Productivity and human capital agglomeration: evidence from indonesian cities

INSTITUTE OF DEVELOPING ECONOMIES

IDE Discussion Papers are preliminary materials circulated to stimulate discussions and critical comments

IDE DISCUSSION PAPER No. 849

Productivity and Human Capital Agglomeration: Evidence from Indonesian Cities

Yoshihiro HASHIGUCHI†
Takayuki HIGASHIKATA ‡

March, 2022

Abstract

Using Indonesian plant-level manufacturing data for 1996 and 2006, this study estimates the external benefits of human capital investment. The external benefits are identified from the relationship between plant-level total factor productivity (TFP) and geographical human capital agglomeration with controlling for workers' skill levels within a plant. The endogeneity problem in the human capital agglomeration is addressed by the instrumental variable (IV) method. We construct the IV by using the geographical distribution of European population in colonial Indonesia. Our IV estimates suggest that human capital agglomeration has a boosting effect on productivity, implying the existence of human capital externalities.

Keywords: Human capital externalities, plant-level data, Population census in colonial Indonesia
JEL classification: E24, J24, J30, L60, O40

† Institute of Developing Economies-JETRO. E-mail: yoshihiro_hashiguchi@ide.go.jp
‡ Institute of Developing Economies-JETRO. E-mail: takayuki_higashikata@ide.go.jp

The Institute of Developing Economies (IDE) is a semigovernmental, nonpartisan, nonprofit research institute, founded in 1958. The Institute merged with the Japan External Trade Organization (JETRO) on July 1, 1998. The Institute conducts basic and comprehensive studies on economic and related affairs in all developing countries and regions, including Asia, the Middle East, Africa, Latin America, Oceania, and Eastern Europe.

INSTITUTE OF DEVELOPING ECONOMIES (IDE), JETRO
3-2-2, WAKABA, MIHAMA-KU, CHIBA-SHI
CHIBA 261-8545, JAPAN

Productivity and Human Capital Agglomeration: Evidence from Indonesian Cities*

Yoshihiro HASHIGUCHI† Takayuki HIGASHIKATA‡

March 2022

Abstract

Using Indonesian plant-level manufacturing data for 1996 and 2006, this study estimates the external benefits of human capital investment. The external benefits are identified from the relationship between plant-level total factor productivity (TFP) and geographical human capital agglomeration with controlling for workers' skill levels within a plant. The endogeneity problem in the human capital agglomeration is addressed by the instrumental variable (IV) method. We construct the IV by using the geographical distribution of European population in colonial Indonesia. Our IV estimates suggest that human capital agglomeration has a boosting effect on productivity, implying the existence of human capital externalities.

Keywords: Human capital externalities, plant-level data, Population census in colonial Indonesia

JEL classification: E24, J24, J30, L60, O40

1 Introduction

Does the presence of workers with higher human capital make other workers more productive? If true, to what extent do the "external" benefits have an effect on a macro economy? The externalities associated with human capital investment have received remarkable attention as a key element in explaining cross-country differences in economic development (Lucas 1988; Romer 1990). In addition, since the degree of the externalities is related to the efficiency of public investment in education, this issue is crucial not only for academics, but also for policy makers.

*This work was supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research (C)) Grant Number 19K01628 and the project "The Impact of Urbanization in Indonesia: Analysis of Firm Productivity and Labor Migration" at the Institute of Developing Economies (IDE-JETRO).

†Institute of Developing Economies-JETRO. E-mail: yoshihiro_hashiguchi@ide.go.jp

‡Institute of Developing Economies-JETRO. E-mail: takayuki_higashikata@ide.go.jp

It is argued theoretically that human capital externalities occur through the sharing of workers' knowledge and skills via social interaction (Lucas 1988; Acemoglu 1996; Duranton and Puga 2004). Many scholars have attempted to find empirical evidence of the externalities using individual wage or firm-level productivity (Rauch 1993; Acemoglu and Angrist 2000; Conley et al 2003; Moretti 2004a; 2004b; Ciccone and Peri 2006; Rosenthal and Strange 2008; Abel et al. 2012; Liu 2013). While most studies have found evidence of positive human capital externalities in developed countries, mainly in the United States and Europe, Ciccone and Peri (2006) found no evidence of such externalities. In addition, only a few studies, including Conley et al. (2003) and Liu (2013), have examined this issue for developing economies. The magnitude of human capital externalities remains controversial and, hence, more empirical studies are required, particularly for developing economies.

In this study, we use Indonesian plant-level manufacturing data for 1996 and 2006 to examine the existence of human capital externalities. Following Moretti's (2004b) approach to the estimation of human capital externalities, we focus on the impact of human capital agglomeration on the plant-level productivity. We hypothesize that plants located in regions with large human capital accumulation are productive more than otherwise similar plants located in regions with smaller human capital accumulation. Our empirical results indicate that human capital agglomeration have a positive effect on productivity for both 1996 and 2006 and the magnitude of the productivity boosting effect tends to increase over the ten-year period for 1996–2006. These results suggest that there exists human capital externalities in the manufacturing industry in Indonesian cities, and its importance has been increasing over time.

Compared to the existing studies, our study has several features. First, to the best of our knowledge, our work is the first to investigate this issue using Indonesian plant-level manufacturing data. The data is based on the annual survey of medium and large manufacturing establishments (*Survei Industri Besar/Sedang*: IBS) conducted by the Indonesian Central Bureau of Statistics (*Badan Pusat Statistik*: BPS). Although several studies have examined the determinants of productivity in Indonesian manufacturing using the IBS plant-level database (Todo and Miyamoto 2006; Suyanto and Bloch 2009; Hallward-Driemeirer and Rijkers 2013; Widodo et al. 2014), few studies explored the relationship between human capital and productivity. Second, we identify cities (urban areas) following the OECD's (2012) methodology of defining urban areas, which is comparable internationally.[1] As Rauch (1993) argued, urban areas within a country are the most appropriate regional units to use to identify human capital externalities, because these areas are presumably at the same stage of economic development.[2] However, the definition of an urban area differs among countries, which means the choice of definition can influence estimations of human capital externalities significantly. For this reason, we refer to the OECD's (2012) methodology of defining urban areas, and apply it to community-level (the lowest administrative unit) map information and population census data. [3]

The third feature of this study, we introduce plant's own human capital accumulation in the regression to control for the internal effect of human capital on the plant-level productiv-

[1] In this paper, we use the terms "cities" and "urban areas" interchangeably.

