Highway Access and Urban Development In the Jakarta Metropolitan Area

Andhika Putra Pratama
(2638506)

Master Thesis

MSc in Spatial, Transport, and Environmental Economics

Vrije Universiteit Amsterdam

July 2019

Abstract

Transport costs have been widely recognized as one of the major drivers of urban development. Using the spatial data from GHSL and ESACCI, this paper aims to find any causation between highway expansion and various forms of urban development in the JMA. Employing historical transport infrastructures data as instruments, the result shows that highway development in the JMA expands the new land developments from 1990 to 2014. One kilometer improvement in highway access increases the urbanized land area in the city of Jakarta by 6.6-7.7%. The impact in the city suburbs is smaller (2.6-3.2%). On urban density, one kilometer improvement in access to highway fosters urban density by 2.6-5.7%. The result for the city of Jakarta, however, remains inconclusive. This paper finds no evidence on the presence of low-density development in the JMA. The results of this paper also indicate that the presence of urban sprawl in the JMA is not evident.

Keywords: urban development, urban expansion, urban density, urban sprawl, transport access, historical transport infrastructures, spatial data.

Table of Contents

Table of Figures

Summary of Tables

1. Introduction

The development of urban area continues to be one of the most important topics in urban economics. Rural-urban migration has been a major component of urbanization (IIED, 2014). This phenomenon is clearly linked to the economic success of cities (IOM, 2015). The bigger the agglomeration benefits of a city, the higher its chance to attract people from rural area to the city. This is no exception for developing countries. Among countries undergo rapid urbanization, developing countries experienced higher urbanization rate, approximately 2.63 – 3.68% annually, higher than the developed countries (0.88%) (UN, 2014).

Jakarta, as one of the biggest megacities in developing countries, has been undergoing rapid urbanization in the last two decades (PRISMA, 2017). The Jakarta metropolitan area (JMA) experienced an annual population growth of 2.8% during the years of 2000 to 2010[1] (Statistics Indonesia, 2019), indicating that the benefits of agglomeration in the JMA are still high and it still attracts people to live and work in the area. The growth, however, it differs throughout the JMA. Since 2000, Jakarta's population grows slower than its surrounding suburbs (Statistics Indonesia, 2019).

The data from Statistics Indonesia shows that almost 6.5 million people work in Jakarta, but only 72% of them live in the city of Jakarta, leaving almost 1.8 million people reside in suburbs and commute daily to the city (National Labor Survey, 2018). This statistic implies that the development of the city of Jakarta should not be studied independently of its surrounding suburbs. One of the most discussed topics of urban development involving core cities and their surrounding suburbs is 'urban sprawl'. As summarized across the variety of literature, urban sprawl is defined as the disperse, scattered and low-density development of an urban area that resulted from market failures in the urban development process (Brueckner, et al. 1983; McGrath, 2005; Burchfield, et al. 2006; Garcia-Lopez, 2015). The study of urban sprawl has incorporated the work of geographers, economists, and urban planners and discussion surrounding the topic continues to grow today.

Of the many indicators of urban sprawl, one major feature is the spreading development of urban areas, mostly known as 'urban expansion'. Research by Bai, et al. (2012) has shown that, in some cases, urban land expansion has been as rapid as the demographic urbanization. Some researchers go as far as to argue that urban expansion may even exceed population growth. A 1% increase of urbanized land area per year, following by less than 1% population growth in the same year results in a less dense urban area for that particular year. A decrease in such urban density may indicate the presence of urban sprawl.

Although numerous studies categorize urban expansion as urban sprawl, not every new urbanized land development could be categorized as urban sprawl. An increase in urbanized land area in the city center in the form of filling up urban space fragmentation may indicate a compact development of urban area. This phenomenon exists in a relatively mature development stage of the city (Wagtendonk, et al, 2019). As such, urban sprawl should not be measured independently using only the expansion of urbanized land area.

Glaeser and Kahn (2003) suggest that a reduction in transport costs plays an important role in explaining urban expansion. This is in line with the classical monocentric model which shows that as commuting costs

[1] Authors calculation using population census data from national bureau of statistics (BPS, 2019). Population growth calculated using compound annual growth rate for 10 years.

fall, cities expand. Subsequently, studies from Burchfield, et al. (2006) and Garcia-Lopez (2018) indicate that transport infrastructures also induce a scatter development of urban areas in the United States and European cities. Other studies by Garcia-Lopez (2012; 2015) and Yudhistira, et al. (2018) confirm the presence of suburbanization through faster growth of population density in suburbs than in the urban center. These studies imply that transport infrastructures may affect various form of urban development in urban areas.

In the case of the Jakarta metropolitan area (JMA), extensive highway development has taken place since the late 1980s. Since 1990, no less than 150 km highways have been built all around the JMA. The national highway authorities plan to further expand the highway networks inside the JMA until 2030, most notably through the construction of the second layer of Jakarta outer ring roads (JORR II) with a total length of 133 kilometers. Figure 2 depicts the proposed highway development plan in the JMA. The green line represents existing highways in 1989, the blue line exhibits the current highways in the JMA, while the red line is the highways development plan until 2030.

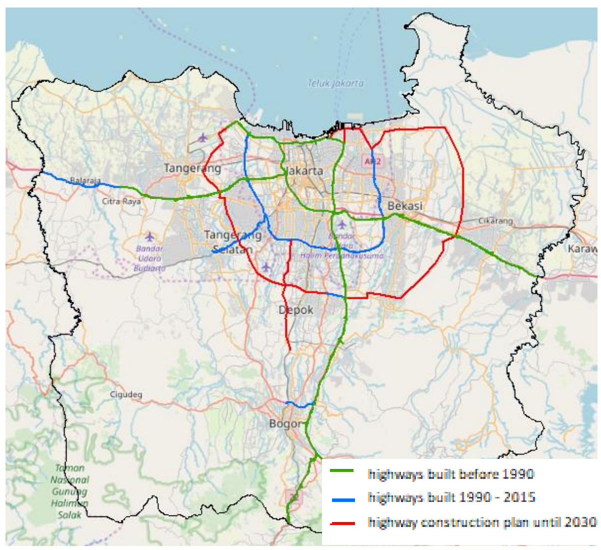

Figure 1. Highways development in the JMA.
Source: National Highway Authority (2019).

Rapid urban development creates numerous challenges for urban policy. It has been widely recognized that a sprawling development of urban areas would create lower agglomeration benefits and higher economic costs. Empirical research in examining the causal evidence of transport infrastructure development and various forms of urban development might extend the existing urban economics literatures in understanding various drivers of urban development.

In general, the main objective of this study is to test the causal evidence between new highway development and urban development in the JMA. This study also aims to test whether urban sprawl can be found in the JMA and whether it can be linked to the new highway development. By taking into account research from Garcia-Lopez (2018) and Yudhistira (2018), and taking advantages of high-resolution spatial data, this study plans to examine the presence of such urban development by testing several indicators, that is: (1) the total urbanized land area, as a measure of urban expansion, (2) the percentage of undeveloped surroundings, as a measure of scattered development, and (3) urban density, as a measure of population per urbanized land area.

2. Literature Review

The presence of unmitigated development of urban area may bring consequences for an urban area. By understanding the general drivers behind urban development and its consequences, policy makers will able to formulate policies to reduce the negative impact and internalize the additional externalities created by urban development. A broad range of literatures have used various measurements of urban development, particularly in examining the urban sprawl. Some basic measurements have been used by McGrath (2005), Burchfield (2006), and Garcia-Lopez (2018), while more complex measures also presented by Frenkel, et al. (2008). At the end of this chapter, I identify several drivers of urban development, including accessibility to transport infrastructures, which is the main talking point in this study.

2.1. Consequences of urban development

A seminal work from Newman and Kenworthy (1989) points out why it is preferred to have a compact and high-density urban area than a sprawling one (Figure 3). Despite having a quite similar population, Atlanta has urban areas of almost 4.230 km^2, way larger than Barcelona by almost 25 times. The carbon emissions from the transportation sector, however, is considerably lower in Barcelona. Atlanta also emits approximately 7.5 tons of CO^2 per person, while Barcelona, on the other hand, only emits 0.7 tons of CO^2 per person. This section provides a general analysis on how urban expansion and sprawling development could increase the economic costs of urban areas.

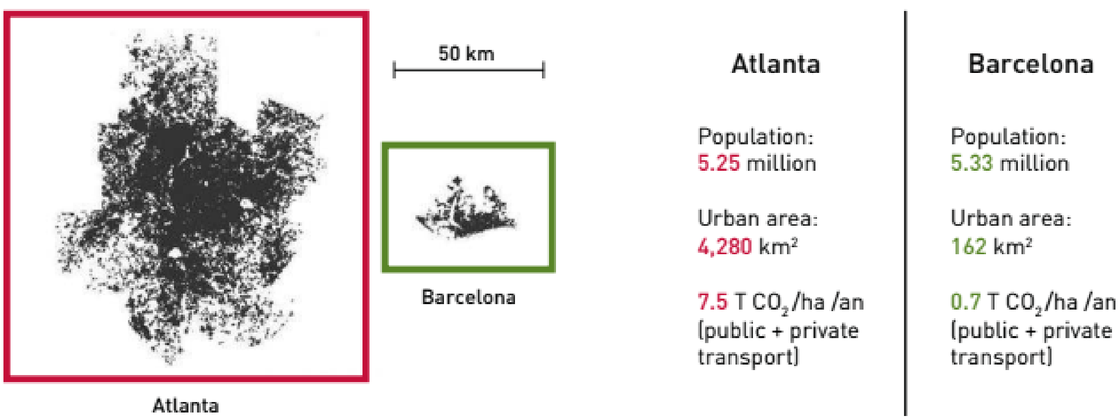

Figure 2. Emission in Barcelona and Atlanta (Newman and Kenworthy, 1999)

2.1.1. Environmental impact.

One of the major contributors to a city's greenhouse gas (GHG) emissions is the transportation sector. An increase in carbon emissions from the transportation sector is perhaps the most pronounced environmental impact of urban development. As urban area expands, more people commute from the suburbs to the city center. This contributes to an increase in vehicle use, simultaneously increasing the rate of carbon emission from transport activities. Prior studies in developed countries have suggested that higher densities and urban mixed land-use reduce commuting trip length and number of motorized trips (Levinson and Kumar 1994; Gordon et al. 1989), thus reducing GHG emissions from the transportation sector. A study by Newman and Kenworthy (1989; 1999) also added that gasoline consumption in cities like Houston or Phoenix could be 20-30 percent lower if their urban structures more closely resembled more compact cities such as Boston or Washington.

In addition to emissions from the transportation sector, increasing energy consumption from housing activities also increases the emission level in metropolitan areas. Glaeser and Kahn (2010) calculated total emissions in cities, not only from transport but also from home heating, and electricity. The result for the transport emission was in line with urban economics theory where less centralized cities produced higher emissions from driving compared to more centralized cities. In terms of electricity consumption, they found statistical evidence that more centralized cities correspond to lower emissions than less centralized cities. This may be due to the difference in average housing space; more centralized cities mostly have smaller houses, and as such the need for electricity is lower than in sprawling cities. The result is different for home heating, as more centralized cities tend to produce higher emissions from home heating (Glaeser and Kahn, 2010).

2.1.2. Socio-economic impact of urban sprawl

Unmitigated urban development also affects socio-economic conditions in urban areas through decreasing productivity. In their study, Fallah, et al. (2010) claim that urban sprawl affects labor productivity, which they explain using the theory of agglomeration. This theory suggests that a low-density setting of urban area diminishes labor productivity due to loss of urban agglomeration benefits. Using the data for metropolitan areas in the United States, they found that urban sprawl was associated with lower average labor productivity, which implies that the implementation of anti-sprawl policy may enhance the economic performance of metropolitan areas in the United States.

In addition to its economic impact, some studies also suggest that urban sprawl give rise to social issues. Glaeser and Kahn (2003) suggest that sprawl corresponds with segregation between the rich and the poor, where the rich, of those which have access to automobiles, prefer to live in suburbs than the city center, generating higher crime rates in a low-income neighborhood (Sole-Olle, 2008). Other social-related impacts of urban sprawl are reduced upward mobility (Ewing, et al. 2016), higher inequality, and worsening financial well-being (Lee, et al. 2018)

2.1.3. Increase in social costs in the provision of public infrastructures

Urban development process also brings several costs associated with it. In the case of urban expansion, when a city expands, the government needs to build more public goods and broaden its services to maintain a certain level of public services to the residents. The larger the area to be covered by the government, the higher the costs of provision of those goods and services. Low-density development also induces higher costs in the way that lower density of individuals in certain areas would undermine the scale economies of public infrastructures provision. Therefore, it induces higher costs than in denser cities (Sole-Olle, 2008).

