Vrije Universiteit Amsterdam

Explaining Urbanization and Airport Connectivity in Spatial Context

Master Thesis

Master of Spatial Transport Regional Environment Economics,
Vrije Universiteit Amsterdam

Nerfertari Pramudhita
2699855

Supervisor: dr. Eric Koomen

Abstract

Indonesia is an archipelagic country where air travel is vital in ensuring the movement of people in between regions hence the impact on the urban population. This research provides further explanations regarding the spatial relationship between accessibility that airports offer and the change in urban population. This thesis explores a unique dataset of origin-destination of air passengers in order to examine the spillover impact of airport size change on the urban population to the first-order neighbor on the flight network. Using an instrumental variable introduced by Sheard constructed with the origin and destination passenger flow dataset. The estimate find that basic OLS highly underestimates the impact of airport size change. Considering spatial correlation using SLX model have only negligible change to the direct impact. We found a negative impact on the one that resulted from the first order of the neighbor, although it was not significant.

Research question: How airport and spatial correlation from connectivity affect urban population of the city?

Table of Contents

Introduction

Transportation infrastructure such as airport has impact to the growth of cities. Its impact on land use and promoting economic activity has widely been researched. The main argument in looking at the impact of transport infrastructure in urban growth is increased accessibility. The accessibility in question is decreasingly easy to move things around. In the case of the airport as a transport infrastructure, accessibility is referred to mainly caused by air travel connectivity opens access to markets, capital, ideas, and people. (Rietveld, 1995).

This ease of travel then has an impact on population changes in urban areas served by the transportation infrastructure. The impact on infrastructure is harder to estimate because of the endogeneity problem between population and the accessibility measure of transportation infrastructure. At airports, the most commonly used measure to measure airport accessibility is the size of the airport as it measures the ease of access one airport has(Brueckner, 2003; Green, 2007; Sheard, 2019). However, this size is very likely to be influenced by the number of populations served by airports, because the larger the population served by the airport, the greater the possibility of travel using air transportation. In Sheard, (2019), using his novel instrumental variable to account for that endogeneity problem it has. the increase in accessibility seen from the number of passengers at the airport caused an increase in the population with an elasticity of 0.04. This estimate also in line with the results of Green, (2007) and Blonigen & Cristea (2012) estimate the effect of airport size on local economic growth and find positive effects.

The impact that is caused by the transportation and its accessibility to urbanization does not come in vacuum, rather it is more about the accessibility that it provides to other urban area. This implies that accessibility will not only affect urban growth in the areas served by the airport but also the areas connected by its existence. The existence of this accessibility effect has long been acknowledged as a network of cities to see how cities influence each other.(e.g. Freeman et al., 1979; Friedmann, 1986; Sassen, 2005). However, the connectivity that caused by the infrastructure itself is rather assumed than it is explored. This is a problem because ignoring spatial aspects on the impact of transportation infrastructure can cause bias in the estimates (Margaretic et al., 2017).

In Indonesian context, the development of transportation infrastructure in the next few years is the flagship program of the Indonesian government. The Indonesian government aims to have a positive impact as shown by the improvement of equitable regional economic conditions.

Transportation infrastructure is divided into land, sea and air transportation. One of the strategic transportations sees the shape of Indonesia which consists of islands is air transportation. Until now, the government has placed an emphasis on the development of transportation infrastructure, especially air transportation[1].

In the domestic sphere, air transportation in Indonesia is the most widely used mode of transportation for intercity travel. This is especially true considering the nature of geographical shape of Indonesia that consist of islands. In 2016, air transportation users reached 160 million people, 5 times greater than train passengers with 31.8 million passengers, and 10 times greater than 14.4 million ship passengers. In terms of growth, air transportation also has a better growth compared to other transportation modes value than train and ship[2].

It is important to see how the influence of this infrastructure on urbanization in Indonesia. Indonesia, as an archipelagic country, air travel is vital in ensuring the movement of people in between regions hence the impact of urban population. Moreover, as a developing country, Indonesia level of urbanization has not yet touched the steady state as most of the region are still highly industrialized. Urbanization determinant in such countries such as still a big question in development. Hofmann & Wan, (2013) have shed light on the positional impact of transportation in urbanization. However, Menashe-Oren & Bocquier, (2021)have begun to argue that migration is not the main factor in urban growth, especially in low-middle income countries but rather natural growth. Moreover, there is still little research on the airport network and its flow at the regional level, especially in Indonesia.

This study, use change of the airport size as the measure of accessibility and change of population as the measure of urbanization. The areas in this study are including 95 airports in Indonesia and their respective served urban area. I used public data that was publicly available in 2015. I overlay the airport against the urban area defined by GHSL SMOD. This is done in order to avoid defining the city in the form of administrative boundaries, since the size of the city is often not depicted with administrative boundaries. I used the instruments introduced by Sheard, (2019), to address potential endogeneity on models of airport size and the number of urban populations. I then looked at the impact of spatial correlation in fixing the model. I utilize

---

[1] Looking at the middle Medium-Term National Development Plan 2020-2024 (in Indonesian: Rencana Jangka Menengah Nasional (RPJMN) 2020-2024)

[2] Based on Statistik Transportasi Indonesia tahun 2016

the SLX model to look at spatial correlation between to the urban population that is bring by the airport. To look at the spatial correlation I employ spatial weight matrix using Origin-Destination (OD).

The main contributions of this research to the existing literature, is to look closely at the impact of accessibility and connectivity in spatial point of view. This thesis tries to understand the spatial impact of the connection of airport make a difference to the estimation that disregard it. Much research that concerned about airport expansion on local employment and other economic outcomes, but not looking into the impact of connectivity it has (e.g. Blonigen & Cristea, 2012; Brueckner, 2003; Green, 2007; Sheard, 2019). This research provides a new method to highlights the importance of considering connectivity in measuring the impact on urbanization as a spatial manifestation of transportation infrastructure while also taking account of the endogeneity problem accessibility and urbanization might have. This thesis also contributes to the growing literature of airport impact in developing country and sparse geography such as Indonesia.

In the next section thesis is derived from several part namely: (1) Models, where I explain the models used; (2) Data source and cleaning, where I explained about the data source, the formation of datasets and the problems that occurred including; (3) Model estimation, where the model is then estimated using the previously created data set; lastly (4) Conclusion, where I summarized the key findings, exposed the weaknesses of the research and provided future opportunities that might be done after this study.

Models

On urbanization.

Urbanization by itself is defined in different way in different literature. It is about population density, the industrialization or at the same time the lack of agriculture, and it is about the concentration of resource. In this thesis I will only explain about the distribution of population in urban area in Indonesia that has direct access to airport. I am trying to understand whether the accessibility does have an impact on the change of the population in Indonesia.

I recognize that the existence of airports is not randomly distributed. As mentioned earlier, airports are often also seen as catalysts for regional growth. I assume that urban populations can be randomly dispersed; agglomerations can occur anywhere, and variances can be explained by our explanatory variable. As done by Dobis et al., (2019) and Sheard, (2019), I use the change of urban population from 2000 to 2015 as a form of urbanization as the rate of urbanization influenced by the size of the city. I controlled the size of the city by using the urban population in 2000 and the size of airport in 2000. The control is designed to absorb the heterogeneity of different level of urbanization in the base year of 2000, as the city becomes oversaturated so that it has slower population growth. As a developing country, especially Indonesia there is a tendency for a population to move to an already densely populated region (Wajdi et al., 2017). The level of urbanization in Java and regions outside Java also has differences. To accommodate this heterogeneity, I introduced the Java-Bali regional dummy and other regions.

The models that are constructed to look at the relationship between airports and urbanization is as follows:

$$\begin{aligned}\log(N_{2015}) - \log(N_{2000}) \\ &= \beta_1 \log(N_{2000}) + \beta_2 \log(A_{2000}) + \theta\,(\log(A_{2015}) - (\log(A_{2000})) \\ &+ factor(Jawa) + \varepsilon\end{aligned} \tag{1}$$

Let A be airport size expressed in total passenger that are going to, from, and transited at the airport, θ as the main coefficient of interest, and N be population size of the urban area that is served by the corresponding airport. As mentioned in the introduction, I am using a novel urban border to accommodate sprawling of urban area beyond the administrative border. The technicality of deriving the population will be discussed in the later section. We are using the log transformation in the model so the estimation will be interpreted as elasticity.

