Urban Agglomeration and City Productivity: the case of Indonesia

Hengky Kurniawan*[1,2] Steven Poelhekke[1] Henri L.F. de Groot[1]

[1]Department of Spatial Economics, Vrije Universiteit Amsterdam
[2]Institute for Economics and Social Research, University of Indonesia (LPEM FEBUI)

This version: July 3, 2017

Abstract

In this paper we examines the relationship between the city size and productivity in Indonesia. Theory in urban economics suggests that worker productivity in a city increases as an effect of agglomeration. However, the increase in productivity at some point in time will reach its peak and then diminishes following the inverted U-shape curve due to congestion. We test this relationship by using Indonesian data for manufacturing and service sectors. The results show that city's productivity is significantly affected by non-agriculture workers and industrial composition. The other factors also positively contributed to city's productivity viz. capital intensity, education, FDI, and market potential. We find that most of cities are smaller than their optimal sizes.

Keywords: Urban agglomeration, City size, Indonesia, Productivity.

1 Introduction

Currently, more than half of Indonesian population live in urban area. According to the World Bank's report, the country has urbanized rapidly and approximately 67.5 percent of the population are expected to live in urban area by 2025.[1] Urban economic theory suggests that agglomeration, as a result of concentration of firms and workers, will increase productivity because of the positive spill over effect. But, the too high concentration could slow down the increase in productivity as negative impact is greater than its benefit. The negative effect of agglomeration such as congestion and over-utilization of production factors.

In other contrasting case, cities could be too small and the optimality is not reached when the its size is too small to benefit from net urban agglomeration. Au and Henderson (2006) by using data from China found that around 51-63 % of Chinese cities are undersizes, depending on the model specifications. Behind of these two reasons, we can observe the rationale behind the creation of new cities in Indonesia. Among other reasons, the proliferation supporters argue that some cities are

*Corresponding author. Email address: h.kurniawan@vu.nl and hengky@lpem-feui.org.

[1]The World Bank's report (2012), title: Indonesia, The Rise of Metropolitan Regions: Towards Inclusive and Sustainable Regional Development

too big (in the sense of area) and split-up is therefore inevitable. In addition, some others see that some regencies need to be upgraded to cities. In the last two decades, economic growth and the increasing number of the middle income group, coupled with the new decentralization policy in 2001 have accelerated the birth of new cities in Indonesia.

In the economic literature, the increasing number of cities can be related to agglomeration economies and the growth of city size as the impact of human capital accumulation and knowledge spillovers (Black and Henderson, 1999; Henderson, 2005). Moreover, the role of institution has also important impact on cities (Henderson and Wang, 2007). Whether the growth of new cities is the consequence of the growth of city size, then it needs to be investigated and if we believe that city size affects its productivity.

The discussion on city size usually based on the common variables used in the literature such as area, population size and density. In this paper we assess the city size in Indonesia by looking at the part of population that is peculiar to cities, *viz.* the number of non-agricultural workers. This optimal or preferred size of workers are those who bring the city's productivity to the optimal level. We use the flexible functional form of theoretical models developed by Au and Henderson (2006). The models relate the city's productivities and number of non-agricultural workers with an inverted-U shaped curve.

Once the relationship is estimated, we are able to assess whether cities are at the peak, on the left, or on the right side of the curve. Both deviations are not optimal, but if a city lies on the left hand side it means that the city is too small and on the right hand side means a city is too big, and productivities can be improved. The larger the deviations (from left or right), the larger productivity losses.

1.1 Research Questions

This paper aims at analyzing the impact of urban agglomeration on the cities's productivities. In the urban economics literature it deals with the inverted U-shaped curve depicting the relationship between the output per worker and number of worker in the city. The peak of the curve is considered as the optimal size of a city (worker), while deviation from the peak (left or right side) resulting the inefficient size.

The research questions are:

1. What is the impact (and magnitude) of urban agglomeration on worker productivity in the city?

2. What are the optimal sizes of Indonesian cities? Are Indonesian cities too big or too small? How much cities will gain or loss if they move to the optimal size?

3. What other important factors that affect the optimal city size?

The general hypothesis is that agglomeration economies will increase worker productivity until a certain level, and decrease afterwards. At low number of workers, the increase in number of workers (and firms) will increase the productivity as an effect of agglomeration. When city grows, it impacts on productivity and wage, but at certain level of size. Agglomeration economies diminish as city grows, worker productivity increases but at diminishing rate. If the growth continues, worker productivity will decrease.

2 The Context: City in Indonesia

Table 1 shows the proliferation of regional units between 1998 to 2013. Province is the first level of local administration (*Daerah Tingkat Satu*), while regency and municipality are the second local administration (*Daerah Tingkat Dua*) and sometimes are called the district level. By political definition, regency (*kabupaten*) refers to less-urbanized region and municipality (*kotamadya*) to more-urbanized area. Between 1998 and 2013 the total number of the second level local government has increased by 62%, while number of cities has increased by almost 51%.

Table 1: Number of Regional Unit

Administrative Level	1998	2000	2013
Province (1^{st} level)	27	32	34
2^{nd} level:	314	341	508
- Municipality (City)	65	73	98
- Regency	249	268	410

Note: In 1999 East Timor gained its independence.
Source: Indonesia's Ministry of Home Affair (2014).

The number of regencies is larger than of municipalities in any province, with exception for Jakarta where the province consists of four municipalities and one regency. In general, the population of municipalities in Indonesia ranges from one hundred thousand to almost three million, and there are only eight municipalities with population less than 100 thousand.

By formal definition, *kota*, literally translated as "city", is more urbanized area than *kabupaten*. But, in some regions, *kabupaten* is more urbanized area than *kota* especially in the island of Java. Figure 1 shows the densities of cities and regencies of percentage of urban population. It shows that the two densities intersect around 60% of urban population. In this study we use the definition of city by its administrative border.

Figure 2 shows the distribution of urbanized area in Indonesia, where every single dot represent an urban area. The urban areas in this figure are those urban areas both in the municipalities and regencies.