2.1.4. Loss of open space due to land conversion

The development of urban areas, whether it is sprawling or not, happen at the cost of various type of open spaces. New urban settlements are created through land conversion of urban spaces inside the city or agricultural land and natural land (forest, grassland, etc) outside the city. An increase of new urban settlements in the city center, for instances, comes in the form of infilling urban patches and shrinkage of existing urban space (Wagtendonk, 2019). The costs associated with the conversion might be different between urban bare lands and urban green space. The loss of urban green space might induce extra environmental or societal costs, such as higher exposure to air pollutants and lower house price due to smaller green spaces (Braubach, et al, 2017; Breunig, 2019). In terms of outward urban development, a report by TRB (1998) points out that as urban areas are sprawling, the agricultural land and forest are decreasing at a similar rate. Loss of these natural lands also comes with numerous consequences, such as loss of employment in agriculture and a higher risk of soil erosion due to deforestation. Nevertheless, these associated costs to urban development should be compensated as loss of benefits due to urban development.

2.2. Measuring urban development

It is an arduous challenge to discuss urban development independently without mentioned urban sprawl. Early studies on urban sprawl even did not use a scattered and low-density development of urban area as the main indicators of urban sprawl but instead using the spatial size of cities. Not until the availability of advanced geospatial data, the definition of urban sprawl is narrowed, not only indicated by an increase in city sizes, but also the sparse and scattered development of urbanized land.

The first study attempting to understand the process of urban sprawl first introduced by Brueckner and Fansler (1983). They use the spatial size of cities in several metropolitan areas as measures of urban sprawl. The study uses a sample of 40 urbanized areas in the United States to empirically test the standard economic theory of urban sprawl. An extension of this study was developed by McGrath (2005) with more comprehensive data and a more extensive set of control variables. His study supports Brueckner and Fansler's arguments that "sprawl is the result of an orderly market process rather than a symptom of an economic system out of control" [McGrath; 2005, p. 482]

Research to explain sprawling development of cities has continued to grow in the last few decades. Most recently, this topic was comprehensively studied by Garcia-Lopez (2018) using three outcome variables to measure urban sprawl. He characterized sprawl into three dimensions: (1) the size of urbanized land, to exhibit spreading urban development, (2) the degree of fragmentation (number of residential lots), to test whether the development is rather compact or more sparse, and (3) the degree of undeveloped surroundings, as an indicator of a scattered and isolated the development of residential land. The availability of high-resolution spatial data, such as the Corine Land Cover (CLC) used in Garcia-Lopez (2018), plays an important role in expanding urban economic literature on urban development. This data allows us to quantify the urbanized land area or the scatteredness of urban areas to measure urban sprawl. As technology advances and better data becomes available, the study of urban development, particularly urban sprawl, will be more comprehensive and more precise in the future.

2.2.1. Urban expansion

Before Garcia-Lopez (2018), other studies have used urbanized land areas as measures of urban development, such as McGrath (2005) who first introduced this measure using urbanized land data made by the US Census Bureau. They defined urbanized land area as an area with a population density of at least 1,000 people per square mile. Using 33 metropolitan areas, they calculated the total square miles of urbanized land for the years of 1950, 1960, 1970, 1980 and 1990. This data was then used in a simple OLS regression to explain the general drivers of urban sprawl.

As the development of satellite images data advanced, various studies began to investigate the presence of urban sprawl by processing and classifying data into urban-settlement and non-urban settlement. This series of data was then used to display the expansion of urban settlement over time. A study by Deng, et al (2008) used Landsat TM scenes data for the years of 1988, 1995, and 2000 to show the expansion of urban core in several counties in China. The Landsat data used in their paper includes three classifications of built-up areas: the urban core, rural settlement, and other built-up areas. They used this continuous urban core as a measure of contiguous urban settlements and aggregated it at county level to calculate the size of the urban core for over 2000 counties in China. A similar approach was then replicated by Oueslati (2015), and Ahrens, et al, (2019).

2.2.2. Urban density

The most conventional indicator used to measure cities density is population density, which is obtained by simply dividing the total population by the total land area. Several notable works on density in urban areas using population density as the main determinant were conducted by Garcia-Lopez (2012; 2015) and Yudhistira, et al, (2018). This measure, however, may not perfectly capture the development of urban area since it only indicates demographic development rather than landscape development (Bai, et al, 2011). Despite having the same population density, areas with larger urbanized land have lower urban density than areas with less urbanized land. Therefore, incorporating developed land into urban density calculations may provide a better representation of density in urban areas.

A further study by Ahrens, et al, (2019) employs another way to measure urban density by calculating building density. The intuition of this concept is, if buildings are clustered in a few areas, the level of sprawl will be, holding other variables constant, low. Contrarily, if buildings are clustered in a large area, the sprawl level is arguably high. Ahrens, et al, (2019) were able to use this indicator due to the availability of building registry data in Ireland. The registry data was obtained from the Ordnance Survey Ireland, which he used to generate building density data of 3.409 Electoral Divisions in Ireland.

2.2.3. Scattered developments

The presence of scattered development indicates the existence of urban sprawl. The first study to develop an indicator for scattered development was conducted by Burchfield, et al, (2006). To do so, they used the 1992 National Land Cover data from the Landsat 5 thematic mapper satellite imagery and the 1976 Land Use and Land Cover data, which was derived primarily from high-altitude serial photographs. The data was then constructed into a 30m x 30m raster cells data classified by its land-use. To measure the scattered development, he started by calculating the percentage of open space in the immediate square kilometer of a residential cell. This calculation, also known as 'sprawl index', is used to test how often residential development goes beyond more than one kilometer away from other residential developments. In this calculation, the result is then averaged for each metropolitan area. An increase in sprawl index over time can be interpreted as an increase in the scattered development of a residential area, which implies a sprawling development of an urban area. Their study motivated other researchers, such as Angel, et al, (2012) who employed a similar approach to calculate for openness index and Garcia-Lopez (2018) who investigated the role of highway infrastructures in explaining urban sprawl in European cities.

2.3. Urban growth and general driving forces

Early research by Brueckner and Fansler (1983) examines spatial sizes of cities in relation to population, income, agricultural land values, and transportation costs to test whether it is a result of an orderly market process. They predict that, had spatial sizes of cities can represent urban sprawl, the population and income will increase the spatial sizes of cities while simultaneously consuming agricultural land. Using Box and Cox non-linear estimation, the study verifies urban economics theory that population and income positively affect spatial sizes of cities, whereas agricultural land value has a negative coefficient in respect to city sizes. The study, however, is unable to conclude any relationship between spatial sizes and transport costs, most likely because the proxies used in this paper do not perfectly capture the actual variation of commuting costs in urban areas.

McGrath (2005) revisits and extends the study by using larger samples of metropolitan areas in the United States. Employing a simple ordinary least squares (OLS) regression and controlling for time fixed effect, the results support Brueckner and Fansler's arguments that population and income positively correspond to urbanized land area and negatively affect the agricultural land values. The effect of transport costs is evident in this study, implying that their role is important in explaining urban sprawl.

Several studies also incorporate other variables into the analysis, such as distance to key places and physical geography. Burchfield, et al, (2006) points out that the physical geography of urban land area plays an important role in shaping cities, for example, the ruggedness of terrain between places. An area which is dominated by flat surface might develop faster than areas in the hillsides since it costs less to build infrastructure in flat surfaces than around rough surfaces.

Another study, by Deng, et al, (2008), adds distance variables, such as distance to port and distance to province capital, in addition to some geographical variables such as rainfalls, slopes, temperatures, and elevation. Their study finds that these variables are statistically different from zero while simultaneously improving the estimation without disrupting the model. As such, controlling for these variables may help in explaining the state of urban development. Other studies also point to several determinants that may affect the development of urban area directly or indirectly, such as crime rates, distribution of population by ethnic or race, political stances, and structure of public finance (Oueslati, 2014; Miguel Gomez, 2014).

2.4. Empirical studies on transport costs and urban sprawl.

Among the various determinants of urban development, transport costs are arguably one of the key driving forces since it directly affects individuals' decisions in choosing a place to live. Lower transport costs, whether in the form of access to private transportation, like owning automobiles, or better access to suburbs due to highway subsidy and shorter commuting distance, will reduce individuals' opportunity costs for living in central cities. This makes the trade-off between commuting costs and having larger space in suburbs become less significant. Therefore, it is likely to attract more people to live in the suburbs. As more people live in suburbs, the urban area expands, and this results in a low-density development in the city center. Lower transport costs may also increase the demand for open spaces which could not be accommodated in the city center. Hence, it induces a more scattered development of urban areas in the suburbs.

Existing literature also contains different proxies of transport costs when examining their impact on urban sprawl. Brueckner and Fansler (1983), for instance, use the percentage of people using public transit and the percentage of people owning automobiles as proxies of transport costs. The study, however, is unable to find statistical evidence of these indicators with respect to urban sprawl. The extension of their study conducted by McGrath (2005) calculates a regionally adjusted private transportation consumer price index (APTCPI) for metropolitan areas as a proxy of transport costs. It finds a negative relationship between transport costs and urban sprawl, which indicates that higher levels of sprawl occurred in the metropolitan area where transport costs are considerably low.

Glaeser and Kahn (2003) used car ownership and gas taxes to estimate how transport costs affect urban density using international data. By controlling for endogeneity, they established that places where it is generally more difficult to own cars have significantly less sprawl. Their study therefore suggests that the presence of urban sprawl is dependent on automobiles. Burchfield, et al, (2006) followed a similar argument by using streetcar passengers as proxy of transport costs. Their study showed that high dependency on automobiles is associated with urban sprawl.

Empirical research to examine transport improvement and urban density under the framework of suburbanization was first introduced by Baum-Snow (2007) under the framework of suburbanization. He used changes in distance to highways as a proxy of transport improvement and estimated this with respect to population density. This is also the first study that established the use of historical roads as instruments to control for the possible endogeneity problem between highways development and population density. Using data for more than 100 metropolitan areas in the United States, the study shows that the development of highways induces suburbanization. Using a similar approach to Baum-Snow (2007), a series of empirical studies on urban spatial structures by Garcia-Lopez (2012; 2015) and Yudhistira, et al, (2018) established a causal relationship between improvement in transport access and the presence of suburbanization in Spain and Indonesia respectively.

Other measurements of transport costs used in literatures about urban sprawl also include length of highways (Oueslati, et al, 2014; Garcia-Lopez, 2018) and time spent commuting and drive time to the nearest motorway (Ahrens, et al, 2019). Unlike the afore-mentioned studies (Garcia-Lopez, 2012; 2015, and Yudhitira, 2018) which describe differences in urban density through the presence of suburbanization, these studies explicitly address the presence of urban sprawl and confirm its relationship with transport costs. The study by Garcia-Lopez (2018) more specifically, confirms the causal link between transport improvement and the presence of spreading and scattered development of urban areas in European cities.

3. Data and Methodology

This study uses the GIS method to process spatial data into several urban development indicators. Using simple descriptive analysis, this study checks whether there is enough variation in the dataset to be used in the estimation. Employing the similar econometrics approach as introduced by Garcia-Lopez (2018) and Yudhistira, et al. (2018), this study tests whether improvement in accessibility, in particular, shorter access to highway ramp, induces urban expansion and scattered development of urban area as well as changes urban density in the JMA.

3.1. Data source

One of the main strengths of this study is the utilization of geographic information system (GIS) data. The Joint Research Center of European Commission (JRC-EC) produces The Global Human Settlement Layer (GHSL) data, consisting of various global spatial information about human presence on the planet over time. This allows the study to examine changes across urbanized land area over time through the presence of built-up area in the JMA. This data is derived from Landsat image collection and available for the years of 1975, 1990, 2000, and 2014. I use the 38m x 38m resolution data since it is the finest resolution that is published by the JRC-EC. For each 38m x 38m grid cell, it contains values of one if the grid cell is considered as a built-up cell and zero if otherwise.

In addition to the GHSL data, this study also generates similar data from European Space Agency Climate Change Initiatives (ESACCI) which produces annual land cover data from 1992 to 2015. The data provides land cover classifications, such as tree coverage, agricultural area, water bodies, and urban settlement. Unlike the GHSL data, the ESACCI data is only available at 300m x 300m resolution, less defined than the GHSL data. I use this data as a comparison against the GHSL data to ensure that the estimation is robust. The ESACCI data has an initial classification of zero if the grid cell is a non-urban settlement cell and one if otherwise.

