emerging countries is highly profitable and that opportunity emerges more likely in cities or specialized regions.

Related studies for Vietnam have so far focused on the spatial patterns of concentration rather than its role in productivity improvement (i.e., Howard et al., 2012; Nguyen & Diez, 2017). The exceptions are Howard et al. (2014) who use data also gathered by

GSO from 2002 to 2007 to study the productivity influences of manufacturing clusters, and Gokan et al. (2019) who employ Vietnamese micro data in 2012 to evaluate the impacts of agglomeration. Both these papers generate regional variables by counting the number of firms in the local region or industry. Rosenthal and Strange (2001) and Martin et al. (2011), nevertheless, point out that those indices capture mainly the learning effects in which firms are the source of knowledge. Moreover, they prove that interactions between employees play a dominant role in determining agglomeration gains. By similar logic, employment-based agglomeration indices used in this paper should be better suited compared to those of Howard et al. (2014) and Gokan et al. (2019) in the context of Vietnam because the economy's main growth engine is its cheap and abundant labor force rather than innovative competence. This features an additional difference between this study and the previous papers for Vietnam.

Considering the postulate on whether inter- or intra-industry relations is more crucial, Henderson et al. (2001) argue that technological progress is geared more towards adoption than towards innovation in developing territories. By facilitating this engine through cross-industry spillover, the urban environment plays a stronger role in comparison with specialized environments that can be seen even in sparse regions. Furthermore, the large magnitude of urbanization economies found in literature for the developing world is another possible indicator of its superior position. Meanwhile, one can argue that regional specialization is more beneficial for firms in developed countries since it often comes with more differentiated labor markets, specialized inputs, and spillovers of industry-specific knowledge, which are all highly important for producing innovative products or at production stages that intensively require high-tech knowledge and advanced skills. Such production is more commonly carried out in the developed world and the underlying agglomeration processes might not be at work in an emerging economy like Vietnam. In addition, Duranton (2008) discusses that stronger regional specialization requires more intensive trade between regions, but transport costs are high in developing countries due to their underdeveloped infrastructure. Thus, intra-industry agglomeration forces are expected to be weaker there, which is also seen from the fact that cities in developing countries are in general less specialized than they are in developed countries (Duranton, 2008).

In terms of economic achievement, institutional system, and management policies of urban and industrialization, Vietnam shares many common features with China (Malesky & London, 2014). Both restrict domestic migration to big cities through the Hukou system or developing industrial zones to promote manufacturing activities and foreign investment. Given these similarities, one might expect several conclusions for Vietnam similar to what are found in Au and Henderson (2006) and Combes et al. (2013). Indeed, in the short-run setting, Howard et al. (2014) find that firms derive no performance benefits from locating next to other same industry firms in Vietnam, which is an additional reason to propose the first hypothesis.

Hypothesis 1. *Urbanization and industrial diversity have positive effects on productivity and dominate over regional specialization in Vietnam.*

The next aspect under examination is influence on productivity of agglomeration in various firm ownerships. Rigby and Brown (2015) make a rare comparison on this topic for a developed country and find no significant difference between domestic-owned enterprises (DOEs) and FOEs in benefiting from spatial proximity. That finding is reasonable due to the fact that the disparities between domestic and foreign firms in technology and their ability to reap productivity advantages from regional conditions are minor in advanced countries compared to the rest of the world including Vietnam. After the 1986 Doi Moi reform, the Vietnamese government implemented many policies to attract foreign direct investment (FDI) in order to boost economic growth, create jobs and gain from technology diffusion pursuing the goal of industrialization. Nowadays, the FDI sector plays a critical role in the Vietnamese economy with an equal position to the domestic sector (Newman et al., 2015). One of the most appealing factors of Vietnam to FDI investors is an abundant, young, and cheap labor force. Hence, when cities become more agglomerated, one might expect FOEs to enjoy a higher productivity premium than domestic firms since FOEs have better opportunities to access the labor market due to higher wages and probably also their international reputation. Lamin and Livanis (2013) report that newborn FOEs save more information cost in a specialized area due to having better access to knowledge and experience related to industry regulation and local institutions from experienced FOEs and domestic local suppliers.

As for domestic firms, since the Doi Moi, Vietnamese policy makers have begun to encourage private business, but regarded state-owned enterprises (SOEs) as a pillar of the national economy with protectionist and preferential treatment. The strategy led to policy-driven rather than profit-maximization behavior of SOEs which made them less productive in comparison with private ownership. Recognizing that mistake, since the early 2000s, the government has intensified privatizing or equitizing SOEs to transform them into private companies or listed companies while also remaining the majority shareholder. This process has altered the behavior of the remaining SOEs to make them more market-oriented which is normally found in private companies. Therefore, among three ownerships, FOEs have been leading in technology and efficiency, while the difference between SOEs and POEs has become narrower. Combes et al. (2012) find the evidence that the most productive firms enjoy the highest productivity premium in highly dense regions. The work of Gokan et al. (2019) is not able to verify that information spreading across or within industries benefits Vietnamese firms in the long run. By using the term "agglomeration", this study henceforth means any of urbanization, diversity, or specialization. Based on above reasons and evidence, the next suggested hypothesis is thus.

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

The last aspect analyzed in this study is the heterogeneous external effects on firms of different sizes and ages. Smaller size or younger age is proposed to be a stroke of fortune in agglomerated areas. Organization theorists point out that to manage a huge number of people, large firms tend to introduce rigid regulations and culture of impersonal relationships which might reduce their flexibility to adapt quickly to regional changes or find effective ways to interact with local business partners (Knoben et al., 2016; McCann & Folta, 2011). This problem is not the case with firms of modest scale. Similarly, the introduction of new organizing principles to replace the typical management pattern of one person ruling them all found in newborn firms is more and more urgent as firms age and expand. If mature firms fail to meet this development requirement, their ability to acquire knowledge disseminated through external effects would be impaired. In contrast, immature firms can be blessed with better capacity to gain new knowledge and higher flexibility to adopt novel routines in dense regions (McCann & Folta, 2011). From a regional science point of view, Nichter and Goldmark (2009) discuss that vertical and horizontal linkages built in the urban area might minimize small-scale weaknesses by boosting the supply through joint production or improving the negotiating position due to the availability of more suppliers. In addition, diversity in supporting services such as consulting, employee training, or shipping enhances the performance of small firms since they might not be able to afford to insource them as large firms do. By similar logic, young firms may perform better in a denser or more specialized region by learning from local experienced predecessors via the formation of larger social networks. Rosenthal and Strange (2010) argue that smaller firms, and thus younger ones, are more dependent on environmental factors, therefore if agglomeration externalities exist in the region, they might be the most advantaged.