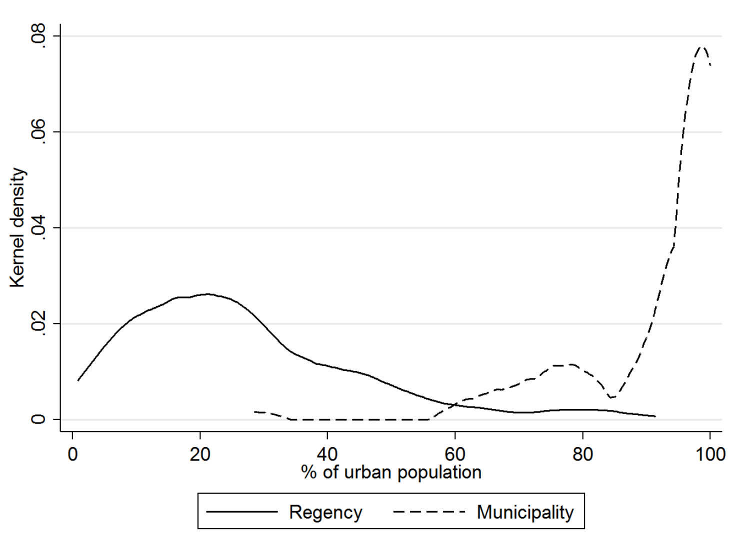

Figure 1: Urbanization rate: cities Vs. regencies

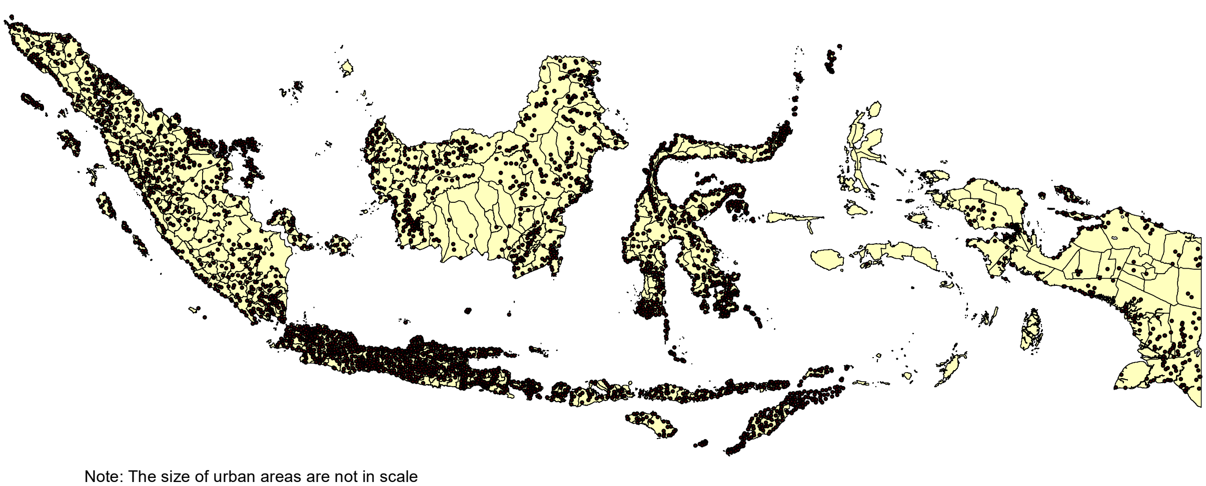

Figure 2: Distribution of urban areas in Indonesia

2.1 City Size and Rank

The relationship between city size and its rank has been studied in a large literature (e.g., Gabaix, 1999; Rosen and Resnick, 1980; Nitsch, 2005). If we rank the city size from the largest (rank 1) and the smallest (rank r), it is hypothesized that the relationship between rank and city size is constant across the cities and follows the relationship of $Rank = C * Size^{-b}$. C is a constant and the exponent b is the parameter to be estimated. If $b = 1$, then the rank-size rule holds or known as a Zipf's law. In the empirical formulation, after taking the logarithmic form and rearranging the equation, the relationship between rank and city size is as follows:

$$ln(Rank) = ln(C) - b * ln(Size) + e \quad (1)$$

where e is the error term. Equation (1) is estimated by OLS.

Figure 3 visualizes the distributions of city size in Indonesia. Panel (a) shows the relationship between city rank and population size, while panel (b) shows the city rank and number of non-agriculture workers. Both panels depict the similar pattern with the slope of figure 5a is steeper than figure 5b.

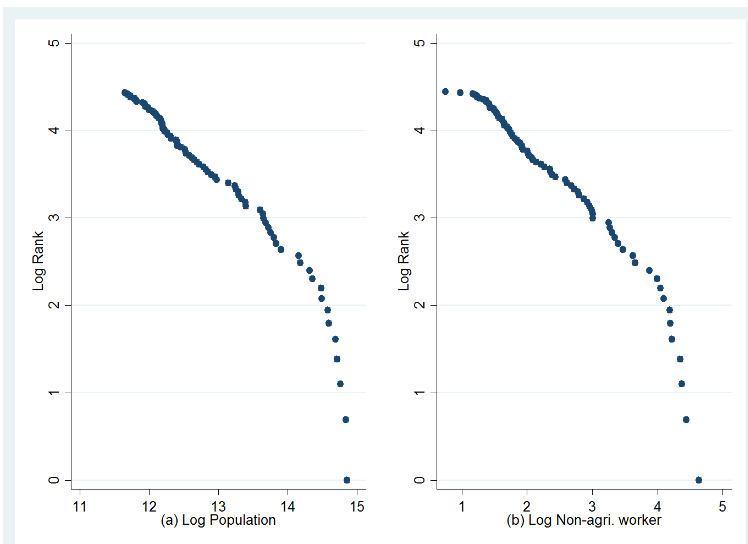

Figure 3: Log Rank against Log Size of Indonesian Cities

The estimations from the data confirm that the relationship between log of city rank and log size are quite constant with beta estimates of 0,94 for log population and 0.90 for log non-agriculture worker. We may conclude that there is no significant difference between coefficient the log rank of population on log population, and log worker rank on log number of non-agricultural worker. The only difference is the magnitude of the constant, where population has 15.57 and worker has 5.59.

Table 2: Regression Results for City Size and Rank

	(1)	(2)
VARIABLES	Log Population Rank	Log Worker Rank
Population(log)	-0.94***	
	(0.06)	
Worker(log)		-0.90***
		(0.05)
Constant	15.57***	5.59***
	(0.70)	(0.10)
Observations	84	85
R-squared	0.92	0.93

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3 Data and Methodology

3.1 Data

The data used in this study are the second tier of Indonesian local administrative districts, municipalities, or also called *kotamadya* or *kota* in Indonesia. The list of variables is available in Appendix A.1.

Data are compiled from the official publication of Indonesian bureau of statistics or BPS. The variables used in our analysis are from year 2012 since those are the most recent and available for all variables and cities we are interested in. The instrumental variables are from year 2005 and if not available we use the data from 2007 and 2008.

Cities's outputs are gross domestic product grouped into three main sectors in the economy, which breaks down into nine more detail sectors. First, primary sector consists only agricultural sector. The secondary sector includes manufacturing; construction; mining and quarrying. The tertiary sector includes utilities; financial service sector; trade, hotel and restaurant; transportation and telecommunication; and other services.

In our analysis, the definition of manufacturing sector is only the manufacturing from the secondary sector and exclude the other two. We choose the specific manufacturing sector that captures the industrial sector. Meanwhile, service sector only includes financial service; trade, hotel and restaurant; and transportation and telecommunication. The service sector excludes the other services sector that are not relevant with business process such as social and voluntary services. The summary are in the table 3.

3.1.1 Market Potential

Market potential of a city is the weighted averaged by distances of all other cities and regencies. We assume that a city sells its output, besides its own, to its neighboring cities domestically and to foreign markets. In spite of the important of foreign market and the fact that Indonesia lies in the archipelago, we drop foreign markets in the calculation. We assume that every city faces the same foreign market potential.

Market potential is calculated from the summation of city's GDP plus other district's GDP weighted by the average distances of city's centroids. Thus, market potential of a city i (MP_i) is:

$$MP_i = GDP_i + \sum_{j=1}^{N-1} \frac{GDP_j}{distance_{ij}} \quad , \quad i \neq j \tag{2}$$

Moreover, for a comparison and robustness check we also simulated the calculation of market potential by group of main island. The calculation is based on equation (2) but the weighting parameters of distances are grouped by main islands or nearest regions. The classification is divided into five group: 1) Sumatra, 2) Java, Bali, Nusa Tenggara, 3) Kalimantan, 4) Sulawesi & Maluku, and 5) Papua. We found that this alternative measurement does not have significant impact on our analysis, except for the coefficient of market potential which is smaller than the previous calculation.

3.2 Descriptive Statistics

This section provides the descriptive statistics of the variables used in our analysis as shown in table 3. The total number of our observations are 98 cities in which the metropolitan Jakarta is broken down into four individual cities, but in our estimations this number has been reduced for the reason of missing observations. First we plot the worker productivity data with other variables as shown in figure 4, 5 and 6.

Figure 4 (on the top-left) shows that correlation between output per worker (in logarithmic form) and number of non-agricultural worker is positive in the cities, meaning that the higher the number of worker the higher the productivity as a result of urban agglomeration. One should be careful that it does not portray the inverted-U curve because of two reasons: first, not all of the cities are at the optimal size, and second, the cities do not have the same manufacturing to service ratios.

Most of the other variables show positive correlation with city productivity with exception for percentage of agricultural GDP. The agriculture data are consistent with findings from much of economic literature, showing that city's productivity declines with agricultural sector. In Indonesian cities, the average of agricultural sector in GDP has declined from 9.05% in 2005 to 7.16% in 2012, and the highest share of agricultural GDP declined from 57.93% to 50.91%. There are 22 cities that agricultural sectors account less than one percent, and six cities which agricultural sectors account more than 20%.

