Macroeconomic Determinants of Urbanization

C. Tudor (SNR: 2029809)

MSc Thesis

Tilburg University

Sustainable Development Track

Supervisor:
B.J.A.M. van Groezen

26/06/2020

Words: 12,627

As the world continues to develop, it is faced with an increased pace of urbanization which places a noticeable strain on a country's resources, particularly for less developed nations. In order to understand and manage these changes, a regression model is proposed to discern the main drivers of urbanization. An instrumental variable approach with lagged regressors is used to solve endogeneity issues. Further, two different truncation conditions allow for analysis of specifically the countries experiencing the largest changes in urbanization, so as to analyze how the relative importance of the determinants changes across stages of development. Overall, basic results are in line with theoretical expectations, but what comes across is the surprising finding that in less developed nations, improvements in the agricultural factors of production have a bigger impact on urbanization than improvements in human capital (through e.g. education). The results of the paper ultimately indicate a dire need for efforts into urban planning, without which the development of growing nations will be significantly hindered.

Table of Contents

1. Introduction

The process of urbanization, while at first glance might seem a desirable step towards the development of any country, could in fact promulgate a range of negative effects, which if a nation is not ready for, may hinder the sustainable long-term growth of a healthy city and populace. For example, accelerated urbanization with limited infrastructure, a lack of jobs, or underdeveloped human capital[1] can prevent the urban center from attaining its status as an engine of growth, and instead become a drag on the country's resources. As Brueckner (2011) discovers for the case of Africa, "increases in the urbanization rate had a significant negative average effect on GDP per capita growth". With this in mind, understanding the causes of urbanization may provide pertinent policy information to manage the flow of people and allow cities to grow sustainably, so as to allow the country itself to benefit and develop. As such, the following research question is proposed:

What are the key macroeconomic factors responsible for the increase in urbanization over the last 50 years across the world? How does the relative importance of these factors change based on the level of urbanization within a country?

In the following section of the paper, I review the relevant literature to help the reader gain a background understanding of urbanization. This firstly includes the history of urbanization, which has evolved dramatically over the last 500 years at an exponential rate. Next, I discuss how urbanization transpires, the way it is modeled, its root causes and other determinants which might impact the pace of its occurrence. Lastly, I review some methods used in analyzing this process, along with their strengths and limitations, in order to motivate my own approach.

In the methodology section, I begin by discussing a framework to understand the decision of rural to urban migration. After this, I present and summarize the data used in the study, as well as delving into how the analysis will be conducted. Specifically, this takes the form of an initial naïve regression, from which statistical success is reviewed, after which an instrumental variable with lagged regressors is taken in order to further improve the statistical accuracy of the study. In order to answer the second part of the research question, I also truncate the data at two different levels of urbanization change to see how the values of the determinants change across this dimension. In the results section, the findings of the paper are presented and analyzed.

In the last section, the results are summarized into conclusions and used towards making policy prescriptions. Briefly, in line with theoretical predictions, I discover that the largest changes in urbanization occur due to increases in wealth, accompanied by a shift away from an agricultural based economy. Further, I show that that for less developed countries, improvements in the agricultural factors

[1] Specifically in skilled services providers (doctors, lawyers, public servants such as urban planners, etc.)

of production have a larger impact on urbanization than improvements in human capital (e.g. education). Lastly, the flaws and the merits of the study at hand are assessed in order to suggest potential improvements and future avenues of inquiry.

Overall, urbanization can be seen as a vital step in the development of any country, especially as it makes the transition away from an agricultural based economy. While some countries have made this shift smoothly, others are struggling, and will continue to struggle as urbanization increases and more pressure is exerted on developing cities' limited resources. In order to understand and manage these changes, a wide scope study is called for, which will ideally shed some light on the issues at hand.

2. Literature Review

2.1 History of Urbanization

The term "urbanization" can loosely be understood to refer to the migration of people from an area with more dispersed locations of living to one where the density of residents is higher. Many classify this migration as a rural to urban one. However, the definition of what an urban center constitutes (particularly in the number of residents, but also in the chief occupation of its residents) varies across countries and governments. Nevertheless, for the purpose of an economic study, urbanization may refer to an area where there are "advantages gained from an urban location", which "include proximity to a market, labor supply, good communications, and financial and commercial services such as auditing, stock broking, advertising, investment, etc." In other words, if there are supranormal returns gained simply from an increased density of living, we may understand that the area is experiencing urban economies; as such, it can be classified as an urban location. As Becker (2007) further points out, "higher densities also mean reduced costs of infrastructure and public service provision" as well as raising "returns to differentiation of labor, and hence the acquisition of a refined set of skills, rather than broad but shallow knowledge" (as the latter would be the case in a rural setting).