The related empirical evidence tends to support the above arguments, but studies remain few and far between for developed countries, and rare or even non-existent for developing countries. Henderson (2003) finds that smaller firms derive more benefits from specialization in the US, which is the same as the conclusion of Martin et al. (2011) for France. The studies of Andersson and Lööf (2011) and Rigby and Brown (2015) using Swedish and Canadian data respectively show no statistical difference in economic gains from local size between firms of various scales, while Combes et al. (2012), with the only article producing a different result, point out the predominance of large firms in harvesting the productivity advantage in densely populated regions of France. Based on Egyptian data, Badr et al. (2019) conclude that firms benefit the most from both intra- and inter- interactions if they are small. As for age factor, McCann and Folta (2011) employs data from biotechnology firms in the US and find that immature firms derive higher productivity premium from industrial specialization. Rigby and Brown (2015) reach the same conclusion for manufacturing Canadian firms. Ciani et al. (2020) notice that on average small firms are five years younger than large firms in low and middle-income countries. Given the behavioral proximity between young and small firms, the findings of evidence of productivity advantages of firms with either of the two characteristics are expected. Nichter and Goldmark (2009) show that market environment is a growth booster for small and medium-sized firms in developing territories, while it is not in their developed counterparts. This may result from the fact that small businesses in developed countries are fewer in number but larger in size (Poschke, 2018; Ciani et al., 2020). In addition, as a result of the earlier development of developed economies, their firms on average are older than ones in developing economies. Those arguments and the empirical evidence mentioned above build up the expectation of finding clearer proof for Vietnam's developing economy, where there is a lion's share of firms being immature or small-scale. Given the fact that young is not equivalent to small as nearly 50 % of newborn firms are large-sized in low-income countries, while the number is about 33% for medium- and high-income countries (Ciani et al., 2020), the deviation in the results when comparing between firms of various sizes and ages is expected. In general, the following are suggested, ceteris paribus.

Hypothesis 3. *Smaller-sized firms benefit more from agglomeration.*

Hypothesis 4. *Younger firms benefit more from agglomeration.*

3. TFP estimation

This section focuses on strategies to obtain consistent productivity indices using firm-level data before conducting the primary estimation for agglomeration economies. To begin with, the value-added based production technology of a firm *i* at time *t* is assumed to take the Cobb-Douglas form as

$$Y_{it} = A_{it} K_{it}^{\beta_k} L_{it}^{\beta_l} \tag{1}$$

where Y_{it} denotes the firm's value added; K_{it} and L_{it} refer to capital and labor inputs; and A_{it} represents the firm's Hickson neutral efficiency level during period t.

The logarithmic transformation of (1), with lower-case letters indexing natural logarithm of variables, is

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + v_{it} + \eta_{it} \tag{2}$$

where $\ln(A_{it}) = \beta_0 + v_{it} + \eta_{it}$; parameters β_k and β_l are capital and labor input elasticities of production function respectively; β_0 is common technological knowledge to all firms in the economy and technically it represents the intercept parameter in the regression model; v_{it} refers to firm's productivity, which is observable to the firm but unobservable to researchers; and disturbance term η_{it} is unobservable and assumed to be independent and identically distributed (i.i.d). A concern raised in literature is likely contemporaneous positive correlation between the unobservable element v_t embodied in the regression's disturbance and both production inputs and outputs. In practical terms, firms optimize their input levels in order for profit-maximization according to their own predetermined productivity. As a consequence, without controlling v_t, the application of the ordinary least squares (OLS) method to (2) could result in an upward bias of elasticities βs estimated, thereby deflating the value of productivity.

This study follows the control function approach to deal with this issue, pioneered by the work of Olley & Pakes (OP for short) (1996) and Levinsohn & Petrin (LP for short) (2003). Their primary strategy is to make v_t explicit through the inversion of demand functions of factors that productivity enters into their decision making process. The OP and LP techniques use functions of investment and intermediate inputs respectively for that purpose under the monotonicity assumption which is proved to be very likely valid in Figure A1 and Figure A2 in Appendix A. In work related to computation of firm-level TFP in Vietnam, Newman et al. (2015) and

Nguyen (2017) employ data from the Census of Enterprises with the control functions of investment and material costs respectively. As noticed by LP, the choice between two procedures should weigh on the volume of null values reported for the potential control factor. The data set used in this paper shows a considerable share of firms recorded with no annual investment that can truncate the data remarkably, thus diminishing the consistency of the estimation. Meanwhile, power consumption, one of common intermediate inputs in manufacturing production, is systematically reported with positive values. In addition, its non-monetary values and non-storable characteristics help to reflect better the annual productivity shocks, and therefore are superior to the features of material costs used in Nguyen (2017). As a result, power consumption is primarily employed in this study as the proxy of intermediate inputs.

To tackle potential simultaneity biases, firm i's demand for intermediate inputs, which is absent from (1), is assumed to depend on its capital stocks and productivity as

$$m_{it} = m_{it}(v_{it}, k_{it}) \tag{3}$$

By assumption, whose validity is demonstrated in Appendix A, (3) is monotonically increasing in v_t, hence can be inverted into the non-parametric function

$$v_{it} = m_{it}^{-1}(v_{it}, k_{it}) = v_{it}(m_{it}, k_{it}) \tag{4}$$

Plugging (4) into (2) gives

$$y_{it} = \beta_0 + \beta_l l_{it} + [\beta_k k_{it} + v_{it}(m_{it}, k_{it})] + \eta_{it} \tag{5}$$

Now the unobserved term v_{it} becomes observable through m_{it} and k_{it}, making (2) estimable without embodying productivity in the error term. LP develop a two-stage procedure to compute the estimators of β_s using (5). The first step is to regress (5) using OLS with fourth-degree polynomials in k_{it} and m_{it} substituted for components inside the square brackets to obtain the estimated coefficient of β_l. Next to do is making use of OLS residuals yielded from the first step to calculate the expected value of v_{it} conditional on its past information with the feature of markovian assumed. Estimated values of β_k and β_m are derived following a routine of optimization to minimize regression residuals. However, Ackerberg et al. (2006) argue that the implementation of OLS during the first stage may be problematic due to the likelihood that firms simultaneously decide on the quantity of workers and intermediate inputs. In other words, l_{it} should be treated similarly to m_{it} as a non-parametric function of state variables, implying an incorrect identification for (5). If this is the case, the estimator of β_l using either the OP or LP procedures is still biased upwards. To solve this potential issue, this paper strengthens the LP model with IV estimation proposed by Wooldridge (2009) who makes use of the GMM framework instead of a two-stage process expressed above. The method is denoted shortly as LP-W to emphasize the use of Wooldridge's method based on the core idea of LP. In addition, Fig. A3 in Appendix A shows that, holding capital constant, LP-W TFP and the power consumption indicate a distinctive positive relationship, which strongly supports the application of the monotonicity condition. Henceforth, this paper considers LP-W to be the most reliable and indeed the standard approach to tackle simultaneity bias. Even though, as a sensitive check on how findings on agglomeration effects change over various productivity estimation methods, TFP is also calculated with the OLS, LP, and OP approaches despite some potential issues. The estimation strategy of LP-W is expressed in greater detail in Appendix A.