Instrumental variable

Following the use of instrumental variable by Sheard (2019). The instrument variable I use in constructed by taking the number of local air traffic that is doing to and from the certain category of the airport, then applying the national growth rate of passenger traffic for that category to create a sum of local traffic of the airport. Category that is used in the instrument designed to capture improvement of airline technology such as distance flown, aircraft class, and flight company as the airport size. This instrument works as an exogenous variable to equation (1), by excluding the local traffic in the variable design eliminating the possibility of local growth to be reflected in the instrument. In Sheard (2019), the instrument is successfully capture the effect of change airport size using period of one year. However, it is reasonable to expect the effect of infrastructure to have a measurable impact on a longer period. The main benefit of this instrumental variable is that I do not need to assume the exogeneity of a certain hand-chosen discontinuity.[3]

In this thesis, I used the airport hierarchy category. Airports have different levels of service. This makes the accessibility provided by the airport also different. This difference also gives a difference in the impact that each airport has (Button et al., 2009). However, the different level of service is solely dictated by the regulation that is not related to the state of urbanization of the city. This chosen category is close to the instrument category aircraft class used by Sheard, as the airport class indirectly dictates the type of aircraft that can fly from and to the airport as embedded by the relevant regulations. The relationship between airport size changes and instrument variables that constructed can be formally shown as follows:

$$\log(A_{2015}) - \log(A_{2000}) = \eta(\log(A_{inst}) - \log(A_{2000})) + \beta_1 \log(N_{2000}) + \beta_2 \log(A_{2000}) \tag{2}$$

To check for the validity of the instrument, it is essential of an instrument are that it satisfies the relevance condition, and exogeneity condition. For relevance condition, which requires that the instruments explain a significant amount of the variation in airport sizes, conditional on the controls. This condition is tested statistically in the first-stage estimation as part of the IV regression estimation to show $\eta \neq 0$. Weak instruments the null hypothesis is: *"All instruments are weak"*. A rule of thumb requires to soundly reject the null hypothesis at a value of the F-statistic greater than 10 or, for only one instrument, t-statistic greater than 3.16, to make sure that an instrument is strong.

Exogeneity condition where it requires instrument to affect change in population only trough change in airport size. I check the exogeneity using Hausman test for endogeneity, where rejecting the null hypothesis indicates the existence of endogeneity and the need for instrumental variables. Moreover, I argue that this instrument is valid because several support evidence. Change of growth in different airport hierarchy that will discussed in later part.

[3] Instrumental variable from other research: (Brueckner, 2003) uses hub status to instrument for airport size, as this implies a larger number of incidental travellers. (Blonigen & Cristea, 2012) use the 1978 deregulation of US air travel to explain variation in air traffic levels.

Airport hierarchy is not corelated to number of flight that the airport has rather just the type of passenger: local (within country), international, transit. With excluding the own size growth, the difference in the overall growth rates within the category, this could not be driven by size of population in the local area.

Sheard use of the controls for the first stage as shown in equation (2). The control is intended to capture the change to corelated to value at the base year. However, it also raises a concern of biased estimation as biased β for the control will also make η bias, hence biased the whole instrument system. The estimate of θ for equation (1) with or without controls to be similar is shown in Appendix.

Spatial Correlation

In looking at the impact of accessibility we also need to look at connectivity spatially. The connectivity referred to in this thesis is the ability to connect with each other due to the existence of an airport. Using the OD data we have, we can see how the flight route occurs as a form of connectivity. This can be captured by implementing spatial correlation on models that did not previously pay attention to spatial relationships. Using the SLX model in spatial correlation, we can see the impact of changes in the size of the airport not only on itself but on the airports connected to it.

Construction of the model, I follow Dubin (2003) for Spatial Lag X model[4]:

$$\log(N_{2015}) - \log(N_{2000}) = \alpha + \theta_1(\log(A_{2015}) - \log(A_{2000})) + W\theta_2 \log(A_{2015}) - \log(A_{2000}) + \beta X + \beta W X + \varepsilon \tag{3}$$

Where in this model, β is the coefficient of direct impact X, and θ is the coefficient to variable of interest, change of airport size, X is the other control variable, W is the spatial weight matrix,

[4] I first consider the full model (Manski Model):

$$Y = \rho W y + X\beta + WX\theta + u, \qquad u = \lambda W u + \varepsilon$$

I also tried to Spatial Lag SAR model ($\theta = \lambda = 0$). I find no significant ρ in SAR model and no spatial autocorrelation in error term on OLS model ($\rho = \theta = \lambda = 0$) as discussed in later part. Considering the finding, equation (3) is based on:

$$Y = \alpha + \beta X + WX\theta + \varepsilon$$

Hence, this thesis looks more carefully into the impact of only first order lag using SLX.

α is a constant term, and ε is an error term. As W in SLX model is not autocorrelated with itself shown in equation (3) the impact of change in size of airport is now also impacting the first order neighbor; every neighbor of one airport (value of W = 1) will also impacted by its size change by θ_2. To see the causation effect on spatial regression I also implement 2SLS for spatial model. This can be achieved with implementing equation (2) to equation (3), the model in 2SLS for spatial regression.

To explore spatial correlation, matrix of spatial weight needs to be constructed. This matrix constitutes of neighbor relation and weight. Study by Margaretic et al., (2017) on urbanization and travel by plane have been carried out using the K nearest neighbor (KNN) as their matrix spatial weight (e.g. However, the use of this spatial weight assumes that the nearest airport will always be connected. In reality, airports are not always connected according to their proximity in distance that is implied in KNN but rather bounded to their flight routes. In consequence, analyzing accessibility of the airport by looking at the proximity to the other airport is not Travel using air transport is more widely preferred for long-distance travel (Suzuki & Audino, 2003). Because the data I use is domestic data, making the distance between airports may be very close and can be replaced with other modes of travel especially for the case of Java Island. For this reason, using KNN for generating spatial weight in our data becomes unappealing for our purposes.

I used OD data to generate spatial weight for SLX model. The limitations of the OD matrix that I have, as well as how I construct the matrix spatial weight for this regression needs are discussed in the next section.

Before doing spatial regression, I do Moran's I test to make sure that whether our model residual of the basic OLS model is significantly different form a value of spatial random distribution. I then take a Lagrange Multiplier diagnostic for spatial dependence in linear models to for spatial dependence in linear models. (Anselin, 1988).

To explore the impact of distance, I use alternative spatial weight matrix by using distance as the weight. This approach taken to identify the impact of inverse distance between airport to the change of population. This becomes relevant because distance indirectly also affects the ticket price for the flight, the farther the flight, the higher the applicable fare. This makes the impact of accessibility different. However, this model will only be limited to the impact of distance as this approach is also coming with problem that the lagged parameter can hardly be interpreted appropriately[5]

Data source and cleaning

Main interest in this thesis is to look at the effect of airport and urbanization. However, urbanization is borderless it is hard to take a complete account of what happen especially using the traditional data. I responded to this problem by creating a unique data set as a combination of airport, urban area and flow. I use data from a variety of sources that I can find. This section attempts to explain how I generate data that I then estimate on the model described in the previous section.

Airports

Based on the model that I described earlier, the airport data that I need namely are number of passengers, location of the airport, and the type of hierarchy within the airport. I gather our data on the number of passengers on the airport through Transportation Statistic 2000 (in Indonesian: Statistik Transportasi tahun 2000) and Air Transportation Statistics 2015 document (in Indonesian: Statistik Transportasi Udara tahun 2015) of the Ministry of Transportation than published by Central Agency on Statistics (in Indonesian: Badan Pusat Statistik). Our airport location data is gathered manually from OpenStreetMaps with a list of airport names contained in the document.