With this broader definition in mind, one may look back and notice that urbanization has been a process churning onwards to varying degrees throughout history. Michaels, Rauch, Redding (2012) point out that "the share of the world's population living in cities grew from less than one tenth in 1300, to around one sixth in 1900 and to more than one half today", undoubtedly forcing humanity to adapt and evolve immensely in order to accommodate these lifestyle changes. Specifically, the highest degrees of urbanization, i.e. areas in which returns from urban economies have peaked relative to their time period, are most likely to have occurred at times of global hegemony. That is, for example, Babylon, Xi'an, Athens, or Rome during their respective ruling periods. And while these urban centers might have produced miracles for their time, the process of urbanization could not become widespread in ancient or

medieval times, simply due to technological constraints. On one hand, a large part of the population needed to be engaged in agricultural labor to provide a stable food supply. And on the other hand, even when modest technological improvements lowered the amount of labor required, the size of cities themselves was constrained through technology by e.g. logistical or sanitary issues. Ultimately however, with the dawn of the industrial age, major improvements in our technology allowed for the sustained increase in the size and number of cities, something persevering and accelerating into the present day.

In more recent times, urbanization has had a mixed progress over the last 50 or so years: while developing countries have seen an increase in the share of their urban populations, some developed countries have experienced a decrease. As Glaeser, Henderson (2017) summarize, "between 2000 and 2020, the United Nations projects a total increase in urban population of 1.48 billion, and of that 1.35 billion will come from less developed regions." As such, we can understand that the most drastic urbanization changes will occur in the areas least fit to deal with it (due to resource constraints). Further, the way in this change comes about is twofold: increases in urban populations occur both because of new children being born in an urban location, but also due to rural urban migration. "Chen, Valente, and Zlotnik (1998) estimate that migration and population reclassification account for about 40% of all developing country urban population growth from the 1950s through the 1980s." In some special cases such as China (due to the one-child policy in cities), "migration accounted for 74% to 80% of urban population growth […] between 1978 and 2000, and that percentage appears to be increasing."

The main challenge of this type of growth is the fact that whereas new children born in urban settings already have a home and means of living in the city, many of the in-migrants from rural areas do not already have an established place of living, and will find it increasingly difficult to settle. Governments must become aware of the pace of urbanization in their respective countries, as well as what are the determinants for a higher or lower inflow, in order to cope with these changes accordingly. This is the main motivation for the study at hand: ascertaining the main drivers of urbanization in the "development" period of a country in order to facilitate and ease the transition.

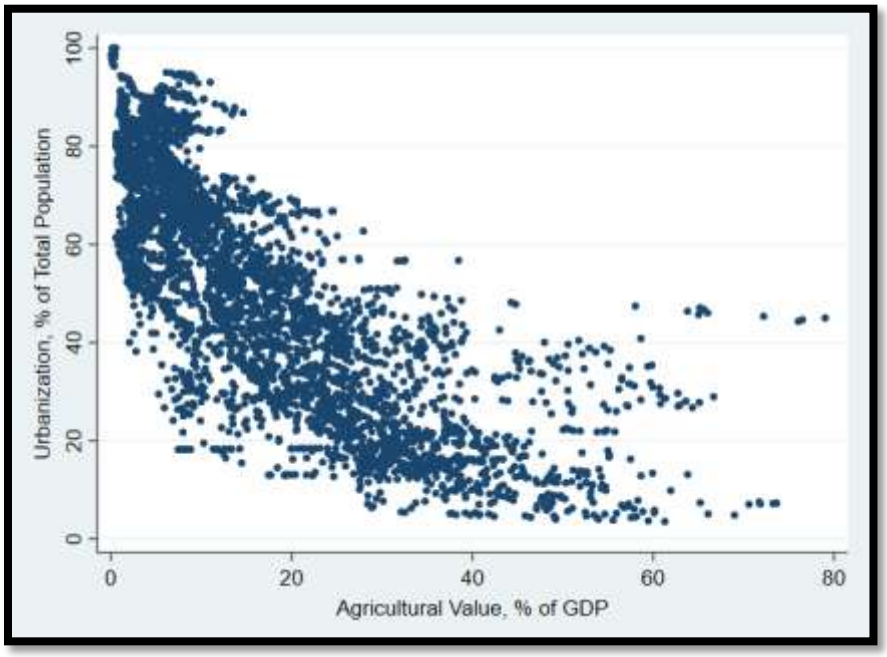
2.2 Models and Determinants of Urbanization

One of the biggest changes that began with the formation of cities was a shift in vocations from farming to manufacturing or service providers (whether that is a blacksmith centuries ago, or a banker in the modern world). As such, a key characteristic of rising urban populations can be understood to be a shift from an agricultural to a manufacturing economy, something Michaels et al. (2012) show: "as other developing countries continue to undergo structural transformation away from agriculture, our findings suggest that it will exert a powerful influence on future patterns of urbanization." This is something that can clearly be seen in Figure 1, showing the relation between urbanization and agricultural value in the

data of the study. As a country becomes less occupied with agricultural labor (in terms of contribution to its GDP), it naturally shifts away into urban vocations:

Figure 1 – Urbanization vs. Agricultural Value

(Source: World Bank Data)



With this in mind, while a simple viewpoint may hold that the chief determinant of urbanization is a shift away from agriculture, the way in which this happens should be analyzed in more detail. In particular, the rate at which structural transformations occur in the economy depends on many factors: for example, relative wages, access to technology, education, and infrastructure all have an influence on the agricultural labor share and thus the rate of urbanization. Therefore, it can be seen that even though the reason urbanization increases is due to a structural shift away from agriculture, the way in which this structural shift comes about depends on many other factors.