[2] Regions at a higher stage of economic development are likely to have a larger and more advanced physical capital stock, which may also be factors that increase wages and productivity. This is why it would be difficult to identify the effects of human capital externalities using regions at different stages of economic development.

[3] Indonesia's administrative divisions are classified as follows: province (*provinsi*), district (*kabupaten/kota*), sub-district (*kecamatan*), and community (*desa/kelurahan*).

ity. This variable can contribute to identifying the external effect of human capital agglomeration. Fourth, the magnitude of human capital agglomeration faced by plants is measured by the number of educated workers in the community-level region (*desa* or *kelurahan*) which is the most detailed administrative division, consisting of more than 78,000 regions. Using the IBS plant-level data and the community-level digital map information (*shapefile*), we construct a distance based index of human capital agglomeration. This index reflects human capital accumulation not only in the community where the plant locates, but also in the other neighboring communities. Compared to the existing studies using larger level of administrative divisions, our approach is more natural to capture the degree of local agglomeration.[4] Finally, to address the endogeneity bias in the impact of human capital agglomeration on plant-level productivity, we exploit the geographical distribution of European population in colonial Indonesia. This historical data, obtained from the population census of 1930 in Netherlands India, includes a paper based map which describes the size and distribution of European population in 1930. We digitize the map and use the historical distribution of European population as an instrument variable for human capital agglomeration. Our identifying assumption is that on the one hand, the past economic activity by European has influences only on a formulation process of human capital agglomeration, but on the other hand the past existence of European is not likely to be correlated with current differences in firm-level productivity. Constructing this historical instrumental variable is also the remarkable feature of this study.

The remainder of the paper is structured as follows. Section 2 describes our empirical model and estimation strategy. Section 3 presents the data sources, and Section 4 reports our empirical results. Lastly, Section 5 concludes the paper.

2 Model and Econometric Strategy

In accordance with Moretti (2004b) and other previous studies, the following equation is used to investigate the extent of human capital externalities:

$$\begin{aligned}\log TFP_{it} = \beta_h \log H_{it} + \beta_e EduYrs_{it} + \beta_e OpYrs_{it} + \beta_p \log CityPop_{rt} + \beta_f Gov_{it} + \beta_f For_{it} \\ + \mathbf{S}'_{it}\boldsymbol{\gamma} + IndustryDm + ProvinceDm + \epsilon_{it}, \quad t \in \{1996, 2006\}\end{aligned} \tag{1}$$

where i and t denote plant and year. The estimation method of total factor productivity (TFP) will be described later. H_{it} is the degree of human capital agglomeration faced by plant i. The coefficient β_h of H_{it} indicates the magnitude of human capital externalities, which is the key parameter of this study. The H_{it} is defined by

$$H_{it} = \sum_{l}^{M} \frac{h_{lt}}{\exp\{\theta d_{cl}\}}, \quad c, l \in M \tag{2}$$

where the M is the number of communities, the subscript (c) denotes the community where plant i locates, and d_{cl} denotes the great circle distance between the community c and l ($d_{cc} = 0$). The

[4] Widodo et al. (2014) uses the province-level regions to capture the degree of local agglomeration.

h_{lt} is the degree of human capital accumulation in community l. We use the following two variables for h_{lt}:

$$h_{lt} \equiv EduDens_{lt} \quad \text{or} \quad EduShare_{lt}, \tag{3}$$

where $EduDens_{lt}$ is the number of educated workers per area (i.e., density) in community $l \in M$, and $EduShare_{lt}$ is the share of educated workers among all manufacturing workers in community $l \in M$. Educated workers are defined as workers who have at least completed senior high school education. For simple notation, we define the human capital agglomeration as follows:

$$\begin{aligned} H^D_{it} &= \sum_l^M \frac{EduDens_{lt}}{\exp\{\theta d_{cl}\}}, \quad c, l \in M \\ H^S_{it} &= \sum_l^M \frac{EduShare_{lt}}{\exp\{\theta d_{cl}\}}, \quad c, l \in M. \end{aligned} \tag{4}$$

The θ is the decay parameter which controls for the rate of decline of distance weight $\exp\{\theta d_{cl}\}$. We choose $\theta = 0.2$ as the bench mark and examine the robustness of estimation results by the use of other values ($\theta = \{0.10, 0.15, 0.20, 0.25, 0.30\}$). Figure 1 demonstrates the relationship between the weight and distance with different values of θ. In the case of $\theta = 0.2$, the effect of human capital accumulation 10 kilometers ahead is reduced to almost 20%.

[– Figure 1 –]

$EduYrs_{it}$ is the average years of education of workers, given by

$$EduYrs_{it} = 0\frac{L^N_{it}}{L_{it}} + 6\frac{L^{PR}_{it}}{L_{it}} + 9\frac{L^{JH}_{it}}{L_{it}} + 12\frac{L^{SH}_{it}}{L_{it}} + 14\frac{L^{D3}_{it}}{L_{it}} + 16\frac{L^{BA}_{it}}{L_{it}} + 19\frac{L^{MD}_{it}}{L_{it}}, \tag{5}$$

where L_{it} is total number of workers, and the other terms denote the number of workers who have not finished primary school (L^N_{it}), and those who have completed primary school (L^{PR}_{it}), junior high school (L^{JH}_{it}), senior high school (L^{SH}_{it}), an associate (*Diploma 3* or *Profesional ahli madya*) (L^{D3}_{it}), a bachelor (L^{BA}_{it}), and a masters/doctoral education (L^{MD}_{it}).