Based on data from the Indonesian Ministry of Transportation in 2015, it has 341 airports. 38 of them are military airports that do not serve commercial flights. 25 of them are international airports managed by PT Angkasa Pura. There are some blanks in the data I had in 2015. I

---

[5] With traditional SLX model, the weight matrix only indicates whether the 2 indices are neighbors hence it will always be a binary value: neighbor or not a neighbor. To introduce distance to the equation I can simply do matrix multiplication between traditional weight and distance matrix. Following SLX model, the new equation for introducing distance will be:

$$Y = \alpha + \beta X + (W_1 \times W_{distance})X\theta + \varepsilon$$

Direct impact, , β, we can still interpret with the same manner. However, the value θ is no longer the value of direct impact to the neighbor but also the distance. This model raises a problem of interpretation as. θ is no longer representing the value of immediate neighbor but rather also the distance between the nodal. As for the interpretation that can be given is that the change in X to the neighbors with a distance of 1000km between them will result on the change of urban population at the amount equal to change in X times $\theta/1000$.

successfully gather 136 airports that can be identified by IATA and serving a civil aviation. In 2000, Indonesia only has 72 airports. 36 of it are military airport that is also serving civil flight. This number of airports is a big change form the number in 2000. As we are interested in the change of airport size to urban population, this thesis only focuses on airports and urban area that are exist in both year 2000 and 2015. Spatial representation of the change of airport between 2000 dan 2015 can be seen on Figure 1.

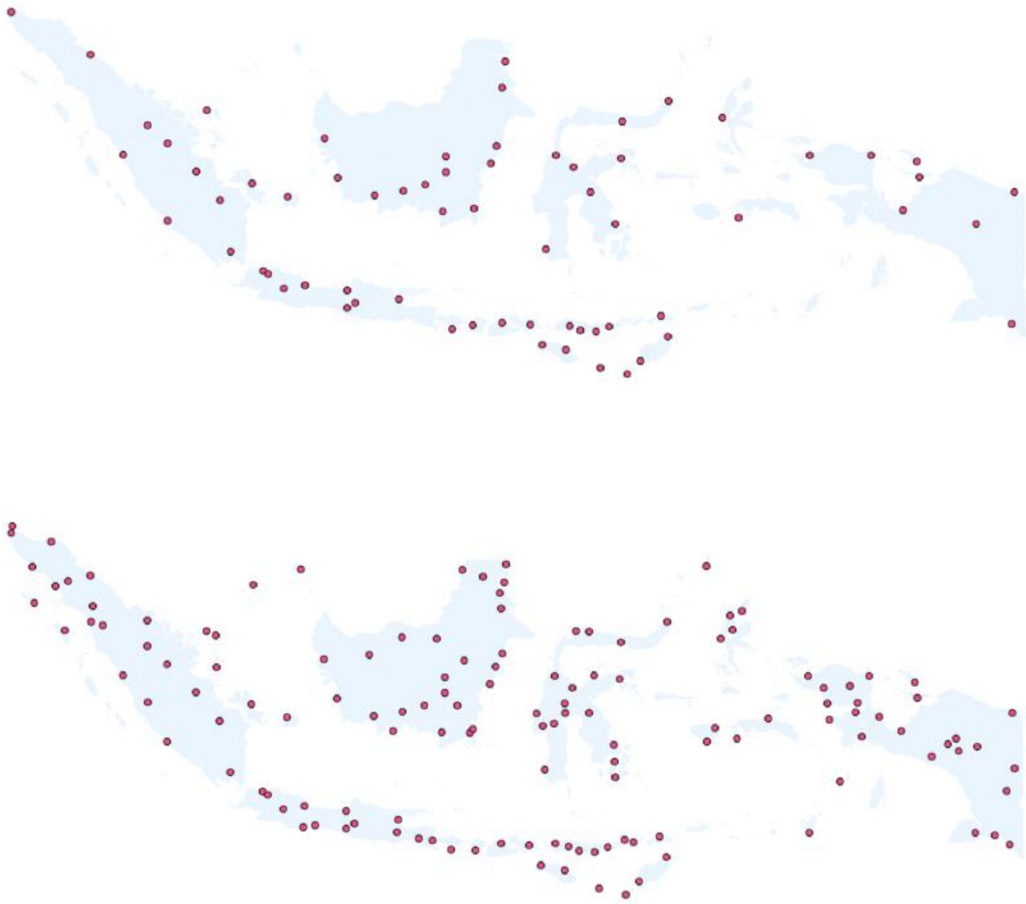

Figure 1. Spatial representation of Airports in 2000 (top) and 2015 (bottom).

Hierarchy in airport analysis is important as it has been highlighted by Button et al., (2009). I provide a unique hierarchy of the airport. I categorize airports in 3 hierarchies: international hubs, domestic hubs, and other airports. This categorization is carried out using passenger data: 1) international hub if there are international passengers, 2) domestic hub if there is data about domestic passengers transiting at the airport, and 3) other airports if the airport does not meet the criteria mentioned above. Through the categorization of this airport, I have 25 international hubs, 33 domestic hubs, and 77 other airports. This categorization regardless of the class of airports embedded by Indonesia there is a classification of airports according to their class according to the Regulation of the Minister of Transportation No. 4 of 1995 for airport class in the year 2000 and Regulation of the Minister of Transportation No. 6 of 2008 for airport class in 2015. The group the hierarchy regulated according to the scale of service, airports in Indonesia are divided into two, namely international/domestic airports and domestic airports. Differences in class and scale of service certainly provide differences in activities within the airport and the surrounding area. As this thesis are concerned about the ability of the airport to serve the passenger flow, I find that employing the legal term with the problem that has been discussed, will be appropriate.

To accommodate the regional heterogeneity in Indonesia, I grouped the airports and cities it serves in the Java-Bali region and other regions. As in the previous study there are striking differences between the Java region and outside Java (Kameswara & Suryani, 2021). Ignoring this may give bias to the estimates made. The spatial representation of the airport distribution used in this study can be seen in the Figure 2.

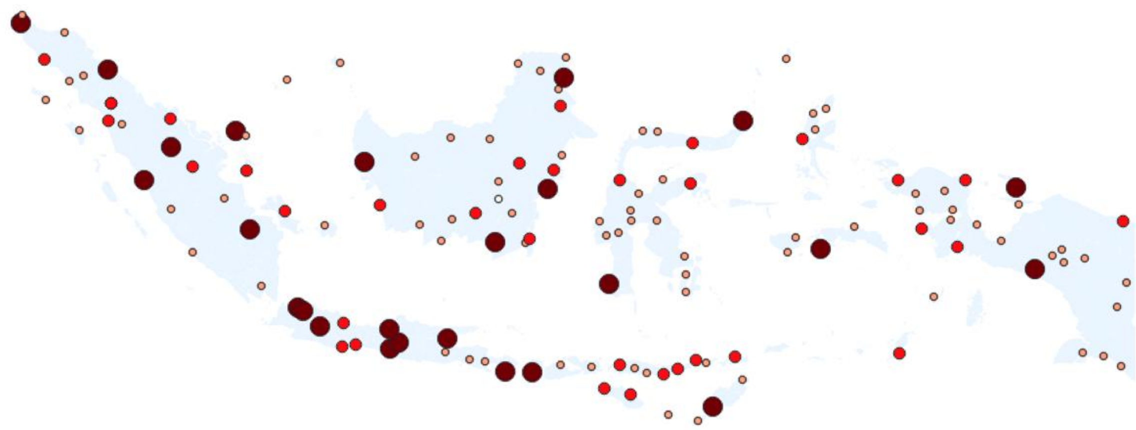

Figure 2. Spatial representation of Airports in 2015

Dark red (largest): 1st Hierarchy; Red: 2nd Hierarchy; Yellow (smallest): 3rd Hierarchy

The Urban Area

To measure the impact of an airport is hard as urban area defies the administrative border more often than not. This cities agglomeration makes it difficult to put the border of the impact of the airport. This presents difficulty in measuring the influence of urbanization precisely because often the data owned by that bound to the administrative border. Our analysis is based on airport and urbanized area that is linked to the airport. The advantage of using this data set is that I are not bounded to the predetermined cities boundary as it most likely to be spilled over the administrative boundary (Iqbal, 2021). I define unit analysis as an urban area that served by an airport. To account for the heterogeneity for the level of urbanization in each city, I introduce population size as the urbanize population that is derived from GHS Settlement Model[6] for the year of 2015 and 2000. This data set is chosen over the traditional population data because several reasons. First, the traditional population data are based on administrative boundaries. This attribute is problematic for our purposes as it could not be compiled in both years due to the large number of administrative border change that occurred. Second problem is the definition of a city using the administrative boundaries used is not sufficient to explain the urbanization that occurs because urbanization occurs across administrative boundaries. I anchored the urban boundaries of the area on the GHS Settlement Model grid data. This model uses the definition of urban center with Cluster Population lower bound of 50000 people and Local Population Density lower bound of 1500 people/km2 (Pesaresi et al., 2019). I used the data of 2000 and 2015 to measure population changes and area build ups that occurred in urban areas in Indonesia.