Meanwhile, the capital intensity variable also has positive correlation with city's productivity. The capital intensity in 2012, measured by private capital per worker, range from 8.87 to 26.22 (in logarithmic form). Compare to 2005 the number is slightly lower but the the highest capital intensity has increased from 23.31 to 26.22.

Population has positive correlation with productivity, but when it comes to limited area than the productivity becomes lower. Figure 6 shows that the correlation of population density is flatter than the total population when it plots against city productivity. This is in-line with the concept of diminishing marginal return of land.

Table 3: Summary statistics for cities

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
City's output (log)	15.68	1.56	12.92	19.22	85
Output per worker (log)	13.34	0.83	11.69	15.84	85
Capital (log)	18.99	3.11	10.49	26.56	82
Worker	16.99	22.26	1.12	102.02	85
Manuf. to service ratio	0.43	0.61	0.02	3.19	85
Manuf. to service ratio in 2005	0.57	0.88	0.02	6.18	82
Capital per worker (log)	16.61	2.46	8.87	23.06	82
Capital/ worker in 2005 (log)	16.61	1.98	12.5	21.94	75
FDI/ worker (log)	5.62	2.44	-1.18	10.74	66
FDI/ worker, 2005 (log)	3.44	2.32	-0.71	8.4	40
Market potential (log)	17.14	1.05	15.38	19.82	81
Market potential in 2005 (log)	16.22	1.05	14.44	18.88	81
Area (ˆ2$)	235.26	353.3	0	2399.5	85
Doctor/capita	0.41	0.22	0.12	1.28	81
% of Agricultural GDP	5.52	5.75	0.01	25.16	85
% of Agricultural GDP in 2005	7.01	7.79	0.02	38.54	82
Dummy for non-FDI in 2005	0.48	0.5	0	1	85
Population in 2005(log)	12.7	0.93	11.55	14.81	80
% of Urban population	91.86	12.61	28.5	100	85
FDI/ capital	21.52	167.03	0	1507.37	82
% of high school graduate	28.11	5.26	14.29	40.41	85
% of college graduate	8.91	2.76	3.96	16.91	85
Dummy(Java=1; Non-Java=0)	0.35	0.48	0	1	85

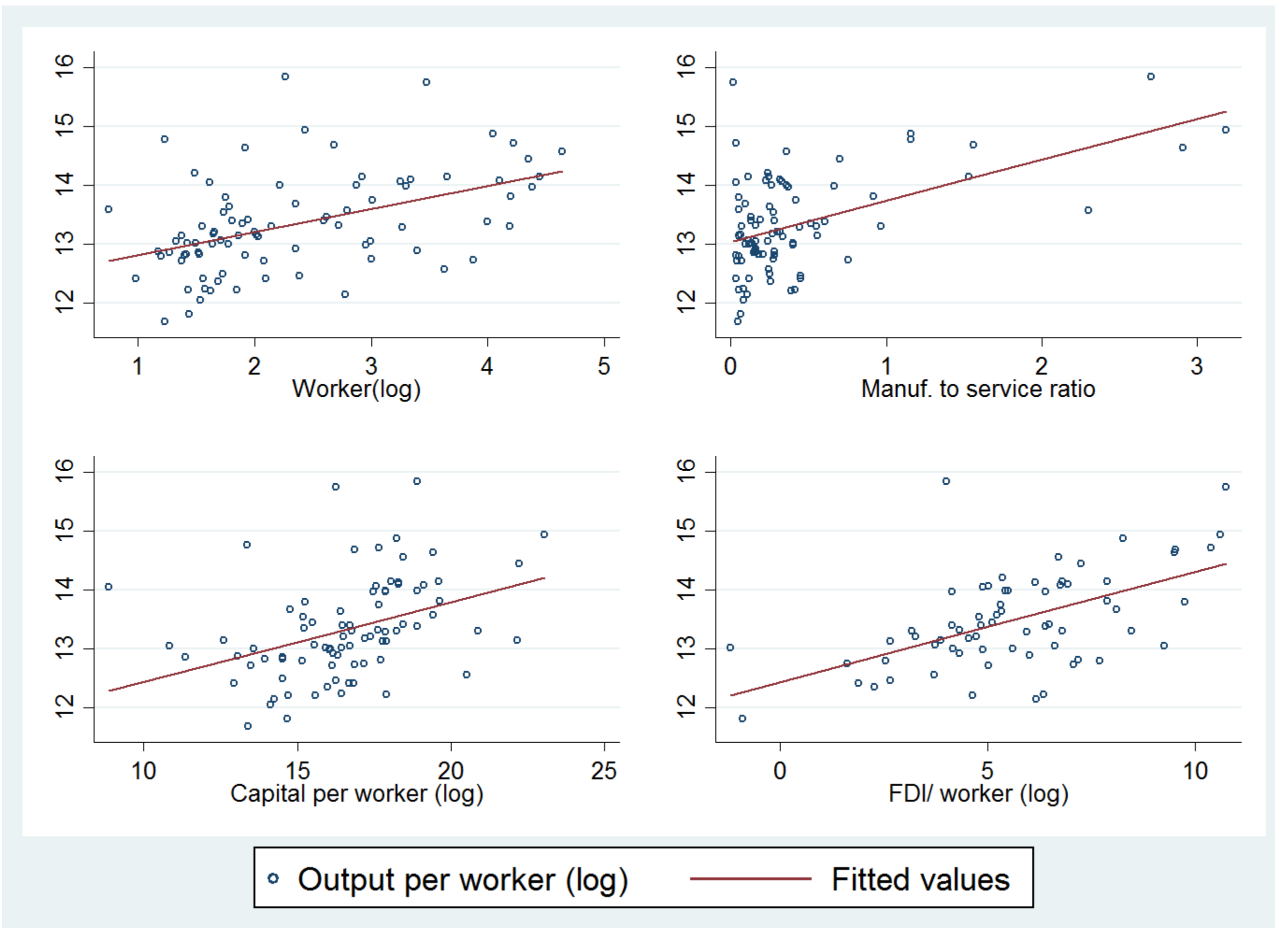

Figure 4: Scatter-plots between cities's output per worker and other variables (part-1)

