LUND UNIVERSITY
School of Economics and Management

Master in Economic Development

# More urban, more unequal?

## Urban agglomeration and income inequality (1955-2015)

by

Julia Listrup

ju4488li-s@student.lu.se

*Abstract:* There is an ongoing debate if urbanization and larger city sizes should be promoted for increasing economic growth at the expense of rising income inequality. The primary purpose of this study is to determine if the relationship between income inequality and average urban agglomeration size is the same for developed and developing countries. Data for this study was combined from different sources including but not limited to the Standardised World Income Inequality database, and the World Urbanization Prospect. The study then ran panel data regression with fixed effects for 96 developing countries and 35 developed countries from 1955 to 2015 to analyze the relationship between income inequality and average urban agglomeration size. The results show that there is a strong association between average urban agglomeration size and income inequality in developed countries, however, this relationship does not hold for developing countries. This indicates that the development trajectory is different for the two sets of countries and that for developing countries today rising average urban agglomeration size does not go hand-in-hand with rising income inequality. The results, give an important insight into how the relationship between income inequality and average urban agglomeration size could affect policy regarding both urbanization and income inequality in different countries in the future.

*Keywords*: Income inequality; urbanization; economic growth; city size

**EKHS21**
Master's Thesis, first year (15 credits ECTS)
June 2019
Supervisor: Anna Missiaia
Examiner: Kerstin Enflo
Word Count: 13434

# Acknowledgements

I would like to express my gratitude to my supervisor for her guidance during this thesis as well as to Lund University for making the research possible. Special thanks to Elsa Kyander, Hannah Leadbetter and Johanna Nilsson for useful suggestions and motivations; Jordan Furlong for correcting my grammar and providing comments. Any remaining errors are entirely mine.

# Table of Contents

# List of Tables

# List of Figures

# 1 Introduction

In 1950, 30 percent of the world's population lived in urban areas, compared to 55 percent of the world's population in 2018 (UN, 2018). Furthermore, in the last 25 years, with-in country income inequality has increased; income shares of the top decile have increased significantly and those of the poorest deciles declined. Chambers & Dhongde (2016) argue that cross-country income distribution became more equally unequal in the past 25 years. An ongoing policy discussion is whether urbanization should be promoted to increase economic growth in countries at the expense of rising national income inequality. It is not concluded on in which way urbanization brings positive and negative externalities to countries development[1]. Studies in three different areas can be related to the phenomena of urbanization, income inequality and economic growth.

The three areas are determinants of income inequality, city size effect on income inequality, and city size and economic growth effect on income inequality[2]. Research on the determinants of income inequality, such as Milancovic (1994) argue that societies choose less income inequality and that it is therefore not only economic factors that make income inequality decrease in richer societies. On the other hand, Gustafsson and Johnsson (1999) highlight that income inequality is lower in developed countries where there is a large public sector and a larger part of the population belongs to a trade union. Furthermore, several authors have tried to confirm the Kuznets curve hypothesis including; Barro (2000), Frazer (2006), Vanhoudt (2000), and Piketty and Saez (2014). Barro (2000) and Vanhoudt (2000) argue that there is evidence for a Kuznets curve between income inequality and income per capita. Contrary to this, Frazer (2006), and Piketty and Saez (2014) found little evidence of a structural decline in income inequality as income per capita rose. Therefore, previous studies does not agree on the actual relationship between income inequality and economic development. Urbanization is

---

[1] The term urbanization will be used solely when referring to the movement of more and more people from the countryside to cities.

[2] The term city size is used here to refer to the number of people living in the same geographic location.

said to be a by-product of economic development; however, several developing countries have experienced increasing urbanization without an increase in income per capita in the last decade (Jedwab & Vollrath, 2019). Therefore, it is important to look at studies that have researched the city size effect on income inequality. Behrens and Robert-Nicoud (2014) argue that increasing returns to skills in cities leads to higher income inequality in cities. Also, Baum-Snow and Pavan (2013) state that changes in the structure of labor demand lead to differences between wages in smaller and larger populated locations in the United States between 1979 and 2007. Finally, when looking at studies that have combined city size and economic growth effect on income inequality, Sulemana et al. (2019) found a positive association between urbanization and income inequality in Sub-Saharan Africa which contributes majorly to the field of research. However, another recent study by Castelles-Quintana (2017) has looked at the relationship between income inequality, city size and economic growth in 131 countries over a 50-year time period. The study argues that there is a u-shaped relationship between income inequality and city size, where income inequality first falls with city size and then increases again after a certain point, but that there is also an n-shaped relationship between income inequality and economic growth. Which is contrary to the findings by Sulemana et al. (2019), where a positive association between urbanization and income inequality was found for Sub-Saharan Africa. Furthermore, the study by Castelles-Quintana (2017) did not address the issue of the possibility that the developed countries could be driving the results upwards due to the large differences in development levels between countries, as both high, middle-and-low income countries are bundled together. The different development levels between developed and developing countries can be highlighted through the difference in average income per capita during the time period of 1950 to 2015 at a stunning 50.3 percent. The migration towards cities that developed countries experienced during their industrial revolution and the increase in average urban agglomeration size in the developing world more recently, might have impacted income inequality in different ways. The world presently has 33 megacities, where 27 of them are located in developing countries, however, the increase in people residing in cities is increasing in all regions of the world (UN,2018)[3].

This study looks at the relationship between income inequality and average urban agglomeration size between developed and developing countries and if it is different. Not only will it cast light on the relationship between income inequality and average urban agglomeration size, but also try to understand if there are different mechanisms driving income inequality in developed countries versus developing countries between 1950 and 2015. The research question asked is, therefore; how does the increase in average urban agglomeration size affect income inequality, and whether there is a difference between developed and developing countries in the mechanism that drives changes income inequality. The hypothesis tested is that there is a different relationship between income inequality and average urban agglomeration size for developed and developing countries due to the different levels of economic development at different points in time.

The two different groups of countries, developed and developing have been classified by the World Economic Situation and Prospects 2018 (Appendix A), which has been developed by the United Nations (UN). To test the hypothesis, the study brings together the most recent data from the Standardized World Income Inequality database (SWIID), the World Urbanization Prospects, the Penn World Table, the Barro and Lee dataset and the World Development Indicators. This quantitative study will estimate the main dependent variable income inequality using the Gini coefficient, looking at independent variables such as average urban agglomeration size, income per capita and average economic growth rates in 5-year intervals from 1950 to 2015. To this end, we will use panel data regression analyzes with fixed effects and different sets of control variables. Due to practical constraints, this study cannot provide a comprehensive review of what is classified as urban, urbanization, urban concentration, city, urban agglomeration and so on. When speaking of average urban agglomeration size or city size, the study refers to areas with more than 300,000 inhabitants, a classification taken from the UN Population Division.

---

[3] Average urban agglomeration size is defined as the average size of cities in a country including cities with at least 300,000 inhabitants. Whereas a megacity is defined as a city with more than 10 million inhabitants.

The contribution of this thesis is to explore, for the first time, the effect of average urban agglomeration size and economic growth on income inequality focusing on the distinction between developed and developing countries. This will be of interest for scholars not only researching in development economics but also looking at growth economics and economic history. Furthermore, it is of interest for policymakers both on a national level, but also for organizations on a supranational level such as the European Union. Finally, it is an important contribution to understanding how to reach the Sustainable Development Goals set by the United Nations (UN) which focuses on reduced inequalities, but also on more sustainable cities. As there is a possibility that city sizes effect income inequality, this could have an impact on the strategies to reach Agenda 2030 for Sustainable Development.

This thesis is organized as follows, in chapter 2, there will be an overview of the relevant theories and previous studies in the area. Chapter 3 discusses the data used, followed by chapter 4 on the methodology. Chapter 5 presents and analyses the results and finally, the study will make some concluding remarks in chapter 6, discussing policy implications and further research needed in the field.

# 2 Literature Review

In this chapter, we first present and analyze the different theories on the relationship between income inequality, economic growth, and urbanization. Furthermore, the theories concerning these three main variables affect the decision of control variables in the models. Secondly, the study will discuss previous empirical research in the area with a focus on three perspectives, income inequality, city size effect on inequality and lastly bringing together the two perspectives to look at research done on city size and economic growth effect on income inequality. Finally, the thesis will present reasoning on the selection of comparing developed and developing countries.

## 2.1 Theoretical background

The three main areas that constitute the theoretical background of the thesis will be theories concerning income inequality, economic growth, and urbanization. Firstly, for income inequality, the Lewis dual-sector model, and the Kuznets curve hypothesis will be discussed. Secondly, for economic growth, theories will include the Solow growth model and the Convergence debate. Finally, for theory on urbanization, economics of agglomeration and the cost of urbanization will be presented to understand the underlying mechanism behind the positive and negative externalities which increasing city sizes brings.

### 2.1.1 Income inequality

Lewis (1954) explained the development of an undeveloped country by a shift from two sectors which existed in an economy. Where income inequality is not an effect of economic growth, but a cause of growth. The dual-sector model is explained by a modern sector and the “other” sector, possibly being the agricultural sector. The agricultural sector has according to Lewis (1954) unlimited amount of labor, and with very low or even zero marginal productivity in the agricultural sector. It is then possible to shift labor from the agricultural sector to the modern sector without any implications for the productivity level in the agricultural sector. Therefore, the supply of labor is elastic, which means that the growth in the modern sector leads to a rising share of profits. Only a small share of this rise in average income goes to the actual labor, where the majority goes to the driving parties in the modern sector, which leads to rising income inequality. Lewis (1954) explains that the distribution of income is skewed towards the high-income earners, as they are able to save the profit earned in the modern sector and also able to re-invest it, which creates further economic growth. According to Lewis (1954), trying to redistribute income among all parties leads to an increased risk for an economic slowdown in an economy.

The theory of Lewis (1954) sparked the idea behind the Kuznets curve hypothesis. Kuznets (1955) meant that income inequality increased with an increasing income per capita. This was due to a shift from the agricultural sector to the modern sector, leading to higher incomes in the first phase of the shift for some. Then as the majority of the population had moved out of the agricultural sector, income inequality and income distribution would equalize (Kuznets, 1955). According to Kuznets (1955), income inequality would first rise with increasing levels of income per capita and then fall when stabilized. This hypothesis was the conventional wisdom in development studies until the late 1980s. Kuznets's (1955) did observe a downward trend in income inequality in the United Kingdom, United States, and Germany after 1918, however, the theory is unable to explain the rise in income inequality post-1980s in the world. Thomas Piketty explains the rise in income inequality post-1980s by that decrease in income inequality after 1918 in for example the United Kingdom was due to special and unusual events such as the political forces and wars at the time. Furthermore, Piketty explains the rising income inequality more recently by the capitalist constellation of the market the world is currently influenced by in for example decision making (Milanovic, 2016, pp. 45-46). Furthermore, today with new and better data collection methods and estimates on income inequality, scholars find little evidence among a large number of countries that income inequality should first rise with rising income per capita, and then decline as development progresses (Bruno et al. 1996).

### 2.1.2 Economic growth

The Solow growth model is one of the basic models explaining the phenomenon of economic growth. The basic economic model is important in the relationship between income inequality and average urban agglomeration size, as previous assumptions among scholars have been that economic growth and increasing urbanization goes hand-in-hand. Capital-output and labor-output ratios vary in the Solow growth model and depend on the factor endowments in the economy and the development of production. The term factor endowment is defined as the amount of labor, land or capital that a country possesses. Due to diminishing returns, a country who has recently begun to economically grow grows quicker than an already developed country. This difference is due to that increasing capital at the beginning of the investment curve raises productivity more than a unit of capital further along the investment curve. A country would according to the Solow growth model grow until it reaches the steady-state level, where the level of investment is the same as the level of depreciation, meaning that there will be a zero-growth rate. In figure 1, where new capital (needed to compensate for the growth in the workforce and depreciation) intercept with new savings. At this point, no new capital would be created in the country. In the same way, human capital can be incorporated into this model, and have the same steady-state level, as there is also diminishing returns in human capital. One way to have growth once a country has reached the steady-state level is to increase the savings ratio in a country, which leads to an outward shift in the capital investment curve. Furthermore, innovation and ideas can shift the curve outwards constantly, leading to a steady growth rate (Perkins et al. 2013, pp. 103-121).