Integrating urban area, airport and other socio-economic data

This research emphasizes more on the relationship between cities or metropolitan areas of rather than flow between airports by itself. This is important to note because the metropolitan area in Indonesia, especially on the island of Java is so large that the boundaries between cities are irrelevant in separating airport service areas. Most of the socio-economic data are based on administrative border. To find out the administrative boundaries, here I used the administrative boundaries of the city area in 2013-2015 which were carto metric survey is carried out in 2016-2017, which are available on the website BIG[7].

I specify airports service area as the urban area within 30 kilometers. This is done to see which urban area are served by the airport. In a perfect world, the boundaries of this service area can be analyzed using the road network as done by Sellner & Nagl, (2010) to see the effect of airport accessibility on its occupancy level. However, I do not have enough information regarding the city's accessibility to the airport; not only using land mode but also sea mode

[6]https://ghsl.jrc.ec.europa.eu/ghs_smod2019.php. The number derived in this model are based on estimation of population. This might cause the numbers that are not round on population variables that I use here.

[7]https://geoservices.big.go.id/portal/apps/webappviewer/index.html?id=9917592df1f24501ae804b7d346c08fb

because the airport is on a different island from urban areas. The boundary of the airport service area was chosen to accommodate the agglomeration that occurred and was not too large to make the urban area too far away which gave inflation to the number of urban residents served by the airport. I use urban areas as the boundaries of the airport service area. It is noteworthy that the spatial data I have are in different projection. I am however pick ESRI:54009 - World_Mollweide projection to analyze the spatial data. The process carried out is as follows:

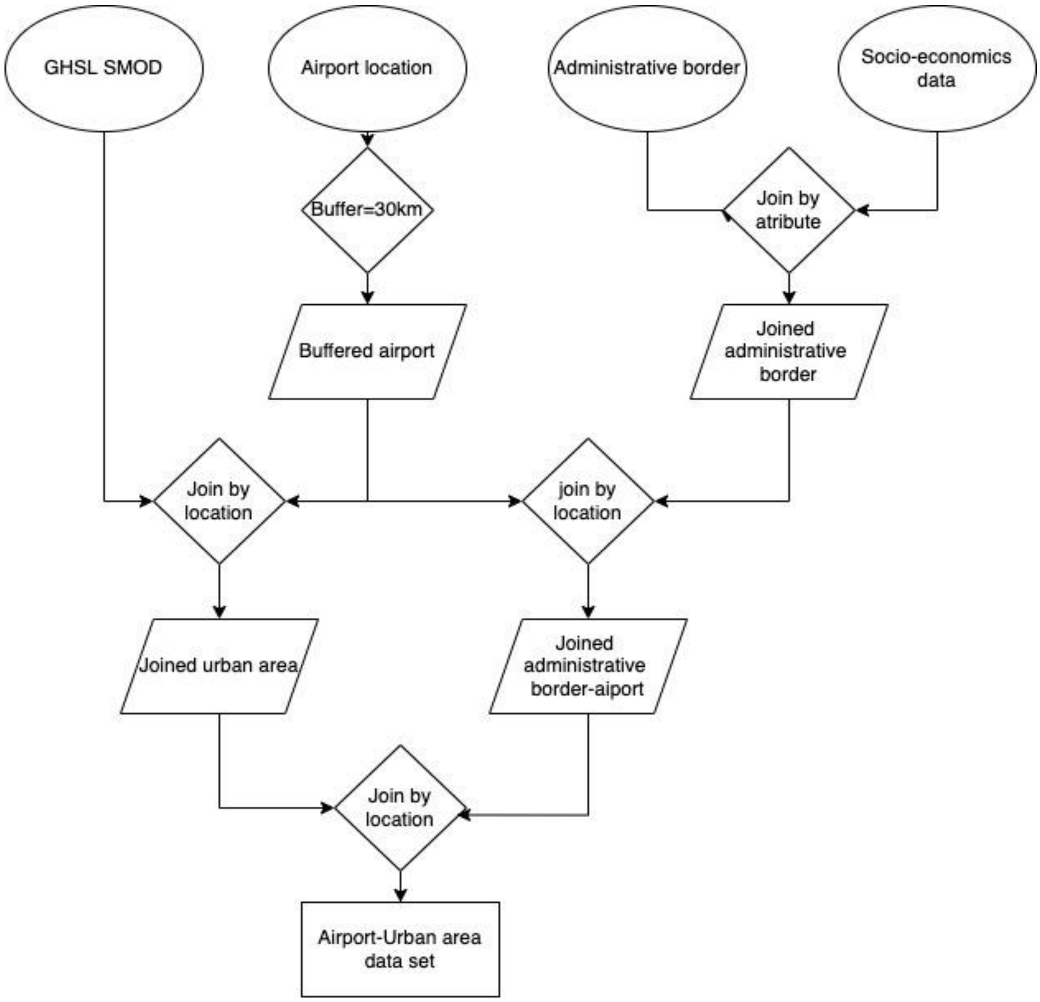

Figure 3. Flow diagram of integrating spatial data and socio-economic data

Problem with Dataset

There are some problems that occur in this process for the completeness of the data set: (1) some airports that are not in urban areas, (2) more than one urban area on one administrative border. (3) one airport on one administrative border or more than one airport in one contagious urban area.

For the first problem I solved it by picking the administrative border that the airport and its buffer are in. This will provide an urban population within the administrative border. This action is taken as we noticed that most airport is not located within an urban area. For that reason, unless the urban area has already sprawl far beyond its original state, the location of the airport will not be within an urban area.

For this second problem, I solved it by performing dissolve tools in any urban area within the airport service area and treat it as one urban area. This action is taken with realization that urbanization does not happen in a flat surface, meaning there might be some place that harder to populate than the others. This will cause a sparse urban area that can as well be one contagious urban area if there is no such obstacle.

To solve this, I pick the highest hierarchy of the airport that is in that particular urban area. This is done because I assume that people are more willing to spend more time to get to the bigger airport as it serves better connectivity. Local airports often become unpopular because of their narrow choice of route, and cause airport leakage: people will still choose a larger airport even though it is 3-4 hours away from where they live (Suzuki & Audino, 2003). Finally, in the case of Jakarta, administratively, the city of Jakarta only has one airport, namely Halim Perdanakusuma airport (HLP) However, the urban area that they have also includes Bandara Soekarno Hatta (CGK) which has the same hierarchy. For this problem I summed the interaction of both airports. I get 134 rows of airport of which 86 of them are in the urban area.