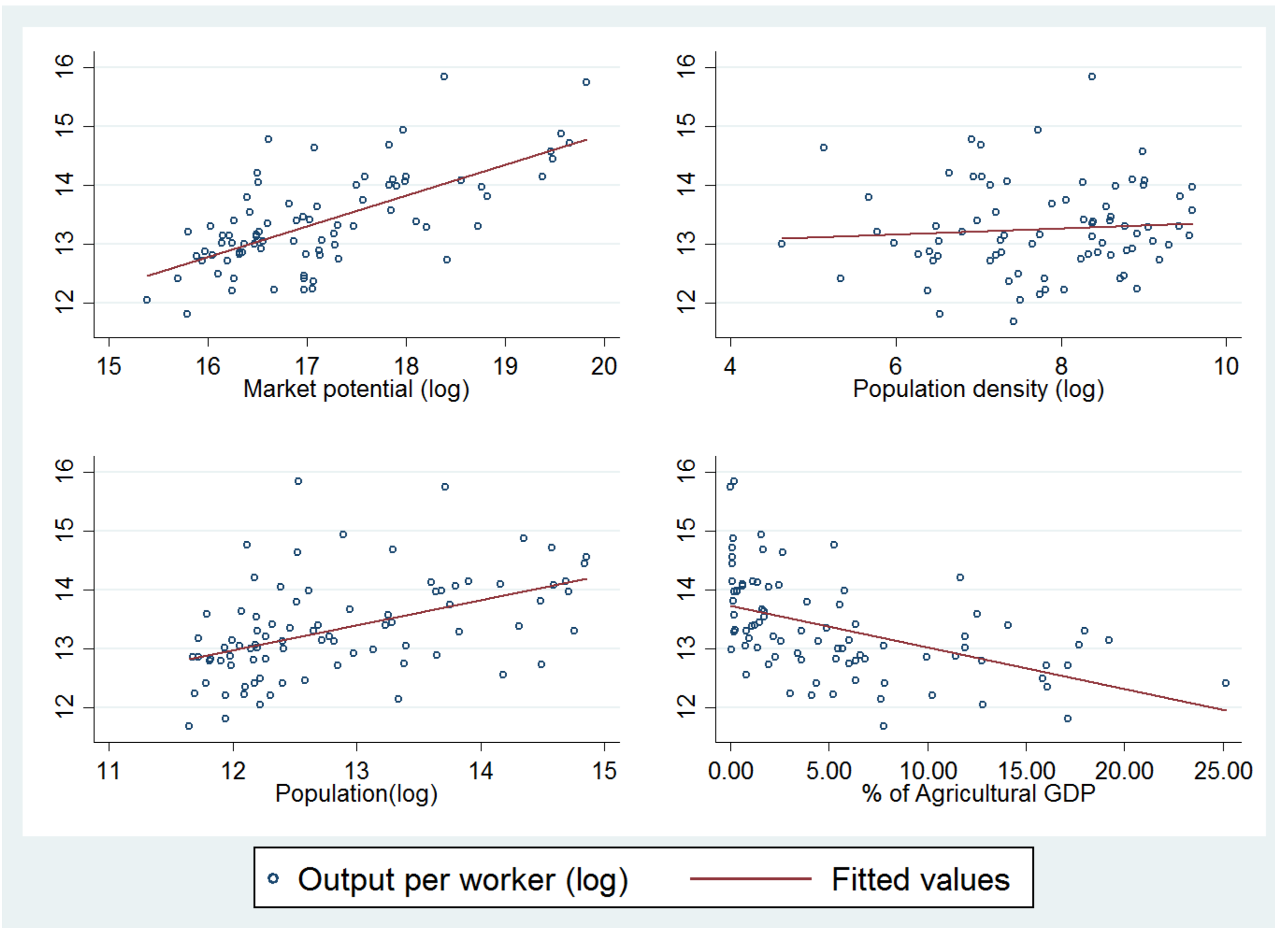

Figure 5: Scatter-plots between cities's output per worker and other variables (part-2)

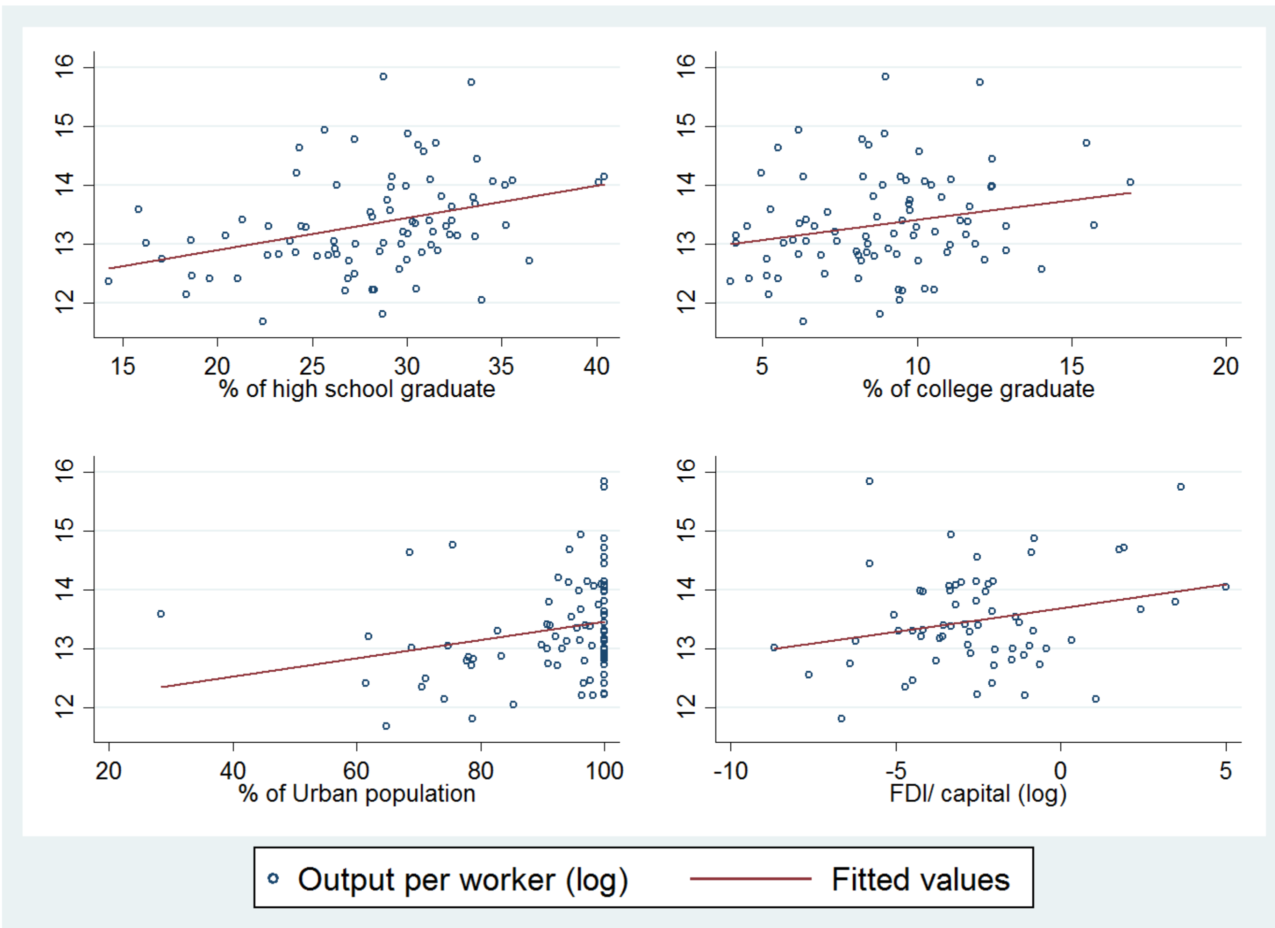

Figure 6: Scatter-plots between cities's output per worker and other variables (part-3)

3.3 Estimation Strategy

In order to understand how urban agglomeration affects its productivity we look at the city size which measured by the number of manufacturing and service workers. City productivity also depends on the city hierarchy in the country, which different hierarchy results in different optimal level of productivity. Manufacturing to service ratio is used to measure the hierarchy of the city. For capital intensity variable we use manufacturing firm's capital per worker. The other control variables come from market potential, human capital stock measured by education attainment, and FDI per worker. In addition, other control variables are also used robustness check.

We follow Au and Henderson (2006) and employ the General Leontief specification with the second order expansion in square roots. This specification is preferred since we are able to calculate the peak points for all cities. The equation for estimation is as follow:

$$\begin{aligned} \ln(\mathrm{GDP/N}) = a_1 N^{0.5} - a_2 N - a_3 (N \times MS)^{0.5} + a_4 MS^{0.5} + a_5 MS \\ + \alpha ln(K/N) + \beta ln(MP) + \gamma X. \end{aligned} \tag{3}$$

The presumption of equation (3) is that $a_1, a_2, a_3 > 0$, and $a_1 - a_3 MS > 0$.

Where GDP/N is city's non-agriculture output per worker, N is the number of non-agriculture worker, MS is manufacture to service ratio which reflects the industrial structure, MP is market potential, K/N is the manufacturing firm's capital per worker, X are the control variables such school enrollment ratio and FDI per worker.

If ratio K/N holds constant, the maximized value-added per worker at peak size (in the square roots terms) as follows:

$$N^* = \left(\frac{a_1 - a_3 MS^{0.5}}{2a_2} \right)^2 . \tag{4}$$

where the term N* denotes the peak size or the "optimal" size of a city.