4. Data

4.1. Handling data

The primary data source of this paper is the Annual Census of Enterprises conducted by GSO in Vietnam for the period 2011–2016. There are two forms of information at firm level merged into the data set. The first type covers all firms with business tax codes in all sectors and regions in Vietnam with data collected from local tax offices and authorities where firms are under their administration. For each firm, available information consists of its 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. The coverage of all firms without any threshold 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.[1] The second type, connected with the first one through the firm identifiers, is extracted from surveys carried out by GSO. Exceptions occurred in 2011 and 2016 as comprehensive censuses were undertaken to collect information on all firms in Vietnam. In the remaining years (2012–2015), firms were sampled based on their industry, ownership, region, and size.[2] More concretely, GSO

[1] One may expect to generate regional variables with a comprehensive establishment-level data set for the highest accuracy. However, based on the data of 2011 when establishment-level information is available to extract, only approximately 2.6 % of firms operating in Vietnam are multi-establishment and have other establishments outside the districts where their headquarters are located. In addition, the workforce in the headquarters commonly makes up the largest share of employment among establishments, thus the disparity in the values of regional variables aggregated from firm-level data compared with those from establishment-level data is negligible.

[2] In 2012, for instance, 168,854 firms completed detailed questionnaires in the GSO survey, while information on location, industry, labor, and revenues is available for a total of 358,558 firms.

investigated all firms that were state-owned, foreign-invested, or belonged to ten of the sparsest provinces of Vietnam. Domestic private-owned enterprises (POEs) were all surveyed if they had at least 20 employees,[3] but sampled randomly if they had less than 20 employees. Information collected consists of the value of assets, investments, energy use, labor costs, interest costs, profits, taxes, and other financial obligations to the authority. A limitation of these surveys is that several variables were gathered in only a few years, such as the qualification structures of workers, and the value of exported and imported goods. Furthermore, no information on hours of work and materials was collected.

To meet the requirements for productivity estimation, data of both types mentioned above is exploited since detailed information on the inputs and outputs of each firm is necessary. Firstly, the measurements of observable variables on the right-hand side of production function should be clarified. They are made up of capital, labor, and intermediate inputs if the LP and LP-W methods are applied, or investments if OP is the case. A firm's labor input is measured as its number of employees at the end of the year. The decision on the measure for capital stock is more problematic since it is related to depreciation and devaluation of assets over time, which varies across firms, industries, regions, and years. If detailed information on a firm's capital such as the specific types of capital inputs and corresponding depreciation methods is available, one can apply the perpetual inventory method (PIM) to compute assets value. However, in the case of GSO data, full information is obtainable for merely a few years. Furthermore, the choice of unbalanced panel data results in difficulties in tracking annual fluctuation of fixed assets. Therefore, this study uses original value of total fixed assets recorded for each firm at the end of the year[4] as a proxy for capital input. Its value is deflated using the deflator of development investment capital of society provided by GSO.

For intermediate inputs, this paper uses annual electricity consumption for production to represent variable m_{it} in Eq. (2). The final measure needed to clarify is value added Y_{it} in (1). It is calculated indirectly due to the absence of data on material cost. Following Ha and Kiyota (2014) who make use of firm-level data from GSO for another subject of study, a firm's nominal value added in a year is computed as the sum of total employment costs, the change in accumulated depreciation between the beginning and end of that year, net operating profit, indirect taxes, and interest costs.[5] The firm's value added is then deflated using a manufacturing value-added GDP deflator, also taken from GSO. The TFP estimation is conducted separately for each of 21 2-digit VSIC manufacturing industries[6] using the Stata command prodest created by Rovigatti and Mollisi (2018).

The final thing to consider is data-cleaning procedures for regression. All observations from the raw data sets, which cover enterprises from all economic sectors, are employed to create regional variables, with approximately 1 % of all observations being dropped because either of their codes of tax, location, or industry, or value of total employment is found missing or negative. For TFP estimation and agglomeration estimation, non-manufacturing firms are removed, leaving 341,116 observations of manufacturing firms over the period 2011–2016 in the regression sample. Next, observations of firms that were out of GSO's survey are removed since their data on many production inputs such as electricity consumption and investment are missing, and firms with under 20 employees are removed to partly reduce the possible effect of random sampling.[7] In addition, the industry code is assigned to the main industry if a firm is multi-industry. The elimination of firms who changed their main industrial sector or their location between 2011 and 2016 is also applied[8] as a solution to potential data input errors and a treatment to mitigate the possible problem of self-selection that is discussed in the next section. Technically, firms with different 3-digit industrial codes or district-level location codes are excluded from the sample. At the step of conducting regression on agglomeration, the remaining sample 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 and correlation matrices of variables for TFP and agglomeration estimation are shown in Tables C1–4 in Appendix C.

4.2. Scale units and summary statistics

Since regional variables are constructed and aggregated based on the choice of geographical boundaries, the identification of their unit type is crucial. Data on regional land area used in this paper is gathered from provincial-level statistical yearbooks published annually in Vietnam. Due to the 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 GSO's 2016 census. There are three administrative units used in the data set including province, district, and commune, with level of the last one being finest. In 2016, officially, there were 63 provincial-level, 793 district-level, and 11,162 commune-level regions in Vietnam. On average, a province is equivalent to a circle 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. On the one hand, the application of a small geographical unit may provide a misleading picture of the agglomeration level due to spatial clusters being broken down into separate areas with borders (Head & Mayer, 2004). On the other hand, external gains from information spillover are solely found at lower levels of geography, say a commune or district. The gains from sharing intermediate inputs or natural sources are found at higher

[3] This firm-size threshold varies across regions and years. Normally, this number is higher in some big cities. For example, in 2012, GSO randomly sampled 20 % of domestic POEs with fewer than 20 employees. Nevertheless, the corresponding figures for the two biggest cities (Hanoi and Ho Chi Minh) were 20 % of firms with 20–50 employees and 10 % of those with under 20 employees.

[4] This choice of timing is appropriate for the characteristic of capital accumulation given in Levinsohn and Perin (2003)'s model.

[5] Nevertheless, the calculation in this paper is more accurate since the factor of interest cost is absent in the Ha and Kiyota (2014)'s computation.

[6] Industry 12 (tobacco) and industry 19 (coke and refined petroleum products) are dropped from the productivity calculation due to their small numbers of observations.

[7] Although this measure helps to better deal with random sampling effects, it greatly reduces the sample size.

[8] This implementation drops about 11.8 % of total observations of the sample at the stage of cleaning.

geographical level, say district or province, meanwhile the gains from labor market pooling are robust across all geographical scopes (Rosenthal & Strange, 2001; Rosenthal & Strange, 2004; and Combes & Gobillon, 2015). To balance those arguments and choose the most suitable unit to identify the external economies, this paper considers districts integrated with 3-digit VSIC codes as the spatial and industrial unit in the baseline analysis, even though the regression is also performed at 2-digit VSIC scale as a sensitivity check.