One way in which this phenomenon was first explained by W. Arthur Lewis (1954, 1955) was through a model of a dualistic economy composed of two sectors: a modern industrialist sector and a traditional agricultural sector. In this framework, "labor migrates across sectors to equate expected utility from each activity" and "economic growth occurs by reinvestment of modern sector profits and drawing workers from the traditional to the modern sector." As such, an implied assumption of this model is that reinvestments from the supranormal gains in the modern sector allow for technological improvements which decrease the amount of labor required in the traditional sector. Ultimately, according to the model, real wages begin rising only at the point where periods of surplus labor end and a standard neoclassical economy is perpetuated in both sectors. While some parts of the model seem reasonable, new evidence shows that "real wages are rising rapidly in growing areas of China and Southeast Asia today, even

though a vast number of low income workers remain in rural areas". One of the ways the model fails to account for this is possibly through the fact that cross-country migration is mostly of an urban nature, i.e. new migrants into a country are more likely to settle in an urban rather than a rural area. Nevertheless, this is a highly country-specific quantity which is hard to model, and more likely better observed on a case by case basis.

Returning to isolated single-country models, one way in which Sabot (1979) attempts to explain the above-mentioned controversy is by introducing education credentials into the framework. This results in the formation of segmented labor sub-markets which then give rise to a "queuing" pattern of migration. "During periods of rapid formal sector employment growth and restricted supply of fresh urban graduates, credentials' requirements decline, providing incentives for migration, especially when rural secondary schooling is widespread. Under the reverse circumstances, migration will be limited, even if average income disparities between urban and rural areas are increasing."

Further, the presence of an informal labor sector and the extent to which it is widespread also plays a role in the migration decision. Todaro (1970) and Harris and Todaro (1969; synthesized in Blomqvist 1978) show that in the presence of an informal sector in cities, migration from rural areas occurs until the ex ante expected utility gain is zero. This is based on the fact that potential benefits from finding a formal sector job will offset losses experienced due to unemployment or employment in the informal sector. Ultimately, this results in wages being highest in the formal sector, while living standards will be lowest for those unemployed or working in the informal sector.

Lastly, another set of models emphasize the fact that this rural-urban migration is highly dependent on age, "with young adults more likely to move than older adults or children" (Becker and Morrison, 1993; Becker and Grewe, 1996). They are able to formulate several relevant conclusions:

- Attractive aspects of an urban settings are strongest for "young, prime age people, especially men"
- Increased rural prosperity is associated with rising education resulting in greater migration rates as secondary school graduates are more likely to make the move
- Increased rural prosperity also affects the rate of migration as the costs of urban migration and job search are more easily handled

Putting everything together, we may understand that the shift away from agricultural jobs and consequential increase in urbanization occurs mostly based on the (expected) differences in payment between the two locations. Increased educational attainment in the rural setting improves the likelihood of a migration decision (based on the increased potential of finding a job in the formal urban sector). Further, the perceived relative sizes of the formal and informal sectors in a city also affect the migration decision as a relatively larger formal sector is likely to attract more migrants, whereas an informal sector

dominated city does not provide as high of an incentive to move (since the expected utility differential is smaller). Something that hasn't been specifically commented on, which I believe also has a tremendous impact is an increase in the technology of rural factors of production. As improved technology allows for less human labor to produce the same amount, we may understand that at least a part of the rural populace will experience obsolescence. This, coupled with the high birth rates in rural areas of developing countries, combines to create an incentive for urban migration based simply on the fact that there is less access to a farming vocation. Understanding the process of urban migration from a macro perspective may allow for policy advice meant to accommodate this drastic lifestyle change in developing economies.

2.3 Methods of Analysis

A large part of the literature on urbanization promotes a micro-approach, with these studies often narrowing down the analysis to a short time span and an isolated case study of a city, region, or country (e.g. Braimoh (2007) – Lagos, Nigeria; Chen (2015) – Central China; You (2017) – 30 Chinese Megacities, Zhang (2018) – 285 Chinese Cities, Fan (2015) – Mongolian Plateau, etc.). While this approach may provide useful information applicable in the short term to the area in question, it lacks the wider scope necessary to understand the common factors that push a country towards an urban transition.