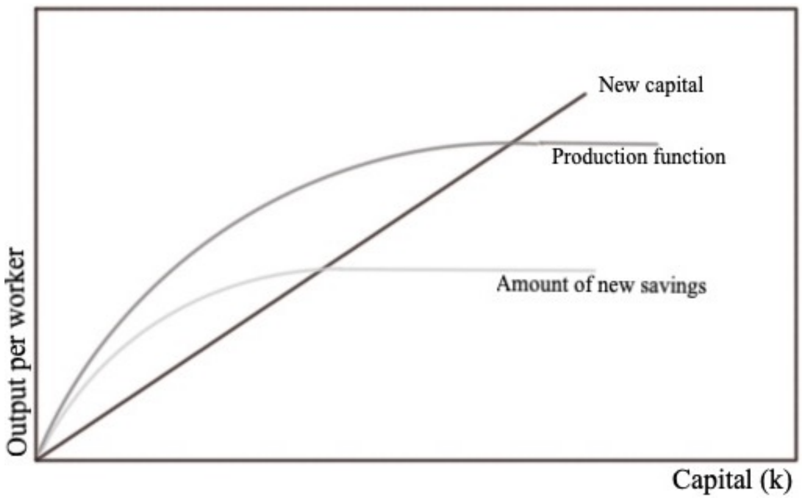


*Figure 1 The Solow growth model, source: author's construction*

According to the Solow growth model, all countries would eventually converge to the same steady-state level, however, this is not possible today due to innovation and ideas differences discussed previously, but furthermore relates to the discussion about the institutional impact on economic growth. The discussion of the impact of institutions on economic growth is outside the scope of this study, and will therefore not be further discussed.

There is relevance to the convergence debate for this study of average urban agglomeration size effect on income inequality, as it explains reasons for why not all countries are reaching the same steady-state level and therefore differences in factors of economic growth. Which matters to the debate about if urbanization always happens with increasing income per capita. This also matters on a country level, rather than a population level as the study investigate the difference between countries. When studying all countries, it is impossible to see a convergence between them, however, there is conditional convergence, meaning countries are grouped depending on certain factors. Sachs and Warner (1995) found that when grouping based on factors such as “open to world trade” they could see convergence among the countries that had an open access approach. However, the conditions in rich and poor countries differ widely, and it is therefore hard with the Solow growth model to predict if poorer countries are catching up to richer. Further studies by for example Caselli et al. (1996) showed that the convergence rate which conventional wisdom has at 2 percent, does not hold, and some of the results showed up to a 10 percent convergence rate towards the steady-state point. A further discussion about the difference between developed and developing countries will take place at the end of the section discussing previous empirical studies done in the research area.

### 2.1.3 Urbanization

The definition of what is urban or what is meant with urbanization varies across scholars. Urban studies analyze cities based on their size and status, but it can range from megacities to small-and medium-sized cities. The broader question which the term urbanization tries to answer is if towns and cities are good for human life (Harding & Blokland, 2014, pp. 1-55). Therefore, this study will discuss the theory and concepts surrounding urban economics as it is the main focus of the study, especially the trade-off between the economics of agglomeration and the cost of urbanization.

The main idea behind economics of agglomeration is that cost savings arise from firms being closer to each other. There is a certain economic mechanism that yields agglomeration, such as the reduced cost generated by the transfer of people, goods, and information. Urban economics shows, contrary to conventional wisdom, that lower transport costs make businesses more sensitive to variation between location, and can, therefore, have a big impact on the distribution of economic activity (Fujita, 2013, pp. 1-25). As businesses cluster, and competition increase, there is an increase in productivity, which leads to economic growth and development. Cities are an engine of growth as it encourages capital accumulation, however, the degree of urbanization can impact growth both positively and negatively (Bertinelli & Black, 2004). Agglomeration can also lead to economic-linkages to rural areas, creating economic growth through consumption linkages (Cali & Menon, 2013). Furthermore, Williamson (1965) indicated that agglomeration matters at early stages of development, and efficiency can be significantly increased by concentrating businesses in the same location, however as expansion happens the development of a city might favor negative externalities.

As an urban population increases due to an incentive for a better-quality life in cities, it also brings negative externalities. These include unemployment, poverty, urban violence, congestion and environmental degradation (Bloom et al. 2008). The increasing productivity and differential in human capital in cities do also affect income inequality, as lower incomes grow proportionally but higher incomes grow quicker (Sarkar, et al. 2018).

## 2.2 Previous Empirical Studies

The strands of studies relating to the question of how economic growth and average urban agglomeration size can affect income inequality can be divided into three sections. Firstly, what determines income inequality, secondly how city size affects income inequality and finally bringing together the two perspectives to look at studies done on how city size and economic growth effect on income inequality.

### 2.2.1 Determinants of income inequality

Recent studies have focused on income inequality as a cause of different variables, but also how regional and country differences impact the determinants of income inequality. For example, Marrero and Rodriguez (2013) argue that the impact of overall inequality on growth is positive, but it is not robust nor significant, that increasing income inequality would be good for economic growth. However, when analyzing inequality of opportunity (the potential to succeed in life determined by birth) it is negative and significant to economic growth. It highlights that the effect of income inequality on economic growth is sensitive to a variety of variables such as regional dummies. Furthermore, Li et al. (1998) agree that the variables that do affect income inequality changes slowly within countries and differs largely between countries. The study also found that income inequality cannot explain the full differences in economic growth rates among countries, and concludes that reducing income inequality does not automatically mean increasing economic growth. Similarly, Ehrhat (2009) found after surveying empirical studies on the relationship between income inequality and economic growth, that initial inequality of assets, had a significant and negative effect on following economic growth. Furthermore, found that asset inequality is a more robust determinant of economic growth than income inequality, and if the initial distribution of income and wealth is less unequal, it is associated with faster long-term economic growth. However, Ehrhat

(2009) offers no explanation on how asset inequality and income inequality could be endogenous of each other. As more assets could mean more opportunities and therefore higher disposable income and the other way around.

On the contrary, this thesis is interested in income inequality as a consequence of urbanization, not a cause and according to Milanovic (1994), two different types of factors can explain income inequality. Firstly, factors that are independent of economic policies in the short term. Secondly, social-choice factors which can be seen in the size of social transfers and state sector employment in a country. Furthermore, Milanovic (1994) mean that it is not only economic factors that make income inequality decrease in richer societies, but that the societies choose less income inequality. However, Milanovic (1994) overlooks how societies choose less income inequality. Gustafsson and Johansson (1999) found in their study looking at 16 industrialized countries between 1966 and 1994 that lower income inequality was found in countries where a large proportion of the population belonged to a trade union and where there was a large public sector. Where Gustafsson and Johansson (1999) made a major contribution to factors leading to differences among developed countries in income inequality.

Finally, several studies have tried to confirm the Kuznets curve hypothesis. Firstly, Barro (2000) argues that income inequality hinders economic growth in poorer countries, but encourage economic growth in richer countries. Stating that there is a different relationship between economic growth and income inequality for developed and developing countries. Additionally, concluded that economic growth tends to rise with income inequality when income per capita is above USD$2000 and that therefore the Kuznets curve is a clear empirical regularity. Secondly, Frazer (2006) used a parametric regression for a cross-country comparison of how income inequalities have developed within countries at different levels of development. The study stated little evidence for the Kuznets curve hypothesis and found examples of low-income countries with decreasing income inequality such as India, and countries with high economic growth rates despite small changes in income inequality such as the Republic of Korea. Thirdly, Vanhoudt (2000) argue that a weak Kuznets curve hypothesis hold. This is due to that higher investment shares and population growth are associated with lower income inequality in industrialized countries, and the opposite occurs in developing countries. Fourthly, Piketty and Saez (2014) found that there has been an inequality reversal between the United States and Europe if looking at before and after WWI. Where the United States now has higher income inequality, but Europe has higher wealth inequality. However, the study then sees either a rise or a flat development pattern of income inequality, not a structural decline in income inequality which Kuznets argue happens with development.

### 2.2.2 City size effect on income inequality

On the question of how city size effect income inequality Alperovich (1995) suggest the factors which affect income inequality among cities are connected through a variety of different mechanism to city size. Moreover, regardless of the different underlying factors, Alperovich (1995) states that city size still turns out highly significant in affecting income inequality.

Baum-Snow and Pavan (2013) argue that for understanding the role of city size in increasing income inequality, changes in the structure of labor demand is important. The study found that 23 percent of the variance in hourly wages can be explained by that income inequality is growing quicker in larger populated locations than smaller in the United States, between 1979 and 2007. However, the study overlooks if there is a difference in the relationship between richer and poorer cities in the United States. As explained earlier the structure of labor demand is important for income inequality. In a study done by Behrens and Robert-Nicoud (2014) tougher selection of laborers was found to increase the returns to skills (returns from for example a college degree) and income inequality in cities. Therefore, larger city sizes which have higher productivity through selection laborers also have higher income inequality. However, higher productivity also creates incentives for movement from rural to urban areas, leading to a vicious cycle in the selection of laborers. Behrens and Robert-Nicoud (2014) have made a major contribution to the understanding of the possible endogeneity issue between income inequality and city size, and the possibility of a reverse causality issue.

In terms of the initial distribution of skills among laborers, a negative association between local income inequality and growth of city-level population is found in a study by Glaeser et al. (2015) when controlling for the initial distribution of skills. However, returns to skills are found more important in explaining the variation of income inequality across American metropolitan areas. This finding can be supported by Haworth et al. (1978) results which show that increases in city size and urban growth will lead to greater income inequality. Through the mechanism that increases economic growth and city size raises monopoly rents earned by those isolated from the competition. The study looked only at the case of the United States in 1970 and concluded that income inequality varied directly with the level of population. Therefore, the study overlooks changes in income inequality over time, due to the changes in city sizes. Nord (1980) argue that income inequality forms a u-shape over increasing city sizes, with smaller and larger cities experiencing the greatest inequality. The smaller cities lack sufficient economies of scale and the larger cities have a growing service sector which employs workers of lower skill sets. This led Nord (1980) to conclude that the greatest income equality is sustained in cities in the United States which range from 10,000 to 50,000 people, however also only looking at data during the1970s.

With respect to research focusing specific countries, Sarkar et al. (2018) found that income distribution in large cities in Australia, where the larger the city, the larger the growth of income at the top income deciles were. Furthermore, highlights that there are fewer people in the higher income categories, compared to the other categories, and conclude that this could lead to a push of people in lower income categories out of the cities. However, a move from concentrated urban areas to suburbanization (movements into suburbs) has been found by Yorukoglu (2002) in the United States to decrease inequality in productivity and therefore partially accountable to contribute to a decline in income inequality in cities. Furthermore, Yorukoglo (2002) argues that this is due to improve transportation technologies today, making the cities wider and less densely populated which makes this transition possible. Finally, Chen et al. (2016) found that urbanization has an immediate mitigating effect on income inequality in China analyzing data between 1978 and 2014, but it also has a lagged exaggerate on the effect of income inequality. Furthermore, this relationship can be linked to the income gap between the rural and urban population in China, and Chen et al. (2016) argue that further urbanization might be able to narrow the current income gap.

### 2.2.3 City size and economic growth effect on income inequality

Few studies have looked at the relationship between income inequality, average urban agglomeration size, and economic growth, on either city level, regional level or by sampling countries. Worth noting is that urbanization and economic growth does in recent studies, no longer goes hand in hand with each other, and poor countries can experience rapid urbanization without changes in income. Urbanization and the level of income are correlated, but urbanization can continue even at times of negative growth (Fay & Opal, 2000). In this section, the main scholars who have studied the mechanism of the relationship between income inequality, urbanization, and economic growth, or tried to relate the three phenomena together are introduced.

Firstly, Castells (2011) argues that developed urbanized countries show lower levels of income inequality, where both patterns of urbanization and level of income inequality seem to influence succeeding economic growth, however, also finally succeeding urbanization and income inequality to a certain extent. Furthermore, they are all driving mechanisms in economic development. However, the study overlooks if the actual size of the city matters to the degree of income inequality. Further Castells-Quintana and Royuela (2015) conclude that concentration of resources both spatial and individual, which occurs from increasing city sizes can be linked to the early stages of development in a country. Where three non-linear patterns arose in the study of economic development; firstly, the Kuznets curve, secondly, an inverted u-shape between spatial concentration and development and finally depending on the level of development a non-linear relationship between income inequality and urbanization. However, overlooks any causal relationships between the variables, and the research only discusses possible linkages.