In terms of spatial structure, the map of Vietnam in Fig. 1 shows 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 the districts with density above 1000 employees per square km are mainly located in or around two biggest cities of Vietnam - Hanoi and Ho Chi Minh city, with a large proportion of the rest typically located along the coastline. Regarding the industrial structure of the final sample, the most influential industries are

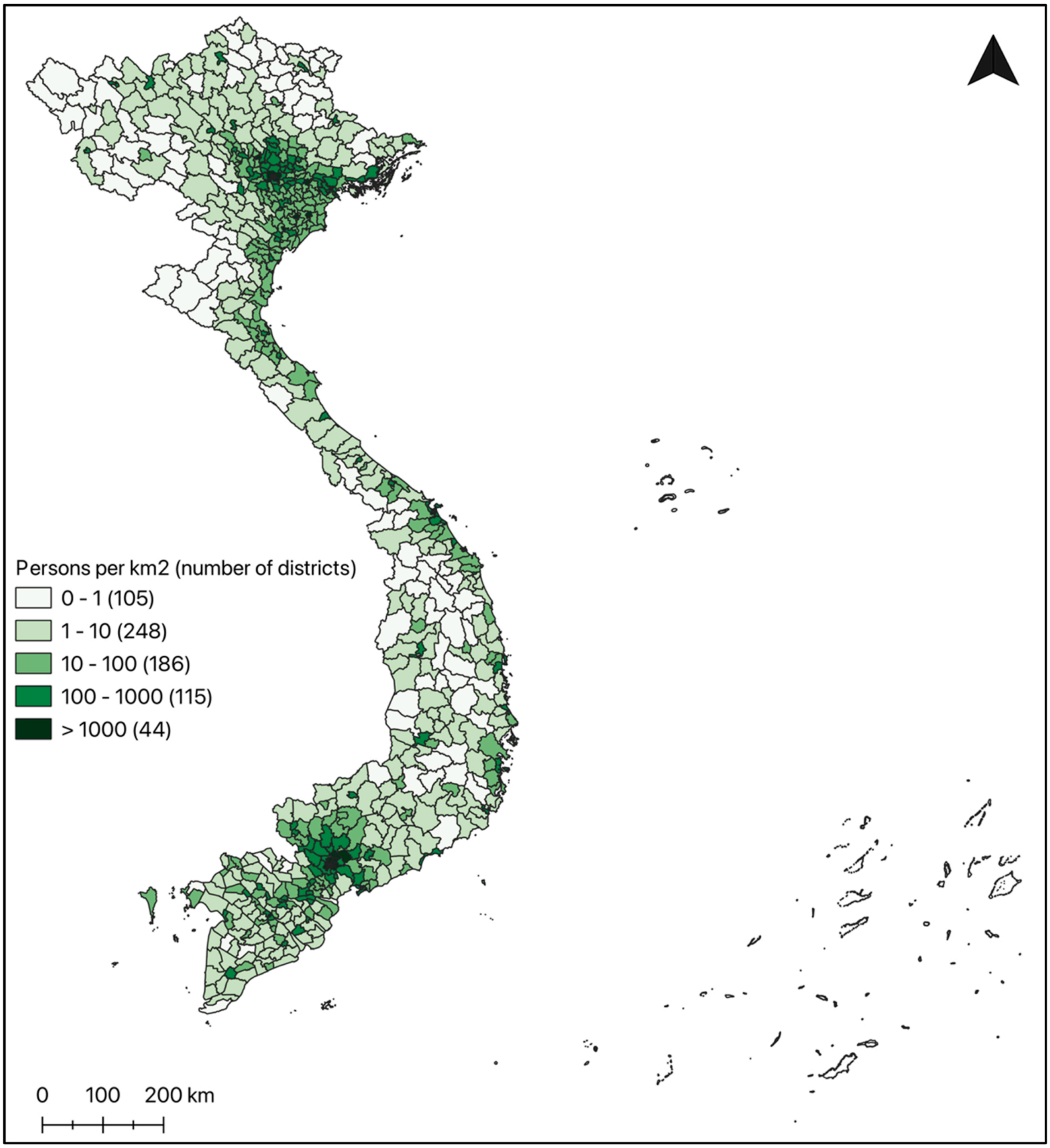

Fig. 1. Map of Vietnam and annual average employment density of district-level regions. *Notes:* The map is drawn using employment data from the censuses of enterprises conducted by General Statistical Office between 2011 and 2016, land-area data from provincial-level statistical yearbooks, and spatial data published by the Ministry of Natural Resources and Environment in Vietnam.

the ones producing food products, clothing, non-metallic mineral products, and fabricated metal products (except machinery and equipment), which are four out of twenty one industries but make up about 46 % of observations in total. According to the statistical classification of economic activities constructed by the European Community (NACE), the four industries are classified as either low-technology or medium-low-technology, which implies that innovation or the pool of highly skilled specialized workers might be less important for the Vietnamese manufacturing sector compared to other economies with the dominance of high-technology industries.

5. Empirical strategy

5.1. Agglomeration proxies

This subsection focuses on regional characteristics as main explanatory variables. The measures used to capture urbanization economies are urbanization and diversity because of their parallel presence in urban regions. However, the former is considered to represent the absolute sense of urbanization, while the latter catches the relative sense, or so-called Jacob externalities. To be specific, urbanization is proxied by a logarithm of employment density, calculated based on micro data as

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

where e_{it}^r is the workforce of firm i located in region r with a_{rt} being its land area at time t. An advantage of this density-based index, compared to other absolute employment measures, is that it is less impacted by zoning idiosyncrasies, thus better reflects the true scale effects and produces more reliable results (Combes et al., 2010). The next index indicates the industrial diversity in region r for firms in industry s, computed as

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

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 total number of workers in region r and time t. $S|\{s\}$ represents the total set of industries with industry s omitted, which is necessary to enable a better interpretation given the presence of a specialization variable in the model (Combes & Gobillon, 2015). The diversity level reaches the minimum value of zero if two industries exist in the region, while it rises with an increase in the number of industries and the equal degree of their regional shares of employment.

Next to mention is the specialization index. This paper uses the logarithm of sum of location quotient (*LQ*) plus one to capture MAR externalities, calculated as

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

where e_t^{sr} and e_t^r are known above, e_t^s is defined as the number of workers in industry s in the whole country, and e_t refers to the total number of employees working for firms in the country in year t. Location quotient is the ratio that compares the employment share of a particular 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 an industry in a region. *LQ* above 1 implies that the regional share of employees in a specific industry is higher than its national share, and therefore the relative specialization strength of that industry is stronger than average. *LQ* is equal to 10 for an industry, for example, meaning that the industry in question is 10 times more concentrated in terms of employment in the local region than on the national average. *SPE* is the log-transformed version of sum of *LQ* plus one, thus has the minimum value of zero. The higher the value of this index shows, the higher the degree of regional specialization it implies.

5.2. Specification

To estimate agglomeration economies, firm productivity is regressed on main regional proxies plus additional controls, with linear relationships assumed, using the following baseline specification

$$\ln TFP_{it}^{sr} = \alpha + \beta_1 DEN_t^r + \beta_2 SPE_t^r + \beta_3 DIV_t^{sr} + \beta_4 X1_{it}^{sr} + \beta_5 X2_t^{sr} + X_{3t} + f_i + \epsilon_{it} \tag{6}$$

where $\ln TFP_{it}^{sr}$ is the natural logarithm of firm *TFP*; *DEN*, *SPE*, and *DIV* are the employment density, specialization, and diversity indices respectively as mentioned; vectors $X1_{it}^{sr}$ and $X2_t^{sr}$ contain time-variant firm-level and regional controls respectively; vector X_{3t} 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 control variables used are foreign ownership and state ownership following Waldkirch (2014) who shows that domestic firms tend to grow faster induced by restructuring after being taken over by foreign investors. Empirical evidence from India, Taiwan, and Kenya confirms that a firm may receive a productivity boost if its percentage of state ownership is lower or its percentage of foreign ownership is higher (Douma et al., 2006; Huang & Shiu, 2009; Ongore, 2011). It is worth emphasizing that foreign ownership in this study 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.