To understand these root causes, the way in which I plan to approach the study is similar to Davis, Henderson (2002) or Hofman, Wan (2013), i.e. conducting a cross-country regression analysis from a macro perspective. Taking this wider approach, it should be possible to observe and determine the common factors in a nation's development, which push it to become more urbanized. As these papers have done, taking note of the technological level of a country is necessary to correctly pinpoint the drivers of urbanization. However, their measures of technology are limited to the quality of factors of production (i.e. various measures of industrialization), which overlook the more recent transformations in personal technology culture. In my opinion, as more and more rurally located citizens gain access to e.g. mobile or broadband connections, this will act as a strong motivator to pursue an urban relocation. Controlling for these factors allows for a better dissection of the channels which push urbanization. Ultimately, with this understanding, policy recommendations can be made to manage the flow of new urban residents. As Henderson (2003) discovers, there is in fact a "best degree of urban concentration" to be attained and straying from this can be costly in terms of productivity growth.

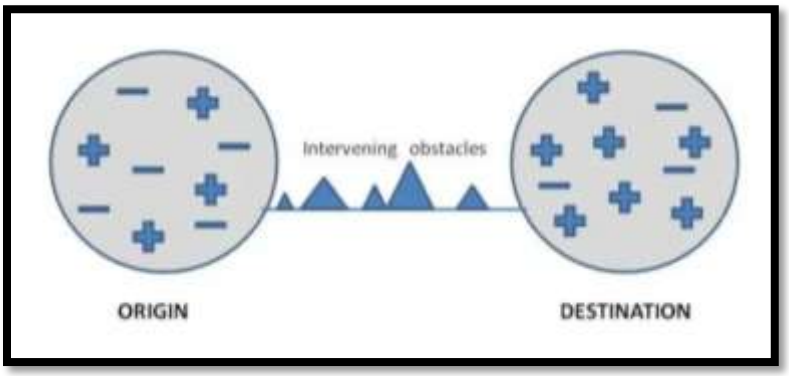
3. Methodology

3.1 Process Analysis

Before delving into the paper's approach, it is useful to assess the mechanisms by which macroeconomic changes to urbanization transpire. What motivates a particular individual or household to make the transition from a rural lifestyle to an urban one? While some models have approached this question through the assumption of homo economicus perfectly optimizing its discounted stream of lifetime income in order to maximize expected utility, this grossly fails to account for the inherent uncertainty associated with any migration decision. A more general framework, purported by Everett Lee in 1966, acknowledges the highly case specific nature of individual migration decisions. His idea was that there are multiple factors, perceived to different levels by each individual, called push and pull factors, which influence and ultimately determine the migration decision of an individual or household. While this was generally applied to cross-country migrations, it can also be applied to a within country migration from a rural to an urban setting.

Put simply, push factors are negative influences, which (in the absence of any positive effects) would prevent an individual from locating themselves in the area in question. For example, push factors in a rural location would be a lack of access to infrastructure, a lack of access to education, etc. On the flipside, pull factors are positive influences, which (in the absence of any negative effects) would motivate an individual to locate themselves in the area in question. As an example, pull factors towards an urban location would be improved access to technology, healthcare, jobs, etc. Lastly, the migration decision is also associated with significant costs such as finding a job, moving or purchasing a new home, illustrated in the theory by intervening obstacles. Ultimately, the decision to migrate boils down to the weight of individually perceived factors, as well as the extent of intervening obstacles, such that a potential migrant makes the move if the benefits outweigh the costs, given his current situation and uncertainty pertaining to the perceived factors.

Figure 2 – Everett Lee's Theory of Migration



In order to formulate a successful model of regression and understand what the determinants influencing rural-urban migration are, I have summarized a list of push and pull factors for both locations. While there certainly are cases where these factors are not necessarily applicable (such as rural areas in developed economies which are in principle as well-kept as urban ones), generally seen, there are major differences in a rural and urban lifestyle which can be commented on.

Table 1 – Summary of Potential Push/Pull Factors

Rural Location		Urban Location	
Push Factor	**Pull Factor**	**Push Factor**	**Pull Factor**
Relatively worse quality of education or access to education	Family located in rural area	Congestion / Overcrowding and associated issues	Relatively better infrastructure
Fewer job opportunities and less diversity of vocations	Simpler lifestyle may be viewed as desirable by some	Relatively higher costs of living	Access to healthcare
Worse infrastructure or complete lack of infrastructure		Likelihood of being forced to accept a job in the informal sector	Access to services
Poverty			Transport Costs Reduction

While the abovementioned factors pertain more towards an individual migration decision, they can be proxied for and translated into a macro perspective through the choice of independent variables in the regression model. In particular, education and infrastructure metrics are most easily identified, whereas poverty or job access has to be proxied for via indirect industrialization metrics as will be seen.