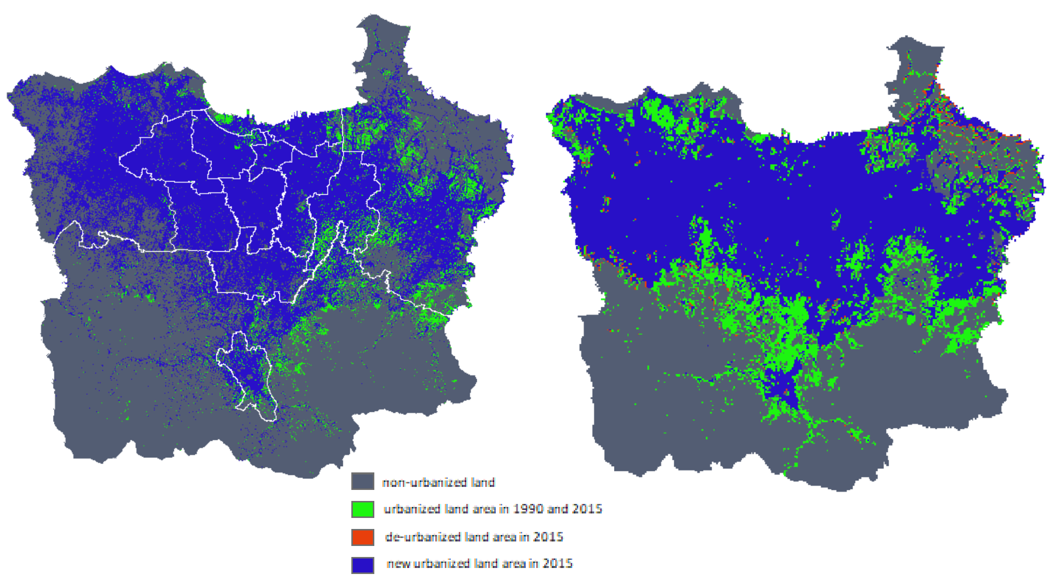

Figure 3. GHSL built-up area data 38m 1990-2014 (left) and ESACCI land cover data 300m 1992-2015 (right)

source: JRC-EC GHSL data (2019), ESACCI data (2019)

To measure accessibility, I use the transportation and road networks data from OpenStreetMaps (OSM) as a baseline. The data is then cross-checked with various sources, such as the National Highways Authority of Indonesia and KAI Commuter Jabodetabek (rail authority), to produce vector maps of highway ramps and railway stations for the year of 1990 and 2014. The OSM maps are also used as the baseline for digitizing other control variables, such as distance to city center, central business district, and local government offices, industrial area, and coastline. As a proxy of distance to the city center, I use distance to the national monument as previously used in Yudhistira, et al, (2018) and generate the centroid of Jakarta using GIS software. I select three main business districts in the JMA, namely the Sudirman area, Mega Kuningan area, and TB Simatupang area as proxy for business district, since most business offices are located in these areas. I also select all industrial areas obtained from the Ministry of Industry[2] as proxy for distance to labor market center and calculate the distance to local government townhall as proxy for district center. Distance to district center is used to control for possible exogenous variation for each district center.

I select two historical transport infrastructures, namely, the old Batavia road and old Batavia tram station. The old Batavia road consists of road networks that is developed around 1930-1940, including the part of Anyer – Panarukan road that is located inside the JMA. The Anyer-Panarukan road is arguably the oldest main road connection built during the governorship era of Herman Willem Daendels (1808-1811) connecting Anyer in the west part of Java and Panarukan in the east. It crosses the JMA through Tangerang to the center of Jakarta, heading south through Bogor and subsequently heading east until Panarukan in East Java. The remaining old Batavia road used in this paper can be seen in the left part of figure 5. The data for the old Batavia road and old Batavia tram line is obtained using historical maps of transport infrastructure of Batavia circa 1930-1934 and then carefully digitized them with GIS software.

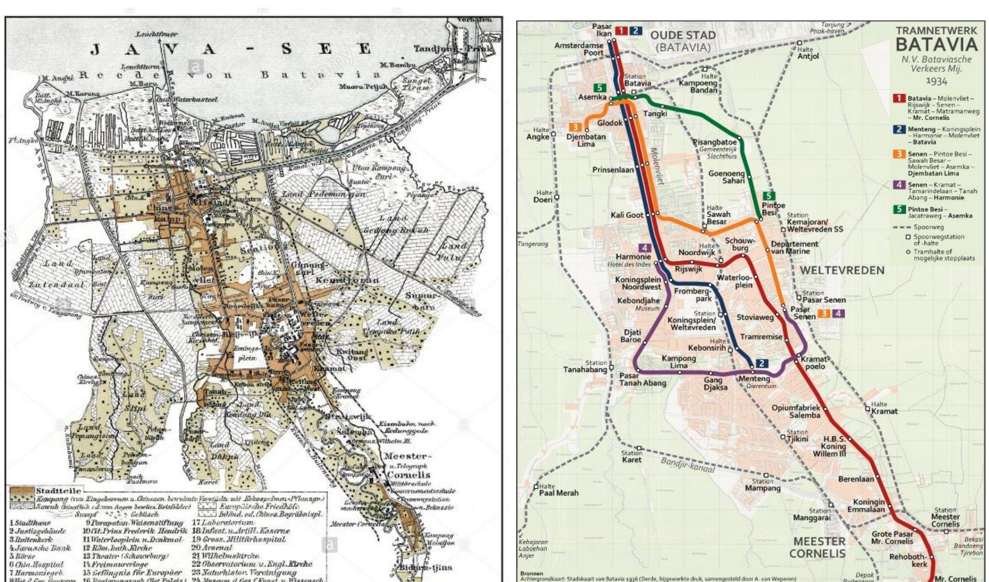

Figure 4. Batavia road network 1930 (left) and Batavia tram network 194 (right)

[2] Retrieved from: http://www.kemenperin.go.id/kawasan on June 27, 2019 (20:22)

To control for geographical variables, this study utilizes the Digital Elevation Model (DEM) data from the Geospatial Information Agency of Indonesia (BIG) with ±12m x 12m grid resolution to calculate elevation level and terrain ruggedness index. I use the administrative boundary vector data on community level from the Central Bureau of Statistics to calculate the area for each community in JMA as well as the main unit of measurement for the study.

3.2. Construction of dataset

3.2.1. Measuring urban expansion

Prior studies have been using aggregated data of urbanized land to measure urban expansion. Garcia-Lopez (2018), for instance, uses the LUZ boundaries for European cities while Deng (2008) in China uses counties level data aggregation. This paper performs aggregation at community level since it is the lowest statistical unit in Indonesia and also considering the vector dataset on this level is available for this study.

First, I construct the data by reclassifying both the GHSL and the ESACCI data into urban settlement cell and non-urban settlement cell. For each community in the JMA, the percentage of urbanized cells is calculated using *zonal statistics* tools in ArcGIS. Separately, the total area for each community in JMA is calculated in square kilometers. By multiplying the percentage of urbanized land and the total area for each community, I calculate the total urbanized land area for community level in the JMA. Figure 6 depicts a complete flowchart for generating the urban expansion dataset.

3.2.2. Measuring urban density

For urban density, I use the population data from the National Population Census of Indonesia for the years 2000 and 2010. The data is then aggregated at community level and divided by the total urbanized land from previous calculation to get the urban density at community level. This method provides a better measurement of urban population rather than using conventional measurements, such as population density, constructed by dividing the total population by the total area of the community. Using this approach, I obtain the level of urban density in number of inhabitants per square kilometers for each community in the JMA. Figure 6 exhibits the complete flowchart of the GIS process that is employed in constructing urban density data.

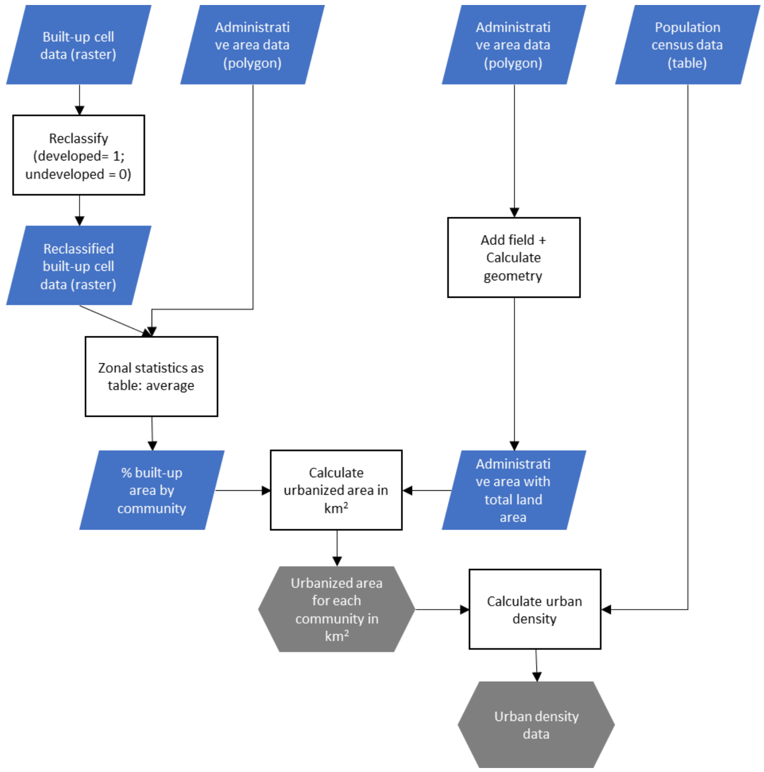

Figure 5. Data generation: urban expansion and urban density

3.2.3. Measuring scattered development

The widely used measurement of scattered development is the sprawl index introduced by Burchfield, et al, (2006). This index is calculated by computing the percentage of undeveloped land surrounding the residential land for a certain radius (for example, one kilometer) and then averaging it across all residential cells in an area. A similar approach is applied in this study. For each urbanized cell in both the GHSL data and ESACCI data, the percentage of undeveloped land surrounding the urbanized cell is calculated in the surrounding square kilometer. The index for each community is computed by averaging the percentage of undeveloped surroundings across all urbanized cells in the community. Figure 7 depicts the complete process of measuring scattered development using GIS software.

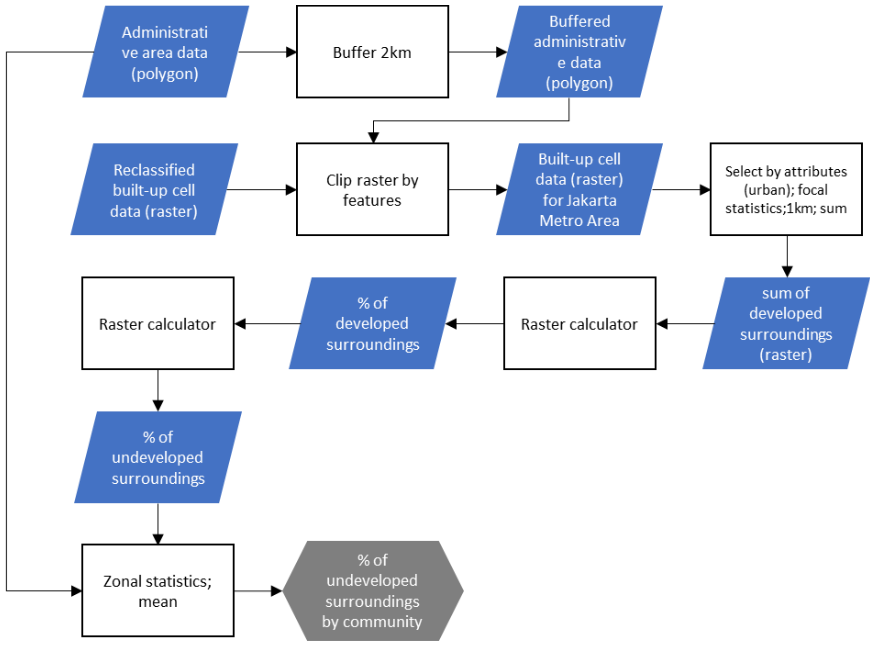

Figure 6. Data generation: Percentage of undeveloped surroundings / Sprawl index

3.2.4. Measuring distances and elevation

This study uses distance to transport infrastructures, such as highway ramps and railway stations as proxy of transport costs. Highway expansions over the years create a shorter distance from some communities to the highway ramps and improve accessibility to highway. Due to limitation of the data, this study is not able to use the network distance to nearest highway ramp and railway stations but instead uses the Euclidean distance. Intuitively, I calculate the straight-line distance and disregard the effect of road congestion in the calculation. This distance is then averaged at community level to get the average distance from each community to the nearest highway ramp and railway station. A similar approach is used for other measurements, including distance to city center, centroid of Jakarta (proxy for city center), local government office (proxy for district center), coastline, industrial area, and historical transport infrastructures.

For geographical variables, this study calculates three variables, the total land area, elevation and terrain ruggedness index (TRI). I calculate this simply by using *calculate geometry* tools in GIS software for each community. To measure elevation level, I average the elevation level for each community in JMA from the DEM raster maps from BIG. The calculation for TRI, however, is slightly more complicated. The method proposed by Riley, et al, (1999) is to calculate the difference in elevation values between each cell and its adjacent cells. This difference is then squared and summed. The TRI is derived by taking the squared root of the calculations. Intuitively, it corresponds to the average elevation change between any point on a cell and its surrounding area. Due to limitations of the geospatial software used in this study, this study unable to perform this calculation, and instead use a simplification of TRI. I calculate the maximum range between the cell and its adjacent cells and then simply average it on community level. This data is still able to capture land ruggedness in each community, where lower TRI implies a relatively plain terrain.