Turning to region-level controls, three variables inserted into the model are industrial competition strength, regional share of number of firms under foreign control, and regional average wages. The first one – competitive pressure – increases with the intensity of intra-industry transactions, employed following Glaeser et al. (1992) who find the positive impact of competition on growth. Its underlying mechanism suggested by Porter (1990) is that free competition gives the impulse for technological improvements and firms' efforts to be more innovative to survive and grow. In the model, the competitive intensity is measured based on the employment share structure of the labor market as

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

In the market with only one firm, the market structure is described as monopoly and the corresponding *COM* gets the minimum value of zero. More firms entering the market or expanding their labor force will eat into others' employment shares, resulting in a rise in *COM*, and therefore implying a tougher degree of competition. The opposite outcome occurs when smaller firms employ more workers, while their bigger rivals reduce their employment scale but the total number of local firms in the industry remains unchanged. Based on the persuasive evidence found in Takii (2005) and Newman et al. (2015), the second regional control is the share of regional firms owned by foreigners, used to control for technological spillovers from advanced foreign companies to both other local FOEs and local domestic firms. The third is regional average wages added to proxy for human capital factor, which is shown to play an important role in local growth in Acs and Armington (2004).

5.3. Econometric issues

Several potential econometric issues may arise at the stage of agglomeration estimation, which are discussed at length in Henderson (2003) and Combes and Gobillon (2015). The first potential issue is reverse causality happening when the reputation of a region packed with many productive firms attracts more firms or workers from the same or different industries in other regions, thus urbanizing and/ or further specializing that region. If agglomeration economies do exist, this issue might lead to an upward bias in OLS estimates. The second possible problem is the endogeneity resulting from missing regional variables correlated to both agglomeration proxies and firm productivity. For instance, an education system with higher quality or a better transportation infrastructure with broad roads, airports, and seaports commonly found in metropolitan or specialized regions is also productivity-enhancing to local firms. Other regional factors such as the ability of local authority, business and life environment, natural and climate advantages, and public amenities may give rise to the similar problems. One may expect an upward bias in agglomeration estimates due to these omitted variables. The third one is sorting behavior, occurring when firms make a location choice based on their own productivity and regional properties that are strictly associated with agglomeration. The final one is selection effects arise 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 productivity-improving effects of agglomeration might be confused with a selection process.

A solution to those issues is to instrument for regional variables with historical or geological information, applied by Ciccone and Hall (1996), Combes et al. (2010), and Duranton (2016). However, this method suffers from several major limitations: it can be used merely to instrument for the scale effects, leaving industrial related variables such as specialization untouched; since instrumental variables (IV) of this kind are time-invariant, practitioners must sacrifice the time dimension of panel data by averaging all variables on both sides of the specification, or make use of data of only one certain year. Another strategy in literature is to first difference data then use two or three-period lagged values of differenced agglomeration measures as instruments, which is the core idea of FD-GMM estimation, employed by Henderson (2003) and Martin et al. (2011). A big advantage of this method over the above one is that it is applicable to all explanatory variables by exploiting their own past information in the data set. Nevertheless, given the unbalanced nature of micro-data used in the strand of literature, the application of this technique reduces sharply the sample size induced by balancing the sample artificially, in other words, keeping mainly survivors throughout most sampling periods. If the reasons behind firms' survivability are linked to both regional characteristics and lagged instruments, the estimated results are not free from biases (Levinsohn & Petrin, 2003). Combes and Gobillon (2015) recommend against the use of FD-GMM in agglomeration estimation.

To tackle potential problems of endogeneity, this paper applies the fixed-effects technique following Combes and Gobillon (2015), which avoids the weakness of the methods mentioned above. By removing observations of firms changing their location or industry during sampling periods, the presence of firm fixed effects as f_i in (6) captures not only time-invariant firm characteristics but also location and industry fixed effects, which solves the problem of missing constant regional variables. Those FE terms in combination with several influential firm-level controls help deal with the sorting problem (Combes & Gobillon, 2015) and with the selection bias. To mitigate the problem of missing time variant regional information, year, year-industry, and year-region fixed effects are also inserted into the model. These terms help to reduce potential biases by taking into account unobserved industrial and local shocks that tend to happen in agglomerated regions.

Besides the above arguments, this study is in favor of the FE technique due to following reasons: when applying the Hausman test to the baseline estimation, random estimates are rejected in favor of the FE method; Henderson (2003) and Combes and Gobillon (2015) reach a conclusion after their efforts to deal with possible econometric problems of agglomeration estimation that the application of fixed-effects at micro level is superior to and more influential than the IV technique and FD-GMM; Combes et al. (2012) put considerable effort towards explicitly the issue of selection taking into account and they receive almost-unchanged estimated results; Gokan et al. (2019), following Combes et al. (2012)'s strategy, find no significant distortion from selection in Vietnam; Finally, as a sensitivity check, the baseline results are compared to the ones from the adjusted sample with two biggest cities in Vietnam (Hanoi and Ho Chi Minh) excluded. If the selection bias is serious, the main results would show a strong or even severe shift in the magnitude and sign of estimated coefficients. However, the results shown in Section 7 indicate that the estimates are stable to the exclusion. On the whole, the combination of multi-level fixed effects and controls is considered to be sufficient to deal with the potential endogenous issues in estimation of agglomeration in this study.

6. Primary results

6.1. Production function results

The coefficients of production input elasticities for 21 separate industries with the LP-W strategy are all positive and significant and vary from industry to industry, detailed in Appendix B. Regarding capital, the lowest capital elasticities are 0.04 and 0.075 belonging to the business lines producing clothing and leather products respectively – the labor intensive industries, while two high numbers are 0.476 for the manufacture of rubber and plastic products and 0.573 for the production of beverages products. Turning to labor, its estimation values lie between 0.503 and 1.026. Among industries, output is the least sensitive to changes in labor quantity in the industry of electrical equipment. In contrast, output responses most strongly to a change in labor force in the industry of clothing, which is unsurprising because of its high demand for manual workforce over machinery. Moreover, it is the sole field with the labor elasticity above 1 and is estimated with the lowest capital estimator as mentioned. The remaining industries show estimated values for two inputs falling into the middle range of the two above extremes. To wrap up, the remarkable heterogeneity of elasticity coefficients highlights the importance of performing separate estimation for various industries, as conducted in this paper, rather than grouping all observations into one single sample to implement regression.