The data being used comes from the World Bank and it is comprised of 122 countries over a 40 year period. The choice of countries was based on the reliability of data they have provided, as well as the size of the country itself. With reference to the latter, the size of a country plays an important role in analyzing changes in the process of urbanization, simply due to the fact that small countries such as Lichtenstein or Monaco are already heavily urbanized due to their political context. As such, in order to explore the macro determinants of urbanization, these smaller countries do not provide a representative picture of the development and accompanying changes in urbanization, and are therefore excluded. Below is a summary of the data used in the analysis:

DATA: World Bank Data, panel of 122 large countries from 1975 to 2015

Dependent Variable:

- Share of Urban Population, % of total population (World Bank data 1975-2015)

Independent Variables (World Bank data 1975-2015)

- Country Specific Fixed Effects (dummy variable for each country – country ID)
- Growth Metrics
    - GDP per Capita (PPP adjusted $)
    - GDP Growth (annual %)
    - Population Growth (annual %)
- Human Capital Metrics
    - Gov. Expenditure on Education (% of GDP)
    - Primary School Enrollment (% of net)
- Industrialization Metrics
    - Agricultural Land (% of land area)
    - Agricultural Value Added (% of GDP)
    - Fertilizer Consumption (kilograms per hectare of arable land)
    - Cereal Yield (kilograms per hectare of arable land)
    - Rural Access to Electricity (% of Rural Population)

A summary of descriptive statistics for the data can be seen in Table 2:

Table 2 – Descriptive Statistics

Variable	Nr. Observations	Mean	Std. Dev.	Min // Max	Data Coverage
Urbanization	4,987	52.5	23.5	3.5 // 100	99.7%
Agricultural Land	4,713	41.7	21.0	1.6 // 90.6	94.2%
Agricultural Value	4,012	16.0	14.2	0.1 // 79.0	80.2%
Cereal Yield	4,713	2704.4	1833.0	132.2 // 21865.5	94.2%
Enrollment (Primary Schooling)	2,531	85.3	17.6	12.6 // 100	50.6%
Electricity Access (Rural)	2,676	70.6	38.1	0 // 100	53.5%
Fertilizer	1,603	232.7	1070.1	0 // 34040	32.0%
GDP Growth	4,398	3.7	6.0	-64.0 // 123.1	87.9%
GDP per Capita	4,449	7757.4	12932.2	94.6 // 102913.5	88.9%
Gov. Expenditure on Education	2,321	4.4	1.9	0 // 44.3	46.4%
Population Growth	4,941	1.7	1.5	-6.8 // 17.5	98.8%

It can immediately be noticed that some variables are not fully covered in the data, i.e. there are missing data points. This is due to the fact that some countries did not keep records of the variables of interest in early periods of the data (around 1975). While this might seem as a troubling issue, it can be seen that besides Fertilizer Consumption, Government Expenditure on Education, Primary Enrollment, and Electricity Access, all other variables have over an 80% representation. Overall, the data representation in the study is 77.8%, so no more than around 22% of data points are missing. Considering, the extent of the study, both in terms of time and number of variables, this is an acceptable representation.

From the above variables, a few are worth discussing in more detail, in order to understand their link to changes in urbanization. Perhaps most important are the indirect measures of technology adoption. Rural Access to Electricity, Fertilizer Consumption and Cereal Yield are all indicative of the level of farming technology present in the country. Put simply, the more advanced farming technology becomes in a particular country, the less human labor is required to generate the same output, ceteris paribus. As such, if there is access to electricity, if more fertilizer is used, or if the yield per hectare is higher (indicating usage of new technology such as farm machines), fewer people are needed to farm a specific area, thus more people are likely to make the rural-urban transition.

Next, Agricultural Value Added is an important indicator of the extent to which a country is engaged in rural agricultural labor. Changes in the agricultural value added come about due to either a lesser reliance on the country's own food supply (potentially due to increased food imports) or increased activity in other labor sectors within the country, minimizing the relative value added by agriculture. Regardless of the case, this indicator presents a direct correlation to the size of the agricultural labor sector and changes in its value refer to a corresponding change in the amount of people engaged in agricultural labor. Overall, if the agricultural value decreases, this can be seen as increased obsolescence of at least part of the agricultural labor pool. This effect will then play a role in increasing urbanization by virtue of lessened job access in the rural setting.

Further, primary school enrollment rates are a sensibly accurate proxy for rural access to education. This relies on the idea that children living in an urban location have a higher likelihood to have ensured primary education, due to the better organized provision of schooling in cities. As such, further changes in the primary school enrollment rate most likely refer to improved access to education in a rural area. As theory suggests, a more educated young population has a higher likelihood of making the rural to urban transition.

Taking a step back, there are expected to be problems regarding endogeneity among the variables above. In particular, there is an extensive literature discussion regarding the muddled relation of urbanization and GDP growth. In my view, this is the case of a feedback loop, whereby in initial periods,

GDP growth causes urbanization which in and of itself has the potential to amplify the GDP growth of future periods. The way in which other papers have dealt with this was ala Henderson (2003): using an instrumental variable approach with lagged regressors to shift the observed relation. In other words, we would see how GDP growth from past periods impacts current urbanization changes, absent of GDP growth effects caused by urbanization in the current period. This should in principle remove the effect of the feedback loop; however the effectiveness of the method in discerning the correct relation is questionable, specifically because the optimal time lag is hard to determine. Perhaps the ideal time lag is the length of the loop, i.e. the time taken for urbanization changes in the current period to impact future GDP growth. This is something that requires testing in the data to find the optimal time lag.