The percentage of losses in net output per worker from moving from peak size or gains of moving to the peak can be calculated as follow:

$$\begin{aligned} \ln(\text{net output/N})^* - \ln(\text{net output/N}) = \\ \frac{1}{1-\hat{\alpha}} \left((\hat{a_1}) - \hat{a_3} MS^{0.5}) [(N^*)^{0.5} - N^{0.5}] - \hat{a_2}(N^* - N) \right) \end{aligned} \tag{5}$$

3.3.1 Endogeneity Problem

The problem in estimating productivity in equation (3) is that the regressors may not fully exogenous, therefore OLS regression will be inconsistent. The productivity of a city may attract number of firms to locate in that city, and thus the amount of capital and number of workers. Moreover, the location choice of foreign direct investment may also depends on the productivity, where investors look for more productive and worker concentrated city. The similar fashion also applies for market potential, education attainment, and industrial concentrations.

One may use instruments to correct for this problem, but choosing for a good instrument maybe tricky. In Indonesia there were two regime changes that could be the sources of good instruments if we use the variables that are not correlated with the errors in our regression but correlated with the regressors. In 2001, a new decentralization policy was introduced in the country and it was called a "big bang" decentralization policy since the massiveness and magnitude of the changes. This decentralization policy gives more power to the cities and municipalities governments. Following that policy, in 2004, the new democratization law was also introduced. In this new democratization process, local government can be directly elected by the people.

Even though those two policies were introduced in 2001 and 2004, but the processes to be implemented fully were gradual for at least four to five years. Based on that notion, we choose the instrumental variables from the year 2005 since the impact of those policies might not fully affect the variables but fulfill the exogeneity assumption. Another important issue to consider is the relevance of the instruments.

For all regressors in the equation 3 we take the same variables in 2005 such as number of worker and manufacturing to service ratio in 2005. Meanwhile, to capture the quality of infrastructure we take percentage of village with asphalt road in 2005. The other instruments are FDI per worker in 2005, area both in 2005 and 2012, number of doctor in 2005, etc. The detail of variables we use are in the appendix A.1. Finally, we use two-stage least square method to estimate equation (3).

4 Results & Discussions

4.1 Estimation Results

Table 4 shows the IV regression results with six specifications (from column (1) to (6)). The first stage regression results and the tests for the instruments are in appendix A.2.

Table 4 shows that coefficients of worker and high school enrollment are statistically significant for full sample. Taking the first specification in column (1) it shows that doubling capital intensity in cities and regions will increase the productivity by around 6% and increased in market potential by 10% will increase the productivity on almost one percent.

However, manufacturing to service ratio shows a better performance than worker in cities regression. The coefficients are significant at specification (2), (3), and (4). FDI per worker shows positive effect on productivity but the coefficient it too low to have an impact. Market potential does not affect productivity in these regressions, while educational variable only has percentage of high school graduate with significant coefficient. In specification (2), doubling the high school graduate will increase the productivity by 4.1%.

If cities in Jakarta are excluded from the sample, the estimations show that manufacturing to service ratio becomes statistically significant. Table 5 shows only manufacturing to services ratio that is statistically significant.

Figure 7: Actual versus Optimal City Size

Table 4: IV-regression of output per worker

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Capital per worker (log)	0.0626	0.0599	0.0604	0.0604	0.0662	0.0662
	(0.0599)	(0.0563)	(0.0638)	(0.0638)	(0.0877)	(0.0877)
$\text{Worker}^{0.5}$	0.4654**	0.1986	0.5178*	0.5178*	0.5058*	0.5058*
	(0.2091)	(0.2401)	(0.2730)	(0.2730)	(0.3046)	(0.3046)
Worker	-0.0354	0.0026	-0.0135	-0.0135	-0.0135	-0.0135
	(0.0320)	(0.0331)	(0.0401)	(0.0401)	(0.0397)	(0.0397)
$(\text{Manuf. to Service ratio x Worker})^{0.5}$	0.1087	-0.1640	-0.1706	-0.1706	-0.1612	-0.1612
	(0.3119)	(0.3259)	(0.3883)	(0.3883)	(0.3996)	(0.3996)
$(\text{Manuf. to service ratio})^{0.5}$	-2.0030	-0.6703	-1.3032	-1.3032	-1.2281	-1.2281
	(1.7875)	(1.9094)	(2.1677)	(2.1677)	(2.1320)	(2.1320)
Manuf. to service ratio	1.1771	1.0593	1.4727	1.4727	1.4217	1.4217
	(0.8140)	(0.8313)	(0.9643)	(0.9643)	(1.0234)	(1.0234)
Market potential (log)	0.0351	0.1175	-0.1629	-0.1629	-0.1646	-0.1646
	(0.2389)	(0.2264)	(0.2869)	(0.2869)	(0.2780)	(0.2780)
FDI/ worker (USD)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
High school enrolment (%)	-0.0060					
	(0.0186)					
% of high school graduate		0.0269*				
		(0.0155)				
% of college graduate			0.0228	0.0228	0.0221	0.0221
			(0.0440)	(0.0440)	(0.0457)	(0.0457)
FDI/ capital					0.0096	0.0096
					(0.0970)	(0.0970)
Constant	11.3107***	9.0388**	13.7320***	13.7320***	13.6689***	13.6689***
	(3.7593)	(3.7162)	(4.3073)	(4.3073)	(4.4678)	(4.4678)
Observations	71	71	71	71	71	71
R-squared	0.5524	0.6558	0.5279	0.5279	0.5496	0.5496

Robust standard errors in parentheses

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

Table 5: IV-regression of output per worker: excluding Jakarta

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Capital per worker (log)	0.0685	0.0716	0.0666	0.0673	-0.0208	-0.0340
	(0.0603)	(0.0538)	(0.0618)	(0.0629)	(0.1459)	(0.1852)
$\text{Worker}^{0.5}$	0.6822***	0.5051*	0.8287***	0.7690**	1.2219**	1.3234**
	(0.2605)	(0.2790)	(0.3158)	(0.3112)	(0.5880)	(0.6046)
Worker	-0.0245	-0.0047	-0.0281	-0.0227	-0.0182	-0.0291
	(0.0260)	(0.0290)	(0.0384)	(0.0383)	(0.0557)	(0.0542)
$(\text{Manuf. to Service ratio x Worker})^{0.5}$	-0.2630	-0.4240	-0.3798	-0.3724	-0.8056	-0.8215
	(0.3632)	(0.3786)	(0.4387)	(0.4311)	(0.7584)	(0.7760)
$(\text{Manuf. to service ratio})^{0.5}$	-0.6778	0.1197	-0.6286	-0.5656	-0.1995	-0.1008
	(1.3183)	(1.3153)	(1.4855)	(1.4797)	(2.0699)	(2.2009)
Manuf. to service ratio	1.3279*	1.1808**	1.5503**	1.5134**	2.3520**	2.1996*
	(0.7356)	(0.5723)	(0.6425)	(0.6369)	(1.1326)	(1.1406)
Market potential (log)	-0.2219	-0.0820	-0.3016	-0.3055	-0.6556	-0.6030
	(0.2301)	(0.2392)	(0.2636)	(0.2652)	(0.4652)	(0.4540)
FDI/ worker (USD)	0.0000					0.0000
	(0.0000)					(0.0001)
High school enrolment (%)	-0.0004					
	(0.0172)					
% of high school graduate		0.0228				
		(0.0159)				
% of college graduate			0.0192	0.0212	0.0383	0.0421
			(0.0467)	(0.0466)	(0.0579)	(0.0599)
FDI/ capital					-0.2360	-0.2564
					(0.3335)	(0.3916)
Constant	14.3316***	11.4806***	15.2178***	15.3397***	21.4748***	20.6094**
	(3.3608)	(4.0544)	(4.0950)	(4.1806)	(8.1868)	(8.1529)
Observations	66	66	66	66	66	66
R-squared	0.4866	0.5855	0.4489	0.4536	-0.2592	-0.2984

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.2 City Size and Productivity Gain

Table 6: Summary of Actual, Optimal Worker, and Percentage Gains

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Worker	17	22.3	1.1	102
Estimated optimal worker	55.6	8.4	45.8	87
Lower bound	-52.3	11.5	-74.5	-29.7
Upper bound	163.5	12.3	157.8	230.3
Percentage gain	75.4	44.3	0	173.7
N		85		

Table 7: Percent Gains from Moving Towards Optimal Level, by Percentile

City ranked by loss	Percent gains
1%	0.0
5%	2.1
10%	6.3
25%	37.4
50%	90.2
75%	106.5
90%	122.5
95%	136.3
99%	173.7

4.3 Robustness check

4.3.1 Sectors Definition

We check whether change in the definition of manufacturing and services sectors affect the estimation results. In our previous estimation, definition of manufacturing and services are the subset of the secondary and tertiary sectors in the cities's output. In the following definition we define the manufacturing sectors as the secondary sector, and services sector as the tertiary sectors. Hereinafter we call it " the broad" definition.