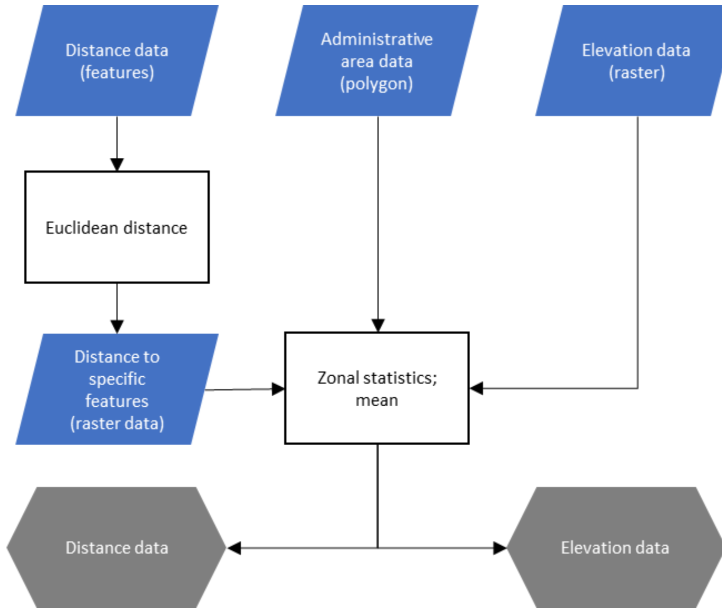

Figure 7. Data generation: distance and elevation features

3.3. Descriptive Statistics

Table 1 illustrates the descriptive statistics for accessibility and other control variables employed in this study. All distance variables are calculated in kilometers and represent the average Euclidean distance from each community to the objects. Highway expansions from 1990 to 2014 shorten the average distance from each community to nearest highway ramp by approximately 1.25 km. For railway stations, improvement in access accounts for almost 8 km throughout 1990-2014. To control for access to city center, where high level of employment is present, this study uses distance to center business district, the national monument (city center of Jakarta), and the centroid of Jakarta (constructed using GIS software). Distance to the district center, using the proxy of local government office, and distance to industrial area are also utilized to control for labor market center on sub-metropolitan area level (districts). I also add distance to coastline as a control variable (Yudhistira, et al, 2018). To control for geographical variables, this study includes average elevation level, TRI, and total land area by each community. Table 1 summarizes the descriptive statistics for distance variables and geographical variables.

I include some historical variables to be used as instrument variables for two-staged least squares (TSLS) regressions, that is old Batavia road and old Batavia tram stations. The average distance to the Batavia tramline is relatively higher than the old Batavia road since the Batavia tramline is located predominantly in the center of JMA, far away from the suburbs. Old Batavia road, on the other hand, has lower average distance since it also including a part of Anyer – Panarukan road which located inside the JMA.

Table 1. Descriptive statistics of accessibility measures and other control variables

Variables	Observations	Mean	SD	Min	Max
Distance to highways ramps 1990 (km)	1,503	9.62	7.99	0.34	42.18
Distance to highways ramps 2000 (km)	1,503	8.57	7.63	0.34	40.92
Distance to highways ramps 2014 (km)	1,503	8.37	7.61	0.34	40.92
Distance to railway stations 1990 (km)	1,503	18.04	13.18	0.37	50.81
Distance to railway stations 2000 (km)	1,503	11.43	9.33	0.37	44.73
Distance to railway stations 2014 (km)	1,503	10.19	8.62	0.37	44.73
Area (km^2)	1,503	4.53	4.68	0.28	61.05
Distance to center business district (km)	1,503	26.35	14.08	0.68	57.36
Distance to coastline (km)	1,503	22.55	15.45	0.17	63.84
Elevation (masl)	1,503	77.11	89.19	0.10	255
Terrain Ruggedness Index	1,503	1.40	0.71	0.00	3.70
Distance to local government office (km)	1,503	10.16	7.01	0.45	38.47
Distance to industrial area (km)	1,503	12.24	6.70	0.66	43.67
Distance to city center / national monument (km)	1,503	32.15	16.13	0.64	69.88
Distance to city center / centroid of Jakarta (km)	1,503	30.88	15.44	0.48	67.37
Distance to old Batavia tramline (km)	1,503	28.46	15.80	0.22	67.62
Distance to old Batavia road (km)	1,503	10.93	9.99	0.12	44.45

Source: OSM (2019), BIG (2019).

Of the 1,503 communities observed in this study, 682 of them experienced improvement in highway access with an average improvement of 2.75 km. Communities in city suburbs benefited the most with an average improvement of 3.21 km from 1990 to 2014. Unlike Jakarta and non-city suburbs, the number of areas that experienced improvement in accessibility in city suburbs was also higher than the area without improvement. For the city of Jakarta, the level of improvement was also smaller (1.78 km) compared to the suburbs. This is possibly due to most areas in Jakarta already being crossed by the inner-city highway and considering that most highways built during 1990-2014 are located on the fringe area of Jakarta and the suburbs, not inside the city. For other suburbs, out of 895 communities, almost 45% experienced improvement in access to highway ramps with average improvement around 2.75 km.

Table 2. Improvement in access to highway ramps 1990 -2014 by regions

	Improvement in access to highway		No improvement
	Number of communities	Average improvement (km)	Number of communities
Jakarta	92	1.78	171
City Suburbs	193	3.21	152
Other Suburbs	397	2.75	498
Total	682	2.75	821

Source: Authors calculation.

3.3.1. Descriptive analysis on urban expansion

As presented in table 3, the urbanized land area in the JMA is increasing over time. Using the 38m GHSL data, the average urbanized land area in the JMA increased from 1.81 km^2 in 1990 to 2.07 km^2 in 2014. Similarly, with 300m ESACCI data, the number increased around 0.43 km^2 from 1990 to 2014. Table 4 compares the development among the city of Jakarta, city suburbs and other suburbs. As expected, the percentage of urbanized land area in Jakarta is already high in the 1990s when almost 92.3% of its area is developed according to GHSL data, or around 95.3% using the ESACCI data. This is in contrast with other suburbs which are located further away from Jakarta, where the average percentage of urbanized land area is around 37–40%.

Table 3. Descriptive statistics - urban expansion measures

Variables	Observations	Mean	SD	Min	Max
Urbanized land area 1990 (%) - GHSL Data	1,503	57.73	36.33	0	100
Urbanized land area 2000 (%) - GHSL Data	1,503	60.12	36.23	0	100
Urbanized land area 2014 (%) - GHSL Data	1,503	63.44	36.09	0	100
Urbanized land area 1990 (km2) - GHSL Data	1,503	1.81	1.54	0.00	14.65
Urbanized land area 2000 (km2) - GHSL Data	1,503	1.92	1.63	0.00	14.66
Urbanized land area 2014 (km2) - GHSL Data	1,503	2.07	1.73	0.00	14.68
Urbanized land area 1992 (%) - ESACCI Data	1,503	59.15	42.31	0.00	100.00
Urbanized land area 2000 (%) - ESACCI Data	1,503	61.94	40.58	0.00	100.00
Urbanized land area 2015 (%) - ESACCI Data	1,503	72.91	37.52	0.00	100.00
Urbanized land area 1992 (km2) - ESACCI Data	1,503	2.02	1.90	0.00	16.09
Urbanized land area 2000 (km2) - ESACCI Data	1,503	2.36	1.98	0.00	16.65
Urbanized land area 2015 (km2) - ESACCI Data	1,503	2.45	2.00	0.00	16.74

Source: JRC-EU (2019), ESACCI (2019). Authors' calculation.

Table 4. Descriptive statistics - urban expansion measures by regions

Variables	Jakarta		City suburbs		Other suburbs	
	Observations	Mean	Observations	Mean	Observations	Mean
Developed area 1990 (%) - GHSL Data	263	92.29	345	83.89	895	37.48
Developed area 2000 (%) - GHSL Data	263	93.41	345	86.39	895	40.21
Developed area 2014 (%) - GHSL Data	263	97.26	345	88.78	895	43.73
Developed area 1990 (km2) - GHSL Data	263	2.18	345	2.04	895	1.60
Developed area 2000 (km2) - GHSL Data	263	2.24	345	2.12	895	1.74
Developed area 2014 (km2) - GHSL Data	263	2.31	345	2.20	895	1.94
Developed area 1992 (%) - ESACCI Data	263	95.28	345	81.41	895	39.96
Developed area 2000 (%) - ESACCI Data	263	95.40	345	84.31	895	43.48
Developed area 2015 (%) - ESACCI Data	263	99.27	345	95.14	895	56.59
Developed area 1992 (km2) - ESACCI Data	263	2.28	345	2.07	895	1.93
Developed area 2000 (km2) - ESACCI Data	263	2.37	345	2.35	895	2.36
Developed area 2015 (km2) - ESACCI Data	263	2.39	345	2.38	895	2.49

Source: JRC-EU (2019), ESACCI (2019). Authors' calculation.

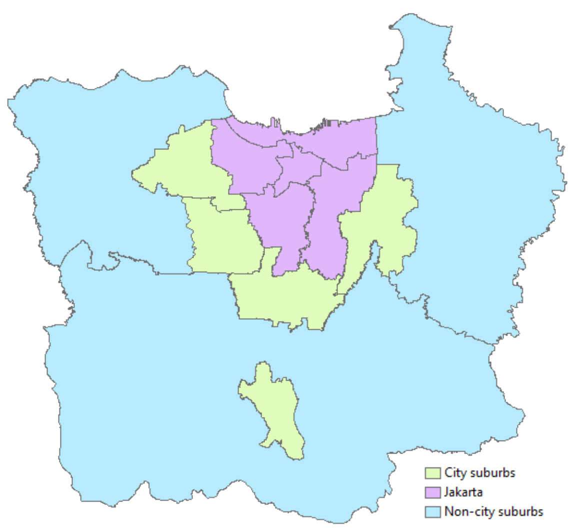

Figure 8. Jakarta Metropolitan Areas by administrative regions

3.3.2. Descriptive analysis on urban density

One of the main characteristics of urban development is the changes of density. On average, the population and population density in the JMA is rising. City suburbs have experienced the highest growth compared to the Jakarta and non-city suburbs. A similar trend occurred when taking urbanized land area into account. Using the urban density calculation as mentioned in the previous section, I found a slight increase in urban density for the city of Jakarta, while city suburbs experienced the highest growth of urban density. This calculation is consistent for both using the GHSL data and ESACCI data. The number for non-city suburbs, however, may be biased by outliers. Due to being located relatively far from Jakarta, the average urbanized land area is very low in some communities. This results in an extremely high urban density in some communities that does not make sense for interpretation. Thus, this study excludes non-city suburbs from the observations when examining the effect of highway access on urban density.

I separately calculated the number of communities which experienced a decreasing rate of urban density. Out of approximately 600 communities (when excluding non-city suburbs from observations), around 15-20% of communities, depending on the data, experienced a decreasing rate in urban density during 1990 to 2014. These figures imply that there is sufficient variation in the data to examine urban density.

Table 5. Descriptive statistics - urban density measures

Variables	Jakarta		City suburbs		Other suburbs	
	Observations	Mean	Observations	Mean	Observations	Mean
Population (inhabitant) 2000	261	31,844	340	16,769	871	8,093
Population (inhabitant) 2010	261	36,731	340	23,539	871	11,577
Population density (inhabitant per sqkm) 2000	261	18,699	340	8,127	871	2,128
Population density (inhabitant per sqkm) 2010	261	20,580	340	10,839	871	3,014
Urban density 2000 (inhabitant per sqkm) - GHSL Data	261	20,848	340	9,277	871	34,116
Urban density 2014 (inhabitant per sqkm) - GHSL Data	261	20,893	340	12,103	871	23,849
Urban density 2000 (inhabitant per sqkm) - ESACCI Data	261	20,346	334	10,922	666	10,347
Urban density 2015 (inhabitant per sqkm) - ESACCI Data	261	20,711	339	11,737	736	11,674

Source: JRC-EU (2019), ESACCI (2019), National Population Census (2010). Authors' calculation.

Table 6. Changes in urban density by regions

	Number of communities with increasing urban density		Number of communities with decreasing urban density	
	GHSL data	ESACCI data	GHSL data	ESACCI data
Jakarta	191	196	70	65
City Suburbs	317	274	23	65
Other Suburbs	663	406	208	330
Total	1171	876	301	460

Source: JRC-EU (2019), ESACCI (2019), National Population Census (2010). Authors' calculation.

3.3.3. Descriptive analysis on scattered development

The average percentage of undeveloped area surrounding the urbanized land, known as sprawl index, has been largely used for measuring sprawl (Burchfield, 2006; Garcia-Lopez, 2018). Prior studies find an increasing sprawl index from some of their observations which indicates the presence of scattered development in the JMA. In the case of the JMA, however, the sprawl index decreased for almost all regions during 2000 – 2014, thus implying an opposite development from other studies. The complete description of sprawl index by regions in the JMA is presented in Table 8. Table 9 presents the number of communities that had an increasing level of sprawl index from 1990 to 2014. According to the ESACCI data, out of approximately 1,500 communities in the JMA, only 24 of them experienced an increase in sprawl index. The variation was even lower when using the GHSL data (13 communities). The small variation in the data compared to other studies by Burchfield (2006) and Garcia-Lopez (2018) implies that the development of urbanized land in the JMA is mostly located inside the urban clusters or next to existing built-up cells rather than leapfrogging away from existing urban clusters. Considering the low variation for this indicator, this study excludes this indicator from the regression analysis and concludes that the scattered development in the JMA is not evident.