The variable $OpYrs_{it}$ is the number of years in operation of plant i. The variable $CityPop_{rt}$ is population size in city r where plant i locates. The city consists of the community-level regions. The definition of a city is described in the next section. The Gov_{it} and For_{it} are the shares of government and foreign capital, respectively. The vector of $\mathbf{S}_{it}$ denotes the set of the plant size quintile dummies. Plants are divided by five groups identifying the quintile of plant output at period t. Specifically, the vector $\mathbf{S}_{it}$ is defined as

$$\mathbf{S}_{it} = \left[Sizedm2_{it} \quad Sizedm3_{it} \quad Sizedm4_{it} \quad Sizedm5_{it} \right]',$$

where $Sizedm2_{it}$–$Sizedm5_{it}$ indicate the firm size dummies from second to fifth quintile ranges. The reference plant size distribution is evaluated by two-digit manufacturing sector and year. Finally, *IndustryDm* and *ProviceDm* denote the set of two-digit industrial dummies and the province level regional dummies, respectively. The numbers of two-digit industries and provinces are 23 and 34, respectively.

The key parameter of Equation (1) is β_h, which captures the degree of human capital externalities. To estimate this parameter, this study applies cross-sectional estimation using Indonesian plant-level manufacturing data for 1996 and 2006. To identify the effect of human capital externalities, we need to deal with the endogeneity of human capital agglomeration. Previous studies have addressed potential correlation between economic agglomeration and the error term because regression analysis may suffer from endogeneity bias resulting from omitted variables and reverse causality between productivity and agglomeration. For example, unobserved local endowments (e.g., local climate, social infrastructure, and natural resources) may increase firm's productivity as well as the degree of human capital agglomeration, leading to omitted-variables bias. Also, high productive firms may self-select to locate their production base in agglomerated area, causing reverse causality.

To deal with this problem, we use an instrumental variable approach. Our instrument is based on the geographical distribution of European population in colonial Indonesia in 1930. European used to live and have a business in colonial Indonesia. However, after the end of Dutch's colonial rule, the influence of European was swept away from Indonesia. Our identifying assumption is that the past economic activity by European has persistent influences only on the preferences of educated workers about the location in which they seek better jobs. That is, the past existence of European should affect a formulation process of human capital agglomeration, thereby producing a positive correlation with a spatial concentration of human capital in the years 1996 and 2006. On the other hand, the past existence of European is not likely to be correlated with current differences in firm-level productivity. The past economic activity by European is likely to correlate with local advantages across regions, which may subsequently account for regional differences in plant-level performance for 1996 and 2006. But we control for such indirect influences through natural advantages by introducing the city-level population and province-level fixed effects. In sum, the past existence of European is likely to be a good predictor of the current human capital accumulation, but is not directory related to current economic performance. These assumptions to use the past data for eligible instrumental variables are similar to the empirical approach in Ciccone and Hall (1996).

The degree of the past European population agglomeration in 1930 is measured by

$$EuAgl_{i,1930} = \sum_{m} \frac{EuPop_{m,1930}}{\exp\{\theta d_{cm}\}}, \quad c \in M \tag{6}$$

where $EuPop_{mt}$ is the size of European population in 1930 at the location m, and d_{cm} is the distance between plant i's location (c) and the location (m) where European used to live. The size and location of European population will be described in the next section.

Finally, the plant-level TFP is estimated by using the method of Ackerberg et al. (2015), based on the value-added based Trans-log production function. Since their methodology requires firm-level panel data, we exploit manufacturing plant-level panel data from 1996 to 2006 to estimate the plant-level productivity.

3 Data description

Our main databases are the 1996 and 2006 plant-level data on the Indonesian manufacturing industry. The plant-level data are obtained from the annual survey of medium and large manufacturing establishments (IBS) conducted by the BPS. The IBS covers all manufacturing plants with 20 or more employees. Because the IBS is published annually, the plant-level TFP is estimated by using the annual unbalanced panel data from 1996 to 2006. However, the regression analysis for human capital externalities uses only the 1996 and 2006 data because they include information on workers' educational attainment which is required to control for internal human capital effects within a plant.

In order to define urban areas (cities, in our empirical specification), we use the community-level population and area data. The population data are obtained from the 2000 population census (*Sensus Penduduk* 2000). The 1999 population data of the Village Potential Data Collection (PODES) database are also used if the 2000 population census data are not available in some places.[5] The area of each community, needed to measure population density, is calculated using the community-level map information of the 2012 *Peta Digital* database in the *shapefile* format. The Peta Digital consists of 506 shapefiles, and each shapefile basically has one district-level map including community-level polygon data. Because almost all shapefiles in the Peta Digital use the format known as WGS84 geographic coordinate system, we covert the coordinates into the Universal Transverse Mercator (UTM) zone projection for Indonesia (DGN95-UTM) to calculate the community-level areas as precisely as possible.[6] The 2012 Peta Digital map data are merged with the 1999 and 2000 population data by using information about historical transition of administrative communities from 1998 to 2013. According to the Peta Digital, the number of communities is 78,934 in 2012. 76,565 (97%) of them are matched with the 1999 and 2000 population database, and thus the matched samples are used to define urban areas.[7]

Referring to the definition of OECD (2012), we compute urban areas as follows: (1) we select communities with a population greater than 1,500 per square km; (2) cluster these communities as an area if they have common borders; and (3) define the computed cluster with a population size greater than 100 thousand as an urban area. The Peta Digital map data is used to identify communities with common borders. As shown in Table 2, the total number of urban areas is 72 in 1996 and 73 in 2006, and the total urban area is 145,345 square km, which covers approximately 7.6% of the total land area in Indonesia. The total urban population is about 62,872 thousand, which is approximately 30.5% of total population in 2000.

To construct the variable of human capital agglomeration, we use the community-level region's area and number of workers with educational background. The IBS database in 1996 and 2006 includes the information about the plant-level number of workers and the worker's educational record. Aggregating them according to the community-level region, we construct the community-level number of educated workers. The geographic area of each community, needed to measure the number of workers area (i.e., worker density), is calculated by using the 2012 *Peta Digital* database. This map data are merged with the 1996 and 2006 plant-level data (the IBS database) by using information about historical transition of administrative commu-

[5]We use the 1999 PODES population data for 4,117 communities. Of these, 3,146 belong to Ache province, and the remaining communities are located in Papua and West Papua provinces (778 units) and North Maluku province (108 units), among others.