The summary statistics for the broad definition are in the appendix 3, while the regression results are in the table 8 and table 9 excluding Jakarta. In the broad definition, the coefficient of capital intensity becomes non statistically significant. By including the sectors that are not relevant with the industrial and business services, the effect of capital intensity becomes smaller.

Table 8: IV-regression of output per worker (Broad definition)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Capital per worker (log)	0.0626	0.0599	0.0604	0.0604	0.0662	0.0662
	(0.0599)	(0.0563)	(0.0638)	(0.0638)	(0.0877)	(0.0877)
$\text{Worker}^{0.5}$	0.4654**	0.1986	0.5178*	0.5178*	0.5058*	0.5058*
	(0.2091)	(0.2401)	(0.2730)	(0.2730)	(0.3046)	(0.3046)
Worker	-0.0354	0.0026	-0.0135	-0.0135	-0.0135	-0.0135
	(0.0320)	(0.0331)	(0.0401)	(0.0401)	(0.0397)	(0.0397)
$(\text{Manuf. to Service ratio x Worker})^{0.5}$	0.1087	-0.1640	-0.1706	-0.1706	-0.1612	-0.1612
	(0.3119)	(0.3259)	(0.3883)	(0.3883)	(0.3996)	(0.3996)
$(\text{Manuf. to service ratio})^{0.5}$	-2.0030	-0.6703	-1.3032	-1.3032	-1.2281	-1.2281
	(1.7875)	(1.9094)	(2.1677)	(2.1677)	(2.1320)	(2.1320)
Manuf. to service ratio	1.1771	1.0593	1.4727	1.4727	1.4217	1.4217
	(0.8140)	(0.8313)	(0.9643)	(0.9643)	(1.0234)	(1.0234)
Market potential (log)	0.0351	0.1175	-0.1629	-0.1629	-0.1646	-0.1646
	(0.2389)	(0.2264)	(0.2869)	(0.2869)	(0.2780)	(0.2780)
FDI/ worker (USD)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
High school enrolment (%)	-0.0060					
	(0.0186)					
% of high school graduate		0.0269*				
		(0.0155)				
% of college graduate			0.0228	0.0228	0.0221	0.0221
			(0.0440)	(0.0440)	(0.0457)	(0.0457)
FDI/ capital					0.0096	0.0096
					(0.0970)	(0.0970)
Constant	11.3107***	9.0388**	13.7320***	13.7320***	13.6689***	13.6689***
	(3.7593)	(3.7162)	(4.3073)	(4.3073)	(4.4678)	(4.4678)
Observations	71	71	71	71	71	71
R-squared	0.5524	0.6558	0.5279	0.5279	0.5496	0.5496

Robust standard errors in parentheses

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

Table 9: IV-regression of output per worker: excluding Jakarta (Broad definition)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Capital per worker (log)	0.0685	0.0716	0.0666	0.0673	-0.0208	-0.0340
	(0.0603)	(0.0538)	(0.0618)	(0.0629)	(0.1459)	(0.1852)
$\text{Worker}^{0.5}$	0.6822***	0.5051*	0.8287***	0.7690**	1.2219**	1.3234**
	(0.2605)	(0.2790)	(0.3158)	(0.3112)	(0.5880)	(0.6046)
Worker	-0.0245	-0.0047	-0.0281	-0.0227	-0.0182	-0.0291
	(0.0260)	(0.0290)	(0.0384)	(0.0383)	(0.0557)	(0.0542)
$(\text{Manuf. to Service ratio x Worker})^{0.5}$	-0.2630	-0.4240	-0.3798	-0.3724	-0.8056	-0.8215
	(0.3632)	(0.3786)	(0.4387)	(0.4311)	(0.7584)	(0.7760)
$(\text{Manuf. to service ratio})^{0.5}$	-0.6778	0.1197	-0.6286	-0.5656	-0.1995	-0.1008
	(1.3183)	(1.3153)	(1.4855)	(1.4797)	(2.0699)	(2.2009)
Manuf. to service ratio	1.3279*	1.1808**	1.5503**	1.5134**	2.3520**	2.1996*
	(0.7356)	(0.5723)	(0.6425)	(0.6369)	(1.1326)	(1.1406)
Market potential (log)	-0.2219	-0.0820	-0.3016	-0.3055	-0.6556	-0.6030
	(0.2301)	(0.2392)	(0.2636)	(0.2652)	(0.4652)	(0.4540)
FDI/ worker (USD)	0.0000					0.0000
	(0.0000)					(0.0001)
High school enrolment (%)	-0.0004					
	(0.0172)					
% of high school graduate		0.0228				
		(0.0159)				
% of college graduate			0.0192	0.0212	0.0383	0.0421
			(0.0467)	(0.0466)	(0.0579)	(0.0599)
FDI/ capital					-0.2360	-0.2564
					(0.3335)	(0.3916)
Constant	14.3316***	11.4806***	15.2178***	15.3397***	21.4748***	20.6094**
	(3.3608)	(4.0544)	(4.0950)	(4.1806)	(8.1868)	(8.1529)
Observations	66	66	66	66	66	66
R-squared	0.4866	0.5855	0.4489	0.4536	-0.2592	-0.2984

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

4.3.2 Alternative Measurement of Market Potential

In this section, we define the alternative measurement of market potential of a city. The calculation is based on equation (2) but the weighting parameters of distances are grouped by main islands or nearest regions. The classification is divided into five group: 1) Sumatra, 2) Java, Bali, Nusa Tenggara, 3) Kalimantan, 4) Sulawesi & Maluku, and 5) Papua.

Using data on cities and regencies, we regress the worker productivity with this alternative market potential with the same classification as equation (3). We found that those two different measurement are interchangeable since the coefficient of estimates are quite similar albeit slightly lower than the previous measurement. For instance, for the first specification, coefficient of log market potential is 0.7614 while from the previous measurement the coefficient was 0.8586. Meanwhile, the capital intensity coefficient is 0.1447 versus 0.1279. Exceptions are for the workers variables, where for the coefficient estimate of $worker^{0.5}$ is -1.6603 (versus -1.3837), worker 0.1451 (versus 0.1177), and $(MS * worker)^{0.5}$ is 0.0736 (versus 0.0836).

5 Conclusion

This paper look at the relationship between urban agglomerations and city's productivity in Indonesia. Urban agglomerations are measured by the number of workers in manufacturing and services sectors, while their product per worker account for the productivity measurements. Moreover, we also test the optimal city size by using Indonesian data.

We find that most of the cities in Indonesia do not operate at the optimal sizes. Around 77 cities are on the left side of the peak, while only six cities are considered to big and two cities are at the optimal sizes. The average of observed number of workers are 170 thousand in our data, while the estimated optimal worker are 556 thousand on average. The average of percentage gain if cities operate at the optimal sizes are about 75.4 percent.

If cities are ranked by the percentage of loss, 10 percent of the lowest could gain as much as 6.3 percent in productivities if they move to the optimal sizes and the cities at the 90 percentile could gain about 122.5 percent in productivities.

The policy implication of this findings is that the city size inequality is considered high in Indonesia. The economic growth is still concentrated in few areas and are not spread evenly across Indonesia. This is the result of long term policy that only focus in the big cities. In the future, the government policy has to favor smaller cities especially from outside Java by creating new economic growth center and providing better infrastructure.

References

Au, C.-C. and Henderson, J. V. (2006). Are chinese cities too small? *The Review of Economic Studies*, 73(3):549–576.