Table 7. Descriptive statistics - percentage of undeveloped surroundings

Variables	Observations	Mean	SD	Min	Max
Undeveloped surroundings 1990 (%) - GHSL data	1,502	42.62	35.46	0	99.90
Undeveloped surroundings 2000 (%) - GHSL data	1,503	40.27	35.39	0	99.97
Undeveloped surroundings 2014 (%) - GHSL data	1,503	36.98	35.34	0	99.95
Undeveloped surroundings 1992 (%) - ESACCI data	1323	32.69	36.51	0	97.96
Undeveloped surroundings 2000 (%) - ESACCI data	1363	31.66	35.06	0	97.96
Undeveloped surroundings 2015 (%) - ESACCI data	1,458	24.79	34.30	0	97.96

Source: JRC-EU (2019), ESACCI (2019). Authors' calculation.

Table 8. Percentage of undeveloped surroundings by regions

Variables	Jakarta		City suburbs		Other suburbs	
	Observations	Mean	Observations	Mean	Observations	Mean
Undeveloped surroundings 1990 (%) - GHSL data	263	8.03	345	17.10	894	62.64
Undeveloped surroundings 2000 (%) - GHSL data	263	6.95	345	14.54	895	59.98
Undeveloped surroundings 2014 (%) - GHSL data	263	3.26	345	12.05	895	56.50
Undeveloped surroundings 1992 (%) - ESACCI data	263	5.08	344	19.38	716	49.22
Undeveloped surroundings 2000 (%) - ESACCI data	263	5.00	344	16.64	756	47.76
Undeveloped surroundings 2015 (%) - ESACCI data	263	1.26	345	5.50	850	39.90

Source: JRC-EU (2019), ESACCI (2019). Authors' calculation.

Table 9. Changes in undeveloped surroundings by regions

<table>
<tr><th colspan="3"></th><th>Jakarta</th><th>City Suburbs</th><th>Other Suburbs</th><th>Total</th></tr>
<tr><td rowspan="6">GHSL data</td><td rowspan="3">Number of communities with increasing undeveloped surroundings</td><td>1990 - 2000</td><td>0</td><td>0</td><td>16</td><td>16</td></tr>
<tr><td>2000 - 2014</td><td>0</td><td>0</td><td>10</td><td>10</td></tr>
<tr><td>1990 - 2014</td><td>0</td><td>0</td><td>13</td><td>13</td></tr>
<tr><td rowspan="3">Number of communities with decreasing undeveloped surroundings</td><td>1990 - 2000</td><td>232</td><td>312</td><td>821</td><td>1365</td></tr>
<tr><td>2000 - 2014</td><td>252</td><td>333</td><td>848</td><td>1433</td></tr>
<tr><td>1990 - 2014</td><td>262</td><td>336</td><td>869</td><td>1467</td></tr>
<tr><td rowspan="6">ESACCI data</td><td rowspan="3">Number of communities with increasing undeveloped surroundings</td><td>1990 - 2000</td><td>81</td><td>120</td><td>149</td><td>350</td></tr>
<tr><td>2000 - 2014</td><td>0</td><td>0</td><td>2</td><td>2</td></tr>
<tr><td>1990 - 2014</td><td>4</td><td>5</td><td>15</td><td>24</td></tr>
<tr><td rowspan="3">Number of communities with decreasing undeveloped surroundings</td><td>1990 - 2000</td><td>68</td><td>159</td><td>565</td><td>792</td></tr>
<tr><td>2000 - 2014</td><td>151</td><td>278</td><td>775</td><td>16</td></tr>
<tr><td>1990 - 2014</td><td>149</td><td>267</td><td>793</td><td>16</td></tr>
</table>

Source: JRC-EU (2019), ESACCI (2019). Authors' calculation.

3.4. Identification and estimation strategy

Under the assumption that the development of urbanized land area and urban density is uncorrelated with transport development, I use OLS as baseline estimation. However, as prior research has pointed out, improvement in transport access is highly correlated with urban development due to simultaneous causation between two of them. It is highly plausible that the transport authority tends to build new highway lines into areas in which density is already high rather than making a new highway line into a desolate area. To address this issue, I employ TSLS regressions with historical transport infrastructures as instruments.

To account for endogeneity when estimating the impact of accessibility to urban development, this study adopts the TSLS regression model introduced by Yudhistira (2018) and Garcia-Lopez (2012) as follows:

$$\Delta \ln(y_i) = \beta_0 + \beta_1 \cdot \Delta\, dist_i^h + \beta_2 \cdot X_i + \varepsilon \qquad \text{(1) structural model}$$

$$\Delta\, dist_i^h = \alpha_0 + \alpha_1 \cdot\ Z_i + \alpha_2 \cdot X_i + \varepsilon \qquad \text{(2) first-stage estimation}$$

$$\Delta \ln(y_i) = \beta_0 + \beta_1 \cdot \Delta\, \widehat{dist}_i^h + \beta_2 \cdot X_i + \varepsilon \qquad \text{(3) second-stage estimation}$$

$$\Delta \ln(y_i) = \beta_0 + \beta_1 \cdot Z_i\ + \beta_2 \cdot X_i + \varepsilon \qquad \text{(4) reduced-form estimation}$$

The first equation shows the structural model for TSLS regression. The main explanatory variable for this study is the changes in distance to nearest highway ramp ($\Delta\, dist_i^h$) from 1990 to 2014 as proxy of improvement in transport accessibility. The outcome variables (y_i) consists of two indicators, the urbanized land area and urban density. For urbanized land area, this study uses both the GHSL data for 1990-2014 and ESACCI data for the period of 1992-2015. For urban density, this study uses the similar data measuring number of urban populations per square kilometer of urbanized land. To control for omitted variable bias, this study uses a set of control variables (X_i), consisting of distance variables and geographical variables. For distance variables, this study uses distance to nearest railway station, city center, central business districts, local government office, industrial area, and coastline. For geographical variables, this study uses elevation level, TRI, and land area.

The second equation depicts the first-stage regression. Prior studies have been using historical infrastructures, such as historical roads (Yudhistira, 2018; Garcia-Lopez, 2018; Garcia-Lopez, 2012), as an instrument for the first-stage estimation. To be valid, instruments need to be relevant and exogenous to the outcome variables. To meet the instrument relevance condition, the instruments used in the second-stage estimation should not be weak. The F-statistics of instruments used in TSLS regressions need to exceed the size and the relative bias critical value to reject the null hypothesis of weak instruments (Stock and Yogo, 2005).

The exogeneity condition requires the instruments to not directly affect the outcome variables but be channeled through the endogenous variables. In this case, historical transport infrastructures were unlikely built anticipating the current urban spatial patterns (Garcia-Lopez, 2012). On the other hand, proximity to old infrastructures would likely affect the cost of construction of new highway as pointed out by Duranton and Turner (2012). However, it is not necessary for the new transport infrastructure to be located near the old ones, due to other factors considered when building new infrastructures (Garcia-Lopez, 2012). In the case of overidentified instruments, where the number of instruments used is more than the endogenous variables, exogeneity of an instrument can be tested by looking at the overidentification p-value. If the p-value is lower than 0.05, it implies that at least one of the instruments used in the estimation is not exogenous.

To select suitable instruments for our TSLS regressions, the first-best option is to choose the instruments that satisfied the following conditions: (1) it significantly affects the endogenous variables, and (2) it significantly affects the outcome variables. The positive or negative sign for the first condition indicates that transport improvements are located either far from or close to instruments, while the negative signs in reduced-form results indicate that proximity to the instruments affects the outcome variables.

Prior studies, however, do not always provide the first-best instruments for analysis due to insignificant results on the reduced-form estimations (Garcia-Lopez, 2012; Yudhistira, 2018). To solve this problem, Garcia-Lopez (2012) decides to prioritize the first-stage estimation results in choosing which instruments to be used. In a less restrictive way, Yudhistira, et al (2018) choose the instruments by looking at which instruments produce a better F-statistics in the second-stage estimations.

In addition to general full-sample estimation, this study also employs a sub-sample regression by estimating the effect for communities in Jakarta, city suburbs, and other suburbs (Yudhistira, 2018), this study can compare the effect of transport improvements between the Jakarta area and its surrounding suburbs. It may also be useful to compare the result to what Yudhistira, et al (2018) found on the presence of suburbanization in their paper. To be noted, due to high numbers of outliers as pointed out in previous section, estimation on urban density will not include the non-city suburbs sample. Therefore

3.5. Hypothesis

The general hypothesis for urban expansion is that shorter average distance to highway ramp results in more urbanized land area in the JMA. The shorter distance to highway ramp implies reduction in transport costs of which induces the development of new built-up areas all around the JMA. Thus, I expect a negative coefficient for changes in distance to highway ramps in the urban expansion estimation. If the coefficient is statistically significant and the instruments is valid, the expansion of urbanized land in the JMA can be linked with to the highways development.

In terms of urban density, I expect a negative coefficient for suburbs, suggesting that a higher density occurred in the suburbs due to transport improvement. For Jakarta, as the metropolitan center, the impact could be positive or negative. A positive effect would imply that improvement in accessibility induces low-density development in the center of the JMA. A negative effect, however, would indicate that improvement in accessibility increases the urban density in Jakarta. If the result produces a significant negative effect, I would expect the magnitude to be less elastic than in the suburbs, supporting previous research by Garcia-Lopez, 2012; 2015; 2018) and Yudhistira, et al. (2018) on suburbanization.

4. Results

This paper divides the analysis into two sections. The first section explains the impact of improvement in highway access on urban expansion, while the second section shows the impact on urban density. For each part, this paper analyzes the first-stage estimation and reduced form results to determine which instrument is suitable for analysis despite its various limitations (Garcia-Lopez, 2012; Yudhistira, 2018). This part is then followed by the analysis of second-stage estimation using various specifications to detect the impact of highway access to urban expansion and urban density. This chapter mainly presents our estimation from the GHSL data and only briefly mention the results of the ESACCI data. The detailed results of ESACCI data are attached in the appendixes.

4.1. Urban expansion

4.1.1. First-stage estimation

To select which instruments to be used in the model, I rely on several approaches introduced by Garcia-Lopez (2012) and Yudhistira, et al (2019), as mentioned in previous chapter. Since I prepare two instruments and has one endogenous variable that needs to be instrumented, I provide three different specifications for the second-stage estimations as presented in column 3, 4, and 5 in the table 11 and 13. I use distance to old Batavia road and old Batavia tram station in the first two specification independently as instrument for distance to nearest highway ramp. Then for the third specification, I use overidentified model by jointly use those two instruments as exogenous variables.

Table 10 presents the complete results of first-stage and reduced-form estimations for improvement in highway access. The first column exhibits the unconditional estimation, while the second and third column show conditional results when controlling for the distance variables and geographical variables. For urban expansion, the distance variables are including distance to nearest rail station, distance to central business district, distance to city center, distance to nearest district center, distance to coastline, and distance to industrial area. For geographical variables, I use land area, average elevation, and terrain ruggedness level for each community.

From the first-stage and reduced-form results in table 10, the conditional model controlling for distance variables and geographical variables produce a statistically better results compared to the unconditional result. Using the preferred estimation in column 3, I select three possible combinations of instrument to be used in the second-stage estimation. For the whole JMA, only old Batavia tram-station that is statistically significant in the first-stage and reduced-form estimation. As such, only this instrument that is satisfied the first-best condition of an exogenous variable.

For sub-sample of Jakarta, on the other hand, the first-best instrument is the distance to old Batavia road. The coefficients of this variable are statistically significant in first-stage and reduced-form estimation. The distance to old Batavia tram-station, on the other hand, can only be considered as an alternative for the first-best instrument, since it only produces a significant coefficient in the first-stage estimation. The results for city suburbs and non-city suburbs are better. The coefficients for both old Batavia road and old Batavia tram-station are significant in first-stage and reduced-form estimation. Therefore, those instruments can be used as the first-best instruments for highway access.