[6]There are 15 UTM zones commonly used for Indonesia. The WGS84 coordinate for each shapefile is converted into the coordinate of a UTM zone in which the center point of the shapefile is located.

[7]The number of unmatched communities is 2,369, and 38% of them belongs to West Sumatra province, followed by North Sumatra province (8.8%). These communities are outside the scope of this study.

nities. The number of communities observed in the plant-level database is 6,185 in 1996 and 7,116 in 2006.

[– Table 1 –]

The geographical distribution of European population in colonial Indonesia is obtained from the Census of 1930 in Netherlands India.[8] This census consists of eight volumes, of which the sixth volume titled "Europeans in Netherlands Indies" provides the historical map of European population. We convert the paper-based map into the digital map (i.e., shapefile) and merge it with the Peta Digital database. The historical map shows the location points where European used to live in 1930. These points describe European population size, which is classified by five towns and three suburb categories. Table 1 shows the categories of European population size. Although some categories overlap with others, we give the class values for each categories (See Table 1, column 2). Using these class values, we quantify the size and distribution of European population in colonial Indonesia.

[– Tables 2 and 3 –]

Table 2 reports the summary statistics of manufacturing plant-level data in 1996 and 2006. The original databases of IBS 1996 and 2006 have 22,997 and 29,468 plants. After dropping outliers, we have 21,055 plants in 1996, and 21,923 plants in 2006.[9] Of which 13,977 and 13,361 plants are located in urban areas in 1996 and 2006, respectively, and the sample of these plants is used for our empirical analysis. For the estimation of plant-level TFP, we need value-added, fixed capital stock, and the number of labor. The value-added data is deflated by GDP deflator, normalized at the year 2000.[10] The fixed capital stock is calculated by the sum of real investment values with the depreciation rate at 0.05. The real investment values are calculated by deflating nominal investment values with the GDP deflator. The number of years in operation is calculated as the observation year minus the start year of operation for each plant. However, we find cases in which the number of years in operation, reported in IBS database, varies among 1996 and 2006 within the same plant. To cope with this situation, we choose the start year of plant's operation which is most frequently reported in the IBS panel data for 1994–2000 and 2006.[11] Table 3 shows the averages of plant-level productivity and human capital agglomeration by the number of years in operation. It is found that younger plants tend to have higher productivity.

[8]Dutch East Indies. Departement van Landbouw, Nijverheid en Handel. (1933-1936) *Volkstelling 1930: Census of 1930 in Netherlands India*. Batavia : Landsdrukkerij.

[9]We remove the following plants: 1) plants with non-positive value of number of workers (L_{it}), value-added (Y_{it}), output values, fixed capital stock (K_{it}); or 2) plants whose Y_{it}/L_{it} and Y_{it}/K_{it} in 2006 is more than 1000 times or less than 0.001 the value in 1996.

[10]We construct the district-level GDP deflator based on the 1996 administrative divisions, where the number of districts is 294.

[11]The IBS plant-level data set for 1994–2000 and 2006 has information about years in operation

4 Estimation results

[– Table 4 –]

Before presenting the estimation results of Equation (1), we report the single regression results between $\log TFP$ and the log of human capital agglomeration ($\log H^D$ and $\log H^S$). As Table 4 shows, the plant-level TFP is positively correlated with the human capital agglomeration across alternative specification.

Table 5 and 6 reports the estimation results of the baseline specification (1), using the cross-section data of 1996 and 2006, respectively. For the first-stage regression of the IV estimation (two-stage least squares: 2SLS), the coefficients of our instrument ($\log EuAgl$) are positive and highly significant, indicating that our instrument is not weak. The number of years in operation (*OpYrs*) has a negative and significant effect on productivity for both OLS and IV estimation in 1996 and 2006, indicating that younger plants tend to be more productive than older plants. The coefficients of average years of education (*EduYrs*) are positive and significant, with the except of 2SLS estimates in 1996, consistent with our prediction. The coefficients of city population ($\log CityPop$) are not significant. The estimates of plant size dummy variables show that productivity in large plants tend to be higher than those in small plants.

The key parameters in this study are the coefficients of $\log H^D$ and $\log H^S$, which indicate the magnitude of human capital externalities. Contrary to our expectation, the OLS estimates of these coefficients shows negative in both 1996 and 2006 and negative and significant in 2006. However, the IV estimates of them are positive and significant in both 1996 and 2006, implying a possibility that the OLS estimates have a positive endogenous bias. Our IV estimates imply that a doubling of human capital agglomeration increases plat-level productivity by around 4.7–5.6% in 1996 and 6.1–8.1% in 2006. These results suggest that there exists human capital externalities for the manufacturing sector in Indonesian cities, and that the magnitude tend to increase over the 10 year-period from 1996 to 2006.

[– Tables 7 and 8–]

To check the robustness of our results, we examine the effect on the log of labor productivity, instead of log TFP. As shown in Tables 7 and 8, the IV estimates of $\log H^D$ and $\log H^S$ for 1996 are 0.029 and 0.035, and those for 2006 are 0.028 and 0.037, respectively. These estimates are significant, indicating that human capital agglomeration has a positive effect on labor productivity, while these coefficients does not differ significantly in the magnitude between 1996 and 2006.

[– Table 9 –]

We check the robustness of our results against different parameter values of θ which is a decay parameter to calculate human capital agglomeration (See Equation (4)). While we choose $\theta = 0.2$ as the benchmark, we examine other values of $\theta = \{0.10, 0.15, 0.20, 0.25, 0.30\}$. Table 9 shows the coefficients of human capital agglomeration estimated by 2SLS estimation. As is shown in this table, our benchmark result is not much different from the results using alternative values of θ.

5 Concluding remarks

Despite its importance to economic development, few studies have examined empirical evidence of human capital externalities for the Indonesian economy. This study analyzed the magnitude of human capital externalities using Indonesian manufacturing plant-level data for 1996 and 2006. The externalities were identified from the relationship between plant-level total factor productivity (TFP) and geographical human capital agglomeration with controlling for workers' skill levels within a plant. The endogeneity problem in the human capital agglomeration was addressed by the instrumental variable method. We constructed the instrument variable by using the geographical distribution of European population in colonial Indonesia.