Overall, the proposed study provides an in-depth discussion of the causes of urbanization, via dissecting the general shift of developing economies from agricultural to knowledge focused. While certain problems are expected to arise in the study, their analysis and consequent discussion can provide a clearer picture of the urbanization process, especially in terms of causation. Lastly, the addition of various technology metrics is a novel point of the study, and controlling for population growth effects; this might reconcile the faster pace of urbanization experienced by current developing economies as opposed to developing economies decades or centuries ago.

3.2 Regression Models and Approach

The simple regression model put forth is as follows:

$$\begin{aligned} Urbanization_{i,t} &= \beta_0 + \beta_1 \ln(GDPCapita)_{i,t} + \beta_2 GDPGrowth_{i,t} + \beta_3 PopGrowth_{i,t} \\ &+ \beta_4 GovExpendEdu_{i,t} + \beta_5 Enrollment_{i,t} + \beta_6 AgriValue_{i,t} + \beta_7 AgriLand_{i,t} \\ &+ \beta_8 Electricity_{i,t} + \beta_9 \ln FertilizerCons)_{i,t} + \beta_{10} \ln(CerealYield)_{i,t} + \gamma_i + u_{i,t} \end{aligned}$$

Where

$$u_{i,t} = Idiosyncratic\ Error$$

$$\gamma_i = Country\ Specific\ Fixed\ Effects$$

The number of independent variables will be varied in the analysis to observe changes in the significance and size of the coefficients. Furthermore, two truncated regression models are ran for the countries with the biggest observed changes in urbanization, i.e. the data set will also be constrained to those countries which experienced a (1) 10% and (2) 20% or more change in urbanization in the period 1975 to 2015. As such, the conditions for data truncation are as follows:

$$urbanization_{i,2015} - urbanization_{i,1975} > 10\%$$

$$urbanization_{i,2015} - urbanization_{i,1975} > 20\%$$

While the above conditions do not capture variability above the threshold within the period of 1975 to 2015, we are ultimately interested in resolved and final changes, as such, the approach still holds. A complete table of the exact countries that are part of the data set under different truncation conditions can be found in the appendix. Furthermore, a summary of descriptive data statistics for the truncated data sets can also be found in the appendix.

Lastly, in order to resolve potential endogeneity issues, an instrumental lagged variables model is presented as such:

$$\begin{aligned} Urbanization_{i,t} &= \beta_0 + \beta_1 \ln(GDPCapita)_{i,t-n} + \beta_2 GDPGrowth_{i,t-n} + \beta_3 PopGrowth_{i,t} \\ &+ \beta_4 GovExpendEdu_{i,t-n} + \beta_5 Enrollment_{i,t-n} + \beta_6 AgriValue_{i,t} + \beta_7 AgriLand_{i,t} \\ &+ \beta_8 Electricity_{i,t} + \beta_9 \ln(FertilizerCons)_{i,t} + \beta_{10} \ln(CerealYield)_{i,t} + \gamma_i + u_{i,t} \end{aligned}$$

Where n is the number of periods that a variable is lagged. The number of periods n should be the same across all lagged dependent variables which may need a similar time to make a causal impact. For example, the variable for primary school enrollment will potentially require a longer lag than the variable for GDP growth, because the resolved time needed for primary schooling to impact urbanization may be longer than that needed for GDP growth to impact urbanization. Finally, once the optimal time lag is determined, the dependent variables are tested for endogeneity issues using a two stage least squares estimation procedure.

4. Results and Analysis

4.1 Data Establishment

Firstly, I perform a correlation analysis to explore possible multicolinearity issues between the regressors:

Table 3 – Correlation Analysis for Regressors

	Electricity Access (Rural)	Agricultural Land	Agricultural Value	Cereal Yield	Fertilizer Usage	GDP Growth	GDP per Capita	Gov. Expenditure on Education	Population Growth	Primary School Enrollment
Electricity Access (Rural)	1	//	//	//	//	//	//	//	//	//
Agricultural Land	-0.0811	1	//	//	//	//	//	//	//	//
Agricultural Value	-0.7892	0.0756	1	//	//	//	//	//	//	//
Cereal Yield	0.5881	-0.0678	-0.5379	1		//	//	//	//	//
Fertilizer Usage	0.0936	-0.1566	-0.0991	0.1665	1	//	//	//	//	//
GDP Growth	-0.2774	0.0055	0.3164	-0.2113	0.1285	1	//	//	//	//
GDP per Capita	0.4536	-0.2163	-0.5465	0.5849	0.1666	-0.2486	1	//	//	//
Gov. Expenditure on Education	0.2900	-0.0982	-0.2699	0.1733	-0.0023	-0.2293	0.4273	1	//	//
Population Growth	-0.4927	-0.1687	0.4058	-0.2295	0.2458	0.3106	-0.0045	-0.1602	1	//
Primary School Enrollment	0.6946	-0.0594	-0.6981	0.4177	0.0478	-0.1736	0.3530	0.2434	-0.4040	1