Table 10. First stage and reduced form estimation on urban expansion (GHSL data)

	First-stage estimates			Reduced-form estimates		
	Δ distance to highway ramp			Δ ln urbanized area (GHSL data)		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. Jakarta Metropolitan Area						
Distance to old Batavia road	-0.010	0.007	0.006	0.002*	-0.006***	-0.005***
	(-1.51)	(1.00)	(0.75)	(1.71)	(-3.91)	(-3.61)
Distance to old Batavia tram-station	-0.004	0.017	0.069*	0.009***	-0.028***	-0.064***
	(-0.94)	(0.55)	(1.87)	(10.40)	(-5.23)	(-7.04)
Distance variables	No	Yes	Yes	No	Yes	Yes
Geographical variables	No	No	Yes	No	No	Yes
N	1503	1503	1503	1498	1498	1498
Adjusted R-sq	0.002	0.493	0.495	0.113	0.419	0.499
Panel B. Jakarta						
Distance to old Batavia road	-0.236***	-0.236***	-0.242***	-0.005	0.015***	0.018***
	(-3.54)	(-6.08)	(-6.12)	(-1.55)	(2.82)	(2.90)
Distance to old Batavia tram-station	-0.103***	-0.120***	-0.123**	-0.001	-0.009	-0.004
	(-4.63)	(-2.78)	(-2.44)	(-0.39)	(-0.89)	(-0.41)
Distance variables	No	Yes	Yes	No	Yes	Yes
Geographical variables	No	No	Yes	No	No	Yes
N	263	263	263	263	263	263
Adjusted R-sq	0.389	0.764	0.763	0.005	0.211	0.233
Panel C. City suburbs						
Distance to old Batavia road	-0.366***	-0.268***	-0.220***	0.005***	0.009***	0.006***
	(-9.61)	(-12.12)	(-9.16)	(3.00)	(4.98)	(3.42)
Distance to old Batavia tram-station	-0.023***	-1.498***	-1.194***	0.009***	0.013	0.024**
	(-2.67)	(-6.04)	(-4.67)	(4.96)	(1.44)	(2.19)
Distance variables	No	Yes	Yes	No	Yes	Yes
Geographical variables	No	No	Yes	No	No	Yes
N	345	345	345	345	345	345
Adjusted R-sq	0.372	0.804	0.830	0.173	0.526	0.572
Panel D. Other suburbs						
Distance to old Batavia road	0.025***	0.020***	0.025***	0.002	-0.011***	-0.009***
	(3.95)	(2.83)	(3.29)	(1.22)	(-6.21)	(-5.02)
Distance to old Batavia tram-station	-0.014**	0.151***	0.132**	0.011***	-0.071***	-0.115***
	(-2.11)	(2.70)	(2.18)	(7.25)	(-5.73)	(-7.84)
Distance variables	No	Yes	Yes	No	Yes	Yes
Geographical variables	No	No	Yes	No	No	Yes
N	895	895	895	890	890	890
Adjusted R-sq	0.014	0.462	0.465	0.053	0.469	0.546

Note: robust standard-error in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level.

4.1.2. OLS and TSLS estimation

Tables 11 provides the regression results for examining the impact of highway improvement on urban expansion using the GHSL data. For descriptive purposes, I provide OLS results in the first two columns. The first column presents the unconditional specification, while the second column shows estimation result conditional on numerous control variables. The preferred specification of OLS estimation controlling for distance and geographical variables in the second column shows that improvement in access to nearest highway ramps is associated with the urban expansion in the JMA. One kilo meter improvement in access to highway corresponds to higher the urbanized land area in the JMA by 2%. The effect is higher for the non-city suburbs (2.3%) than in the city suburbs (1.12%). The impact is consistent with the result using the ESACCI data. Although the impact is slightly different in magnitude, especially for non-city suburbs (Appendix 2). These results, however, are only valid under the assumptions that improvement in highway does not simultaneously affect urbanized land area. Due to this endogeneity issue, the results may still suffer from causality bias.

For the whole JMA, the preferred estimation using distance to old Batavia tram station as first-best instrument presented in column (4) of table 11 shows that the F-statistics do not exceed the 10% of critical value (16.38). The estimation for non-city suburbs produces similar result. Using various specifications presented in column (3), (4), and (5) the result does not satisfy the instruments relevance condition. The other results, on the other hand, confirm that the expansion of highway is indeed expand the urbanized land area for both the city of Jakarta and the city suburbs. One kilo meter improvement in highway access in the city of Jakarta increases the urbanized land area by 6.6-7.7%. The impact is somewhat smaller in the city suburbs. One kilo meter improvement in highway access causes an urbanized land area to increase by 2.6 – 3.2% in city suburbs, depending on the specification that is used. The result is robust for the city suburbs using the ESACCI data for city suburbs, although in higher magnitude most likely due to different resolution between these data.

These results in line with findings from Garcia-Lopez (2018) which found a causal evidence of highways expansion and new land developments in European cities. Although in the context of this study, this urban expansion may not imply that the JMA is experiencing an urban sprawl. One of the reasons is in study by Garcia-Lopez (2018) the analysis is conducted in a larger spatial level (cities), while our estimation is in a smaller spatial level (communities). The main interpretation may also differ when considering that this paper distinguishes the impact in city center and surrounding suburbs of a metropolitan area. A greater magnitude in city center might indicate that the process of infilling urban spaces in city center is happening in a larger extent compared to new land expansion in the suburbs.

Table 11. Regression results for urban expansion (GHSL data)

	Δ ln urbanized area (GHSL data)				
	OLS		TSLS		
	(1)	(2)	(3)	(4)	(5)
Panel A. Jakarta Metropolitan Area					
Δ distance to highway ramp	0.039***	-0.020***	-1.010	-0.898*	-0.915*
	(12.11)	(-6.61)	(-0.61)	(-1.87)	(-1.91)
distance to railway station	0.004***	0.001	-0.133	-0.118*	-0.120*
	(5.01)	(0.39)	(-0.59)	(-1.80)	(-1.83)
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	1498	1498	1498	1498	1498
Adjusted R-sq	0.052	0.476			
Cragg-Donald F-statistics			0.492	2.936	1.745
Overidentification p-value					0.9432
Panel B. Jakarta					
Δ distance to highway ramp	0.010***	-0.011	-0.077***	0.086	-0.066**
	(3.49)	(-1.08)	(-2.59)	(0.59)	(-2.45)
distance to railway station	0.001	-0.018***	-0.030***	0.000	-0.028***
	(0.33)	(-2.64)	(-2.94)	(0.01)	(-2.94)
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	263	263	263	263	263
Adjusted R-sq	0.002	0.213			
Cragg-Donald F-statistics			44.947	1.889	25.192
Overidentification p-value					0.203
Panel C. City suburbs					
Δ distance to highway ramp	0.010***	-0.012**	-0.032***	0.004	-0.026***
	(4.60)	(-2.56)	(-2.66)	(0.12)	(-3.33)
distance to railway station	-0.005***	0.010**	-0.003	0.021	0.001
	(-4.34)	(2.30)	(-0.33)	(0.86)	(0.20)
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	345	345	345	345	345
Adjusted R-sq	0.042	0.571			
Cragg-Donald F-statistics			49.352	3.235	55.759
Overidentification p-value					0.459
Panel D. Other suburbs					
Δ distance to highway ramp	0.052***	-0.023***	-0.365***	-0.880**	-0.507***
	(10.02)	(-5.34)	(-2.85)	(-2.23)	(-3.74)
distance to railway station	-0.002	-0.006**	-0.038***	-0.085**	-0.051***
	(-1.13)	(-2.18)	(-2.92)	(-2.26)	(-3.67)
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	890	890	890	890	890
Adjusted R-sq	0.058	0.498			
Cragg-Donald F-statistics			10.610	4.017	7.359
Overidentification p-value					0.087

Note: robust standard-error in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level

4.2. Urban density

4.2.1. First-stage estimation

Table 12 presents the first-stage and reduced-form estimation for urban density using the GHSL data. Column 1 exhibit the unconditional specification, while column 2, 3, and 4 shows conditional specification using several control variables, such as, lagged urban density in 2000, distance variables and geographical variables. For distance variables, I include distance to nearest railway station, distance to center of Jakarta, district center, coastline, and industrial area. While for geographical variables, I include land area, elevation, and terrain ruggedness level.

As mentioned in the the descriptive statistics part, I only use communities in Jakarta and city suburbs as observations due to high outliers in non-city suburbs data. Using similar approach when determining instruments used for estimating urban expansion, I select either one instrument or a combination of two instruments as exogenous variable depending on the first-stage and reduced-form results. The results show that both instruments are statistically significant in either first-stage or the reduced-form estimation for full-sample as well as sub-sample of Jakarta and city suburbs. This result allows me to use both instruments in the second-stage estimation.

Table 12. First-stage and reduced-form estimation for urban density using GHSL data

	First-stage estimates				Reduced-form estimates			
	Δ distance to highway ramp				Δ ln urbanized area (GHSL data)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Panel A. Jakarta Metropolitan Area								
Distance to old Batavia road	-0.334***	-0.323***	-0.276***	-0.256***	0.025***	0.014***	0.011***	0.012***
	(-11.15)	(-9.92)	(-11.51)	(-10.29)	(10.97)	(5.82)	(4.34)	(4.58)
Distance to old Batavia tram-station	-0.004	-0.001	-0.491***	-0.419***	0.004***	0.001**	0.026***	0.048***
	(-1.61)	(-0.49)	(-10.91)	(-8.32)	(5.75)	(2.01)	(5.18)	(7.28)
Lagged urban density 2000	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Distance variables	No	No	Yes	Yes	No	No	Yes	Yes
Geographical variables	No	No	No	Yes	No	No	No	Yes
N	608	601	601	601	601	601	601	601
Adjusted R-sq	0.397	0.399	0.552	0.556	0.244	0.421	0.455	0.496
Panel B. Jakarta								
Distance to old Batavia road	-0.236***	-0.234***	-0.393***	-0.339***	-0.011**	-0.012***	-0.018***	-0.019***
	(-3.54)	(-3.51)	(-7.65)	(-6.76)	(-2.47)	(-2.99)	(-3.03)	(-2.96)
Distance to old Batavia tram-station	-0.103***	-0.112***	-0.411***	-0.246***	0.030***	0.023***	0.034***	0.036***
	(-4.63)	(-4.86)	(-10.61)	(-3.91)	(10.32)	(7.99)	(4.64)	(3.43)
Lagged urban density 2000	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Distance variables	No	No	Yes	Yes	No	No	Yes	Yes
Geographical variables	No	No	No	Yes	No	No	No	Yes
N	263	261	261	261	261	261	261	261
Adjusted R-sq	0.389	0.392	0.669	0.690	0.293	0.371	0.388	0.404
Panel C. City suburbs								
Distance to old Batavia road	-0.366***	-0.349***	-0.102***	-0.078**	0.015***	0.008**	0.006**	0.008**
	(-9.61)	(-9.03)	(-3.24)	(-2.45)	(4.27)	(2.51)	(2.16)	(2.52)
Distance to old Batavia tram-station	-0.023***	-0.025***	-1.718***	-1.570***	-0.004***	-0.003**	0.017	0.031*
	(-2.67)	(-2.80)	(-7.07)	(-6.48)	(-2.85)	(-2.12)	(1.16)	(1.96)
Lagged urban density 2000	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Distance variables	No	No	Yes	Yes	No	No	Yes	Yes
Geographical variables	No	No	No	Yes	No	No	No	Yes
N	345	340	340	340	340	340	340	340
Adjusted R-sq	0.372	0.380	0.692	0.696	0.139	0.336	0.400	0.426

Note: robust standard-error in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level.

4.2.2. OLS and TSLS estimation

Table 13 presents the regression result for urban density using the GHSL data. Column 1 and 2 exhibit the OLS estimation using unconditional and conditional specification respectively, while column 3 to 5 present the TSLS estimation using three different combination of instruments. The OLS estimation using the GHSL data for the full sample provides statistically significant results, while the sub-sample estimation provides no coefficient that is statistically different from zero. The negative coefficient in full sample estimation indicates that one kilo meter improvement in access to highway corresponds to an increase urban density in the JMA by 1.5%. The results are substantially different for the ESACCI data (Appendix 4). The full specification in column 2 of Appendix 4 shows that the result is only significant in city suburbs. The results, however, provide a positive coefficient, implying that improvement in access to highways associated with reduction in urban density by 2.4%. Nevertheless, the OLS results still suffer from causality bias.

Column 3 to 5 in table 13 depict the TSLS regression results for urban density. For the full-sample estimation, the result shows that improvement in highway access fosters urban density in the JMA in general. One kilo meter improvement in highway access increase the urban density by 5.4-10.1% depends on the instrument used in the model. However, the result using both distance to old Batavia road and old Batavia tram station produces low overidentification p-value, indicating that one of the instruments is not exogenous. Nevertheless, the result confirms the presence of highway-led urban density growth in the JMA.

The similar result happened for the Jakarta sample, both instruments provide a low overidentification p-value, despite being the first-best option of instruments. Using only distance to old Batavia tram-station as instruments (column 4), the result produces a low F-statistics, indicating a weak instrument is used in the estimation. The other specification using old Batavia road as instrument (column 3) produces high F-statistics and a positive sign. It indicates that highway development leads to reduction in urban density. The magnitude implies that one kilo meter improvement in highway access reduces urban density by 6%. This result, however, may not hold should the distance to old Batavia road is not an exogenous variable, which may be the case considering the result in column 5. Hence, the impact of improvement in highway access in Jakarta is statistically inconclusive.

The result is somewhat more vivid for city suburbs. The instrument relevant condition is satisfied for all specifications. The negative coefficient indicates that improvement in highway access lead to an increase in urban density. It implies that one kilo meter improvement in access to highway in the JMA fosters urban density in city suburbs by 2.6-5.7%. This results in line with the study by Garcia-Lopez (2012) and Yudhistira (2018) about the highway-led population density growth in suburbs.