We considered that plants located in cities with large human capital accumulation are productive more than otherwise similar plants located in regions with smaller human capital accumulation. The magnitude of human capital agglomeration is measured by educated workers who have at least completed senior high school education. Controlling for plant's own human capital accumulation, years in operation, plant size and city population size, our results of the instrumental variable estimation showed that human capital agglomeration has a significant positive impact on plant-level productivity and that a doubling of human capital agglomeration increases plat-level productivity by around 4.7–5.6% in 1996 and 6.1–8.1% in 2006. These results suggest that there exits human capital externalities in the manufacturing sector in Indonesian cities, and that the magnitude tends to increase during the period 1996–2006.

Note that these findings are preliminary, and our empirical analysis has at least two problems that need to be solved. First, since our index of human capital agglomeration was based on the annual survey of medium and large manufacturing establishments, human capital stock in non-manufacturing sector and of unemployed workers was not considered for the community-level human capital accumulation. Second, we did not examine the non-linear relationship between productivity and human capital agglomeration. Using more flexible functional forms for estimation can bring about different results. We need to address these remaining issues.

References

Abel, J. R., Dey, I., Gabe, T. M. (2012) Productivity and the density of human capital. *Journal of Regional Science* 52(4), 562–586.

Acemoglu ,D. (1996) A microfoundation for social increasing returns in human capital accumulation. *Quarterly Journal of Economics* 111(3), 779–804.

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

Ackerberg, D. A., Caves, K., Frazer, G. (2015) Identification properties of recent production function estimators. *Econometrica* 83(6), 2411–2451.

Ciccone, A., Giovanni, P. (2006) Identifying human-capital externalities: Theory with applications. *Review of Economic Studies* 73(2), 381–412.

Ciccone, A., Hall, R. E. (1996) Productivity and the density of economic activity. *American Economic Review*, 86(1), 54–70.

Combes, P., Duranton, G., Gobillon, L. (2008) Spatial wage disparities: sorting matters! *Journal of Urban Economics* 63(2), 723–742.

Combes, P., Gobillon, L. (2015) The empirics of agglomeration economies. In: G. Duranton, J. V. Henderson and W. C. Strange, Eds, *Handbook of Regional and Urban Economics*, Elsevier, Vol. 5, 247–348.

Conley, T. G., Flyer, F., Tsiang, G. R. (2003) Spillovers from local market human capital and the spatial distribution of productivity in Malaysia. *Advances in Economic Analysis & Policy* 3(1),

Duranton, G., Puga, D. (2004) Micro-foundations of urban agglomeration economies. In: Henderson JV, Thisse JF (eds) *Handbook of regional and urban economics*, vol. 4. North-Holland, New York.

Gandhi, A., Navarro, S., Rivers, D. (2016) On the identification of production functions: How heterogeneous is productivity?

Gobillon, L. (2004) The estimation of cluster effects in linear panel models. Processed, INED.

Hallward-Driemeier, M., Rijkers, Bob. (2013) Do crises catalyze creative destruction? Firm-level evidence from Indonesia. *Review of Economics and Statistics* 95(5), 1788–1810.

Liu, Z. (2013) Human capital externalities in cities: Evidence from Chinese manufacturing firms. *Journal of Economic Geography* ibt024.

Lucas, R. E. (1988) On the mechanics of economic development. *Journal of Monetary Economics* 22(1), 3–42.

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

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

Moulton, B. R. (1990) An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics* 72(2), 334–338.

OECD (2012) *Redefining "Urban": A new way to measure metropolitan areas*. OECD Publishing, Paris.

Rauch, J. E. (1993) Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics* 34(3), 380–400.

Romer, P. R. (1990) Endogenous technological change. *Journal of Political Economy* 98(5) part 2, S71–S102.

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

Suyanto, R. A. S., Bloch, H. (2009) Does foreign direct investment lead to productivity spillovers? Firm level evidence from Indonesia. *World Development* 37(2), 1861–1876.

Todo, Y., Miyamoto, K. (2006) Knowledge spillovers from foreign direct investment and the role of local R&D activities. *Economic Development and Cultural Change* 55(1), 173–200.

Widodo, W., Salim, R., Bloch, H. (2014) Agglomeration economies and productivity growth in manufacturing industry: Empirical evidence from Indonesia. *Economic Record* 90(s1), 41–58.

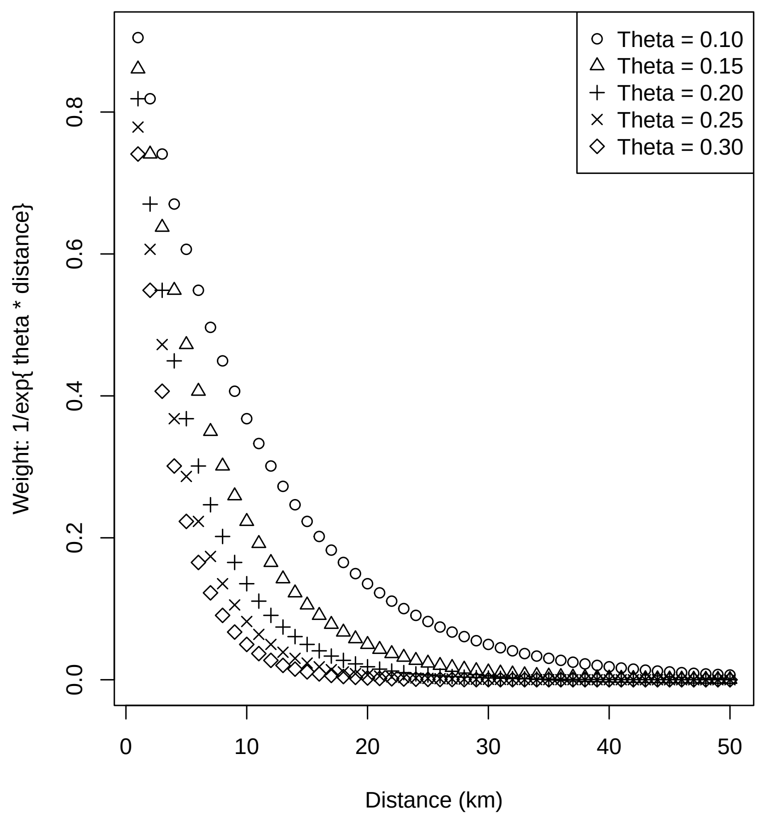

Figure 1: Plots of $\exp\{-\theta d_{cl}\}$ with different values of θ

Notes: The vertical axis is the distance weight $\exp\{\theta d_{cl}\}$ used in Equation (6). The horizontal axis is distance measured by kilometer.