Black, D. and Henderson, J. V. (1999). A Theory of Urban Growth. *Journal of Political Economy*, 107(2).

Gabaix, X. (1999). Zipf's law and the growth of cities. *The American Economic Review*, 89(2):129–132.

Henderson, J. V. (2005). URBANIZATION AND GROWTH. volume 1, chapter 24. Elsevier, Amsterdam.

Henderson, J. V. and Wang, H. G. (2007). Urbanization and city growth: The role of institutions. *Regional Science and Urban Economics*, 37(3):283–313.

Nitsch, V. (2005). Zipf zipped. *Journal of Urban Economics*, 57(1):86–100.

Rosen, K. T. and Resnick, M. (1980). The size distribution of cities: An examination of the pareto law and primacy. *Journal of Urban Economics*, 8(2):165–186.

A Appendix

A.1 Data Description

Variable	Definition
lnyna	ln(output of city) [VA manufacture and services sectors]
lna	Employment in 2nd and 3rd sector [10,000s]
lnrkl05	ln(capital)
ryms05	manufacturing to service ratio in 2005
ryms	manufacturing to service ratio [value added] in 2012
ner05	Net enrolment ratio in 2005 (senior high school)
ner12	Net enrolment ratio in 2012 (senior high school)
doctor05	Number of doctors per 1000 people in 2005
rfdlna05	FDI / (employment*10000) in 2005
rfdlnan	Cumulative FDI since 2005 / (employment*10)
area	City area in 2012 (in km2)
urban	ratio of people living in urban area in 2011
rural	ratio of people living in rural area in 2011
lmp05	ln(Domestic Market potential) in 2005
lmp	ln(Domestic Market potential) in 2012
lnrkl05	ln(capital/employment) in 2005
road05	% of village with asphalt road
agn05	ratio of agriculture sector to output of district in 2005
agn12	ratio of agriculture sector to output of district in 2012
djava	dummy for in Java island (1=java; 0=non-java)
nofdi05	dummy for zero FDI in 2005
nofdi	dummy for zero FDI in 2005-2012
fdcap	FDI per capital
hschool	% of population with high school graduate
college	% of population with college graduate

A.2 First Stage Regressions

A.2.1 Full sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	lnrklna	lna5	lna	ms5lna5	ryms5	ryms	lmp	rfdlnan	ner12
area	0.00124	-0.00226	-0.0216	-0.00165	0.0000359	0.000316	0.000461	-6.748	-0.0203
	(0.34)	(-1.27)	(-0.80)	(-0.74)	(0.18)	(0.73)	(1.92)	(-0.75)	(-1.52)
area5	0.0197	0.0822	0.539	0.0587	-0.00318	-0.0136	-0.00794	-326.6	-0.187
	(0.20)	(1.38)	(0.61)	(1.12)	(-0.70)	(-1.37)	(-1.67)	(-1.18)	(-0.52)
doctor05	-0.720	0.239	3.480	-0.283	0.0260	0.0659	-0.0170	-2018.3	0.714
	(-0.94)	(0.41)	(0.45)	(-0.64)	(0.79)	(1.05)	(-0.29)	(-0.84)	(0.17)
ms5area5	-0.0591	0.00219	0.189	-0.0126	0.00110	0.00283	-0.00228	548.0*	0.650*
	(-0.88)	(0.04)	(0.29)	(-0.23)	(0.27)	(0.28)	(-0.40)	(2.38)	(2.17)
road05	-0.0549	-0.0183	-0.0854	-0.0113	-0.00241	-0.00557	-0.00158	-61.06	-0.0267
	(-1.84)	(-0.81)	(-0.32)	(-0.57)	(-1.54)	(-1.41)	(-0.61)	(-0.81)	(-0.16)
agn05	0.0390	0.000590	0.207	0.0157	-0.00101	-0.000748	0.000492	29.81	-0.0570
	(0.84)	(0.03)	(1.06)	(0.97)	(-0.70)	(-0.25)	(0.24)	(0.52)	(-0.31)
lnrkl05	0.556**	-0.0112	-0.379	-0.0279	0.0133	0.0222*	-0.00628	-320.6	1.115
	(3.11)	(-0.16)	(-0.42)	(-0.38)	(1.75)	(2.05)	(-0.47)	(-0.98)	(1.93)
ryms05	-0.478	-2.514**	-30.91**	-1.360	0.183*	1.232***	0.00799	11058.1	-2.512
	(-0.38)	(-3.06)	(-3.38)	(-1.47)	(2.45)	(6.35)	(0.12)	(1.65)	(-0.42)
ryms055	2.554	3.987*	47.35*	5.940**	0.587***	-0.600*	0.00669	-23909.5	0.112
	(0.88)	(2.46)	(2.24)	(3.26)	(4.55)	(-2.28)	(0.05)	(-1.67)	(0.01)
lmp05	0.542	2.145***	23.83***	1.350***	-0.0155	-0.0283	1.003***	3313.1	-3.012
	(1.74)	(7.69)	(5.94)	(5.60)	(-0.80)	(-0.68)	(55.76)	(1.68)	(-1.52)
ner05	0.0425	-0.0340	-0.297	-0.0173	0.000994	0.00521	0.00175	206.1	0.227
	(1.37)	(-1.99)	(-1.37)	(-0.88)	(0.48)	(0.98)	(0.78)	(1.41)	(1.51)
nofdi05	-0.176	-0.0631	0.919	-0.139	-0.0190	-0.0594	-0.0480	-1323.7	0.974
	(-0.38)	(-0.18)	(0.22)	(-0.46)	(-0.73)	(-1.13)	(-1.67)	(-0.98)	(0.33)
rfdln05	0.000191	-0.000196	-0.00207	-0.000523**	0.00000210	0.00000947	-0.0000207	7.038**	0.000858
	(0.74)	(-0.77)	(-0.60)	(-3.23)	(0.12)	(0.18)	(-1.23)	(2.68)	(0.50)
djava	-0.205	-0.750*	-8.153*	-0.366	-0.0157	-0.0734	-0.0526	-4241.3	0.979
	(-0.32)	(-2.17)	(-2.07)	(-1.04)	(-0.40)	(-0.80)	(-1.38)	(-1.80)	(0.28)
urban	0.0469	-0.0282	-0.452*	-0.0258	-0.000305	0.00153	0.00177	-2.531	-0.0539
	(1.12)	(-1.64)	(-2.01)	(-1.60)	(-0.21)	(0.49)	(0.78)	(-0.03)	(-0.30)
hschool	-0.0757	0.0502	0.607	0.0539	-0.00885*	-0.0259*	-0.0105	-340.0	-0.0289
	(-0.99)	(1.21)	(1.20)	(1.00)	(-2.02)	(-2.21)	(-1.88)	(-0.98)	(-0.07)
college	-0.00889	0.171*	1.369	0.0933	0.00564	0.0145	0.0217*	-201.1	-0.0251
	(-0.08)	(2.28)	(1.28)	(1.32)	(0.93)	(1.19)	(2.28)	(-0.46)	(-0.04)
_cons	-1.708	-29.27***	-343.3***	-20.17***	0.578	1.084	1.026*	-28377.1	87.76**
	(-0.35)	(-6.80)	(-5.33)	(-5.53)	(1.79)	(1.56)	(2.55)	(-1.30)	(3.02)
N	71	71	71	71	71	71	71	71	71