Secondly, Henderson (2003) looked at the optimal level of urbanization, and whether urbanization promotes economic growth. The study argues that there is little support for that urbanization should drive economic growth and that it is rather a by-product of economic development. Furthermore, concludes that there is an optimal level of urbanization, which balances knowledge accumulation against the negative externalities related to crowded cities. The main limitation of the study is that it does not link the two phenomena of urbanization and economic growth to income inequality, even though the concept of structural transformation is discussed. However, Royuela et al. (2014) focused on urbanization as a transmission channel of income inequality on economic growth in OECD (Organisation for Economic Co-operation and Development) countries. Furthermore, Royuela et al. (2014) argue a negative relationship between urbanization and income inequality, leading to the conclusion more equal societies are more resilient to economic shocks. The main contribution of the study being that income inequality is lower in smaller city sizes within the OECD group.

Thirdly, when Castelles-Quintana and Royuela (2012) differentiated depending on the level of urbanization, the research found that high levels of urbanization and increasing levels of income inequality harms economic growth in some countries but also that income inequality harms economic growth in countries where there is persistent unemployment, but low levels of urbanization. The study argues that increasing income inequality harms economic growth especially in countries with high levels of urbanization. However, offers no explanation of different patterns between countries. Furthermore, Fallah and Partdige (2007) looked at US data over the 1990s on the transmission channels of which income inequality affects economic growth, where one of the identified channels was urbanization. The study overlooks how income inequality over time is affected by changes in urbanization and economic growth. However, the study's main argument is that there is a difference between metropolitan and non-metropolitan areas on the relationship between initial income inequality and economic growth.

Today, little is known about the negative externalities of living in large cities and how much of development policy should be focused on cities (Henderson, 2010). The one study that has looked at the relationship between size and distribution of cities and income inequality is Castells-Quintana (2017), which looked at 131 countries over a time period of approximately 50 years. The study states support for the Kuznets curve being n-shaped rather than inverted u-shaped and that there is also a u-shaped relationship between average urban agglomeration size and income inequality, which has previously been overlooked by empirical literature. Furthermore, a recent study by Sulemana et al. (2019) argues that there is a positive association between urbanization and income inequality in Sub-Sharan Africa when studying the time period between 1996 and 2016. This is contrary to Castells-Quintana's (2017) findings of a non-linear relationship. Furthermore, all of the studies discussed in this section overlooks the fact that there could be a different relationship between income inequality and average urban agglomeration size for developed versus developing countries.

There are several reasons to assume the relationship between income inequality and average urban agglomeration size to be different for developed and developing countries. Firstly, in the past, the largest cities in the world have been located in high-income countries. Today, the largest cities in the world are located in developing countries. Indicating that urbanization and growth of cities, do not always go hand-in-hand with economic growth (Jedwab & Vollrath, 2019). This has also been concluded earlier by Opal and Fay (2000) who studied determinants of urbanization over 40 years. Urbanization was found to continue to grow, even at times of negative economic growth. Furthermore, Fay and Opal (2000) highlight that the relationship between urbanization and the economic reason is weaker in countries with less civil and political choice.

Secondly, the negative externalities of urbanization are much higher in developing countries compared to developed due to the inefficiency in the process, and the lack of sufficient investment in line with increasing city size (Bertnelli & Black, 2004). As urbanization has been proven to happen without economic growth in developing countries, it is uncertain if the relationship between income inequality, average urban agglomeration size, and economic growth holds for both developed and developing countries, as the processes of increasing city sizes seem to be due to different mechanisms (Adams & Klobodu, 2018).

Thirdly, Piketty and Saez (2014) argue that the reason for that income inequality is rising today is due to the race between education and technology. Where the supply of skills in the form of education, has to match the demand for skills, which arise from technology development. As a result of globalization, there has been a rise in a global competition for skills, which could indicate the rise in income inequality and an unbalanced relationship between education and technology. However, Piketty and Saez (2014) conclude that this race is not sufficient to explain variations between countries. Therefore, the contribution of this thesis is to explore the effect of urbanization on income inequality, focusing on the difference between developed and developing countries. For policymakers and decision makers to be able to understand the dynamics of urbanizations. This for influencing coming strategies to reach Agenda 2030 on Sustainable Development in a timely manner, and with the right approaches. The following chapter will be about the methodology which will be used in the study.

# 3 Methodology

This section is divided into two parts. Firstly, the empirical model is presented and reasoning surrounding using average urban agglomeration size as a measurement for urbanization is presented. In the second part, the method and the challenges with the procedure are discussed.

## 3.1 Empirical model

Previous research by Castelles- Quintana (2017) has shown a relationship between income inequality and average urban agglomeration size that is independent of the effect of income changes on income inequality. However, the study overlooks if there is a difference in the relationship between developed and developing countries. The hypothesis of this thesis as outlined in the introduction, that developed and developing countries do have different relationships between income inequality and average urban agglomeration size. This difference is assumed because of the different levels of development at different time periods. Furthermore, the study uses panel data from 1950 to 2015 to run a cross-country statistical analysis of the expected relationship between the two samples of countries.

This thesis aims at testing a theoretical model to explain the relationship of income inequality and economic growth through the transmission channel of average urban agglomeration size. The estimation will rely on *Zipf's law* to motivate the decision to pick average urban agglomeration size as the measurement of city size or urban concentration, and therefore urbanization. *Zipf's law* indicates that the size of cities is inversely proportional to their rank developed by George Kingsley Zipf in the early 1940s (Sanford, 1942). Where $Log(P) = Log(K) - q. Log(R)$ explains the formal relationship, where q is referred to as *Zipf's* exponent, and where the exponent $q = 1$ the *Zipf's* law is followed by the city size distribution. In other words, this law entails that in a structure of cities the largest city is approximately twice the size of the second largest city and so on (Arshad et al. 2017). Furthermore, the thesis ability to use average urban agglomeration size indicates an ability to draw a conclusion on the distribution of cities across countries, rather than if a ranking method would have been used the average urban agglomeration size gives a better ability for a cross-country comparison of the scale of cities within countries.

The empirical analysis covers 65 years (1950-2015) in 5-year intervals for a total of 131 countries, which is divided into two different groups. One includes the 35 developed countries and another one which has 96 developing countries or economies in transition (Appendix A). The classification which has been used is the UN country classification from the World Economic Situation and Prospects 2018. Where the classification is based on basic economy country conditions. Where high-income countries are classified as developed and middle-and-low income countries are classified as developing countries or economies in transition. From previous research, the empirical model is developed to fit both groups of countries. The study argues that income per capita and average urban agglomeration size is the two main independent variables that will affect income inequality in the long-run. Which leads to the specification to be the following (adapted from Castelles- Quintana, 2017):

$$inequality\iota\tau = \alpha 1 income\iota\tau - 1 + \alpha 2 income^2\iota\tau - 1 + \beta AveAggSize\iota\tau - 1 + \psi X\iota\tau - 1 + \varepsilon\iota\tau \quad (1)$$

Where inequality$\iota\tau$ is income inequality in country $\iota$ in time $\tau$, income is income per capita (in logs), X is potential factors influencing income inequality, and ε$\iota\tau$ is a country-time specific shock. To capture the Kuznets inverted u-shape income per capita is both considered linear and squared in the model. AveAggSize (average urban agglomeration size) is the main independent variable for the model and is also considered in country $\iota$ in time period $\tau$, and takes on both a linear and squared form to be able to distinguish a u-shaped relationship between income inequality and average urban agglomeration size which was originally found by Castelles- Quintana (2017). All the variables on the right-hand side will be lagged one time period, to capture the possible lagged effect on income inequality, and therefore the actual time period being1955 to 2015. It is assumed that past estimates of right-hand side variables would predict future income inequality.

This thesis will use the Gini coefficient as the measurement for income inequality, as it is a widespread measurement of income inequality among research today and has the ability to be standardized in a cross-country study. The Gini coefficient is derived from the Lorenz curve, which is a graphical representation of a country's size distribution. Size distribution describes the total share of consumption or income acquired by different groups of households, arranged according to their consumption or income status (Perkins et al. 2013, p. 169). Furthermore, the Lorenz curve arranges income earners from the lowest to the highest on the horizontal axis in cumulative percent and the share of cumulative consumption or income on the vertical axis. Perfect equality would be all income earners lying along the 45-degree line.

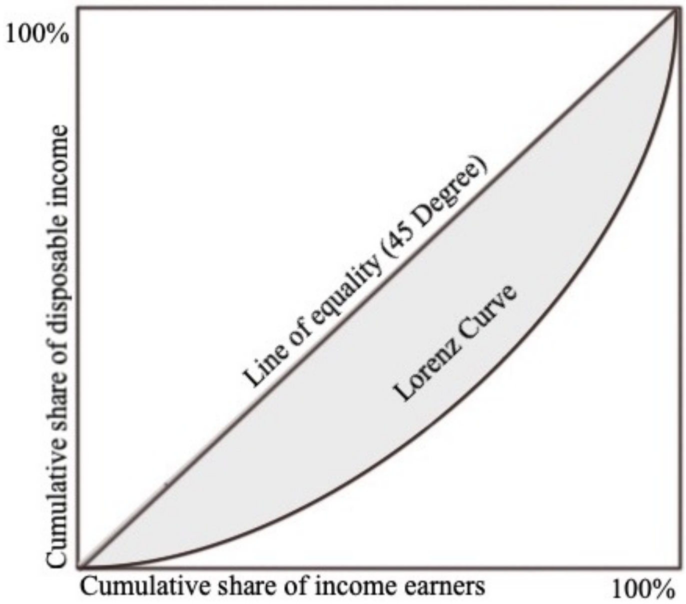


*Figure 2 Lorenz curve, source: author's construction*

From the Lorenz curve, it is possible to derive a ratio of the top 20 percent of households to the share received by the bottom 20 or 40 percent. This is known as the Gini coefficient. The larger the area is between the 45-degree line and the Lorenz curve, the higher the value of the Gini coefficient will be (Perkins et al. 2013, pp. 171 – 172).

Furthermore, the study will use the ordinary least square (OLS) method as a starting point for the statistical analysis of the panel data. OLS estimates the unknown parameters in a linear regression model (Kennedy, 2008, pp. 40-41). However, to extend the possibility of the regression to be non-linear, the study has added the square transformation of average urban agglomeration size and income per capita. The OLS method does have the analytical ability to draw a conclusion from the model explained above but is limited due to at least four reasons:

Firstly, to be able to have the average effect of the variables included in the model, it is important to control for time-specific shocks and country-specific conditions through the use of fixed effects. Otherwise, certain events in history or certain countries could bias the result. This will be done through the use of fixed effects, both for the time variable of years and the country variable. Fixed effects in the model are introduced as dummy variables. However, an issue with using fixed effects when estimating results could be that fixed effects create increasing noise in the statistical data. Meaning that the irregularities in real life data which exist in the datasets, as they are on a country level, increases (Kennedy, 2008, pp.284-285). In the end, it could mean that the increase of noise can lead to estimates not being statistically significant even though with less noise the variables could have been.

Secondly, the study will cluster the standard error by country as it assumed that residuals of income inequality could be correlated with country-specific conditions. There is a risk that there could be unexplained variation in the Gini coefficient that is correlated across time which known as heteroskedasticity, even if having country-and -time fixed effects already in place. The assumption is that the observation in the datasets are independent and identically distributed meaning that observation in the same country should be closer together than those in other countries, and to correct for this event the study has used clustering of the standard errors at country level. This ensures that there are homoscedasticity and the sequence for each clustered country has the same variance across the estimations (Kennedy, 2008, pp. 115).

Thirdly, the OLS estimate does not control for the sample being of a small set in both datasets. This can lead to estimates being weak in power and increases the margin of error. This can also indicate at type II error where the results may be able to confirm the hypothesis, but where an alternative hypothesis is the true hypothesis (Kennedy, 2008, p.67). This would skew the result in the direction of the original hypothesis, leading to weak results. To try to correct for this the study will use a small sample correction technique in the statistical software Stata when doing an instrumental variable estimation. Stata will be used as the main software to estimate the statistical results. The sample in this study is small due to the lack of resources to collect sufficient and complete data for all countries, in all time periods and for all variables.

Finally, OLS is unable to handle the possible endogeneity issue of that larger cities attract people with different sets of skills and therefore raises income inequality and therefore could have reverse causality (Castelles- Quintana, 2017). This is problematic to be able to establish a possible causal or a strong association between income inequality and average urban agglomeration size. The study will use first differences with an instrumental variable estimation to show on the one-way association of that average urban agglomeration size lagged both in second and third level as instruments for that average urban agglomeration size in the past affect income inequality today, but that income inequality today cannot affect the average urban agglomeration size of the past. The following chapter will be about the data used in the study.