Table 1: Categories of European population size

Categories of European population size	Class values
Town: Population size is less than 50	47.5
Town: Population size is between 50 and 100	75
Town: Population size is less than 500	475
Town: Population size is between 500 and 1000	750
Town: Population size is over 1000	On the map*
Suburb: Population size is between 45 and 50	47.5
Suburb: Population size is between 95 and 100	97.5
Suburb: Population size is between 495 and 500	497.5

* The number of European population (over 1000) is shown on the map.

Table 2: Summary statistics

Sample size				
The original number of plants	22,997 in 1996		29,468 in 2006	
Without outliers (1996)	21,055 in 1996		21,923 in 2006	
Surviving plants during 1996–2006	9,867			
Exit plants during 1996–2006	11,188			
Entry plants during 1996–2006	12,059			
Number of urban areas	74			
where sample plants are located	72 in 1996		73 in 2006	
Number of plants in urban area (without outliers)	13,977 in 1996		13,361 in 2006	
1996	Mean	Std. Dev.	Min	Max
Log of TFP	0.105	0.949	−6.083	6.007
Log of labor productivity	3.031	0.463	1.484	5.365
Log of human capital agglomeration (H)				
Log of educated worker density	1.102	5.658	−88.513	7.484
Log of educated worker share	−5.332	4.994	−92.005	−0.652
Worker's average years of education	8.324	2.409	0.000	15.129
Log of city population	15.042	1.533	11.555	16.742
Capital share of government	1.444	11.597	0.000	100.000
Capital share of foreign	4.153	17.420	0.000	100.000
2006	Mean	Std. Dev.	Min	Max
Log of TFP	0.102	1.139	−6.404	8.753
Log of labor productivity	3.039	0.490	1.541	6.281
Log of human capital agglomeration (H)				
Log of educated worker density	1.808	4.671	−71.511	8.265
Log of educated worker share	−4.378	3.961	−74.511	−0.112
Worker's average years of education	9.459	2.439	0.000	15.900
Log of city population	15.016	1.469	11.555	16.742
Capital share of government	2.180	14.341	0.000	100.000
Capital share of foreign	6.093	22.502	0.000	100.000

The definition of human capital agglomeration (H^D and H^S) is described in Equation (4).

Table 3: Summary statistics by years in operation

1996

Years in operation	Number of plants	Mean log TFP	Mean log LP	Mean log H^D	Mean log H^S
[0, 5]	4,320	0.162	3.069	1.418	−5.013
[6, 10]	3,230	0.142	3.024	1.236	−5.273
[11, 15]	2,258	0.065	3.035	0.852	−5.602
[16, 20]	1,663	0.073	3.010	1.002	−5.489
21+	2,458	0.016	2.982	0.634	−5.642
NA	48	0.102	3.014	2.900	−4.103

2006

Years in operation	Number of plants	Mean log TFP	Mean log LP	Mean log H^D	Mean log H^S
[0, 5]	2,159	0.272	3.143	0.945	−4.950
[6, 10]	2,150	0.114	3.084	0.906	−5.035
[11, 15]	2,708	0.133	3.029	2.241	−4.066
[16, 20]	2,162	0.073	3.013	2.244	−4.135
21+	4,182	0.004	2.981	2.211	−4.074

The definition of human capital agglomeration (H^D and H^S) is described in Equation (4). "NA" indicates plants with data not available or minus years in operation.

Table 4: Single regression

	Estimate	Std. Error
Intercept	0.0750**	0.0081
log H^D in 1996	0.0274**	0.0014
Intercept	0.0405**	0.0105
log H^D	0.0341**	0.0021
Intercept	0.2688**	0.0116
log H^S in 1996	0.0307**	0.0016
Intercept	0.2768**	0.0146
log H^S in 2006	0.0399**	0.0025

This table shows the results of single regression between log TFP and log of human capital agglomeration. log H^D is the log of human capital agglomeraton based on educated worker density, and log H^S is the log of human capital agglomeraton based on educated worker share. The definition of these variables is described in Equation (4). The asterisks * and ** denote 5%, and 1% significance levels.

Table 5: Estimation results: 1996

	Manufacturing plant-level data in 1996			
	1: OLS	2: OLS	3: 2SLS	4: 2SLS
$\log H^D$	−0.0023		0.0474**	
	(0.0019)		(0.0178)	
$\log H^S$		0.0007		0.0564**
		(0.0021)		(0.0203)
OpYrs	−0.0037***	−0.0037***	−0.0040***	−0.0039***
	(0.0013)	(0.0013)	(0.0014)	(0.0013)
EduYrs	0.0167**	0.0156**	−0.0012	−0.0022
	(0.0061)	(0.0062)	(0.0072)	(0.0077)
log *CityPop*	0.0225	0.0196	−0.0282	−0.0265
	(0.0203)	(0.0199)	(0.0294)	(0.0280)
Gov	0.0007	0.0007	0.0001	−0.0001
	(0.0009)	(0.0009)	(0.0009)	(0.0009)
For	0.0019*	0.0019*	0.0020*	0.0021*
	(0.0010)	(0.0010)	(0.0011)	(0.0010)
Sizedm2	0.2881***	0.2860***	0.2527***	0.2470***
	(0.0182)	(0.0181)	(0.0175)	(0.0181)
Sizedm3	0.3949***	0.3916***	0.3390***	0.3348***
	(0.0257)	(0.0262)	(0.0436)	(0.0441)
Sizedm4	0.4854***	0.4797***	0.3862***	0.3856***
	(0.0608)	(0.0612)	(0.0759)	(0.0750)
Sizedm5	0.6976***	0.6897***	0.5562***	0.5770***
	(0.0773)	(0.0783)	(0.0825)	(0.0788)
2-digit industrial dummies	Yes	Yes	Yes	Yes
Province dummies	Yes	Yes	Yes	Yes
First stage regression				
log *EuAgl*			0.6414***	0.5388***
			(0.0285)	(0.0262)
N	13,929	13,929	13,929	13,929