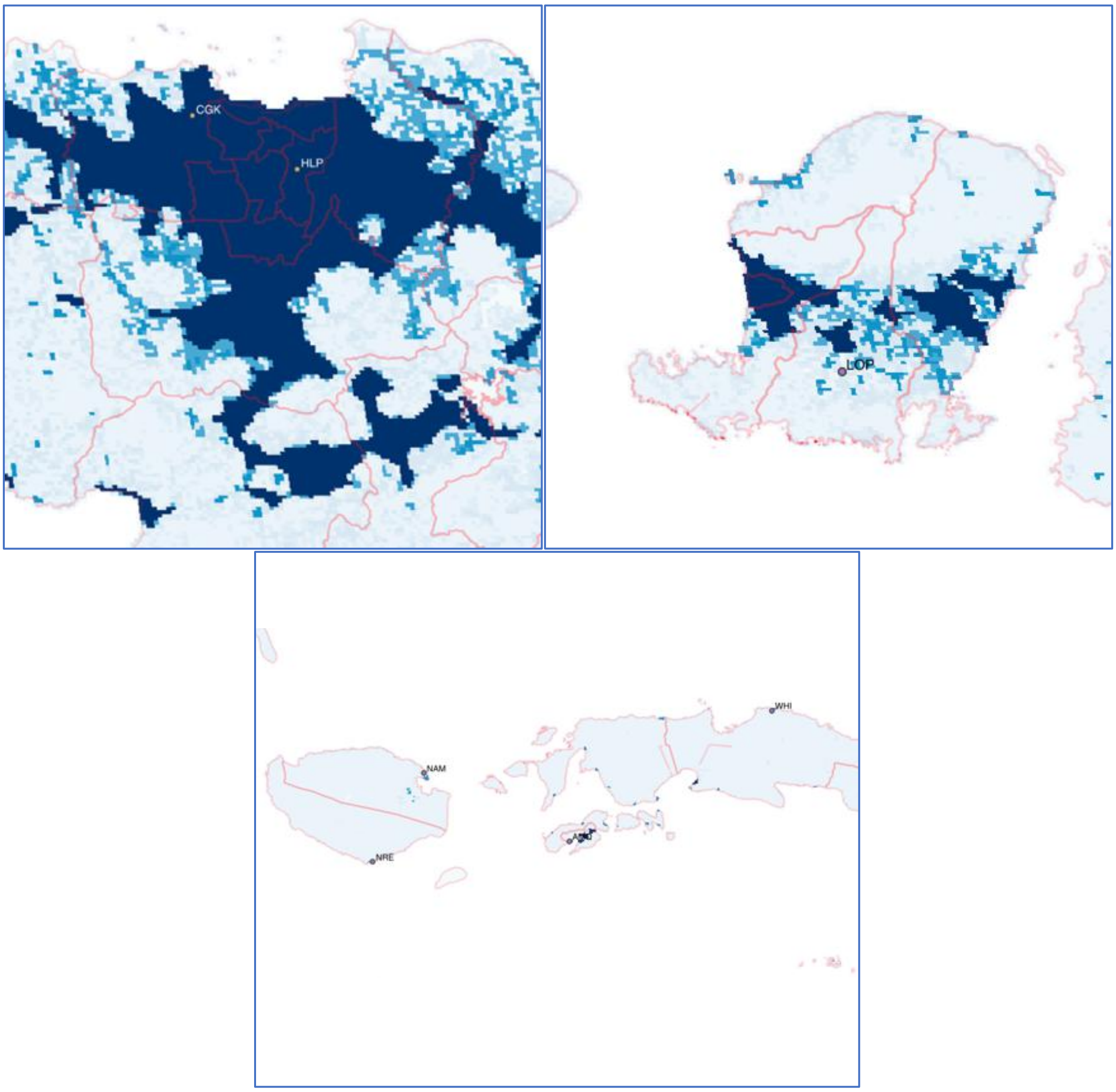

Figure 4. Example case in urban area and airport

Case for Jakarta where it has more than 1 airport. And the urban area (in dark blue)is beyond the administrative border (in red). Case of Lombok, the airport is not on any urban area. Case of Maluku the nearest urbanized place is Ambon, but the other airport is not urbanized.

Instrumental Variable

Instruments variable used in this research are constructed using categories from airport hierarchy. Using Sheard (2019) methods, the instruments are constructed by dividing up the air traffic in a metropolitan area by the airport hierarchy categorization using the regulation that we have discussed in the last part. We then calculating what the traffic the end of the period which in our case is 2015. Local flow of the airport is not included to the overall growth of the category to prevent the changes resulted from an increase in the population in the urban areas served by the airport. Formally, the instrument for the change in air traffic in airport is shown in equation (3), let the n be the airport, m is other airport in the time span, and c be hierarchy as the category I used.

$$A_{n,instrument} = \sum_{c} A_{c,n,2000} \left(\frac{\sum_{m\neq n} A_{c,2015}}{\sum_{m\neq n} A_{c,2000}}\right)$$

(3)

Based on the number of urban populations, I make a threshold for airports with an urban population above 30000 urban population in their service area. After unifying all the data and purging the incomplete data of airport and their respective urban area, I thus have a data set of 95 airport. Table 1 shows the description of the dataset used for analysis in the later part of this thesis.

Table 1. Descriptive data of airport and served urban area

Statistic	Mean	St. Dev.	Min	Max
Size_2015	1,568,421	4,762,589	90	41,889,868
Size_2000	177,186.	602,181	0	4,488,019
Hierarchy	2.227	0.823	1	3
POP_2000	785,899	2,467,451	0.000	21,224,596
POP_2015	1,108,108	3,920,819	54,656	36,337,674

Origin destination (OD)

The data used is data on the movement of aircraft passengers based on the origin and destination of the flight. Flight data was obtained from air transport traffic statistics in 2010 and 2015 published by Ministry of transportation through the official website of the statistical data of the Ministry of Transportation of the air section. The data I have does not provide an explanation of the route taken by passengers but only information about the origin airport and destination airport. The data on passengers who transited was not clearly known the airport of origin and destination There is no way of knowing whether the trip has a transit or not. This makes our data assume that any trip recorded on this document is a direct flight without transit. With the limitations already mentioned, I have 626 pairs of airports with nonzero flow for 2015 and 254 for the year 2000. Spatial representation can be seen in Figure 5.

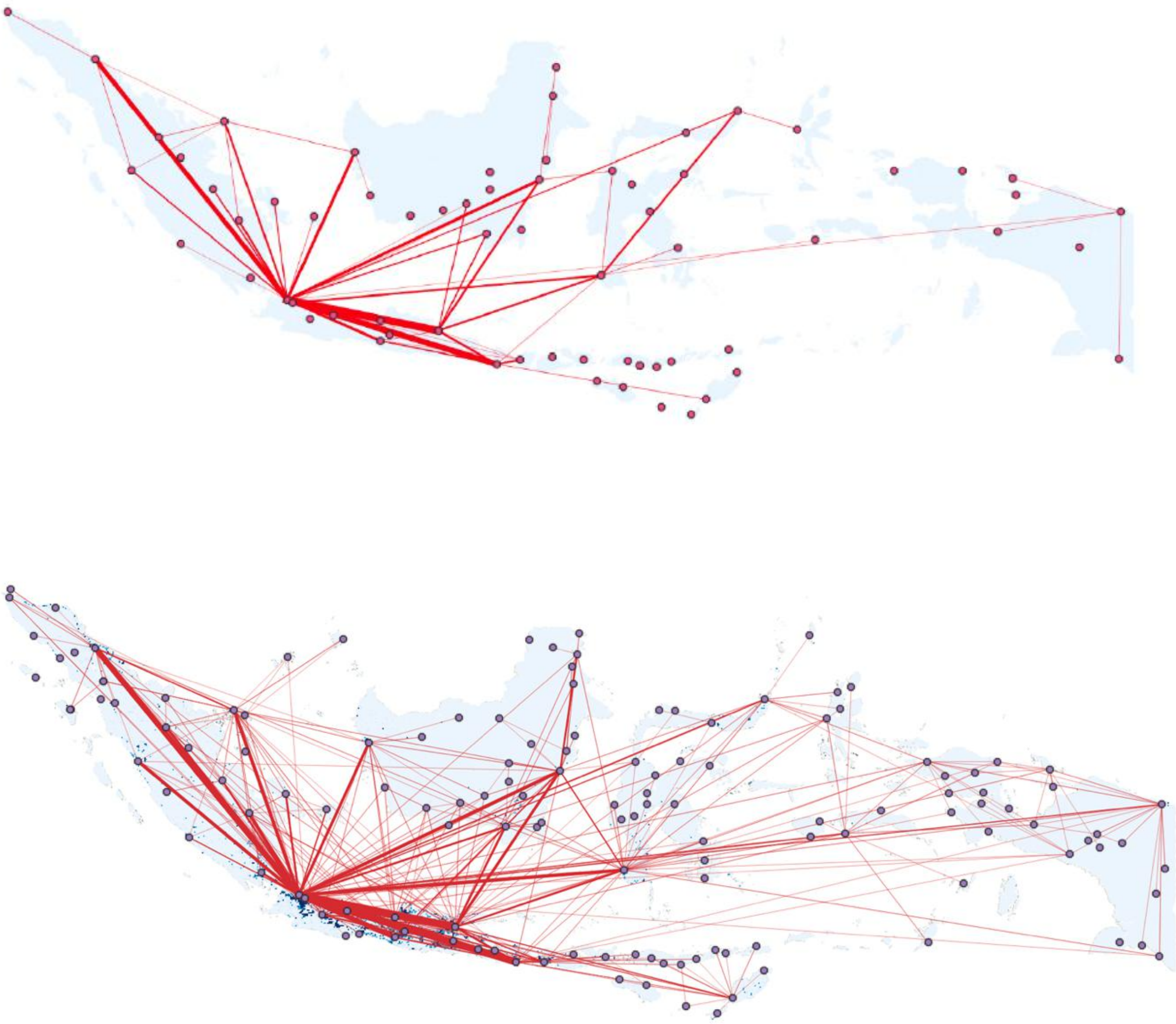

Figure 5. Passenger flow in OD data.