# 4 Data

This chapter consists of four parts. The first part introduces the sources used to construct the two comparative datasets of developed and developing countries. The second part discusses the different variables used in the estimations, and how they have been constructed. The third part analyzes limitations to the data which should be considered when interpreting the results. Finally, the fourth section gives a descriptive overview of the data in the two samples.

## 4.1 Datasets

The thesis brings together five different datasets to be able to collect and construct all relevant variables, needed to understand to what extent average urban agglomeration size affect income inequality in developed compared to developing countries. Furthermore, this creates two new datasets with a selected number of variables discussed in the variables section below, followed by an overview of the data of the two samples. However, below is an introduction to the five different datasets which the data for the thesis has been collected from.

Firstly, *The Standardized World Income Inequality Database (SWIID)* developed by Solt (2019) incorporates income inequality data from several datasets to give the widest possible coverage over time and across countries. This includes but is not limited to OECD Income Distribution Database, Eurostat and the World Bank. The dataset contains comparable Gini coefficients for 196 countries from 1960 to the present. The income inequality data in the SWIID is estimated using multi-imputation, to reach as many estimates as possible for income inequality, therefore having the greater range compared to other datasets freely available.

Secondly, *The World Urbanization Prospects* (2018) is issued by the Population Division of the Department of Economic and Social Affairs of the United Nations. The prospects include several different sets of data with estimates and projections of the urban and rural populations of countries in the world and of their major urban agglomerations. The different estimates of urban and rural populations are derived from available census and official population estimates in each individual country. The estimates use the term urban agglomeration not referring to administrative boundaries but to the inhabited population at urban density levels.

Thirdly, *Penn World Table* version 9.1 is a dataset that covers 182 countries from 1950 to 2017. It compiles information on relative levels of income, output, input, and productivity. The dataset converts GDP at national currency to a common currency of US dollars through using the International Comparison Programs (ICP) collection of prices across countries in benchmark years. This to be able to construct purchasing-power-parity (PPP) in US dollars for cross-country comparison (Feenstra et al. 2015).

Fourthly, *the Barro and Lee* dataset covers 146 countries over the time period of 1950 to 2010. It looks in 5-year intervals on education attainment data in all available countries of the world. However, it does also present knowledge about the distribution of educational attainment across the adult population in a specific country. The estimates are constructed using the most recent census and survey observation in each country presented in the dataset (Barro & Lee, 2013).

Finally, *The World Development Indicators* covers 217 countries and 1,600 different time series, and include a lot of the indicators going back at least 50 years in time. It comprises of different types of measurements to give comparable statistics about global development and poverty eradication. The database is constructed from primary data from primarily national statistical agencies, central banks, and customer services agencies, but also other international organizations.

The samples used in this thesis is restricted to countries where the study has been able to match data from the SWIID to the urban agglomeration dataset of more than 300,000 inhabitants from the World Urbanization Prospects. This due to the dependent variable being income inequality and the main independent variables of interest being average urban agglomeration size. Furthermore, this restricts the sample to countries which have at least one urban agglomeration with more than or equal to 300,000 inhabitants between 1950 and 2015.

## 4.2 Variables

In this section three aspect of variables are discussed; the dependent variable, main independent variables and control variables. It will be highlighted how the variables have been derived if they are constructed from combinations of variables and from what datasets they originate.

### 4.2.1 Dependent variable

*Gini coefficient*

The SWIID has combined comparable Gini coefficients over a 59-year time period. As discussed previously the Gini coefficient is derived from the Lorenz curve and ranges from 0-100. With 0 being total income equality and 100 being total income inequality. The decision to use the Gini coefficients from the SWIID dataset was made because the coefficients are standardized across countries. This made it possible to perform cross-country comparisons, then if the study were to use Gini coefficients which had not been standardized across countries. Furthermore, that would also have created less reliable estimates and therefore given biased results. The Gini coefficient selected from the SWIID database is the Gini coefficient that measures disposable income differences. This because it is a comparison of living standards the thesis focuses on and to give an actual reality of how different populations in different countries live.

### 4.2.2 Main independent variables

*Average urban agglomeration size*

The average urban agglomeration size variable is constructed using the World Urbanization Prospect dataset on urban agglomerations above 300,000 inhabitants. From the dataset, an average for each five-year interval has been calculated for each individual country. The dataset list cities which by 2018 from 1950 at some point reached a population of at least 300,000 inhabitants. The average is calculated by adding up the total inhabitants in all cities in a set time period and country and dividing by the number of cities in the same time period. However, for example, if a country did not have a city with 300,000 inhabitants until 1960, this city will still be used to calculate the average in 1950.

*Income*

The income variable is per capita GDP (Gross domestic product) in logarithmic transformation. It is constructed from the Penn World Table version 9.1. The variable is calculated using expenditure-side real GDP at chained PPPs in millions of 2011$USD. This data is collected at a five-yearly interval and divided by the total population at the same time period. After being calculated at per capita GDP level, the sum is logarithmically transformed and merged into the two samples. The expenditure-side real GDP at chained PPP is a better estimate to compare living standards across countries and to get a view of the reality in the different countries, compared to data which combines the expenditure and output side. Also, if looking at market income, the study would be unaware of the actual amount of income which the population received in their pockets due to differences in taxation across countries (Feenestra et al. 2015).

*Growth*

The economic growth variable is the cumulative annual average per capita GDP growth rate. It is also constructed from the Penn World Table version 9.1. The variable is calculated using expenditure-side real GDP at chained PPPs in millions of 2011$USD. The per capita GDP is calculated for each individual year by dividing expenditure side GDP by total population in that year. Using the per capita GDP for each individual year in each individual country, a five-year cumulative average is calculated. Adding five time periods growth rates and then dividing it by the five time periods to get the average growth rate during the five-year interval.

*Investment*

The investment variable is collected from the Penn World Table dataset version 9.1. It measures gross capital formation as a percentage share of GDP for each country annually. From this, every five-year period is collected to the new datasets for the study starting from 1950 to 2015.

*Government expenditure*

The government expenditure variable is also collected from the Penn World Table dataset version 9.1. It measures gross government expenditure as a percentage share of GDP for each country annually. To the new datasets, every five-year period is collected from 1950 to 2015 to construct the new datasets.

*Schooling*

The schooling variable is constructed using data from the Lee and Barro dataset. It looks at the average years of secondary and tertiary schooling by the adult population in each country at five-year intervals. The average years of secondary schooling by the adult population is added with the average years of tertiary schooling by the adult population. The adult population is classified as anyone over the age of 15. As it is from this age you either stay in education or enter the workforce as mandatory schooling regulated by the law in some countries end.

### 4.2.3 Control variables

In this section the control variables used in the statistical analysis are described, they are selected as important variables that are either related with average urban agglomeration size or is believed to be variables which also influences income inequality in the long-run.

*Poplargest is* the total population of people who live in the country's largest city. This is an absolute measurement and is collected from the World Urbanization Prospect 2018, constructed by the UN.

*Urb1m* is the total population living in cities with more than 1 million inhabitants, which has been calculated as a percentage of the total population in a country. This variable ranges from 0 to 100. It has been collected from the World Development Indicators, which has collected data from several databases both nationally and multilaterally to distribute it on the same platform.

*Primacy* is the total population living in the largest city of the country, however as a percentage of the total urban population of the country. This variable ranges from 0 to 100. It has been collected from the World Development Indicators, which has collected data from several databases both national and multinational organizations to distribute it on a joint platform.

*Poptotal* is the total population in the country, seen in thousands. This variable has been collected from the World Development Indicators which is a platform developed by the World Bank, which is a collection of datasets and databases from national as well as multinational organizations.

*Urbrate* is the population living in urban areas, as a percentage of the total population. This variable has been collected from World Urbanization Prospects 2018. The World Urbanization Prospect uses the definition of urban set by each individual country. The national statistical offices who carried out the census are the deciding factor in the definition behind urban. However, if the definition has changed between the census being carried out, the UN has tried to adjust for this in their estimates (UN, 2018).

The *fertility* variable is the fertility rate. *Coal* is coal rents as a percentage of GDP. *Exports* are total exports as a percentage of GDP and *agriculture* is value added in the agricultural sector, as a percentage of GDP. All of these variables have been collected from the World Development Indicators derived by the World Bank.

## 4.3 Critical discussion of the data

As with any empirical study, there are limitations to the data used in this thesis. There are three aspects which need to be discussed here, which may affect the overall significance of results. The availability, the construction of variables and the collection of data.

Firstly, several of the variables in the two constructed datasets have missing values. This means that there is not a full range of observations of all variables in all time periods. This is due to several reasons, for example, independence year of a specific country has taken place after 1950, or that actual collection and calculation of variables did not start until 1960. The main dependent variable, the Gini coefficient, does therefore not consist of a full set of observations, which can be summarised in the descriptive statistical section. The missing observations do lead to that certain years consist of fewer countries, and that during the statistical analysis observations are dropped as more variables are added to the models. However, the data has been collected from sources with the most available data possible, even though its missing observations, it would not be possible to increase the number of observations.

Secondly, the construction of for example the Gini coefficient has been done using multi-imputation by SWIID. Meaning that some of the Gini coefficients for certain countries are constructed rather than observed estimates. The selection of the SWIID has been done to get the most observations of the Gini coefficient over the longest time period possible, however, there should be caution taken in regards to the correctness of the estimates, and should be kept in mind when analyzing the results.

Finally, the collection of data by for example the UN and the population division on the average urban agglomeration size is derived from different census from for example national statistical offices as discussed previously. However, as these offices are not independent of the government in each country, the observations could be biased to show what the country wants the rest of the world see of perceiving the country and this also goes for several of the other estimates.

## 4.4 Descriptive statistics

Below are descriptive statistics related to the study. In table 1 observations and variables for developed countries is presented, followed by the same information but for developing countries in table 2.

*Table 1 Descriptive statistics - Developed countries*

| **Developed** | | | | | | |
|---|---|---|---|---|---|---|
| **Variable** | **Definition** | **Mean** | **St.dev** | **Min** | **Max** | **Freq. of obsv** |
| **Inequality** | Income inequality measued by the gini coefficient | 29.5 | 5.7 | 17.3 | 52.8 | 327 |
| **AveAggSize** | Average agglomeration size, in terms of population (thousands) | 801.8 | 466.9 | 106 | 2776 | 429 |
| **Income** | Per capita GDP (thousands) | 19952.3 | 12578.1 | 568.1 | 64274.4 | 384.0 |
| **Growth** | Cummulative annual average per capita GDP growth rate | 2.4 | 2.5 | -12.4 | 11.3 | 350 |
| **Investment** | Investment share (% of GDP) | 25.9 | 6.9 | 8.6 | 48.9 | 384 |
| **Gov Spend** | Goverment consumption (% of GDP) | 17 | 6.5 | 3.3 | 42.3 | 384 |
| **Schooling** | Average years of seconday and teritiary schooling of adult population | 2.9 | 1.7 | 0.3 | 7.5 | 455 |

Table 1 shows the main dependent and independent variables; however, the control variables are not included. Two observations stand out, firstly, average urban agglomeration size has a very large standard deviation, due to the range between the minimum and maximum value. This creates a mean which could possibly be biased upwards or downwards from the true population. Secondly, in the developed dataset there is good comparability across observations, with the smallest frequency of observation being the one for the Gini coefficient at 327 and the largest the one of schooling at 455 observations.

*Table 2 Descriptive statistics - Developing countries*

| Developing | | | | | | |
|---|---|---|---|---|---|---|
| **Variable** | **Definition** | **Mean** | **St.dev** | **Min** | **Max** | **Freq. of obsv** |
| **Inequality** | Income inequality measued by the gini coefficient | 40.3 | 7.4 | 21.4 | 62.3 | 667 |
| **AveAggSize** | Average agglomeration size, in terms of population (thousands) | 577.2 | 655.8 | 3 | 7025 | 1247 |
| **Income** | Per capita GDP (thousands) | 5758.4 | 10094.3 | 265.8 | 144340.4 | 1015 |
| **Growth** | Cummulative annual average per capita GDP growth rate | 1.4 | 4.4 | -28.2 | 24.3 | 919 |
| **Investment** | Investment share (% of GDP) | 17.1 | 9.8 | 0.8 | 65.6 | 1016 |
| **Gov Spend** | Goverment consumption (% of GDP) | 19.5 | 11.5 | 1.5 | 81.5 | 1016 |
| **Schooling** | Average years of seconday and teritiary schooling of adult population | 1.5 | 1.5 | 0 | 7.6 | 1027 |

However, for the developing dataset, the thesis does not have the same comparability across observations. This is especially true when comparing the frequency of observations for the Gini coefficient with the main independent variables. However, the developing dataset does include more countries than the developed dataset, contributing to this increasing variation. Furthermore, also includes countries where statistical censuses only recently started taking place on a regular basis. As mentioned before, the data collected for this study is the best available and has been used by other authors such as Castelles-Quintana (2017). Another stand out observation between the developed and developing dataset is the mean of the Gini coefficient, where the mean between 1950 and 2015 in the developed countries is 29.5, whereas in the developing countries is at 40.3. It is to be noted that the largest average urban agglomeration size in the developing dataset is three times bigger than the largest average urban agglomeration size in the developed.