Next to notice are the different outcomes between regression using LP-W and the rest, in particular OLS. In comparison with results from LP-W, OLS input estimators show a distinct disparity pattern in their magnitude. As discussed above, uncontrolled productivity embodied in the error term might bias input parameters of the production function upwards, and consequently underestimate the firms' real performance. The results yielded in this paper confirm the prediction as, on average, the OLS estimators of capital and labor are about 10.8% and 6.9% higher respectively compared with the LP-W versions. Consequently, it is anticipated that using OLS productivity will underestimate the magnitude of agglomeration economies. For LP and OP, the degree of difference is lower, but the direction might be positive or negative depending on specific industries. In terms of LP and LP-W, the remedy against potential simultaneity problem of LP-W leads to a pattern: the latter tends to show lower $\widehat{\beta}_l$ but higher $\widehat{\beta}_k$. Such an outcome is reasonable, given that the LP-W method is applied to correct the potential upward bias of $\widehat{\beta}_l$ in LP. This sort of compensation results in the trivial disparity of productivity values using each of the trio LP, OP and LP-W. However, as analyzed in previous sections, the LP-W technique is considered to produce the most consistent estimates of productivity, and thus a consistent agglomeration estimation.

6.2. Baseline estimates

As a starting point for giving results on agglomeration economies, this subsection focuses on the productivity effects of spatial concentration on a general firm, displayed in Table 1. Columns (1) to (6) present estimation results with controls added in succession. Standard errors are clustered at a provincial and 2-digit VSIC level following the guide of Moulton (1990) and Cameron & Miller (2015) as far as the regression of micro-level explained variables on regional regressors is concerned. It is well recognized that estimated coefficients for agglomeration proxies remain stable with the addition of various controls. Two parameters of urbanization are always positive and significant whereas that of specialization remains statistically insignificant. Regarding magnitude, taking into account control factors leads to merely minor changes in the strength of computed agglomeration impacts.

Next to examine is the results with all controls included in columns (7) and (8). The difference between the two columns is that the FE terms of time-industry and time-region are dropped from (7) in order to emphasize the importance of controlling unobserved industrial shocks which tend to happen in agglomerated areas and are likely positive in a fast-growing economy. 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 *DEN*'s estimated coefficient is in line with the anticipation, and thus underlines the presence of upward bias if regional shocks are not taken into consideration. The results in column (8) indicate the existence of urbanization economies with a statistically significant coefficient for the local density. 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. The interpretation is consistent with the reports on density externalities of other authors in literature, such as Rosenthal and Strange (2004) and Combes et al. (2010), despite differences in the specification and construction of variables. Noticeably, the elasticity is a bit larger for Vietnam as an emerging country when compared with its US and European counterparts where productivity elasticities to urbanization vary between 3 % and 8 % (Rosenthal & Strange, 2004). This disparity, again, supports the observation of Combes and Gobillon (2015). In the case of industrial variety, another aspect of urbanization, the estimated results confirm its positive influence on productivity, which echoes Jacobs (1969)'s ideas and the findings of Glaeser et al. (1992).

As for the impacts of specialization, its coefficient is statistically insignificant, implying an ambiguous role of industrial

Table 1
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 provincial 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 %

concentration on local productivity, which might be driven by the dominance of low-technology manufacturing industries in Vietnam. To sum up, the results verify the existence of agglomeration economies in Vietnam and suggest inter-industry rather than intra-industry relationships as the main source of external benefits, and thus approves Hypothesis 1. This is in line with the findings of Au and Henderson (2006) and Combes et al. (2013) on the dominance of urbanization economies in China, another emerging country. When compared to available evidence found in developed countries during their industrialization period in the past, the results are consistent with Hanlon and Miscio (2017), who also find an insignificant effect of regional specialization among various agglomeration forces in English cities between 1851 and 1911, but inconsistent with Klein and Crafts (2020), who confirm the presence of the positive externalities due to both specialization and diversity in the US between 1880 and 1930.[9] This mixed consistency implies that to correlate Hypothesis 1 to the industrialization stage of countries to which Vietnam belongs, further studies and empirical evidence in the future are needed.

The last point to focus on in this subsection is the influence of control variables. Regarding firm ownership, its estimate is positive and significant for foreign ownership but insignificant for state ownership. As expected, it indicates that a firm performs better with the more intense presence of foreign investors, while its performance is not enhanced with the control level of the state. In terms of regional controls, the regression shows all positive signs and significant impacts as predicted, except for fraction of FOEs in the region. These results support the findings of Glaeser et al. (1992) on the existence of Porter externalities and the findings of Acs and Armington (2004) on positive influence of local human capital.

6.3. *Interaction between agglomeration effects and firm characteristics*

This part aims at revealing who gains the most from agglomeration, which is achieved technically by adding interaction terms between three proxies of agglomeration and firm characteristics of interest one after the other. The results are expressed in Table 2. First to consider is the heterogeneity of agglomeration economies across various ownerships. To yield the outputs found in column (1), a categorical variable of ownership type is added to the baseline model, plus its product terms with *DEN*, *SPE*, and *DIV*, setting domestic private-owned status as the benchmark group. Estimators show positive sign with statistical significance purely for interactions of foreign-owned status with *DEN* and *SPE*, whereas interacting state- owned status with regional variables gives all insignificant parameters. It implies that FOEs benefit more from local urbanization and specialization in comparison with both SOEs and POEs. The external effects on SOEs and POEs, however, are insignificantly differentiated. To make a more meaningful comparison, the variable of

[9] It is worth noticing that the manufacturing sector's structure during the reference time of the two studies is starkly different from the present time, and they report the evidence of dynamic effects rather than static effects as in this paper.

Table 2
Agglomeration effects with interaction terms of agglomeration and firm characteristics, District/3-digit VSIC.

		(1)	(2)	(3)	(4)
Variables of interest	*DEN*	0.075***	0.075***	0.128***	0.153***
		(0.028)	(0.028)	(0.028)	(0.029)
	SPE	-0.024	-0.022	0.010	0.053**
		(0.021)	(0.021)	(0.019)	(0.022)
	DIV	0.079***	0.079***	0.086***	0.064***
		(0.022)	(0.022)	(0.020)	(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***	0.186***	-	-
		(0.024)	(0.024)		
	SPE * Ownership type = Foreign-owned	0.119***	0.117***	-	-
		(0.033)	(0.033)		
	DIV * Ownership type = Foreign-owned	-0.021	-0.021	-	-
		(0.036)	(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***	-1.046***	-	-
		(0.191)	(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 provincial and 2-digit VSIC level. Each column of this table replicates column (8) of Table 1 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 the product terms. Controls 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%

ownership status is transformed into a binary variable which is one for domestic-owned status as the base group and two for foreign-owned. The new results are shown in column (2) after the regression is rerun. The estimated parameters are almost similar to what they are in column (1), suggesting the productivity advantage of FOEs over DOEs in general in a more agglomerated region. Concretely, FOEs are predicted to be 18.6 % points more productive relative to DOEs when local employment density doubles. Graph A in Fig. 2 presents these disparities intuitively with a more inclined line for FOEs in the relationship between density and productivity.[10] Interestingly, the graph indicates that DOEs perform better than FOEs in low-density regions, but they lose to their foreign counterparts in high-density regions. In brief, the empirical results support Hypothesis 2 and echo the conclusion of Howard et al. (2014) and Gokan

[10] Although the estimate of the interaction between regional specialization and foreign-owned status is statistically significant, its graph is not plotted since the insignificant coefficient of the main term of specialization might result in lines with less reliable slopes.