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.2.2 Excluding Jakarta

	(1) lnrklna	(2) lna5	(3) lna	(4) ms5lna5	(5) ryms5	(6) ryms	(7) lmp	(8) rfdlnan	(9) ner12
area	-0.000350	-0.00339	-0.0389	-0.00499	-0.000375	-0.000295	0.000499	-20.90	0.0219
	(-0.09)	(-1.14)	(-1.05)	(-1.82)	(-1.45)	(-0.50)	(1.23)	(-1.59)	(1.24)
area5	0.0726	0.115	1.069	0.117	0.00833	0.00464	-0.00855	333.0	-1.422*
	(0.56)	(1.36)	(1.08)	(1.63)	(1.17)	(0.30)	(-0.90)	(0.93)	(-2.44)
doctor05	-0.467	0.314	4.854	-0.445	0.0434	0.0899	-0.0204	778.5	-2.228
	(-0.64)	(0.50)	(0.63)	(-0.95)	(1.04)	(1.11)	(-0.31)	(0.42)	(-0.55)
ms5area5	-0.0427	0.0184	0.414	0.0662	0.00721	0.0110	-0.00320	539.4*	0.0671
	(-0.64)	(0.27)	(0.48)	(1.00)	(1.48)	(0.86)	(-0.41)	(2.61)	(0.21)
road05	-0.0603	-0.0193	-0.116	-0.0139	-0.00268	-0.00703	-0.00190	-61.48	-0.0910
	(-1.92)	(-0.77)	(-0.38)	(-0.61)	(-1.36)	(-1.40)	(-0.70)	(-0.79)	(-0.56)
agn05	0.0299	-0.00130	0.170	0.00316	-0.00224	-0.00286	0.000553	1.031	0.0240
	(0.63)	(-0.08)	(0.93)	(0.22)	(-1.77)	(-1.02)	(0.26)	(0.02)	(0.14)
lnrkl05	0.569**	-0.0339	-0.711	0.00396	0.0176*	0.0275*	-0.00699	37.13	0.645
	(2.88)	(-0.56)	(-1.01)	(0.06)	(2.35)	(2.68)	(-0.47)	(0.17)	(1.29)
ryms05	0.183	-2.053*	-23.80**	-0.266	0.253**	1.340***	-0.00337	7814.1	-6.782
	(0.16)	(-2.55)	(-2.86)	(-0.36)	(3.48)	(7.65)	(-0.04)	(1.56)	(-1.05)
ryms055	0.844	2.930	30.99	2.266	0.363*	-0.953**	0.0366	-14709.1	12.83
	(0.30)	(1.99)	(1.90)	(1.74)	(2.24)	(-2.94)	(0.17)	(-1.87)	(0.89)
lmp05	0.537	2.148***	23.86***	1.405***	-0.0390	-0.0608	1.004***	209.6	0.694
	(1.18)	(6.51)	(5.18)	(4.52)	(-1.50)	(-1.06)	(39.34)	(0.23)	(0.34)
ner05	0.0470	-0.0256	-0.173	-0.0106	0.00191	0.00658	0.00166	212.1	0.119
	(1.57)	(-1.45)	(-0.77)	(-0.58)	(1.01)	(1.42)	(0.71)	(1.64)	(0.78)
nofdi05	-0.0270	-0.0991	0.756	0.0255	-0.00107	-0.0125	-0.0428	-498.7	1.150
	(-0.06)	(-0.28)	(0.18)	(0.09)	(-0.04)	(-0.32)	(-1.42)	(-0.57)	(0.39)
rfdln05	0.000962	-0.000399	-0.00272	-0.0000188	0.0000667	0.000240	0.0000193	10.81	0.00655
	(1.19)	(-1.07)	(-0.60)	(-0.07)	(0.82)	(0.91)	(0.48)	(1.25)	(1.77)
djava	-0.0835	-0.614	-6.150	-0.187	0.0143	-0.0333	-0.0570	-2972.7	-2.840
	(-0.13)	(-1.53)	(-1.38)	(-0.49)	(0.36)	(-0.37)	(-1.47)	(-1.57)	(-0.85)
urban	0.0528	-0.0227	-0.365	-0.0250	0.000625	0.00297	0.00170	65.97	-0.177
	(1.17)	(-1.23)	(-1.52)	(-1.60)	(0.43)	(0.98)	(0.70)	(0.86)	(-0.90)
hschool	-0.0828	0.0542	0.637	0.0570	-0.0105*	-0.0298*	-0.0110	-559.3	0.0719
	(-1.01)	(1.40)	(1.35)	(1.37)	(-2.34)	(-2.49)	(-1.91)	(-1.85)	(0.18)
college	-0.0370	0.126	0.772	0.0675	0.00518	0.0205	0.0248*	164.7	0.495
	(-0.32)	(1.48)	(0.66)	(0.97)	(0.68)	(1.38)	(2.28)	(0.55)	(0.77)
_cons	-1.641	-29.48***	-345.5***	-20.62***	0.793*	1.447	1.040*	318.7	59.75*
	(-0.32)	(-6.73)	(-5.25)	(-5.13)	(2.40)	(1.97)	(2.58)	(0.03)	(2.22)
N	66	66	66	66	66	66	66	66	66

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.2.3 Test of Excluded Instruments

Table 10: Excluded Instruments Test

	lnrklna	lna5	lna	ms5lna5	ryms5	ryms	lmp	rfdlnan	ner12
rmse	1.33	0.91	11.87	0.85	0.08	0.17	0.10	5432.54	7.04
sheapr2	0.41	0.40	0.18	0.17	0.46	0.50	0.40	0.31	0.22
pr2	0.70	0.88	0.81	0.88	0.96	0.95	0.99	0.68	0.29
F	8.85	29.55	15.92	17.05	121.50	45.31	1825.13	7.86	1.48
df	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00
df r	53.00	53.00	53.00	53.00	53.00	53.00	53.00	53.00	53.00
pvalue	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14
SWF	5.52	6.13	2.22	1.11	2.92	3.71	4.52	1.23	2.09
SWFdf1	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
SWFdf2	53.00	53.00	53.00	53.00	53.00	53.00	53.00	53.00	53.00
SWFp	0.00	0.00	0.03	0.37	0.01	0.00	0.00	0.30	0.05
SWchi2	66.52	73.85	26.82	13.37	35.18	44.67	54.48	14.82	25.17
SWchi2p	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.10	0.00
SWr2	0.39	0.30	0.10	0.11	0.32	0.19	0.20	0.13	0.18
APF	6.47	1.01	0.53	1.41	3.33	6.72	135.42	13.80	2.06
APFdf1	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
APFdf2	53.00	53.00	53.00	53.00	53.00	53.00	53.00	53.00	53.00
APFp	0.00	0.44	0.84	0.21	0.00	0.00	0.00	0.00	0.05
APchi2	78.06	12.21	6.43	17.00	40.20	80.99	1632.68	166.41	24.79
APchi2p	0.00	0.20	0.70	0.05	0.00	0.00	0.00	0.00	0.00
APr2	0.40	0.07	0.03	0.08	0.27	0.29	0.91	0.30	0.21

Table 11: Excluded Instruments Test: excluding Jakarta

	lnrklna	lna5	lna	ms5lna5	ryms5	ryms	lmp	rfdlnan	ner12
rmse	1.27	0.85	10.72	0.75	0.07	0.16	0.10	3714.14	6.90
sheapr2	0.45	0.41	0.21	0.25	0.62	0.64	0.40	0.48	0.29
pr2	0.72	0.86	0.79	0.90	0.97	0.96	0.99	0.66	0.32
F	9.33	18.56	8.22	24.24	132.84	53.87	647.45	8.54	1.59
df	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00
df r	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00
pvalue	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
SWF	6.65	4.95	2.61	2.06	8.04	4.54	4.70	2.18	2.16
SWFdf1	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
SWFdf2	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00
SWFp	0.00	0.00	0.02	0.05	0.00	0.00	0.00	0.04	0.04
SWchi2	82.30	61.28	32.24	25.53	99.44	56.23	58.11	26.92	26.67
SWchi2p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SWr2	0.41	0.35	0.16	0.20	0.45	0.34	0.26	0.31	0.24
APF	8.74	0.69	0.53	2.18	4.43	6.08	115.17	5.37	2.46
APFdf1	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
APFdf2	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00	48.00
APFp	0.00	0.71	0.85	0.04	0.00	0.00	0.00	0.00	0.02
APchi2	108.17	8.54	6.52	26.94	54.81	75.28	1425.20	66.43	30.46
APchi2p	0.00	0.48	0.69	0.00	0.00	0.00	0.00	0.00	0.00
APr2	0.44	0.08	0.04	0.10	0.37	0.38	0.88	0.39	0.27