Below shows the correlation between the variables for both developed and developing countries. In table 3, it is worth noting is the correlation between income per capita and average years of secondary and tertiary schooling is the most correlated of all variables in developed countries. This relationship is to be expected as more years of schooling, would increase an individual's average income.

*Table 3 Correlation of variables - Developed countries*

| Correlation of variables (Developed) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Gini** | **AveAggSize** | **Log (income)** | **Growth** | **Investment** | **Gov spend** | **Schooling** |
| **Gini** | 1 | | | | | | |
| **AveAggSize** | 0.0379 | 1 | | | | | |
| **Log (income)** | -0.1947 | 0.1305 | 1 | | | | |
| **Growth** | -0.0379 | 0.0614 | -0.119 | 1 | | | |
| **Investment** | -0.3057 | 0.0579 | 0.1482 | 0.3065 | 1 | | |
| **Gov spend** | -0.2251 | -0.0147 | -0.1598 | -0.1979 | -0.4524 | 1 | |
| **Schooling** | -0.0873 | 0.0304 | 0.7212 | -0.1482 | -0.1594 | 0.0870 | 1 |

In table 4, which looks at the correlation of variables for the developing countries, there is also a quite strong correlation between income per capita, and average years of secondary and tertiary schooling, however not as strong as for developed countries. Worth noting for developing countries is that there is a quite strong correlation between income per capita and investment as a percentage share of GDP, also a quite strong correlation between average urban agglomeration size and income per capita. This can be seen as a population movement to cities could lead to a change in income per capita in developing countries. The following chapter will be about the empirical analysis and the main results of the thesis.

*Table 4 Correlation of variables - Developing countries*

**Correlation of variables (Developing)**

| | Gini | AveAggSize | Log (income) | Growth | Investment | Gov spend | Schooling |
|---|---|---|---|---|---|---|---|
| **Gini** | 1 | | | | | | |
| **AveAggSize** | -0.0726 | 1 | | | | | |
| **Log (income)** | -0.0166 | 0.4252 | 1 | | | | |
| **Growth** | -0.0572 | 0.1487 | 0.2283 | 1 | | | |
| **Investment** | 0.0153 | 0.2509 | 0.4558 | 0.2706 | 1 | | |
| **Gov spend** | -0.1580 | -0.2592 | -0.0685 | -0.2318 | -0.1149 | 1 | |
| **Schooling** | -0.2392 | 0.3461 | 0.5795 | 0.0274 | 0.2172 | 0.1212 | 1 |

# 5 Empirical Analysis

As a starting point to interpret the results, two comparative graphs have been created. One for comparing the development of average urban agglomeration size in developed and developing countries from 1955 to 2015 (due to average urban agglomeration size being lagged one time period). Figure 3 shows a clear advantage in average urban agglomeration size among the developed countries, which can be observed until around 2010 where the average urban agglomeration size between developed and developing countries equalize.

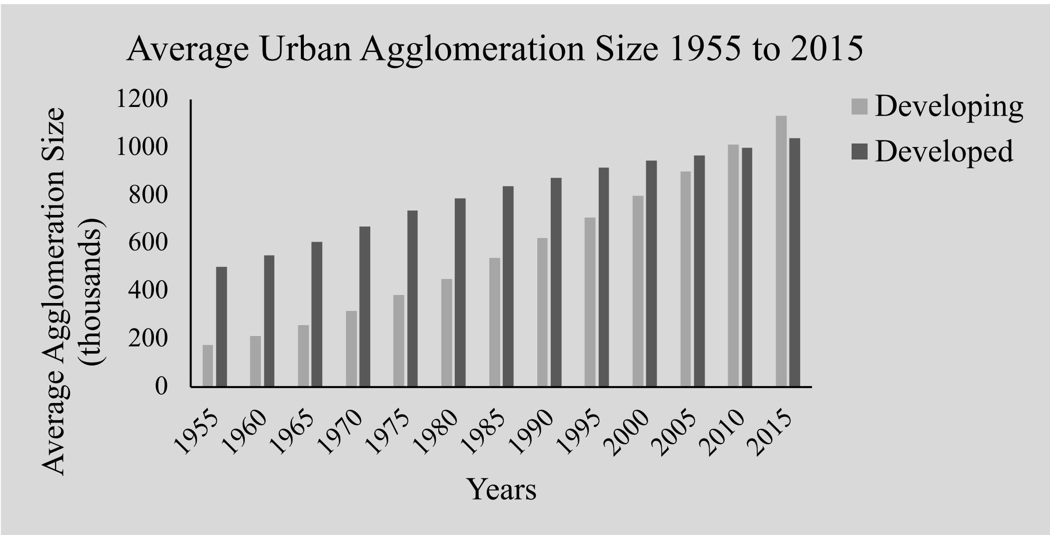


*Figure 3 Average urban agglomeration size from 1955 to 2015, source: author's construction*

However, from looking at figure 4, which is a comparison of per capita income between developed and developing countries, there is still in 2015 a gap between the two groups of countries. This simple illustration would indicate that the increase in average urban agglomeration size in the developing countries, does not go hand in hand with economic development. Which can be compared to the industrialization of the developed countries, where the main theories of the Kuznets curve hypothesis and Lewis dual-sector model are derived from. This provides motivation to distinguish how the relationship of average urban agglomeration size and per capita income does affect income inequality, where the hypothesis is that this relationship between income inequality and average urban agglomeration size is different between developed and developing countries due to different types of development trajectories. To be able to answer the research question of how the relationship is different for developed and developing countries, and what mechanism drives income inequality in the different groups of countries, the study ran several panel data regressions with the discussed set of main independent variables and control variables.

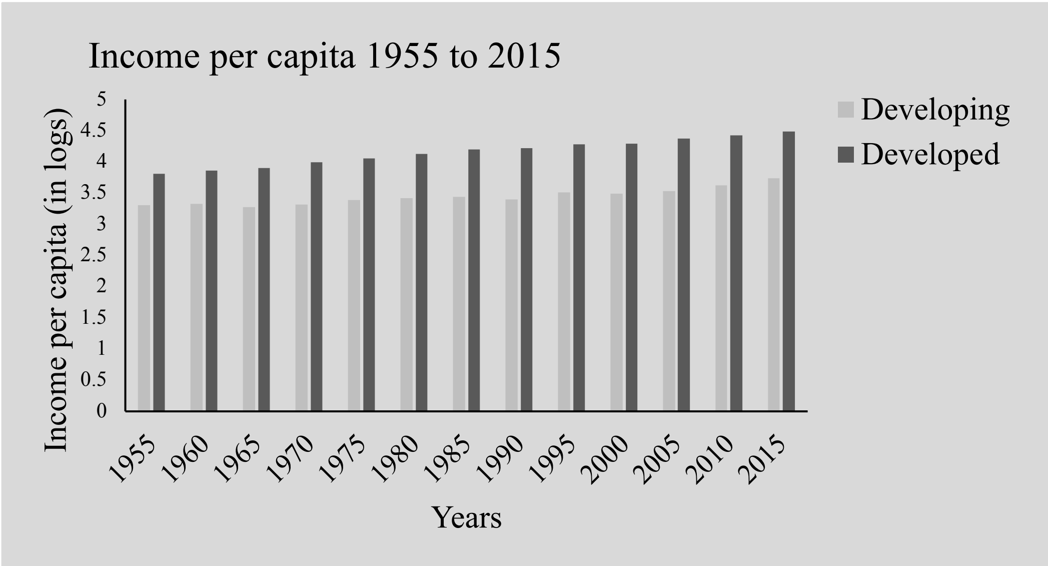


*Figure 4 Comparison of income per capita from 1955 to 2015, source: author's construction*

# 5.1 Main results: fixed effect model

Simple statistical analysis was used to compare the difference between developed and developing countries. Panel data regressions were used to predict the relationship between income inequality and average urban agglomeration size together with a set of control variables.

*Table 5 Main results- Developed countries*

**Relationship Between Inequality and City Size (Developed)**
**Dependent variable is Gini coefficient**

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| **AveAggsize** | -0.001 | -0.009** | -0.008* | -0.009* | -0.009* | -0.004 |
| | (-0.002) | (-0.003) | (-0.005) | (-0.005) | (-0.006) | (-0.007) |
| **AveAggsize^2** | | 0.001*** | 0.001** | 0.001** | 0.001** | 0.001 |
| | | (-0.001) | (-0.001) | (-0.001) | (-0.001) | (-0.001) |
| **Log (income)** | | | 21.205 | 19.485 | 13.553 | -16.410 |
| | | | (-23.163) | (-26.981) | (-31.323) | (-39.1) |
| **Log (income)^2** | | | -2.812 | -2.496 | -1.741 | 2.052 |
| | | | (-2.876) | (-3.413) | (-3.847) | (-4.845) |
| **Growth** | | | | | 0.036 | 0.015 |
| | | | | | (-0.074) | (-0.057) |
| **Investment** | | | | | 0.022 | -0.015 |
| | | | | | (-0.063) | (-0.037) |
| **Gov spend** | | | | | -0.041 | 0.056 |
| | | | | | (-0.089) | (-0.058) |
| **Schooling** | | | | | -0.515 | -0.526 |
| | | | | | (-0.490) | (-0.457) |
| **Beta coefficient** | 30.799 | 34.896 | -5.339 | -1.758 | 10.994 | 71.792 |
| | (-2.873) | (-3.151) | (-44.511) | (-50.436) | (-60.078) | (-81.485) |
| **Year FE** | YES | YES | YES | YES | YES | YES |
| **Country FE** | NO | NO | NO | YES | YES | YES |
| **Additonal Controls** | NO | NO | NO | NO | NO | YES |
| **Observations** | 311 | 311 | 284 | 284 | 277 | 174 |
| **No. Of counties** | 33 | 33 | 33 | 33 | 33 | 33 |

Robust standard errors (clustered by country) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Additional controls include *poptotal, urbrate, fertility, coal, exports* and *agriculture*

Table 5 compares the results obtained from running regression analysis on six different models for developed countries, whereas table 6 compares the same models but for developing countries. As can be seen from table 5 average urban agglomeration size and its squared variable is a driving mechanism of income inequality in developed countries, until a set of additional controls enters in model 6. In model 1 only the relationship between income inequality and average urban agglomeration size is captured, however as the relationship is assumed to be nonlinear between income inequality and average urban agglomeration size, average urban agglomeration size squared is added in model 2. In model 2 to 5 in table 5, a non-linear and u-shaped relationship can be detected between income inequality and average urban agglomeration size and its squared variable. In model 3 when income and it's squared is added into the model, the study shows that the relationship between income inequality and average urban agglomeration size is independent of a non-linear relationship between income inequality and income per capita. This is also true when fixed effects are used for country-specific conditions. Growth, investment, gov spend and schooling does not have a statistically significant relationship to income inequality when entered in model 4. The ultimate city size in developed countries derived from model 4 in table 5 is 3.2 million inhabitants, where the Gini coefficient would be as close to zero as possible for a specific country. Finally, model 5 is the selected model to do comparisons between the two groups of countries. The reasons for this are that it has more observation than model 6, and it is believed the additional controls in model 6 makes the model more demanding then model 5. This can be further seen from the size of the beta coefficient in model 5. In model 3 and 4, it is significantly small indicating a small effect, whereas the value does increase in model 5, however, the independent variables are also still statistically significant in model 5 compared to model 6.

To analyze if the relationship between income inequality and average urban agglomeration size is different between developed and developing countries, the same six models are run on the two different datasets. This is in accordance with the previous study done by Castells-Quitana (2017), to see if a certain group of countries are driving the result in the already tested model. In the developed dataset the study only sees statistical significance of average urban agglomeration size on income inequality, even when adding on important main independent variables such as years of schooling, government expenditure, investment, economic growth rates, and income per capita.