Notes: The explained variable is the log of plant-level TFP. $\log H^D$ is the log of human capital agglomeraton based on educated worker density, and $\log H^S$ is the log of human capital agglomeraton based on educated worker share. The definition of these variables is described in Equation (4). $OpYrs_{it}$ is the number of years in operation. $EduYrs_{it}$ is the average years of education of workers. $CityPop_{it}$ is city-level poplation. For_{it} is the share of foreign capital. $Sizedm2_{it}$–$Sizedm5_{it}$ are the firm size dummies. log *EuAgl* is European population size in colonial Indonesia which is the instrument variable for the human capital agglomeration. The asterisks *, **, and *** denote 10%, 5%, and 1% significance levels. Figures in parentheses are standard errors, clustered at the 2-digit industrial classification.

Table 6: Estimation results: 2006

	Manufacturing plant-level data in 2006			
	1: OLS	2: OLS	3: 2SLS	4: 2SLS
$\log H^D$	-0.0197^{***}		0.0610^{**}	
	(0.0046)		(0.0246)	
$\log H^S$		-0.0155^{***}		0.0807^{**}
		(0.0047)		(0.0327)
OpYrs	-0.0082^{***}	-0.0084^{***}	-0.0099^{***}	-0.0099^{***}
	(0.0009)	(0.0010)	(0.0011)	(0.0011)
EduYrs	0.0328^{***}	0.0306^{***}	0.0003	−0.0049
	(0.0082)	(0.0079)	(0.0081)	(0.0087)
log *CityPop*	0.0359	0.0306	−0.0247	−0.0286
	(0.0358)	(0.0363)	(0.0472)	(0.0466)
Gov	0.0026^{***}	0.0026^{***}	0.0023^{**}	0.0023^{**}
	(0.0009)	(0.0009)	(0.0010)	(0.0010)
For	−0.0009	−0.0009	−0.0011	−0.0009
	(0.0006)	(0.0006)	(0.0007)	(0.0007)
Sizedm2	0.4993^{***}	0.4956^{***}	0.4605^{***}	0.4597^{***}
	(0.0377)	(0.0382)	(0.0431)	(0.0435)
Sizedm3	0.7361^{***}	0.7325^{***}	0.6951^{***}	0.6929^{***}
	(0.0610)	(0.0609)	(0.0623)	(0.0632)
Sizedm4	0.8556^{***}	0.8469^{***}	0.7808^{***}	0.7872^{***}
	(0.0725)	(0.0741)	(0.0949)	(0.0957)
Sizedm5	1.0181^{***}	0.9981^{***}	0.8639^{***}	0.8886^{***}
	(0.1319)	(0.1331)	(0.1434)	(0.1464)
2-digit industrial dummies	Yes	Yes	Yes	Yes
Province dummies	Yes	Yes	Yes	Yes
First stage regression				
log *EuAgl*			0.6208^{***}	0.4692^{***}
			(0.0244)	(0.0217)
N	13,361	13,361	13,361	13,361

Notes: The explained variable is the log of plant-level TFP. $\log H^D$ is the log of human capital agglomeraton based on educated worker density, and $\log H^S$ is the log of human capital agglomeraton based on educated worker share. The definition of these variables is described in Equation (4). $OpYrs_{it}$ is the number of years in operation. $EduYrs_{it}$ is the average years of education of workers. $CityPop_{it}$ is city-level poplation. For_{it} is the share of foreign capital. $Sizedm2_{it}$–$Sizedm5_{it}$ are the firm size dummies. log *EuAgl* is European population size in colonial Indonesia which is the instrument variable for the human capital agglomeration. The asterisks *, **, and *** denote 10%, 5%, and 1% significance levels. Figures in parentheses are standard errors, clustered at the 2-digit industrial classification.

Table 7: Estimation results of log labor productivity: 1996

	Manufacturing plant-level data in 1996			
	1: OLS	2: OLS	3: 2SLS	4: 2SLS
$\log H^D$	-0.0031^{**}		0.0290^{***}	
	(0.0013)		(0.0092)	
$\log H^S$		-0.0003		0.0345^{***}
		(0.0014)		(0.0102)
OpYrs	-0.0026^{***}	-0.0026^{***}	-0.0028^{***}	-0.0027^{***}
	(0.0007)	(0.0007)	(0.0008)	(0.0007)
EduYrs	0.0063^{**}	0.0053^{*}	-0.0053	-0.0059
	(0.0026)	(0.0026)	(0.0036)	(0.0036)
$\log \mathit{CityPop}$	0.0117	0.0087	-0.0211	-0.0201
	(0.0120)	(0.0119)	(0.0142)	(0.0136)
Gov	-0.0003	-0.0003	-0.0006	-0.0008^{*}
	(0.0003)	(0.0003)	(0.0004)	(0.0004)
For	-0.0002	-0.0002	-0.0001	-0.0001
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Sizedm2	0.0029	0.0009	-0.0200	-0.0235
	(0.0158)	(0.0157)	(0.0156)	(0.0152)
Sizedm3	-0.0960^{***}	-0.0993^{***}	-0.1322^{***}	-0.1347^{***}
	(0.0295)	(0.0298)	(0.0357)	(0.0353)
Sizedm4	-0.2571^{***}	-0.2629^{***}	-0.3213^{***}	-0.3216^{***}
	(0.0525)	(0.0530)	(0.0560)	(0.0551)
Sizedm5	-0.5829^{***}	-0.5913^{***}	-0.6743^{***}	-0.6616^{***}
	(0.0471)	(0.0481)	(0.0448)	(0.0425)
2-digit industrial dummies	Yes	Yes	Yes	Yes
Province dummies	Yes	Yes	Yes	Yes
First stage regression				
$\log \mathit{EuAgl}$			0.6414^{***}	0.5388^{***}
			(0.0285)	(0.0262)
N	13,929	13,929	13,929	13,929

Notes: The explained variable is the log of labor productivity. $\log H^D$ is the log of human capital agglomeraton based on educated worker density, and $\log H^S$ is the log of human capital agglomeraton based on educated worker share. The definition of these variables is described in Equation (4). OpYrs_{it} is the number of years in operation. EduYrs_{it} is the average years of education of workers. $\mathit{CityPop}_{it}$ is city-level poplation. For_{it} is the share of foreign capital. $\mathit{Sizedm2}_{it}$–$\mathit{Sizedm5}_{it}$ are the firm size dummies. $\log \mathit{EuAgl}$ is European population size in colonial Indonesia which is the instrument variable for the human capital agglomeration. The asterisks *, **, and *** denote 10%, 5%, and 1% significance levels. Figures in parentheses are standard errors, clustered at the 2-digit industrial classification.