We may notice that the most problematic relationship occurs between Agricultural Value Added and Electricity Access (Rural), with a correlation of -0.789. From this we may understand that as a country gains increased access to electricity in its rural environment, this jointly occurs with a decrease in the agricultural value added. In other words, as the rural environment is industrialized (and labor is focused in other sectors of the economy), the agricultural value added decreases. This joint correlation should be kept in mind when making statements about the causation of urbanization through these variables.

To further explore the issue of multicolinearity I check the variance inflation factors for the variables. This is first checked in the simple OLS regression (OLS (3)) as well as the instrumental variable regression (IV (3)), both of which will be later presented.

Table 4 – Variance Inflation Factors

Variable	Variance Inflation Factor OLS (3)	Variance Inflation Factor IV (3)
GDP per Capita	5.68	4.96
Agricultural Value	5.23	5.17
Rural Access to Electricity	4.57	4.78
Fertilizer Consumption	3.21	3.01
Cereal Yield	3.04	2.78
Primary School Enrollment	2.30	4.22
Population Growth	1.78	1.75
Government Expenditure on Education	1.28	1.34
GDP Growth	1.23	1.16
Agricultural Land	1.16	1.18

From this we may understand that the variance of particularly GDP per Capita and Agricultural Value are significantly exacerbated by multicolinearity issues. It can also be seen that this effect is lowered when an instrumental variable approach is implemented, however it still remains large. This issue should be kept in mind when commenting both on the size of the coefficient as well as the significance of these two variables.

Next, in order to ascertain the reliability of OLS regressions, I perform the Im Pesaran and Shin (IPS) unit root test:

Table 5 – Im Pesaran and Shin Unit Root Tests

	IPS URT	First Differences IPS URT
GDP per Capita	Non-stationary	Stationary
GDP Growth	Stationary	//
Population Growth	Stationary	//
Gov. Expenditure on Education	Insufficient Obs.	Insufficient Obs.
Primary School Enrollment	Insufficient Obs.	Insufficient Obs.
Rural Access to Electricity	Insufficient Obs.	Insufficient Obs.
Fertilizer Consumption	Non-stationary	Stationary
Cereal Yield	Non-stationary	Stationary
Agricultural Land	Non-Stationary	Stationary
Agricultural Value	Stationary	//

As it can be seen, all indicators are stationary in first differences or no differences, besides Government Expenditure on Education, Primary School Enrollment and Rural Access to Electricity. The test was unable to find unit roots for the latter three due to insufficient observations. However, given the stationarity of the other variables similar to these 3, I make the assumption that these 3 would also be

proven to be stationary if enough observations were present. As such, it can be concluded that the standard assumptions for asymptotic analysis remain valid and we can rely on the OLS results (not taking into account potential endogeneity issues).

Before delving into the regression results, the choice of fixed effects over random effects should be motivated. The assumption of fixed effects is that country specific effects are correlated with the independent variables. Given the presence of variables such as agricultural land (highly specific to each country and not subject to much in-country variation), it seems natural to make use of fixed effects. This was tested by utilizing a Hausman test. The value of chi2 in the test is 129, with a probability > chi2 being 0.0, thus allowing us to reject the null hypothesis that the random effects model is more appropriate. As such, a fixed effects model will be used.

4.2 OLS Regressions

Moving on, the results of the initial regression can be seen in the following table. This regression does not truncate the data, so all 122 countries are studied. Standard errors are in the parenthesis and two ** represent a significance level of 95% and one * of 90%.

Table 6 – OLS Regression for Complete Data Set

(Urbanization)	OLS (1)	OLS (2)	OLS (3)
ln(GDP per Capita)	8.2112 ** (0.2921)	6.5258 ** (0.5027)	2.6545 ** (0.6107)
GDP Growth	-0.1270 (0.0797)	0.05511 (0.1186)	0.0947 (0.1105)
Population Growth	0.2734 (0.2662)	1.4507 ** (0.3740)	2.2008 ** (0.3558)
Gov. Expenditure on Education	0.0950 (0.2145)	1.0202 ** (0.3011)	1.5176 ** (0.2844)
Primary School Enrollment	0.3626 ** (0.02985)	0.3671 ** (0.05682)	0.2006 ** (0.05511)
Rural Access to Electricity		0.1519 ** (0.02780)	0.09405 ** (0.02642)
ln(Fertilizer Consumption)		0.03794 (0.4680)	-0.04041 (0.4477)
ln(Cereal Yield)		-1.5412 (1.1332)	0.02833 (1.0753)
Agricultural Land			0.04169 * (0.02221)
Agricultural Value			-0.9578 ** (0.09096)
Country Specific F.E.	YES	YES	YES
Countries	122	122	122
Adjusted R-Squared	0.6558	0.6583	0.7037

In OLS (1) I introduce only the indicators for general population metrics such as wealth, population growth and education. Here we can see that the biggest determinant of urbanization is GDP per capita, with a 1% point increase in GDP per capita leading to an 8.2% point increase in urbanization. Further, Primary School Enrollment also has a significant effect with a 1% point increase in enrollment leading to a 0.36% point increase in urbanization.