The estimation results using the ESACCI data, on the other hand, can only confirm an increase in urban density for the full-sample of JMA. The estimation results for all sub-sample is statistically inconclusive due to weak instruments issue. The detailed estimation of the ESACCI data is presented in Appendix 4.

Table 13. Regression results for urban density using GHSL data

	Δ ln urbanized area (GHSL data)				
	OLS		TSLS		
	(1)	(2)	(3)	(4)	(5)
Panel A. Jakarta Metropolitan Area					
Δ distance to highway ramp	-0.033***	-0.015***	-0.054***	-0.101***	-0.067***
	(-7.20)	(-2.78)	(-5.11)	(-5.86)	(-6.93)
distance to railway station	0.011***	-0.006	-0.018***	-0.032***	-0.022***
	(6.88)	(-1.46)	(-3.28)	(-4.63)	(-4.10)
Lagged urban density 2000	No	Yes	Yes	Yes	Yes
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	601	601	601	601	601
Adjusted R-sq	0.190	0.427			
Cragg-Donald F-statistics			174.877	67.742	121.156
Overidentification p-value					0.007
Panel B. Jakarta					
Δ distance to highway ramp	-0.021**	0.004	0.060***	-0.170**	0.022
	(-2.53)	(0.46)	(2.75)	(-2.23)	(1.22)
distance to railway station	0.023***	0.011*	0.024***	-0.029	0.015**
	(5.84)	(1.92)	(3.04)	(-1.41)	(2.28)
Lagged urban density 2000	No	Yes	Yes	Yes	Yes
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	261	261	261	261	261
Adjusted R-sq	0.169	0.367			
Cragg-Donald F-statistics			70.505	10.358	45.788
Overidentification p-value					0.000
Panel C. City suburbs					
Δ distance to highway ramp	-0.026***	0.002	-0.057***	-0.026**	-0.029***
	(-5.02)	(0.23)	(-2.59)	(-2.38)	(-2.69)
distance to railway station	0.001	-0.024***	-0.036***	-0.030***	-0.030***
	(0.47)	(-4.38)	(-4.61)	(-5.20)	(-5.25)
Lagged urban density 2000	No	Yes	Yes	Yes	Yes
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	340	340	340	340	340
Adjusted R-sq	0.076	0.406			
Cragg-Donald F-statistics			31.696	160.962	86.457
Overidentification p-value					0.079

Note: robust standard-error in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level.

5. Discussion and conclusions

5.1. Highways access and urban development in the JMA

The general objective of this study is to examine whether highway accessibility represents an exogenous variable that induces various forms of urban development, including urban expansion, urban density, and scattered development of urban area. On scattered development of urban area, our descriptive statistics shows that the location of new urban settlement in the JMA is more often infilled the urban spaces or developed between urban settlement rather than leapfrogged the existing urban settlement. It is indicated by the decreasing level of undeveloped surrounding (sprawl index) in the JMA from 1990 to 2014. Our calculation is robust using two different datasets, the GHSL data and ESACCI data.

Using the availability of pre-existing data on historical transport infrastructures in the JMA, I have been able to draw on these data as instruments, along similar lines as in research previously conducted by Garcia-Lopez (2014; 2018) and Yudhistira (2018) to identify any causation between improvement in highway access and urban development, in the form of urban expansion and urban density.

On urban expansion, this paper evidently shows that highway expansion fosters urbanized land area in the city of Jakarta and its city suburbs. Interestingly, the estimated magnitude is higher in Jakarta than in the city suburbs. It indicates that the conversion of land from non-urban settlement to urban settlement caused by highway expansion is faster in Jakarta rather than in the city suburbs from 1990 to 2014. It also implies that the process of infilling urban spaces in city center is happening in a larger extent compared to spreading development in the suburbs. Although the results are more apparent using the GHSL data, the presence of increasing urbanized land area in city suburbs is also confirmed from the ESACCI data. Nevertheless, the result implies the existence of highway-led urban expansion in the JMA.

On urban density, this paper finds no evidence of low-density urban development in the JMA. The result of this paper, however, in line with the study from Garcia-Lopez (2012) and Yudhistira, et al (2018) in showing an increase in densities in the suburbs. This paper also is unable to confirm the presence of slower growth of urban density in the city center (Jakarta), which may indicate the presence of low-density development of city center as well as the presence of suburbanization in the JMA. The empirical estimation for the city of Jakarta is statistically inconclusive mainly due to limited instruments that are available and compatible for the estimations, as we often encountered in urban economics literatures.

The empirical results of this study provide some important insights regarding urban development in the JMA. Causal evidence of the impact of better access to highways on urban expansion suggests that the JMA authority should prepare for the additional costs of provision of public goods and services associated with an increase in new land development and higher urban density. They should also take the various externalities associated with such development into account when developing highway expansion plan.

This paper also found no evidence of urban sprawl in the JMA as reflected in decreasing level of scattered development throughout the years. It may indicate that the JMA is arguably in mature development stage of the city, where the development of city center is more likely to infill the urban spaces rather than scattered from the existing urban settlement (Wagtendonk, et al, 2019). This paper also found no statistical evidence of low-urban density development whether in the city center or in the whole JMA. It may also imply that the agglomeration benefits from the JMA is still high, as such people still choose to live in the JMA.

One of the novels of this study is the utilization of a fine resolution of satellite images from the GHSL and ESACCI to generate various indicators of urban development and aggregate it at community level to provide a detailed analysis of urban development while also subsequently analyze the presence of sprawl in the JMA. Employing multi-sources data is also important to ensure that results are consistent. The utilization of both GHSL and ESACCI data, to the best of my knowledge, is also still limited for a study in large developing countries with megacities like the JMA.

This paper adds an interesting perspective on urban development in large developing countries, something that remains under-studied in the existing empirical literatures for this topic. This paper uses urban density instead of population density to account for changes in urbanized land area in the JMA, something that is not considered in previous literatures from Garcia-Lopez (2012) and Yudhistira, et al, (2018). This paper also measures urban sprawl through multiple indicators instead of only using a single indicator urban sprawl. It allows us to ensure the absence of urban sprawl during urban development process of the JMA.

5.2. Avenues for future research

Some may argue that urban sprawl is more apparent in a rather urban area rather than in a large metropolitan area. Considering the numerous consequences of unmitigated urban sprawl, it may be important to extend the analysis to other cities in Indonesia. The absence of urban sprawl in the JMA does not imply the similar condition happened in other cities, considering other cities are probably not yet in the same development stage as Jakarta. Having numerous car-dependent cities might be a tempting reason to examine the presence of transport-led urban sprawl in other cities in Indonesia.

Unmitigated urban development may lead to lower agglomeration benefits and higher economics cost for urban areas. Following the causal evidence between improvement in highway access and urban development. It might be better to extend the research by calculating the costs associated with urban expansion and changes in urban density due to highway development. The calculation may include the environmental and socio-economic impact of the development of new highway. It may also include the loss of agricultural and natural resources area due to conversion of land into urban settlement. The research might be beneficial to denote how much social costs that entail to urban development in the future.

References

Ahrens, A., & Lyons, S. (2019). Changes in land cover and urban sprawl in Ireland from a comparative perspective over 1990-2012. *Land 8*.

Angel, S., Parent, J., & Civco, D. L. (2012). The fragmentation of urban landscapes: global evidence of a key attribute of the spatial structure of cities 1990-2000. *Environment and Urbanization 24*, 249-283.

Bai, X., Chen, J., & Shi, P. (2012). Landscape urbanization and economic growth in China: Positive feedbakcs and sustainable dilemmas. *Environmental Science and Technology 46*, 132-129.

Baum-Snow, N. (2007). Did highways cause suburbanization? *Quarterly Journal of Economics 122 (2)*, 775-805.

Braubach, M., Egorov, A., Mudu, P., Wolf, T., Ward Thompson, C., & Martuzzi, M. (2017). Effects of urban green space on environmental health, equity and Resilience. In N. Kabisch, H. Korn, J. Stadler, & A. Bonn, *Nature-based solutions to climate change adaptation in urban areas. Theory and practice of urban sustainablity.* Springer.

Breunig, R., Hasan, S., & Whiteoak, K. (2019). Value of playgrounds relative to green spaces: Matching evidence from property prices in Australia. *Landscape and Urban Planning 190*.

Brueckner, J. K. (2000). Urban sprawl: Diagnosis and remedies. *International Regional Science Review 23*, 160-171.

Brueckner, J. K., & Fansler, D. A. (1983). The economics of urban sprawl: Theory and evidence on the spatial sizes of cities. *The Review of Economics and Statistics 65*, 479-482.

Burchfield, M., Overman, H. G., Puga, D., & Turner, M. A. (2006). Causes of sprawl: A portrait from space. *The Quarterly Journal of Economics*, 588-632.

Corbane, C., Florczyk, A., Pesaresi, M., Politis, P., & Syrris, V. (2019, 05 01). *GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), R2018A.* Retrieved from European Commission, Joint Research Centre (JRC): http://data.europa.eu/89h/jrc-ghsl-10007

Deng, X., Huang, J., Rozelle, S., & Uchida, E. (2008). Growth, population and industrialization, and urban land expansion of China. *Journal of Urban Economics 63*, 96-115.

Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, 1-36.

Engelfriet, L. (2016). *The impact of urban form on commuting in Chinese megacities.* Amsterdam: Vrije Universiteit Amsterdam Thesis.

European Space Agency. (2019, 05 01). *European Space Agency - Landcover Project*. Retrieved from European Space Agency, Climate Change Initiatives: https://www.esa-landcover-cci.org/?q=node/164

Ewing, R., Hamidi, S., Grace, J. B., & Wei, Y. D. (2016). Does urban sprawl hold down upward mobility? *Landscape and Urban Planning 148*, 153-169.

Fallah, B. N., Partridge, M. D., & Olfert, M. R. (2011). Urban sprawl and productivity: Evidence from US metropolitan areas. *Papers in Regional Science 90*, 452-472.

Frenkel, A., & Ashkenazi, M. (2008). Measuring urban sprawl: How can we deal with it? *Environment and Planning B: Urban Analytics and City Science 35*, 56-79.

Garcia-Lopez, M.-A. (2012). Urban spatial structure, suburbanization and transportation in Barcelona. *Journal of Urban Economics 72*, 176-190.

Garcia-Lopez, M.-A. (2018). All roads lead to Rome and to Sprawl? Evidence from European cities. *IEB Working Paper 02*.

Garcia-Lopez, M.-A., Holl, A., & Viladecans-Marsal, E. (2015). Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics 85*, 52-67.

Geospatial Information Agency. (2019, 05 01). *Seamless Digital Elevation Model (DEM) and Batimetri Nasional*. Retrieved from Badan Informasi Geospatial: http://tides.big.go.id/DEMNAS/

Glaeser, E. L., & Kahn, M. E. (2004). Sprawl and urban growth. In J. V. Henderson, & J. F. Thisse, *Handbook of Regional and Urban Economics* (pp. 2481-2527). Elsevier.

Glaeser, E. L., & Kahn, M. E. (2010). The greenness of cities: Carbon dioxide emissions and urban development. *Journal of Urban Economics 67*, 404-418.

Gomez-Antonio, M., Hortas-Rico, M., & Li, L. (2014). The causes of urban sprawl in Spanish urban areas: a spatial approach. *GEN Working Paper B*.

Gordon, P., Kumar, A., & Richardson, H. W. (1989). The influence of metropolitan spatial structure on commuting time. *Journal of Urban Economics 26*, 138-151.

Guastella, G., Oueslati, W., & Pareglio, S. (2019). Patterns of urban spatial expansion in European cities. *Sustainability 11*.

International Institute for Environment and Development. (2014, 09 02). *Introduction to urbanisation and rural-urban linkages*. Retrieved from International Institute for Environment and Development: https://www.iied.org/introduction-urbanisation-rural-urban-linkages

International Organization for Migration. (2015). *World Migration Report.* Geneva: International Organization for Migration.

Jabodetabek Railway Authority. (2019, 05 01). *Kereta Commuter Indonesia*. Retrieved from PT. Kereta Commuter Jabodetabek: http://www.krl.co.id/

Kahn, M. E. (2000). The environmental impact of suburbanization. *Journal of Policy Analysis and Management Vol. 19, No.4*, 570-586.

Lee, W. H., Ambrey, C., & Pojani, D. (n.d.). How do sprawl and inequality affect well-being in American cities? *Cities*.

Lefevre, B. (2009). Urban transport energy consumption: Determinants and strategies for its reduction. *SAPIENS 2*.

Levinson, D. M., & Kumar, A. (1994). The rational locator: Why travel times have remained stable. *Journal of the American Planning Association 60*, 319-332.

McGrath, D. T. (2005). More evidence on the spatial scale of cities. *Journal of Urban Economics 58*, 1-10.

Mieszkowski, P., & Mills, E. S. (1993). The causes of metropolitan suburbanization. *Journal of Economic Perspectives Vol. 7, No.3*, 135-147.