*Table 6 Main results-Developing countries*

| | Relationship Between Inequality and City Size (Developing)<br>Dependent variable is Gini coefficient | | | | | |
|---|---|---|---|---|---|---|
| | **Model 1** | **Model 2** | **Model 3** | **Model 4** | **Model 5** | **Model 6** |
| **AveAggsize** | -0.001 | -0.003** | -0.001 | -0.001 | 0.000 | -0.001 |
| | (-0.001) | (-0.002) | (-0.002) | (-0.002) | (-0.002) | (-0.003) |
| **AveAggsize^2** | | 0.001*** | 0.001 | 0.001 | 0.001 | 0.001 |
| | | (-0.001) | -0.001 | (-0.001) | (-0.001) | (-0.001) |
| **Log (income)** | | | 2.839** | 2.586** | 2.543** | 5.452*** |
| | | | (-1.176) | *-1.415) | (-1.531) | (-2.077) |
| **Log (income)^2** | | | -0.288 | -0.210 | -0.205 | -1.005* |
| | | | (-0.300) | (-0.372) | (-0.421) | (-0.605) |
| **Growth** | | | | | 0.007 | 0.034 |
| | | | | | (-0.036) | (-0.042) |
| **Investment** | | | | | 0.069** | 0.048* |
| | | | | | (-0.030) | (-0.026) |
| **Gov spend** | | | | | -0.003 | 0.027 |
| | | | | | (-0.026) | (-0.031) |
| **Schooling** | | | | | 0.052 | 0.333 |
| | | | | | -0.711 | (-0.689) |
| **Beta coefficient** | 40.287 | 40.578 | 34.785 | 31.906 | 31.194 | 32.395 |
| | (-1.152) | (-1.397) | (-1.652) | (-2.174) | (-2.976) | (-6.015) |
| **Year FE** | YES | YES | YES | YES | YES | YES |
| **Country FE** | NO | NO | NO | YES | YES | YES |
| **Additonal Controls** | NO | NO | NO | NO | NO | YES |
| **Observations** | 667 | 667 | 624 | 624 | 537 | 418 |
| **No. Of counties** | 96 | 96 | 96 | 96 | 79 | 75 |

Robust standard errors (clustered by country) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Additional controls include *poptotal, urbrate, fertility, coal, exports* and *agriculture*

In table 6, where the study uses the sample with developing countries, the main results are different. This indicates that the mechanisms driving income inequality in developing countries are not the same as in developed countries. In table 6, model 2, the study does see a statistically significant result of average urban agglomeration size squared on income inequality. However, when including income per capita in model 3, the study can assume that the statistical significance seen in model 2 on average urban agglomeration size was possibly driven by the increasing income per capita. Where income per capita then after both adding main independent variables and control variables is statistically significance throughout the analyzes from model 3 to 6 in table 6. Furthermore, when the investment share of GDP is included in the regression this also shows as a mechanism that is impacting income inequality in the long-run. The economic significance of income per capita does increase with additional variables in the models, however, investment does decrease. To try to compare ultimate city size with developed countries, even though there is no significance of the relationship in table 6, the study uses the estimations from model 4 to calculate the ultimate city size with the lowest possible income inequality. Among the developing countries, it would not be able to get close to a zero Gini coefficient, but the lowest possible would be around 31.33 and at that point, the city size would be just over 2 million inhabitants. From interpreting the beta coefficient in the different models, it can be observed that all the models show a strong effect from analyzing the beta coefficient. As the larger the value of the beta coefficient the stronger the effect of the model. Furthermore, as previously mentioned when discussing developed countries, model 5 has been the selected model from the main results. This can also be concluded on from the main results of the developing countries, as it is the model with not only the greatest number of variables in combination with statistically significance but also with a large beta coefficient with a rather small standard error.

### 5.1.1 Comparison of the main result of developed and developing countries

If looking at a comparison between results of model 5 for both developed and developing countries the following can be concluded upon:

*Table 7 Comparison of main results*

**Comparison Developed and Developing countries**
**Dependent variable is Gini coefficient**

| | Model 5.1 | Model 5.2 |
|---|---|---|
| **AveAggsize** | -0.009* | 0.000 |
| | (-0.006) | (-0.002) |
| **AveAggsize^2** | 0.001** | 0.001 |
| | (-0.001) | (-0.001) |
| **Log (income)** | 13.553 | 2.545** |
| | (-31.323) | (-1.531) |
| **Log (income)^2** | -1.741 | -0.205 |
| | (-3.847) | (-0.421) |
| **Growth** | 0.036 | 0.007 |
| | (-0.074) | (-0.036) |
| **Investment** | 0.022 | 0.069** |
| | (-0.063) | (-0.030) |
| **Gov spend** | -0.041 | -0.003 |
| | (-0.090) | (-0.026) |
| **Schooling** | -0.515 | 0.052 |
| | (-0.491) | (-0.711) |
| **Beta coefficient** | 10.994 | 31.194 |
| | (-60.978) | (-2.976) |
| **Year FE** | YES | YES |
| **Country FE** | YES | YES |
| **Additonal Controls** | NO | NO |
| **Observations** | 277 | 537 |
| **No. Of counties** | 33 | 79 |

Robust standard errors (clustered by country) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Additional controls include *poptotal, urbrate, fertility, coal, exports and agriculture*

The study does not see a u-shaped relationship between income inequality and average urban agglomeration size in developing countries, model 5.2 in table 7. However, this relationship can be observed in developed countries, model 5.1 in table 7. Furthermore, the variables of importance for income inequality in developing countries are from this result income per capita and investment share of GDP. When looking at the economic differences, the study observes much higher coefficients in the developed countries dataset, compared to the developing. Furthermore, even though the amount of observations for developed countries is half of those of developing, average urban agglomeration size still has a statistical significance. Another economical difference which is a surprising result is the effect of schooling on income inequality in the two different sets of countries.

Assumingly, in developed countries, model 5.1 in table 7, the study sees a negative association between the Gini coefficient and schooling, meaning that with additional years of secondary or tertiary education, there is a decrease of 0.515 in the Gini coefficient, meaning lower income inequality. On the other hand, developing countries show a different story, where an extra year of secondary or tertiary schooling, leads to a 0.0523 increase in the Gini coefficient, indicating rising income inequality. Finally, when comparing the beta coefficient of the two groups of countries, a higher absolute value is observed in the set of developing countries. This would indicate that the effect of the selected variables in model 5 has a stronger effect on income inequality than for the developed countries. However, the developed countries show still quite a strong beta coefficient if compared with the other models seen in table 6. The difference between the two sets of countries could be due to the difference in the amount of countries studied and therefore the amount of observations between the two groups.

## 5.2 Robustness checks

To be able to conclude that average urban agglomeration size is the driving mechanism of income inequality in developed countries, but not in developing countries, a robustness check bringing in three different control variables, including the population in the largest city, percentage of population living in cities with more than 1 million inhabitants and primacy, the population living in the largest city as a share of total urban population. This is to be able to indicate that the results are robust for developed countries and to test the possibility of income per capita being the driving mechanism of income inequality for developing countries.

*Table 8 Robustness check -Developed countries*

| | Robustness Check (Developed) Dependent variable is Gini coefficient | | | |
|---|---|---|---|---|
| | **Model 1** | **Model 2** | **Model 3** | **Model 4** |
| **AveAggsize** | -0.010* | 0.014* | -0.010* | -0.010* |
| | (-0.005) | (-0.008) | (-0.006) | (-0.005) |
| **AveAggsize^2** | 0.001*** | -0.001* | 0.001 *** | 0.001*** |
| | (-0.001) | (-0.001) | (-0.001) | (-0.001) |
| **Log (income)** | 245.005 | 136.533 | 234.732 | 118.238 |
| | (-177.618) | (-109.474) | (-202.227) | (-156.366) |
| **Log (income)^2** | -60.218 | -28.492 | -57.996 | -28.347 |
| | (-46.178) | (-28.436) | (-50.944) | (-40.330) |
| **Log (income)^3** | 4.897 | 1.848 | 4.740 | 2.316 |
| | (-3.979) | (-2.465) | (-4.277) | (-3.463) |
| **Pop largest city** | | -0.001*** | | |
| | | (-0.001) | | |
| **Pop largest city^2** | | 0.001*** | | |
| | | (-0.001) | | |
| **Urb 1m** | | | -0.090 | |
| | | | (-0.258) | |
| **Urb 1m^2** | | | 0.001 | |
| | | | (-0.002) | |
| **Primacy** | | | | -0.698* |
| | | | | (-0.366) |
| **Primacy^2** | | | | 0.012*** |
| | | | | (-0.005) |
| **Year FE** | YES | YES | YES | YES |
| **Country FE** | YES | YES | YES | YES |
| **Controls** | YES | YES | YES | YES |
| **Observations** | 277 | 272 | 252 | 272 |
| **No. Of counties** | 33 | 33 | 28 | 33 |

Robust standard errors (clustered by country) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Additional controls include *poptotal, urbrate, fertility, coal, exports* and *agriculture*

In table 8 there is still a significant negative relationship between income inequality and average urban agglomeration size in model 1, 3 and 4. Indicating that even when bringing in additional variables with similar characteristics, the average urban agglomeration size does influence income inequality in the long-run. Following the addition of population in the largest city, there is a decrease in the significance of average urban agglomeration size, seen in model 2. This could be due to that in a developed country the largest city derives rising income inequality due to a rural-urban divide, but also relates to the discussion around Zipf's law. Furthermore, strong evidence was also found in model 4 when adding on the percentage of people living in the largest city as the share of the total urban population, that the largest city in developed countries does impact income inequality negatively, leading to a rise in the Gini coefficient.

*Table 9 Robustness check -Developing countries*

| Robustness Check (Developing)<br>Dependent variable is Gini coefficient | | | | |
|---|---|---|---|---|
| | **Model 1** | **Model 2** | **Model 3** | **Model 4** |
| **AveAggsize** | 0.001 | -0.001 | 0.001 | 0.001 |
| | (-0.002) | (-0.002) | (-0.003) | (-0.003) |
| **AveAggsize^2** | 0.001 | 0.001 | 0.001 | 0.001 |
| | (-0.001) | (-0.001) | (-0.001) | (-0.001) |
| **Log (income)** | 1.425 | 4.569 | -3.404 | -0.279 |
| | (-6.841) | (-6.014) | (-9.439) | (-7.363) |
| **Log (income)^2** | 0.340 | -0.785 | 2.737 | 1.418 |
| | (-3.371) | (-3.084) | (-4.839) | (-3.626) |
| **Log (income)^3** | -0.069 | 0.011 | -0.390 | -0.233 |
| | (-0.414) | (-0.396) | (-0.635) | (-0.446) |
| **Pop largest city** | | 0.001 | | |
| | | (-0.001) | | |
| **Pop largest city^2** | | -0.001 | | |
| | | (-0.001) | | |
| **Urb 1m** | | | 0.187 | |
| | | | (-0.263) | |
| **Urb 1m^2** | | | -0.002 | |
| | | | (-0.002) | |
| **Primacy** | | | | 0.008 |
| | | | | (-0.232) |
| **Primacy^2** | | | | -0.001 |
| | | | | (-0.002) |
| **Year FE** | YES | YES | YES | YES |
| **Country FE** | YES | YES | YES | YES |
| **Controls** | YES | YES | YES | YES |
| **Observations** | 537 | 524 | 460 | 524 |
| **No. Of counties** | 79 | 78 | 65 | 78 |

Robust standard errors (clustered by country) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Additional controls include *poptotal, urbrate, fertility, coal, exports* and *agriculture*

When the study did the same robustness check on developing countries, the study would have assumed that income per capita would hold a significant value throughout the different control variables, however, this was not the case. In the robustness check, income in a quadratic term was added, to test for an n-shaped relationship between income inequality and economic growth, to look for potential for a Kuznets wave rather than a Kuznets curve. No statistical significance was found in any of the different models in table 9 when analyzing developing countries. The most striking observation to emerge from the data comparison is that in model 3 in table 9, the economic signage of nearly all variables did change. The thesis can conclude from this observation that the percentage of the population living in cities with more than one million inhabitants does have an association with income inequality. This does logically make sense when considering that the majority of megacities located in developing countries.