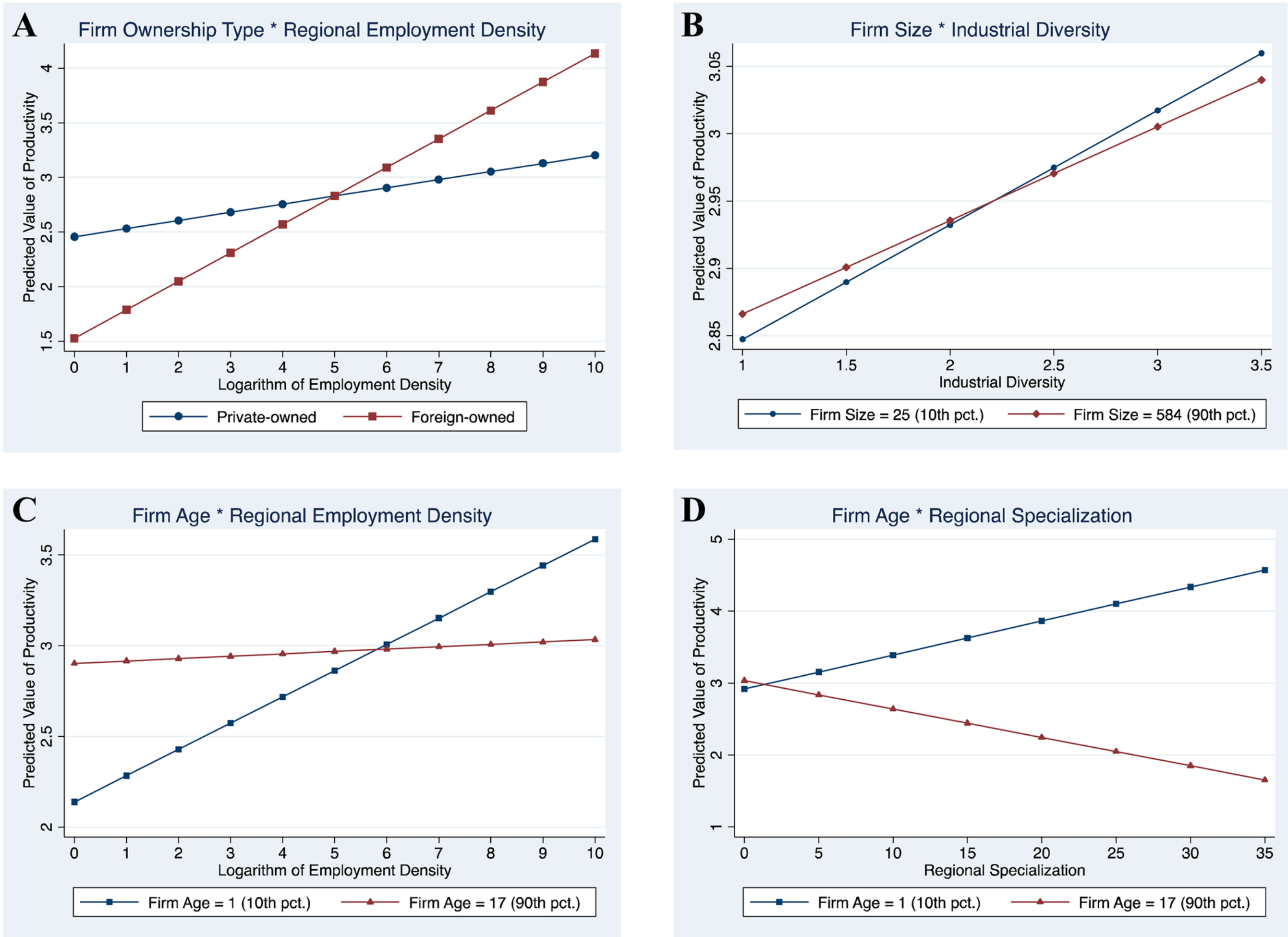

Fig. 2. Graphs of several interactions between agglomeration and firm characteristics. *Notes:* To generate each figure, the predicted value of the logarithm of total factor productivity measured with the method of Woodridge (2009) on the y-axis is produced using estimated coefficients and the mean value of the remaining explanatory variables. This value is then plotted separately for two groups of comparison with their different values accordingly against an agglomeration variable on the x-axis. Graphs A and B are produced following the regression results shown in column (2) of Table 2, while graphs C and D are based on columns (3) and (4) respectively. Abbreviation; pct., percentile.

et al. (2019) for Vietnam.

Next to analyze is how external benefits change across firm sizes. Column (3) shows the estimates from the baseline regression with the addition of firm employment size and its product terms with agglomeration proxies. Of three terms, only the one of employment scale and diversity is statistically significant, with the coefficient of − 0.00003 at 95% level of confidence, indicating the superiority of smaller firms in more diverse industrial regions. These results are consistent with the findings of Rosenthal and Strange (2010) that firms benefit more from agglomeration if their size is small. The disparity is illustrated in Graph B of Fig. 2, which shows productivity gains from diversity between firms of the 10^{th} and 90^{th} percentiles of the size distribution. The upward-sloping line with slightly higher gradient belongs to small-scale firms, reflecting the negative estimate of the product term. This proves Hypothesis 3, though the productivity differential is modest and only Jacobs externalities matter.

Finally, the results in column (4) exhibit the heterogeneity in agglomeration influences over different groups of firm age. The negative and significant coefficients of interaction terms imply at the 99% confidence level that younger firms derive more benefits from both density and specialization than older firms do. Intuitively, adding one year old to a firm's age reduces its external benefits by 0.8 % and 0.5 % if local employment density and location quotient plus one double respectively. Regarding density, graph C of Fig. 2 confirms the marginal cost of aging with the upward line that is steeper for younger firms. It is noteworthy from the graph that, despite the higher marginal external benefits, immature firms remain overwhelmed by their local mature fellows' performance in less dense regions. As regards regional specialization, graph D of Fig. 2 continues to indicate the productivity advantage of younger firms with an upward-sloping line. Somewhat surprisingly, the slope turns into downward for older firms, meaning that they suffer from diseconomies as their industries become more and more specialized in the region. These opposite impacts on firms in the two age groups might explain why specialization appears to have no effect on a generic firm. To summarize, the results support Hypothesis 4, and are in line with the findings of McCann and Folta (2011) and Rigby and Brown (2015). Interestingly, smaller firms are found to be better off in a more diverse environment, while younger firms enjoy that premium in a denser or more specialized environment. This divergence is expected due to the overlap between small firms and young ones being just partial.