The thickness of lines constitutes of the amount of passenger that is in traveling in between the pair of airports that is shown as the nodes for the year 2000 (top) and 2015 (bottom). For the sake of clarity only flow bigger than 10000 passengers are shown

Spatial weight matrix

Our main interest is to explore different spatial interaction that is formed with airline interaction help to describe the spatial formation of the cities in Indonesia. To generate neighbors and spatial weight for spatial analysis, I use the OD of airline passenger data explained in the previous part. Most studies dealing with geographical contingency adopt a binary contiguity matrix with the value of one if two units share a common border and zero otherwise. I adopt this value into our OD data. Since our OD data is assumed to be direct flights, all airport pairs that have a non-zero flow value are assumed to be neighbors. This makes the matrix spatial weight symmetrical. I will call this matrix without interaction value as neighbor matrix later in this thesis.

As it mentions before, traditionally spatial weight constitutes of binary unit of one and zero. The first matrix I are generating using the traditional weight matrix using the neighbor relation. I employ a binary value for the weight for this matrix as 1 is the neighbor and 0 for non-neighbor. To explore the impact of distance, I generate is using distance decay inspired by Newton's law of gravity (Elhorst & Vega, 2017). I am calculating the distance of airport of the non-zero pair of neighbors using geometry calculator between two points of in QGIS. I will call this matrix without interaction value as inverse distance matrix later in this thesis.

In order to estimate the model, I are using R-package spdep (Bivand & Wong, 2018). This the package requires a row-normalize for spatial weight matrix. For both matrix I then normalize the data flow according to its origin. It is worth to notes that this treatment of row normalizing will makes a node with higher count of neighbor to have less impact than nodes with lower count of neighbor as can be seen in the construction of model (2)

Figure 4. Passenger neighbor relation in spatial weight matrix

This figure constitutes all the neighbor relationship of the airport. The yellow line is the neighbor relationship. I assume that all the data I have are direct flight. This neighbor relationship that is used for spatial weight does not differentiate between origin and destination as the matrix is symmetrical.

Model Estimation

OLS for urbanization.

The results from the ordinary least squares (OLS) estimation of (1) are presented in Table 2. These results show how changes in airport size are correlated with changes in urban population.

To solve the zero-value problem, I add an arbitrary value of one. This value convenient as the result of $\log(1) = 0$ that will nullify the parameter effect in the later estimated model.

Table 2. OLS estimation

	Dependent variable:			
	(log(1 + POP_2015) - log(1 + POP_2000))			
	(1)	(2)	(3)	(4)
log((1 + Size2015) - log(1 + Size2000))	0.252^{**}	-0.053^{*}	0.233^{***}	0.165^{***}
	(0.101)	(0.030)	(0.057)	(0.046)
log(1 + POP_2000)		-0.849^{***}	-0.882^{***}	-0.923^{***}
		(0.026)	(0.023)	(0.019)
log(1 + Size2000)			0.244^{***}	0.197^{***}
			(0.043)	(0.035)
factor(Jawa)Jawa - Bali				1.855^{***}
				(0.249)
Constant	0.536	11.307^{***}	8.468^{***}	9.356^{***}
	(0.704)	(0.380)	(0.601)	(0.490)
Observations	95	95	95	95
R^2	0.062	0.928	0.946	0.967
Note:. Standard error in parenthesis				*p**p***p<0.01

Column (1) shows the estimates with no fixed effects or controls, column (2) uses urban size in the start period of the year 2000, column (3) uses airport size in the start period of the year 2000, and column (4) is my preferred specification, which includes the control for region Java-Bali.

In Table 2 the parameter of $\log(Size_{2015}) - \log(Size_{2000})$ clearly showing a positive correlation throughout the model. This finding is not surprising as the larger the urban population meant a larger probability for air traveler. However, as mentioned in the introduction this correlation makes the impact estimation of changing accessibility become problematic by the existence of reverse causality it has. However, it is interesting to see the difference in the column (1) and (2) where the correlation of the change of size for airport to change of population when I controlled with number of urban populations in the base year of 2000. The correlation to change of airport size remains in a positive and significant relationship in the remaining models. Finally, the airport size variable becomes less significant when the dummy variable Java-Bali is introduced. The Java-Bali variable has a very high and positive value and is statistically significant. Together with the urban percentage variable may tell us that Indonesia is still not in steady state of urbanization and the cities are still growing. The finding is in line with Wajdi et al., (2017) who stated that urbanization in Indonesia is still concentrated in areas that has been developed.

2SLS IV Regression

In this section we are trying to look at the impact of airport size in the change in population. We do 2SLS using the instrument variables discussed in the previous section. First stage of the 2SLS regression follow equation (2) can be seen in Table 3

Table 3. First stage regression

	Dependent variable:			
	log(1+Size2015)- log(1+Size2000)			
	(1)	(2)	(3)	(4)
log(1+Size_inst) - log(1+Size2000)	0.770***	0.770***	0.643***	0.639***
	(0.033)	(0.035)	(0.122)	(0.124)
log(1 + POP_2000)		-0.001		0.009
		(0.036)		(0.037)
log(1 + Size2000)			-0.117	-0.123
			(0.108)	(0.112)
Constant	-0.107	-0.088	1.599	1.580
	(0.294)	(0.559)	(1.606)	(1.616)
Observations	95	95	95	95
R^2	0.854	0.854	0.856	0.856
Adjusted R^2	0.853	0.851	0.853	0.853
Note:. Standard error in parenthesis				*p**p***p<0.01

Column (1) is a first stage regression without control. Column (2) and column (3) are first stage regression with one control. Lastly column (4) is the preferred first stage regression stated in equation (2). In general, adding control to the first stage equation reduces the parameter value of the instrument variable. Population in 2000 as control had almost absolutely no impact on this equation seen in column (2). Adding airport size control in 2000 provides additional power for instrument variables in explaining variables of interest, changes in airport size. In this table, the correlation between variables of interest and instrument variables remains significant with the addition of variable control. We then proceed to use the column (4) estimation to the next step of 2SLS following the equation (1). The results from OLS compared to IV regression estimation are presented in Table 4.

Table 4. Comparing OLS and 2SLS

	Dependent variable:	
	(log(1 + POP_2015) - log(1 + POP_2000))	
	OLS	*2SLS*
log(1+Size2015) - log(1+Size2000)	0.165^{***}	0.468^{***}
	(0.046)	(0.124)
log(1 + POP_2000)	-0.923^{***}	-0.926^{***}
	(0.019)	(0.023)
log(1 + Size2000)	0.197^{***}	0.403^{***}
	(0.035)	(0.086)
factor(Jawa)Jawa - Bali	1.855^{***}	1.527^{***}
	(0.249)	(0.326)
Constant	9.356^{***}	6.388^{***}
	(0.490)	(1.240)
R^2	0.967	0.951
Weak instrument test p-value		6.93e-06
Hausman test p-value		0.0006

Note: the calculation results are obtained using R package ivdep. Number of observation 95. Standard error in parenthesis. $^{*}p\ ^{**}p\ ^{***}p<0.01$

I first need to check whether our instrument is valid. Weak instruments test using the F-statistics are sufficiently large to rejects the null hypothesis hence it can be considered a relevant instrument for airport size. Hausman test for endogeneity: rejects the null that the variable of concern is uncorrelated with the error term, indicating that our instrument is exogenous. I can say with confidence that our chosen instrument is valid.

In comparison to OLS, I can see that I underestimate the impact of change in airport size by more than 2 times over. The increase in magnitude in this finding is common in studies that use instrumental variables to estimate the effects of transport infrastructure. Duranton & Turner, (2012) explain this phenomenon with two reason: (1) missing variable such as ‘amenities’ and; (2) there is a possible reverse causality. I also suspect that the reason is due to missing spatial correlation with from the network that airport has.

The magnitude of the effect 0.468 in the measure of passenger in that using the airport. The elasticity of 0.468 for the effect of airport size on urban population can be expressed in terms of the urban population in a urban area of a particular size. For a typical urban area with one million residents would have 4680 population increase from a 1% increase in the size of the local airport

Checking for Spatial correlation

In looking at spatial correlation, I test the residual of the preferred OLS model and its depended variable, $\log(Size_{2015}) - \log(Size_{2000})$, with Moran's I test using the spatial with all weight matrix I have.