### 5.2.1 Comparison of robustness checks developed and developing countries

If the study compares the robustness check of the developed and developing countries where the population in the largest city is set as the control variable, we can conclude the following:

*Table 10 Comparison of robustness checks*

**Comparison Developed and Developing countries**
**Dependent variable is Gini coefficient**

| | Model 2.1 | Model 2.2 |
|---|---|---|
| **AveAggsize** | 0.014* | -0.001 |
| | (-0.008) | (-0.002) |
| **AveAggsize^2** | -0.001* | 0.001 |
| | (-0.001) | (-0.001) |
| **Log (income)** | 136.533 | 4.569 |
| | (-109.4737) | -6.014 |
| **Log (income)^2** | -28.492 | -0.785 |
| | (-28.436) | (-3.084) |
| **Log (income)^3** | 1.848 | 0.011 |
| | (-2.465) | (-0.396) |
| **Pop largest city** | -0.001*** | 0.001 |
| | (-0.001) | (-0.001) |
| **Pop largest city^2** | 0.001*** | 0.001 |
| | (-0.001) | (-0.001) |
| **Urb 1m** | | |
| **Urb 1m^2** | | |
| **Primacy** | | |
| **Primacy^2** | | |
| **Year FE** | YES | YES |
| **Country FE** | YES | YES |
| **Controls** | YES | YES |
| **Observations** | 272 | 524 |
| **No. Of counties** | 33 | 78 |

Robust standard errors (clustered by country) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Additional controls include *poptotal, urbrate, fertility, coal, exports and agriculture*

The non-linear relationship between average urban agglomeration size and income inequality, is different for developed and developing countries, with the economic sign of the values being different at average urban agglomeration size and average urban agglomeration size squared. Furthermore, as concluded before in the comparison of the two sets of countries in the main result section, the economic significance is much greater for the developed countries than in developing countries. In conclusion, the robustness check holds for the developed countries but does not hold for the developing countries.

## 5.3 Endogeneity issue

To be able to conclude that there is a strong association between average urban agglomeration size and income inequality or even a possibility of causation in developed countries the study must consider the possibility of an endogeneity issue. Meaning that the increase of average urban agglomeration size does not only affect income inequality, but that income inequality does also affect the average urban agglomeration size. The possible issue has to some extent already been addressed, when all variables on the right-hand side of the equation (1) have been lagged in one time period.

*Table 11 Instrumental variable estimations- Developed countries*

| | Instrumental variable estimations (Developed)<br>Dependent variable is Gini coefficient | | |
|---|---|---|---|
| | **Model 1 (FD)** | **Model 2 (FD-IV)** | **Model 3 (FD-IV)** |
| **ΔAveAggsize** | -0.010*** | -0.012** | -0.009* |
| | (-0.004) | (-0.005) | (-0.005) |
| **ΔAveAggsize^2** | 0.001*** | 0.001** | 0.001** |
| | (-0.001) | (-0.001) | (-0.001) |
| **ΔLog (income)** | 229.726*** | 231.862** | 203.680 ** |
| | (-79.430) | (-101.589) | (-103.568) |
| **ΔLog (income)^2** | -54.556*** | -58.292 ** | -51.762** |
| | (-19.827) | (-25.586) | (-26.238) |
| **ΔLog (income)^3** | 4.244** | 4.758** | 4.245* |
| | (-1.671) | (-2.138) | (-2.205) |
| **Year FE** | YES | YES | YES |
| **Country FE** | YES | YES | YES |
| **Controls** | YES | YES | YES |
| **Observations** | 244.000 | 239.000 | 230.000 |
| **No. Of counties** | 33.000 | 33.000 | 33.000 |
| **AP first-stage F-stat P-value** | | 0.0000; 0.0000 | 0.0000; 0.0000 |
| **Kleibergen-Paap F-stat** | | 79.350 | 50.310 |
| **Kleibergen-Paap LM-stat** | | 62.16*** | 57.71*** |
| **Hansen J stat P-value** | | 0.375 | 0.696 |

Robust standard errors (clustered by country) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Controls include G*rowth, Investment, Gov spend* and *schooling*

However, endogeneity can arise due to omitted or unobserved variables when confounding independent variable, but also possibly dependent variables. To solve for this possibility, the study firstly ran a first-difference estimation, and then a first-difference instrumental variables estimation, with two sets of different instruments.

The first-difference result for developed countries, which can be seen in table 11, model 1, shows great statistical significance to all variables in the model. First-differences is important to use in panel data as every country has its own specific unobserved effects, and there is not a common intercept for all countries, if assuming the intercept is the same, the result becomes biased and inconsistent. However, there is also the possibility of reverse causality leads to bias estimates of the results. In model 2, the instrument is a second lagged level and third lagged level of average urban agglomeration size, meaning 10- and 15-years lags. In model 3, there is a third and fourth lagged level of average urban agglomeration size, meaning 15-and-20-year lags. What stands out in table 11, is that most of the variables increased in significance both statistically and economically when running first differences, and adding on an instrumental variable to control for possible endogeneity. This shows that the first difference instrumental variable estimator confirms our main results that average urban agglomeration size is strongly associated with income inequality, and that it is the size of the urban agglomeration that affects income inequality and not the other way around in developed countries.

*Table 12 Instrumental variable estimation- Developing countries*

| | **Instrumental variable estimations (Developing)<br>Dependent variable is Gini coefficient** | | |
|---|---|---|---|
| | **Model 1 (FD)** | **Model 2 (FD-IV)** | **Model 3 (FD-IV)** |
| **ΔAveAggsize** | 0.002 | -0.001 | -0.001 |
| | (-0.001) | (-0.002) | (-0.002) |
| **ΔAveAggsize^2** | -0.001 | 0.001 | 0.001 |
| | (-0.001) | (-0.001) | (-0.001) |
| **ΔLog (income)** | 5.282 | 4.777 | 4.646 |
| | (-3.298) | (-2.929) | (-2.905) |
| **ΔLog (income)^2** | -1.201 | -0.626 | -0.562 |
| | (-1.624) | -1.473 | (-1.460) |
| **ΔLog (income)^3** | 0.112 | -0.015 | -0.022 |
| | (-0.197) | (-0.186) | (-0.184) |
| **Year FE** | YES | YES | YES |
| **Controls** | YES | YES | YES |
| **Observations** | 453.000 | 453.000 | 444.000 |
| **No. Of counties** | 79.000 | 79.000 | 79.000 |
| **AP first-stage F-stat P-value** | | 0.0000; 0.0000 | 0.0000; 0.0000 |
| **Kleibergen-Paap F-stat** | | 79.620 | 45.910 |
| **Kleibergen-Paap LM-stat** | | 81.23*** | 64.7*** |
| **Hansen J stat P-value** | | 0.672 | 0.705 |

Robust standard errors (clustered by country) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Controls include G*rowth, Investment, Gov spend* and *schooling*

From the data presented in table 12, the study shows that the results for the developing countries first difference estimations and then first difference instrumental variable estimations are not the same. This was to be expected from the estimations in the main result section and robustness check. The assumption here would be that income per capita would hold as statistically significant through the endogeneity estimator which it did not. When adding on income per capita to the power of three, to assume an n-shaped relationship between income per capita and income inequality, the statically significance found in the main result, disappear. There is no evidence that there is a strong non-linear relationship in developing countries between average urban agglomeration size and income inequality, but neither one for income per capita and income inequality in the long-run. To be able to understand the developing countries mechanism for income inequality, the next section moves on to discuss the results from splitting the developing countries into regional sub-groups.

## 5.4 Further extension

In table 13 the developing countries have been divided into four continent groups, which are Africa, America, Asia and Europe (classified by geographic location). This to understand further if there are differences between continents in the mechanism that is driving changes in income inequality.

*Table 13 Further extension estimations*

| | **Relationship Between Inequality and City Size (Developing)**<br>**Dependent variable is Gini coefficient** | | | |
|---|---|---|---|---|
| | **Africa** | **America** | **Asia** | **Europe** |
| **AveAggsize** | 0.009 | -0.002 | 0.001 | 0 .064*** |
| | (-0.008) | (-0.004) | (-0.005) | (-0.018) |
| **AveAggsize^2** | -0.001 | -0.001 | 0.001 | 0.001 |
| | (-0.001) | (-0.001) | (-0.001) | (-0.001) |
| **Log (income)** | -6.857 | 2.340 | 4.481 | 123.112*** |
| | (-13.609) | (-2.546) | (-18.650) | (-32.936) |
| **Log (income)^2** | -1.198 | -0.084 | -0.510 | -16.606 *** |
| | (-1.846) | (-0.724) | (-2.540) | (-4.458) |
| **Growth** | 0.001 | 0.035 | -0.005 | -0.061*** |
| | (-0.058) | (-0.108) | (-0.043) | (-0.013) |
| **Investment** | 0.001 | 0.053 | 0.077* | -0.111* |
| | (-0.031) | (-0.046) | (-0.041) | (-0.064) |
| **Gov spend** | -0.055 | 0.110 | 0.007 | -0.117*** |
| | (-0.034) | (-0.072) | (-0.037) | (-0.017) |
| **Schooling** | 0.616 | -0.488 | -1.012 | -2.616*** |
| | (-0.908) | (-0.591) | (-1.207) | (-0.562) |
| **Year FE** | YES | YES | YES | YES |
| **Country FE** | YES | YES | YES | YES |
| **Additonal Controls** | NO | NO | NO | NO |
| **Observations** | 167.000 | 158.000 | 193.000 | 19.000 |
| **No. Of counties** | 27.000 | 20.000 | 27.000 | 5.000 |

Robust standard errors (clustered by country) in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Additional controls include *poptotal, urbrate, fertility, coal, exports* and *agriculture*

From table 13 the study can conclude on the following results: Europe is the main driving region of the statistically significant results, even though being the lowest number of countries and therefore the lowest number of observations. There is a statistically significant investment coefficient for Asia, which is to be expected due to the inflow and outflow of foreign direct investment since the 1980s. However, no significance was found for neither Africa nor America. This indicates that there are other variables that could be the mechanism driving changes in income inequality in these country groups. Taken together, the results show that there is an association between average urban agglomeration size and income inequality in developed countries, however, that this does not hold for the developing countries. The result in this chapter, therefore, indicates that there are different driving mechanisms for developed and developing countries, therefore, the next section moves on to discuss the results in relation to previous studies in the same area to analyze further where the difference could be.

## 5.5 Discussion

As mentioned in the literature review Castells-Quitana (2017) found a u-shaped relationship between average urban agglomeration size and income inequality, which indicated that income inequality first fell with urban agglomeration size and then rose. However, very little of the previous studies focused on the difference between countries and did rather look at cities within a country or a panel of mixed countries. The present study was designed to determine the difference between developed and developing countries and their relationship between income inequality and average urban agglomeration size. The most important finding was that developed countries were driving the statistical significance of the results in Castells-Quitana (2017) study. When splitting the two groups of countries, developing countries showed very little association between average urban agglomeration size and income inequality, and when adding on first difference and instrumental variable estimations very little association between income per capita and income inequality was found for developing countries. Comparison of the findings with those of other studies confirms that when looking at developed countries, such as the United States, which Fallah and Partdige (2007) studied, indicated that there should be a strong association between average urban agglomeration size and income inequality. However, the outcome that developing countries do not have this association is contrary to finding by Castells-Quitana (2017) but is also contrary to the recent study done by Sulemana et al. (2019) looking at Sub-Saharan Africa. The findings that none of the variables for developing countries are significant is rather disappointing.

A possible explanation for this result is that there are other variables that are important in developing countries, this could potentially include the type of governance and foreign direct investment share of GDP. However, these results are also likely to be related to the different stages of development and the different time periods they have occurred in. As Jedwab & Vollrath (2019) have indicated urbanization without growth is occurring today, which has not been observed during the industrialization of the developed countries in the past. This has also been previously observed by Fay and Opal (2000). A notion of causation is due here due to the scarcity of the data but also as it has tried to control for country differences and time differences, there could be unobserved aspects affecting the results. These findings may help us understand the reasons for urbanization happening without growth, but also the different development trajectories countries take today compared to in the past. Furthermore, some of the issues emerging from these findings relate specifically to how to consider sustainable city development in developed countries in the future with rising income inequality on a national level in mind. The last chapter will conclude and give some final remarks to the findings of the thesis.

# 6 Conclusion and Final Remarks

The aim of the thesis is to examine the relationship between income inequality and average urban agglomeration size and to further qualify this relationship with respect to developed and developing countries. The second aim was to understand what mechanism other than average urban agglomeration size could be driving changes in income inequality in developed and developing countries. Through panel data regressions with fixed effects, the study looked at several different models for both developed and developing countries, with different control variables to estimate the relationship between average urban agglomeration size and income inequality. One of the more significant findings to emerge from this study was that average urban agglomeration size does have a strong association with income inequality in developed countries, however, this relationship between average urban agglomeration size and income inequality does not hold for developing countries. This indicates that urbanization as a policy process could be used to enhance economic growth without the risk of increasing income inequality in developing countries.