Ministry of Industry. (2019, 05 01). *Daftar Kawasan Industri*. Retrieved from Kementerian Perindustrian Republic of Indonesia: https://kemenperin.go.id/kawasan

National Highway Authorities. (2019, 05 01). *Jalan Tol Beroperasi di Indnesia*. Retrieved from Kementerian Pekerjaan Umum dan Perumahan Rakyat, Badan Pengatur Jalan Tol: http://bpjt.pu.go.id/konten/progress/beroperasi

Newman, P. W., & Kenworthy, J. R. (1989). Gasoline consumption and cities: A comparisan of US cities with a global survey. *Journal of the American Planning Association 55 (1)*, 52-60.

Newman, P., & Kenworthy, J. R. (1999). *Sustainability and cities: Overcoming automobiles dependence.* Washington DC: Island Press.

OpenStreetMap contributors. (2019, 05 01). *OSM Data Indonesia*. Retrieved from OpenStreetMap: http://download.geofabrik.de/asia/indonesia.html

Oueslati, W., Alvanides, S., & Garrod, G. (2014). Determinants of urban sprawl in European cities. *GRANEM Paper*, 1-26.

Sole-Olle, A., & Hortas-Rico, M. (2008). Does urban sprawl increase the costs of providing local public services? Evidence from Spanish municipalities. *IEB Working Paper*.

Statistics Indonesia. (2018, 05 01). *Jabodetabek Commuting Survey 2014*. Retrieved from Badan Pusat Statistik: https://www.bps.go.id/

Statistics Indonesia. (2019, 05 01). *National Labor Survey 2018*. Retrieved from Badan Pusat Statistics: https://www.bps.go.id/

Statistics Indonesia. (2019, 05 01). *National Population Census 2000*. Retrieved from Badan Pusat Statistik: https://www.bps.go.id/

Statistics Indonesia. (2019, 05 01). *National Population Census 2010*. Retrieved from Badan Pusat Statistik: https://www.bps.go.id/

Transportation Research Board, National Research Council. (1998). *The costs of sprawl - Revisited.* Washington D.C.: National Academy Press.

United Nations, Department of Economic and Social Affairs, Population Division. (2014). *World Urbanization Prospects: The 2014 revision.* United Nations.

Wei, Y. D., Li, H., & Yue, W. (2017). Urban land expansion and regional inequality in transitional China. *Landscape and Urban Planning 163*, 17-31.

Yudhistira, M., Indriyani, W., Pratama, A. P., Sofiyandi, Y., & Kurniawan, Y. R. (2018). Transportation network and changes in urban structure: Evidence from the Jakarta Metropolitan Area. *Research in Transportation Economics*.

Appendix

Appendix 1: First stage and reduced form estimation on urban expansion: ESACCI data

	First-stage estimates			Reduced-form estimates		
	Δ distance to highway ramp			Δ ln urbanized area (ESACCI data)		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. Jakarta Metropolitan Area						
Distance to old Batavia road	-0.010	0.007	0.006	0.007***	0.009***	0.007***
	(-1.51)	(1.00)	(0.75)	(2.98)	(4.45)	(3.27)
Distance to old Batavia tram-station	-0.004	0.017	0.069*	0.014***	0.053***	0.083***
	(-0.94)	(0.55)	(1.87)	(10.20)	(4.01)	(4.36)
Distance variables	No	Yes	Yes	No	Yes	Yes
Geographical variables	No	No	Yes	No	No	Yes
N	1503	1503	1503	1211	1211	1211
Adjusted R-sq	0.002	0.493	0.495	0.106	0.311	0.325
Panel B. Jakarta						
Distance to old Batavia road	-0.236***	-0.236***	-0.242***	-0.009**	0.007	0.009
	(-3.54)	(-6.08)	(-6.12)	(-2.26)	(1.40)	(1.33)
Distance to old Batavia tram-station	-0.103***	-0.120***	-0.123**	0.004	0.022**	0.027**
	(-4.63)	(-2.78)	(-2.44)	(1.43)	(2.41)	(2.08)
Distance variables	No	Yes	Yes	No	Yes	Yes
Geographical variables	No	No	Yes	No	No	Yes
N	263	263	263	263	263	263
Adjusted R-sq	0.389	0.764	0.763	0.006	0.106	0.102
Panel C. City suburbs						
Distance to old Batavia road	-0.366***	-0.268***	-0.220***	0.013***	0.015***	0.026***
	(-9.61)	(-12.12)	(-9.16)	(3.97)	(2.76)	(3.88)
Distance to old Batavia tram-station	-0.023***	-1.498***	-1.194***	0.022***	0.035	0.145***
	(-2.67)	(-6.04)	(-4.67)	(5.98)	(1.06)	(3.13)
Distance variables	No	Yes	Yes	No	Yes	Yes
Geographical variables	No	No	Yes	No	No	Yes
N	345	345	345	327	327	327
Adjusted R-sq	0.372	0.804	0.830	0.188	0.322	0.397
Panel D. Other suburbs						
Distance to old Batavia road	0.025***	0.020***	0.025***	0.003	0.008**	0.004
	(3.95)	(2.83)	(3.29)	(1.31)	(2.34)	(0.96)
Distance to old Batavia tram-station	-0.014**	0.151***	0.132**	-0.001	-0.090***	-0.075**
	(-2.11)	(2.70)	(2.18)	(-0.29)	(-3.16)	(-2.40)
Distance variables	No	Yes	Yes	No	Yes	Yes
Geographical variables	No	No	Yes	No	No	Yes
N	895	895	895	621	621	621
Adjusted R-sq	0.014	0.462	0.465	-0.001	0.376	0.387

Note: robust standard-error in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level.

Appendix 2: Regression result on urban expansion: ESACCI data

	Δ ln urbanized area (ESACCI data)				
	OLS		TSLS		
	(1)	(2)	(3)	(4)	(5)
Panel A. Jakarta Metropolitan Area					
Δ distance to highway ramp	-0.016*	-0.022**	-0.502	0.958	0.270
	(-1.79)	(-2.29)	(-1.23)	(1.60)	(1.48)
distance to railway station	-0.001	-0.026***	-0.090	0.104	0.013
	(-0.45)	(-4.30)	(-1.63)	(1.29)	(0.53)
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	1211	1211	1211	1211	1211
Adjusted R-sq	0.002	0.311			
Cragg-Donald F-statistics			2.475	2.762	2.487
Overidentification p-value					0.001
Panel B. Jakarta					
Δ distance to highway ramp	0.007**	-0.012	-0.030	-0.302	-0.047
	(2.00)	(-1.00)	(-1.15)	(-1.36)	(-1.57)
distance to railway station	0.004	-0.005	-0.009	-0.059	-0.012
	(0.95)	(-0.86)	(-1.00)	(-1.24)	(-1.27)
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	263	263	263	263	263
Adjusted R-sq	0.000	0.084			
Cragg-Donald F-statistics			44.947	1.889	25.192
Overidentification p-value					0.044
Panel C. City suburbs					
Δ distance to highway ramp	0.010*	-0.034**	-0.116***	-0.145	-0.121***
	(1.70)	(-2.22)	(-2.77)	(-0.82)	(-3.52)
distance to railway station	-0.019***	-0.067**	-0.120***	-0.139	-0.123***
	(-5.12)	(-2.46)	(-3.50)	(-1.15)	(-3.65)
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	327	327	327	327	327
Adjusted R-sq	0.077	0.377			
Cragg-Donald F-statistics			43.358	2.659	48.760
Overidentification p-value					0.876
Panel D. Other suburbs					
Δ distance to highway ramp	-0.066***	-0.070***	0.154	-2.822	0.090
	(-6.01)	(-5.67)	(0.88)	(-0.38)	(0.57)
distance to railway station	-0.035***	-0.055***	-0.037**	-0.273	-0.042**
	(-12.43)	(-6.39)	(-2.00)	(-0.47)	(-2.49)
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	621	621	621	621	621
Adjusted R-sq	0.230	0.406			
Cragg-Donald F-statistics			5.886	0.099	3.024
Overidentification p-value					0.017

Note: robust standard-error in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level.

Appendix 3: First stage and reduced form estimation on urban density: ESACCI data

	First-stage estimates				Reduced-form estimates			
	Δ distance to highway ramp				Δ ln urbanized area (ESACCI data)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Panel A. Jakarta Metropolitan Area								
Distance to old Batavia road	-0.334***	-0.321***	-0.272***	-0.256***	0.023***	0.007**	0.004	0.008**
	(-11.15)	(-10.05)	(-11.38)	(-10.26)	(8.24)	(2.34)	(1.35)	(2.30)
Distance to old Batavia tram-station	-0.004	-0.004	-0.475***	-0.391***	-0.006***	-0.008***	0.005	0.003
	(-1.61)	(-1.59)	(-10.65)	(-7.97)	(-3.57)	(-5.62)	(0.77)	(0.27)
Lagged urban density 2000	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Distance variables	No	No	Yes	Yes	No	No	Yes	Yes
Geographical variables	No	No	No	Yes	No	No	No	Yes
N	608	595	595	595	595	595	595	595
Adjusted R-sq	0.397	0.406	0.555	0.558	0.120	0.362	0.379	0.419
Panel B. Jakarta								
Distance to old Batavia road	-0.236***	-0.234***	-0.394***	-0.341***	-0.008*	-0.010**	-0.010*	-0.009
	(-3.54)	(-3.51)	(-7.67)	(-6.87)	(-1.91)	(-2.31)	(-1.91)	(-1.62)
Distance to old Batavia tram-station	-0.103***	-0.112***	-0.410***	-0.243***	0.026***	0.020***	0.023***	0.023**
	(-4.63)	(-4.85)	(-10.67)	(-3.87)	(9.39)	(6.46)	(3.62)	(2.00)
Lagged urban density 2000	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Distance variables	No	No	Yes	Yes	No	No	Yes	Yes
Geographical variables	No	No	No	Yes	No	No	No	Yes
N	263	261	261	261	261	261	261	261
Adjusted R-sq	0.389	0.392	0.669	0.691	0.247	0.310	0.317	0.322
Panel C. City suburbs								
Distance to old Batavia road	-0.366***	-0.357***	-0.106***	-0.084**	0.005	-0.003	-0.007	-0.003
	(-9.61)	(-9.52)	(-3.38)	(-2.59)	(1.06)	(-0.71)	(-1.60)	(-0.69)
Distance to old Batavia tram-station	-0.023***	-0.039***	-1.692***	-1.567***	-0.019***	-0.014***	0.023	-0.018
	(-2.67)	(-3.84)	(-7.04)	(-6.55)	(-6.62)	(-6.79)	(1.15)	(-0.88)
Lagged urban density 2000	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Distance variables	No	No	Yes	Yes	No	No	Yes	Yes
Geographical variables	No	No	No	Yes	No	No	No	Yes
N	345	334	334	334	334	334	334	334
Adjusted R-sq	0.372	0.393	0.694	0.697	0.244	0.457	0.503	0.548

Note: robust standard-error in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% level.

Appendix 4: Regression result on urban density: ESACCI data

	Δ ln urbanized area (GHSL data)				
	OLS		TSLS		
	(1)	(2)	(3)	(4)	(5)
Panel A. Jakarta Metropolitan Area					
Δ distance to highway ramp	-0.026***	-0.007	-0.028**	-0.012	-0.024*
	(-4.82)	(-1.12)	(-2.16)	(-0.58)	(-1.78)
distance to railway station	0.013***	-0.007	-0.013**	-0.008	-0.011*
	(7.63)	(-1.27)	(-2.16)	(-1.04)	(-1.92)
Lagged urban density 2000	No	Yes	Yes	Yes	Yes
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	595	595	595	595	595
Adjusted R-sq	0.112	0.416			
Cragg-Donald F-statistics			174.270	63.231	117.312
Overidentification p-value					0.3923
Panel B. Jakarta					
Δ distance to highway ramp	-0.020**	0.000	0.030*	-0.108*	0.008
	(-2.56)	(0.05)	(1.72)	(-1.81)	(0.46)
distance to railway station	0.020***	0.005	0.012*	-0.019	0.007
	(5.47)	(0.99)	(1.87)	(-1.24)	(1.15)
Lagged urban density 2000	No	Yes	Yes	Yes	Yes
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	261	261	261	261	261
Adjusted R-sq	0.153	0.309			
Cragg-Donald F-statistics			72.004	9.947	46.413
Overidentification p-value					0.010
Panel C. City suburbs					
Δ distance to highway ramp	-0.028***	0.024***	0.024	0.014	0.015
	(-4.25)	(2.80)	(0.92)	(1.03)	(1.08)
distance to railway station	0.014***	0.003	0.003	0.001	0.002
	(4.93)	(0.41)	(0.37)	(0.18)	(0.21)
Lagged urban density 2000	No	Yes	Yes	Yes	Yes
Distance variables	No	Yes	Yes	Yes	Yes
Geographical variables	No	Yes	Yes	Yes	Yes
Instruments:					
Distance to old batavia road	No	No	Yes	No	Yes
Distance to old batavia tram-station	No	No	No	Yes	Yes
N	334	334	334	334	334
Adjusted R-sq	0.087	0.561			
Cragg-Donald F-statistics			32.300	156.711	85.009
Overidentification p-value					0.620