Table 5. Moran's I value for change of population and OLS residual

	Moran I	Std. Devian
log(1+Size2015) - log(1+Size2000)		
Neighbor matrix	0.113*	0.005
Inverse distance matrix	0.151*	0.006
OLS Residual		
Neighbor matrix	0.017	0.005
Inverse distance matrix	0.014	0.006
Note: the calculation results are obtained using R package spdep		*p<0.1

The interaction weight is not significant. With Moran's I value not statistically significant for the residual of our OLS model, I can accept the null hypothesis that there is no spatial correlation in our OLS residual. OLS. On the other hand, the change of population appears to have some spatial correlation for spatial weight both neighbor matrix and invers distance matrix. This result motivates the choice of not using error-based model for spatial regression as it will not improve

Spatial regression

In order to see the impact of spatial accessibility, spatial regression is used following the equation (2) discussed in the previous section. In this section, hope to shed a light on the importance of connectivity that airport give using the spatial regression. I also applied 2SLS with the IV setup that has been discussed before. As the spatial weight matrix is explaining about the connectivity of cities using air travel, the difference between OLS and SLX is the difference caused by that connectivity. Moreover, I also explore the distance effect of the spatial correlation by using the invers distance matrix in the SLX model. The result comparison of SLX, OLS and presented in Table 6.

Table 6. Model comparison OLS, IV regression, and SLX to change of urban population

	Dependent variable: log (1 + POP_2015) – log (1 + POP_2000)		
	OLS	SLX (Neighbor matrix)	SLX (Invers distance matrix)
	(1)	(2)	(3)
log(1+Size2015)- log(1+Size2000)	0.165***	0.183***	0.172***
	(0.046)	(0.047)	(0.046)
log(1+Pop_2000)	-0.923***	-0.942***	-0.944***
	(0.019)	(0.021)	(0.021)
log(1+Size2000)	0.197***	0.221***	0.210***
	(0.035)	(0.037)	(0.036)
factor(Java)Java - Bali	1.855***	0.221***	0.210***
	(0.249)	(0.037)	(0.036)
Spatial Lag	**N**	**Y**	**Y**
Constant	9.356***	9.714***	9.747***
	(0.490)	(0.593)	(0.597)
R^2	0.967	0.969	0.969
Adjusted R^2	0.965	0.966	0.967

Note: Number of observations 95. Standard error in parenthesis. The calculation results are obtained using R package spdep and ivreg — *p<0.1 **p<0.05 ***p<0.01

Implementing spatial weight provides additional explanations to the model created. In the SLX model with the neighbor matrix in column (2) we can see that implementing spatial weight also increases the parameter value in the variable $\log(Size_{2015}) - \log(Size_{2000})$. This reinforces that estimation by ignoring spatial correlation can give rise to biases in estimates as found by Margaretic et al., (2017). In column (3) we can see that the impact of the airport size change when implementing the inverse distance matrix in the SLX model is smaller than the one seen in column (2) which does not consider the distance in it. This is not surprising given the distance of determining the price which indirectly also certainly reduces accessibilities as discovered by Jung et al., (2008) on highways in Korea.

2SLS with IV

To address endogeneity issue, we implement the 2SLS with the same IV setup and first stage estimation discussed before. By implementing the residual of first stage we have the estimate endogeneity issues. The comparison between the regressions used can be seen in Table 7.

Table 7. Model comparison OLS, IV regression, and SLX to change of urban population

	Dependent variable: log(1 + POP_2015) - log(1 + POP_2000)			
	OLS	2SLS	2SLS (OLS+SLX Without weight)	2SLS (OLS+ SLX With inverse distance weight)
	(1)	(2)	(3)	(4)
log(1+Size2015)-log(1+Size2000)	0.165^{***}	0.468^{***}	0.470^{***}	0.458^{***}
	(0.046)	(0.124)	(0.094)	(0.094)
log(1+Pop_2000)	-0.923^{***}	-0.926^{***}	-0.949^{***}	-0.953^{***}
	(0.019)	(0.023)	(0.020)	(0.020)
log(1+Size2000)	0.197^{***}	0.403^{***}	0.408^{***}	0.402^{***}
	(0.035)	(0.086)	(0.065)	(0.065)
factor(Java)Java - Bali	1.855^{***}	1.527^{***}	1.513^{***}	1.257^{***}
	(0.249)	(0.326)	(0.269)	(0.362)
Spatial Lag	**N**	**N**	**Y**	**Y**
Lagged log(1+Size2015)-log(1+Size2000			-0.043	-0.050
			(0.069)	(0.054)
Constant	9.356^{***}	6.388^{***}	6.489^{***}	6.597^{***}
	(0.490)	(1.240)	(1.064)	(1.069)
R^2	0.967	0.982	0.972	0.972
Adjusted R^2	0.965	0.982	0.969	0.970

Note: Number of observations 95. Standard error in parenthesis. $^{*}p<0.1^{**}p<0.05^{***}p<0.01$
The calculation results are obtained using R package spdep and ivreg

Column (1) and (2) are showing OLS and 2SLS IV regression model that I have discussed above. Column (3) is SLX model with neighbor matrix and column (4) is SLX model with invers distance matrix. The complete impact of direct, indirect and total of SLX model can be found on the Appendix.

The coefficients on $\log(Size_{2015}) - \log(Size_{2000})$ in Table 7 demonstrate a clear, positive relationship between changes in airport size and urban population in all the model that I investigate. As we can see in the column (2) and (3), SLX model increasing the parameter value of the change of airport size. However, the correction experienced by the airport size parameter is not large even almost negligible. Similar to what was found before performing the second regression on the SLX model, it can be seen in the difference in the airport size parameter in column (4) when the distance introduced in the model gives a negative change.

Lastly, it appeared that the lagged estimate parameter for is negative $\log(Size_{2015}) - \log(Size_{2000})$ although it is not statistically significant. Contrary to finding of the previous research about highlight the importance of connecting other region to Java-Bali region (Kameswara & Suryani, 2021; Wajdi et al., 2017), this result is certainly not in line with expectations that neighborly relations do not have an impact on changes in the number of urban populations. This finding remarks a the question whether increase of accessibility is good for the growth of smaller urban area with slow growth as connecting it to the larger urban area with rapid change in population, meaning making them neighbors in our matrix, will reduce the rate of growth in that area is not plausible.

Conclusion

This thesis we are trying to include connectivity to old long question of how transportation infrastructure changes the urban population. To estimate the impact, this thesis use instrumental variable that is introduced by Sheard, (2019), using airport hierarchy and city pair origin destination flow data. To improve finding in this field, this thesis tries to implement spatial correlation to the model. Estimation in this thesis implementing 2SLS method to take account for endogeneity problem that change of population and accessibility has.

Going back to the original question "How airport and spatial correlation from connectivity affect urban population of the city?", there are several things to mention. First, Throughout the model that we estimate, we found that the impact of airport size changes on urban populations had a positive impact. This is finding rather not surprising as many of the previous study also shown the same positive result (Blonigen & Cristea, 2012; Button et al., 2009; Green, 2007; Sheard, 2019). Second, Model 2SLS uses variable instrument to avoid endogeneity problems that may occur between airport size and population growth, find that OLS is underestimating the impact of airport size change to urbanization. Third, after we consider spatial correlation provide a better explanation power over basic OLS. Interestingly, after controlling the endogeneity issue using the same IV setup, implementing spatial correlation does little to change the direct impact of the airport size change. Fourth, implementing distance in spatial weight matrix for spatial correlation reduces the value of the direct effect parameter from the airport size change, indicating that the distance heterogeneity needs to take account of as it also hampers accessibility, naturally. Lastly, we found that explaining connectivity gives a negative value although it is not significant, raising the question to the assumption of connectivity will increase the urbanization of the lagging area proven to be not proven.

The main limitation in this study is data. This study only looks at the impact in the extended period. The instrumental variables we use are used by Sheard, (2019) to see the change in the short-term period. However, our dataset only allows us to do research in the long-term period as there is no data available to fill our purposes between 2000-2015. As Indonesia continues to grow as a country, the administrative boundaries change very quickly to accommodate this rapid growth. Therefore, data that uses the boundaries of administrative areas, cannot be compared within extended period of time. The use of data on a smaller scale such as "villages" may be able to overcome this problem because changes on that scale occur less frequently. However, the data is not open to the public. Another problem with the data, as explained in the data section, the OD data used in this study is that a lot of data is lost due to the incompleteness of the attributes of origin-destination for our purpose. This makes the instrument variables we generate very fragile because they only rely on data that is not complete.