Table 8: Estimation results of log labor productivity: 2006

	Manufacturing plant-level data in 2006			
	1: OLS	2: OLS	3: 2SLS	4: 2SLS
$\log H^D$	-0.0116^{***}		0.0281^{**}	
	(0.0030)		(0.0114)	
$\log H^S$		-0.0083^{**}		0.0372^{**}
		(0.0032)		(0.0154)
OpYrs	-0.0027^{***}	-0.0028^{***}	-0.0035^{***}	-0.0035^{***}
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
EduYrs	0.0202^{***}	0.0186^{***}	0.0042	0.0018
	(0.0030)	(0.0029)	(0.0039)	(0.0041)
$\log CityPop$	0.0256^{*}	0.0220	−0.0042	−0.0060
	(0.0139)	(0.0143)	(0.0195)	(0.0193)
Gov	0.0008^{***}	0.0008^{***}	0.0007^{**}	0.0007^{*}
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
For	-0.0013^{***}	-0.0013^{***}	-0.0014^{***}	-0.0013^{**}
	(0.0004)	(0.0004)	(0.0005)	(0.0005)
Sizedm2	0.0804^{***}	0.0780^{***}	0.0614^{**}	0.0610^{**}
	(0.0197)	(0.0201)	(0.0232)	(0.0238)
Sizedm3	0.0497	0.0473	0.0296	0.0285
	(0.0370)	(0.0376)	(0.0428)	(0.0438)
Sizedm4	-0.1116^{**}	-0.1171^{**}	-0.1483^{**}	-0.1454^{**}
	(0.0456)	(0.0467)	(0.0583)	(0.0587)
Sizedm5	-0.4602^{***}	-0.4728^{***}	-0.5359^{***}	-0.5246^{***}
	(0.0397)	(0.0406)	(0.0466)	(0.0473)
2-digit industrial dummies	Yes	Yes	Yes	Yes
Province dummies	Yes	Yes	Yes	Yes
First stage regression				
$\log EuAgl$			0.6208^{***}	0.4692^{***}
			(0.0244)	(0.0217)
N	13,361	13,361	13,361	13,361

Notes: The explained variable is the log of labor productivity. $\log H^D$ is the log of human capital agglomeraton based on educated worker density, and $\log H^S$ is the log of human capital agglomeraton based on educated worker share. The definition of these variables is described in Equation (4). $OpYrs_{it}$ is the number of years in operation. $EduYrs_{it}$ is the average years of education of workers. $CityPop_{it}$ is city-level poplation. For_{it} is the share of foreign capital. $Sizedm2_{it}$–$Sizedm5_{it}$ are the firm size dummies. $\log EuAgl$ is European population size in colonial Indonesia which is the instrument variable for the human capital agglomeration. The asterisks *, **, and *** denote 10%, 5%, and 1% significance levels. Figures in parentheses are standard errors, clustered at the 2-digit industrial classification.

Table 9: Coefficients of human capital agglomeration with different values of θ.

	Estimated coefficients of $\log H^D$ (2SLS estimates)	
θ	1996	2006
0.1	0.0831**	0.1034**
	(0.0318)	(0.0419)
0.15	0.0601**	0.0764**
	(0.0227)	(0.0309)
0.2	0.0474**	0.0610**
	(0.0178)	(0.0246)
0.25	0.0392**	0.0509**
	(0.0146)	(0.0205)
0.3	0.0334**	0.0438**
	(0.0124)	(0.0176)

	Estimated coefficients of $\log H^S$ (2SLS estimates)	
θ	1996	2006
0.1	0.1137**	0.1659**
	(0.0408)	(0.0681)
0.15	0.0750**	0.1081**
	(0.0270)	(0.0439)
0.2	0.0564**	0.0807**
	(0.0203)	(0.0327)
0.25	0.0453**	0.0646**
	(0.0164)	(0.0262)
0.3	0.0378**	0.0539**
	(0.0137)	(0.0218)

Notes: The explained variable is the log of plant-level TFP. $\log H^D$ is the log of human capital agglomeraton based on educated worker density, and $\log H^S$ is the log of human capital agglomeraton based on educated worker share. This table shows the coefficients of $\log H^D$ and $\log H^S$, separately estimated by different θ, using the two-stage least squares estimator. The θ is a parameter to calculate human capital agglomeration (See Equation 6). The instrumental variable for $\log H^D$ is European population size in colonial Indonesia ($\log EuAgl$). Figures in parentheses are standard errors, clustered at the 2-digit industrial classification.

Table 10: Industrial classification

ISIC	Description
15	Food and drink
16	Tobacco
17	Textiles
18	Clothes
19	Leather and leather goods
20	Wood, articles of wood, and wicker
21	Paper and paper products
22	Publishing, printing and reproduction
23	Coal, oil and natural gas, and nuclear fuel
24	Chemical and goods from chemicals
25	Rubber and plastic goods
26	Non metallic mineral goods
27	Base metals
28	Goods of metal and crockery
29	Machines and equipment
30	Hardware office, accounting and data processing
31	Other electric machines and equipment
32	Radio, television and communication equipment
33	Medical devices, measuring instruments, navigation, optical and hour
34	Motor vehicles
35	Other transport equipment
36	Furniture and other processing industries
37	Recycling