The findings of this research provide insight into the development trajectory that developing countries are experiencing is not exactly the same as those that have occurred in the developed countries. Therefore, the main hypothesis to why the results are different in developed and developing countries is the one of that the industrialization which happened in the developed countries happened without increasing globalization and an increasing service sector which can be observed in the developing countries today. This hypothesis is in line with the findings done by Yorukoglo (2002) that improved technology and increased service sector today make cities wider, and this leads to decreasing income inequality. Furthermore, these findings have significant policy implications for using urbanization not being an effective policy tool to increase economic growth and development, but also to sustain long-term economic growth at the expense of more unequal societies. However, if the policy is suitable for developing countries the study is unable to conclude on, and that it should proceed with caution. The most important limitation in the study lies in that is has been unable to conclude on what the driving mechanism of income inequality is in the developing countries. The study has opened a window which shows that income per capita could be important for the development trajectory of developing countries to explain the future development of income inequality. However, the findings are unable to explain how this relates to the phenomena of that more and more people live in cities in developing countries. In spite of its limitations, the study certainly adds to our understanding of the relationship between average urban agglomeration size and income inequality in different types of countries. The question raised by this study is what does affect income inequality in developing countries? And how will the increasing amount of people living in cities in developing countries affect income per capita and long-run income inequality? However, further work is also needed to be done to establish a strong causal effect between income inequality and average urban agglomeration size in developed countries. Also, further research should be carried out to establish if the relationship is different regionally between all type of countries, not just a region split of developing countries, to understand if a certain continent is driving results also for the developed countries. If true, further research should be carried out to establish what is particular about that region compared to other regions in the world. The findings of this study suggest caution to policymakers and government officials when considering a push or pull strategy for populations towards cities to increase economic growth. This suggestion is in line with Agenda 2030, focusing on reduced inequalities and sustainable cities. To be able to reach sustainable development both in developed and developing countries, policymakers need to consider what vision they see for peoples' living situation to not only to increase equality, but also to have sustained economic prospects and improving living standards.

# References

Adams, S. & Klobodu, E. (2018). Urbanization, Economic Structure, Political Regime, and Income Inequality, *Social Indicators Research*, vol. 142, no. 3, pp. 971-995

Alperovich, G. (1995). The Relationship between Income Inequality and City Size: A General Equilibrium Model for an Open System of Cities Approach, *Urban Studies*, vol. 32, no. 6, pp. 853-862

Arshad, S., Hu, S. & Ashraf, B. (2018). Zipf's law and the city size distribution: A survey of the literature and future research agenda, *Physica A: Statistical Mechanics and its Applications*, vol. 492, pp. 75-92

Barro, R. (2000). Inequality and Growth in a Panel of Countries, *Journal of Economic Growth*, vol. 5, no. 1, pp. 5-32

Barro, R. & Lee, J. (2013). A New Data Set of Educational Attainment in the World, 1950-2010, *Journal of Development Economics*, vol. 104, pp.184-198

Baum-Snow, N. & Pavan, R. (2013). Inequality and city size, *The Review of Economics and Statistics*, vol. 95, no. 5, pp. 1535-1548

Behrens, K. & Robert-Nicoud, F. (2014). Survival of the fittest in cities: urbanisation and inequality, *The Economic Journal*, vol. 124, no. 581, pp. 1371-1400

Bertinelli, L. & Black, D. (2004). Urbanization and growth, *Journal of Urban Econo*mics, vol. 56, no. 1, pp. 80-96

Bloom, D., Canning, D. & Fink, G. (2008). Urbanization and the Wealth of Nations, *Science*, vol. 319, no. 5864, pp. 772 – 775

Bruno, M., Ravallion, M. & Squire, L. (1996). Equity and growth in developing countries: old and new perspectives on the policy issues, *policy research paper 1563*, the World Bank

Cali, M. & Menon, C. (2013). Does Urbanization Affect Rural Poverty? Evidence from Indian Districts, *The World Bank Economic Review*, vol. 27, no. 2, pp. 171-201

Caselli, F., Esquivel, G. & Lefort, F. (1996). Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics, *Journal of Economic Growth*, vol. 1, no. 3, pp. 363-389

Castells, D. (2011). Agglomeration, Inequality and Economic Growth: Cross-section and panel data analysis, *Working Paper 2011/14*, Research Institute of Applied Economics

Castells-Quintana, D. & Royuela, V. (2012). Unemployment and long-run economic growth: The role of income inequality and urbanization, *Investigaciones Regionales*, vol. 24, pp. 153-173

Castells-Quintana, D. & Royuela, V. (2015). Are increasing urbanisation and inequalities symptoms of growth? *Applied Spatial Analysis and Policy*, vol. 8, no.3, pp. 291-308

Castells-Quintana, D. (2017). Beyond Kuznets: Inequality and the size and distribution of cities, *Journal of Regional Science*, vol. 58, no. 3, pp. 568-580

Chambers, D. & Dhongde, S. (2016). Are countries becoming equally unequal*? Empirical Economics*, vol. 53, no. 4, pp. 1323 – 1348

Chen, G., Glasmeier, A., Zhang, M. & Shao, Y. (2016). Urbanization and income inequality in post-reform China: a causal analysis based on time series data, *PloS one*, vol. 11, no. 7

Ehrhat, C. (2009). The effects of inequality on growth: a survey of the theoretical and empirical literature, *working paper series*, ECINE, Society for the Study of Economic Inequality

Fallah, B. & Partridge, M. (2007). The elusive inequality-economic growth relationship: are there differences between cities and the countryside? *The Annals of Regional Science*, vol. 41, no. 2, pp. 375-400

Fay, M. & Opal, C. (2000). Urbanization without Growth: A Not-So-Uncommon Phenomenon, *Policy Research Working Paper 2412*, The World Bank

Feenstra, R., Inklaar, R. & Timmer, M. (2015). The Next Generation of the Penn World Table, *American Economic Review*, vol. 105 no. 10, pp. 3150-3182

Frazer, G. (2006). Inequality and Development Across and Within Countries, *World Development*, vol. 34, no. 9, pp. 1459-1481

Fujita, M. (2013). Economics of Agglomeration, Cambridge University Press

Glaeser E., Resseger, M. & Tobio, K. (2009). Inequality in cities, *Journal of Regional Science*, vol. 49, no. 4, pp. 617-646

Gustafsson, B. & Johansson, M. (1999). In Search of Smoking Guns: What makes income inequality vary over time in different countries? *American Sociological Review*, vol. 64, pp. 585-605

Harding, A. & Blokland, T. (2014). Urban Theory: a critical introduction to power, cities and urbanism in the 21st century, SAGE

Haworth, C., Long, J. & Rasmussen, D. (1978). Income Distribution, City Size, and Urban Growth, *Urban Studies*, vol. 15, no. 1, pp. 1-7

Henderson, V. (2003). The Urbanization Process and Economic Growth: The So-What Question, *Journal of Economic Growth*, vol. 8, no. 1, pp. 47-71

Henderson, V. (2010). Cities and Development, *Journal of Regional Science*, vol. 50, no. 1, pp. 515-540

Jedwab, R. & Vollrath, D.  (2019). The Urban Mortality Transition and Poor-Country Urbanization, *American Economic Journal: Macroeconomics*, vol. 11, no. 1, pp. 223-275

Kennedy, P. (2008). A guide to econometrics. 6th ed. Malden, Mass (U.S), Wiley-Blackwell

Kuznets, S. (1955). Economic growth and income inequality, *American Economic Review*, vol. 45, pp. 1-28

Lewis, A. (1954). Economic development with unlimited supply of labor, *Manchester School of Economics and Social Studies*, vol. 22, pp. 139 -191

Li, H., Squire, L. & Zou, H. (1998). Explaining international and intertemporal variations in income inequality, T*he Economic Journal*, vol. 108, no. 446, pp. 26-43

Marrero, G. & Rodriguez, J. (2013). Inequality of opportunity and growth, *Journal of Development Economics*, vol. 104, pp. 107 – 122

Milanovic, B. (1994). Determinants of Cross-Country Income Inequality, *Policy Research Working Paper 1246*, The World Bank

Milanovic, B. (2016). Global Inequality: A New Approach for the Age of Globalization, Harvard University Press

Nord, S. (1980). An Empirical Analysis of Income Inequality and City Size, *Southern Economic Journal*, vol. 46, no. 3, pp. 863-872

Perkins, D., Radelet, S., Lindauer, D., & Block, S. (2013). Economics of Development. 7th edition, New York: W.W. Norton & Company

Piketty, T., & Saez, E. (2014). Inequality in the long-run, *Science*, vol. 344, no. 6186, pp. 838-843

Royela, V., Veneri P. & Ramos, R. (2014). Income Inequality, Urban Size and Economic Growth in OECD Regions, *OECD Regional Development Working Papers*

Sachs, J. & Warner, A. (1995). Economic Convergence and Economic Policies, *working paper*, National Bureau of Economic Research

Sanford, F.H. (1942). Review of National unity and disunity: The nation as a bio-social organism, Journal of Abnormal Psychology, vol. 37, no. 1, pp. 151

Sarkar, S., Phibbs, P., Simpson, R. & Wasnik, S. (2018). The scaling of income distribution in Australia: Possible relationships between urban allometry, city size, and economic inequality, *Environment and Planning B: Urban Analytics and City Science*, vol. 45, no. 4, pp. 603-622

Solt, F. (2019). Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database. SWIID Version 8.0, February 2019

Sulemana, I., Nketiah-Amponsah, E., Codjoe, E. & Andoh, J. (2019). Urbanization and income inequality in Sub-Saharan Africa, *Sustainable Cities and Society*, vol. 48

UN. (2018). World Urbanization Prospects: The 2018 Revision. Department of Economic and Social Affairs, Population Division. UN DESA Press.

Vanhoudt, P. (2000). An assessment of the macroeconomic determinants of inequality, *Applied Economics*, vol. 32, no. 7, pp. 877-883

Williamson, J. (1965). Regional Inequality and the Process of National Development: A Description of the Patterns*, Economic Development and Cultural Change*, vol. 13, no. 4, pp. 1-84

World Development Indicators (2019). The World Bank

Yorukoglu, M. (2002). The Decline of Cities and Inequality, *American Economic Review*, vol. 92, no. 2, pp. 191-197

# Appendix A

| Developing Countries | | Developed Countries |
|---|---|---|
| Albania | Malawi | Australia |
| Algeria | Malaysia | Austria |
| Angola | Mali | Belgium |
| Argentina | Mauritania | Bulgaria |
| Armenia | Moldova | Canada |
| Azerbaijan | Mongolia | Croatia |
| Bangladesh | Morocco | Czech Republic |
| Belarus | Mozambique | Denmark |
| Bolivia | Myanmar | Estonia |
| Brazil | Namibia | Finland |
| Burkina-Faso | Nepal | France |
| Burundi | Nicaragua | Germany |
| Cambodia | Niger | Greece |
| Cameroon | Nigeria | Hungary |
| Central Africa | Pakistan | Ireland |
| Chad | Panama | Italy |
| Chile | Paraguay | Japan |
| China | Peru | Republic of Korea |
| Colombia | Philippines | Latvia |
| Costa Rica | Qatar | Lithuania |
| Cote d'Ivoire | Russia | Mexico |
| Djibouti | Rwanda | Netherlands |
| Dominican Republic | Senegal | New Zeeland |
| Ecuador | Serbia | Norway |
| El Salvador | Sierra Leone | Poland |
| Ethiopia | Singapore | Portugal |
| Gambia | South Africa | Romania |
| Georgia | Sri Lanka | Slovakia |
| Ghana | Sudan | Slovenia |
| Guatemala | Syria | Spain |
| Guinea | Taiwan | Sweden |
| Guinea-Bissau | Tajikistan | Switzerland |
| Haiti | Tanzania | Turkey |
| Honduras | Thailand | United Kingdom |
| Hong-Kong | Trinidad & Tobago | United States |
| India | Tunisia | |
| Indonesia | Turkmenistan | |
| Iran | Uganda | |
| Israel | Ukraine | |
| Jamaica | Uruguay | |
| Jordan | Uzbekistan | |
| Kazakhstan | Venezuela | |
| Kenya | Vietnam | |
| Kyrgyzstan | Yemen | |
| Laos | Zambia | |
| Lebanon | Zimbabwe | |
| Macedonia | | |
| Madagascar | | |