7. Robustness check

Table 3 presents the regression results conducted to assess the robustness of estimated results against various samples. Columns (1) – (4) show that agglomeration benefits are highly robust to the alternatives of productivity. It is worth pointing out that the estimators' magnitude of all three agglomeration proxies yielded from the specification using OLS TFP are lower compared to the rest of TFP measures. The estimate of specialization using OLS TFP even displays the negative sign. These are not surprising given that OLS tends to overestimate input elasticities, thus underestimating productivity records. Another remark is that the application of LP-W in calculating TFP generates almost identical results of agglomeration estimates to the use of LP or OP, though the estimated production input elasticities are different among the three methods. Next, as comparing column (7) with (4), there are only small changes in the magnitude of three regional variables of interest, featuring the stability of results to different choices of industrial units. Similarly, the removal of multi-plant firms from the sample results in the only subtle difference between the estimates expressed in column (5) and the baseline results in column (4). Finally, column (6) shows a very minor shift in estimated parameters as observations recorded in the two biggest cities in Vietnam are dropped from the estimation, implying that the sorting issue should not be a concern in this study. In

Table 3
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
Unit	District & 3-digit VSIC						District & 2-digit VSIC
DEN	0.089***	0.133***	0.129***	0.129*;**	0.128***	0.144***	0.119***
	(0.029)	(0.028)	(0.028)	(0.028)	(0.030)	(0.029)	(0.027)
SPE	-0.041**	0.015	0.015	0.015	0.020	0.014	0.031
	(0.020)	(0.019)	(0.019)	(0.019)	(0.020)	(0.021)	(0.021)
DIV	0.077***	0.080***	0.077***	0.077***	0.078***	0.084***	0.063***
	(0.020)	(0.019)	(0.020)	(0.020)	(0.023)	(0.022)	(0.022)
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

Olley & Pakes, dices are estimated with the ordinary least square in column (1), and with the methods of Levinsohn & Petrin and Olley & Pakes (2003, 1996) in columns (2) and (3) respectively; in column (5), only observations from single-plant firms are employed; in 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 %

summary, the agglomeration estimates are demonstrated to be very robust across various productivity measures, industrial levels, and samples.

8. Conclusions

This paper contributes to the literature with the focus on verifying the existence of agglomeration economies, finding out which one between inter- and intra-industry relationships is dominant, and which firms reap most external benefits in Vietnam - an emerging and developing country. It provides discussions and empirical proof about the contrasting backgrounds between developing and developed countries in terms of geographical concentration, dominant industries, and firm characteristics. To yield estimators with the high accuracy, this study exploits the power of control function and instruments to obtain the consistent measure of productivity. Furthermore, it makes advantage of micro data and within-variation to construct explanatory variables, remove firms with less than 20 employees or with different main industry or location over time, deal with unobserved information and industrial shocks, and estimate the agglomeration effects. The results reveal the crucial role of urbanization and cross-industry relations in agglomeration externalities, and that firms gaining the most from agglomeration tend to be foreign-owned, small, or young.

The findings may give local firms and authorities guidance when making development policies. It supports the current policies of the Vietnamese government in encouraging urbanization as a productivity stimulant for local businesses. In addition, it offers a warning to domestic, large, and old firms about their underperformance in 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.

It is worth pointing out that the explanatory power of the model in this study is limited due to a low value of R-squared, though this is identical to many influential papers in literature. Afterward studies with richer data sets at firm-level or even establishment-level might exploit other time-variant factors that better explain time variations of TFP. According to the review on determinants of firm productivity by Syverson (2011), such potential factors may be: details about worker quality, technological level, research and development investment, and level of innovation in products.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to sincerely thank Professor Susanne Soretz of Greifswald University for her helpful comments and valuable remarks to this manuscript. We thank Georgios Tsiachtsiras as well as the participants at the “13th RGS Doctoral Conference in Economics” held at Bochum University, Bochum, Germany, in February 2020, who commented on the earlier draft of this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.asieco.2023.101620.

References

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.
Andersson, M., & Lööf, H. (2011). Agglomeration and productivity: evidence from firm-level data. *The Annals of Regional Science, 46*(3), 601–620.
Arrow, K. J. (1962). The economic implications of learning by doing. *The review of economic studies, 29*(3), 155–173.
Au, C. C., & Henderson, J. V. (2006). Are Chinese cities too small? *The Review of Economic Studies*.
World Bank. (2011). *Vietnam urbanization review: technical assistance report (English)*. Washington, DC: World Bank.
Badr, K., Rizk, R., & Zaki, C. (2019). Firm productivity and agglomeration economies: evidence from Egyptian data. *Applied Economics, 51*(51), 5528–5544.
Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372.
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.
Combes, P.P., Démurger, S., & Li, S. , (2013). Urbanisation and migration externalities in China.
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. *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,.
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.
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. (2016). Agglomeration effects in Colombia. *Journal of Regional Science, 56*(2), 210–238.

OECD. (2018). *OECD urban policy reviews: Viet Nam*. Paris: OECD Publishing,.
Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics* (vol. 4, pp. 2063–2117). Elsevier.
Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy, 100*(6), 1126–1152.
Gokan, T., Kuroiwa, I., & Nakajima, K. (2019). Agglomeration economies in Vietnam: A firm-level analysis. *Journal of Asian Economics, 62*, 52–64.
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.
Hanlon, W. W., & Miscio, A. (2017). Agglomeration: A long-run panel data approach. *Journal of Urban Economics, 99*, 1–14.
Head, K., & Mayer, T. (2004). The empirics of agglomeration and trade. In *Handbook of regional and urban economics* (vol. 4, pp. 2609–2669).
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.
Howard, E., C. Newman, & F. Tarp, (2012) Measuring Industry Agglomeration and Identifying the Driving Forces. WIDER Working Paper, No. 2012/84.
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*. New York, NY: Vintage.
Klein, A., & Crafts, N. (2020). Agglomeration externalities and productivity growth: US cities, 1880–1930. *The Economic History Review, 73*(1), 209–232.
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.
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.
Malesky, E., & London, J. (2014). The political economy of development in China and Vietnam. *Annual Review of Political Science, 17*, 395–419.
Marshall, A. (1920). *Principles of economics*. London: Macmillan.
Martin, P., Mayer, T., & Mayneris, F. (2011). Spatial concentration and plant-level productivity in France. *Journal of Urban Economics, 69*(2), 182–195.
McCann, B. T., & Folta, T. B. (2011). Performance differentials within geographical clusters. *Journal of Business Venturing, 26*(1), 104–123.
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.
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.
Newman, C., Rand, J., Talbot, T., & Tarp, F. (2015). Technology transfers, foreign investment and productivity spillovers. *European Economic Review, 76*, 168–187.
Nguyen, H. Q. (2017). Business reforms and total factor productivity in Vietnamese manufacturing. *Journal of Asian Economics, 51*, 33–42.
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.
Nichter, S., & Goldmark, L. (2009). Small firm growth in developing countries. *World Development, 37*(9), 1453–1464.
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.
Porter, M. E. (1990). The competitive advantage of nations. *Competitive Intelligence Review, 1*(1), 14-14.
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.
Rigby, D. L., & Brown, W. M. (2015). Who benefits from agglomeration? *Regional Studies, 49*(1), 28–43.
Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy, 98*(5), 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. (2010). Small establishments/big effects: Agglomeration, industrial organization and entrepreneurship. *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.
Syverson, C. (2011). What determines productivity? *Journal of Economic Literature, 49*(2), 326–365.
Takii, S. (2005). Productivity spillovers and characteristics of foreign multinational plants in Indonesian manufacturing 1990–1995. *Journal of Development Economics, 76*(2), 521–542.
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.
Waldkirch, A. , (2014). Foreign firms and productivity in developing countries. Department of Economics, Colby College. Available at [Accessed 2 July 2018].
Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters, 104*(3), 112–114.
World Bank, 2019, accessed on 19 August 2019, https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=VN.