This study only looks at the urbanization impact of the movement of people. However, urbanization can also be defined as the change of economic activity into goods and services. This makes research on urbanization through the movement of goods also very relevant. Not only using aircraft, but this research can also be done for other modes of transportation such as sea transportation. Furthermore, this study used only 2000 and 2015. In research by Sheard, (2019), the instrument variable used in this study can also be used capture impact in short-term period, (e.g. 1 year period). This can provide insight into infrastructure impacts through short-term estimates that were previously difficult to perform (Rietveld, 1995)

Reference

Anselin, L. (1988). Model Selection in Spatial Econometric Models. In L. Anselin (Ed.), *Spatial Econometrics: Methods and Models* (pp. 243–252). Springer Netherlands. https://doi.org/10.1007/978-94-015-7799-1_14

Bivand, R. S., & Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *TEST*, *27*(3), 716–748. https://doi.org/10.1007/s11749-018-0599-x

Blonigen, B. A., & Cristea, A. D. (2012). *Airports and Urban Growth: Evidence from a Quasi-Natural Policy Experiment*. 38.

Brueckner, J. K. (2003). Airline Traffic and Urban Economic Development. *Urban Studies*, *40*(8), 1455–1469. https://doi.org/10.1080/0042098032000094388

Button, K., Doh, S., & Yuan, J. (2009). *The role of small airports in economic development*. *4*(2), 13.

Dobis, E. A., Delgado, M. S., Florax, R. J. G. M., & Mulder, P. (2019). Hierarchy and spatial contagion: Population in American cities between 1990 and 2010. *Journal of Economic Geography*, lbz012. https://doi.org/10.1093/jeg/lbz012

Dubin, R. (2003). Robustness of Spatial Autocorrelation Specifications: Some Monte Carlo Evidence. *Journal of Regional Science*, *43*(2), 221–248. https://doi.org/10.1111/1467-9787.00297

Duranton, G., & Turner, M. A. (2012). Urban Growth and Transportation. *The Review of Economic Studies*, *79*(4), 1407–1440. https://doi.org/10.1093/restud/rds010

Elhorst, J. P., & Vega, S. H. (2017). *The SLX model: Extensions and the sensitivity of spatial spillovers to W*. 27.

Freeman, L. C., Roeder, D., & Mulholland, R. R. (1979). Centrality in social networks: Ii. experimental results. *Social Networks*, *2*(2), 119–141. https://doi.org/10.1016/0378-8733(79)90002-9

Friedmann, J. (1986). The World City Hypothesis. *Development and Change*, *17*(1), 69–83. https://doi.org/10.1111/j.1467-7660.1986.tb00231.x

Green, R. K. (2007). Airports and Economic Development. *Real Estate Economics*, *35*(1), 91–112. https://doi.org/10.1111/j.1540-6229.2007.00183.x

Hofmann, A., & Wan, G. (2013). Determinants of Urbanization. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2295736

Iqbal, L. M. (2021). *DEMYSTIFYING THE GEOGRAPHY OF URBANIZATION IN INDONESIA*. 16.

Jung, W.-S., Wang, F., & Stanley, H. E. (2008). Gravity model in the Korean highway. *EPL (Europhysics Letters)*, *81*(4), 48005. https://doi.org/10.1209/0295-5075/81/48005

Kameswara, B., & Suryani, T. A. (2021). Regional Economic Growth and Air Transportation in Indonesia especially outside Java: A Backward Method in Multiple Regression Model. *TATALOKA*, *23*(2), 190–200. https://doi.org/10.14710/tataloka.23.2.190-200

Margaretic, P., Thomas-Agnan, C., & Doucet, R. (2017). Spatial dependence in (origin-destination) air passenger flows: Spatial dependence in air passenger flows. *Papers in Regional Science*, *96*(2), 357–380. https://doi.org/10.1111/pirs.12189

Menashe-Oren, A., & Bocquier, P. (2021). Urbanization Is No Longer Driven by Migration in Low- and Middle-Income Countries (1985–2015). *Population and Development Review*, *47*(3), 639–663. https://doi.org/10.1111/padr.12407

Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M., & Maffenini, L. (2019). *GHS-SMOD R2019A - GHS settlement layers, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015)*. https://doi.org/10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218

Rietveld, P. (1995). Infrastructure and spatial economic development. *The Annals of Regional Science*, *29*(2), 117–119. https://doi.org/10.1007/BF01581801

Sassen, S. (2005). The Global City: Introducing a Concept. *The Brown Journal of World Affairs*, *11*(2), 27–43.

Sellner, R., & Nagl, P. (2010). Air accessibility and growth – The economic effects of a capacity expansion at Vienna International Airport. *Journal of Air Transport Management*, *16*(6), 325–329. https://doi.org/10.1016/j.jairtraman.2010.04.003

Sheard, N. (2019). Airport Size and Urban Growth. *Economica*, *86*(342), 300–335. https://doi.org/10.1111/ecca.12262

SUZUKI, Y., & AUDINO, M. J. (2003). The Effect of Airfares on Airport Leakage in Single-Airport Regions. *Transportation Journal*, *42*(5), 31–41.

Wajdi, N., Adioetomo, S. M., & Mulder, C. H. (2017). Gravity Models of Interregional Migration in Indonesia. *Bulletin of Indonesian Economic Studies*, *53*(3), 309–332. https://doi.org/10.1080/00074918.2017.1298719

Appendix

1

	Dependent variable:			
	log(1 + POP_2015) - log(1 + POP_2000)			
	(1)	(2)	(3)	(4)
log(1+Size2015)- log(1+Size2000)	0.447***	0.444***	0.371***	0.371***
	(0.093)	(0.092)	(0.077)	(0.077)
log(1 + POP_2000)	-0.931***	-0.927***	-0.927***	-0.927***
	(0.018)	(0.018)	(0.018)	(0.018)
log(1 + Size2000)	0.388***	0.385***	0.333***	0.333***
	(0.065)	(0.064)	(0.054)	(0.054)
factor(Jawa)Jawa - Bali	1.764***	1.764***	1.764***	1.764***
	(0.239)	(0.239)	(0.239)	(0.239)
Constant	6.625***	6.621***	7.364***	7.370***
	(0.922)	(0.923)	(0.772)	(0.771)
Observations	95	95	95	95
First-stage Control				
Pop_2000	Y	N	Y	N
Size2000	Y	Y	N	N
R^2	0.970	0.970	0.970	0.970
Adjusted R^2	0.968	0.968	0.968	0.968
Note:				*p **p ***p<0.01

The changes that occur in the coefficient variable of interest, log(1+Size2015)- log(1+Size2000) have not changed much with the presence of control in the first stage regression. For this basis we use the estimate on the column (1) as our 2SLS estimate.

	Dependent variable: log(1 + POP_2015) - log(1 + POP_2000)					
	Impact SLX (Neighbor matrix)			Impact SLX (Invers distance matrix)		
	Direct	Indirect	Total	Direct	Indirect	Total
log(1+Size2015)-log(1+Size2000)	0.469***	-0.043	0.427***	0.457***	-0.050	0.407***
	(0.094)	(0.070)	(0.114)	(0.093)	(0.054)	(0.107)
log(1+Pop_2000)	-0.949***	0.081	-0.867***	-0.953***	0.077**	-0.876***
	(0.020)	(0.057)	(0.052)	(0.020)	(0.038)	(0.035)
log(1+Size2000)	0.407***	-0.108*	0.299***	0.402***	-0.096**	0.306***
	(0.065)	(0.059)	(0.084)	(0.065)	(0.044)	(0.078)
factor(Java)Java - Bali	1.512***	0.385	1.897***	1.257***	0.590	1.848***
	(0.268)	(0.385)	(0.370)	(0.362)	(0.444)	(0.303)

Note: the calculation results are obtained using R package spdep. Standard error in parenthesis. *p<0.1**p<0.05***p<0.01

Some things I think noteworthy, First, indirect impact of airport size is only significant in the model without distance weight. This indicates when the distance nullifies the indirect impact of airport size. This raises argument of the importance of connecting the less urbanize area in remote places as it shows that distance dilute the impact of airport size.