In OLS (2) I add the indicators referring to rural industrialization, i.e. Rural Access to Electricity, Fertilizer Consumption and Cereal Yield. It can be noted that the additional explanatory variables reduce the coefficient for GDP per capita, while significantly increasing the coefficient values of Population Growth and Government Expenditure on Education. Namely, a 1% point increase in Population Growth leads to a 1.45% point increase in urbanization, indicating that the largest population increases occur in cities. Furthermore, a 1% point increase in Government Expenditure on Education also increases urbanization by 1% point. Primary school enrollment retains its significance and size of the coefficient and of the additional variables added in OLS (2), only rural access to electricity is significant such that a 1% point increase in access leads to a 0.15% point increase in urbanization.

Lastly, in OLS (3), I control for the size of the agricultural sector in a country by adding the variables of Agricultural Land and Agricultural Value. Once again, the coefficient value of GDP per capita is significantly decreased such that a 1% point increase in GDP per Capita results in a 2.6% point increase in urbanization (down from 8.2% in OLS (1)). Further, the coefficient values of Population Growth and Gov. Expenditure on Education increase, while those of Primary School Enrollment and Rural Access to Electricity decrease. Most notably, Agricultural Value has a strong significant effect, such that a 1% point increase in Agricultural Value leads to a 1% point decrease in urbanization. This indicates the natural result that the more a country is engaged in agricultural labor, the less urban associated economic activity exists. As such, a lowering of the agricultural value added within a country leads to an increase in urban activity and therefore urbanization, ceteris paribus.

The first truncated data results show the OLS regression for countries experiencing a 10%+ change in urbanization over the period 1975-2015. This truncation reduced the data set from 122 countries to 86 countries. The introduction of additional explanatory variables is performed in the same fashion as the previous regression.

Table 7 – OLS Regression for Truncated Data Set (10% +)

(Urbanization)	OLS (4)	OLS (5)	OLS (6)
ln(GDP per Capita)	10.0702**	8.1151**	4.8053 **
	(0.3723)	(0.6540)	(0.7390)
GDP Growth	-0.1376	0.1613	0.2455 *
	(0.09228)	(0.1460)	(0.1251)
Population Growth	0.18533	1.3922**	2.1756 **
	(0.2928)	(0.4175)	(0.3626)
Gov. Expenditure on Education	-0.4033 *	0.2109	0.6017 *
	(0.2347)	(0.3573)	(0.3087)
Primary School Enrollment	0.2492 **	0.2347 **	0.09104
	(0.03494)	(0.07081)	(0.06198)
Rural Access to Electricity		0.1684**	0.06286 *
		(0.03382)	(0.02993)
ln(Fertilizer Consumption)		-1.0149 *	0.02425
		(0.5829)	(0.5110)
ln(Cereal Yield)		0.7889	3.2697 **
		(1.3686)	(1.1913)
Agricultural Land			0.1896 **
			(0.02766)
Agricultural Value			-0.9862 **
			(0.09227)
Country Specific F.E.	YES	YES	YES
Countries	86	86	86
Adjusted R-Squared	0.6617	0.6677	0.7569

In OLS (4) we may notice similar results to OLS (1), except that the coefficient for Government Expenditure on Education is negative and significant. This is most likely an issue of the data due to a lack of explanatory variables and should be taken with a grain of salt. In OLS (5) a similar issue arises with fertilizer consumption and should also be disregarded. In OLS (6), new results emerge contributing an interesting idea. Most notably, the coefficient for Cereal Yield indicates that a 1% point increase in Yield results in a 3.3% point increase in urbanization. This could be explained by the fact that a higher cereal yield indicates the usage of better agricultural practices or tools. As such, improved factors of production corresponding to increased productivity in the agricultural sector allows for the same amount to be farmed, but utilizing less labor, ceteris paribus. The obsoleteness of at least a part of the agricultural labor pushes agents to find new employment opportunities, most of which are located in cities, therefore increasing urbanization.

The last truncated data set looks at countries experiencing a 20% or more increase in urbanization from 1975 to 2015. This reduces the data set from 122 countries to only 44 countries.

Table 8 – OLS Regression for Truncated Data Set (20% +)

(Urbanization)	OLS (7)	OLS (8)	OLS (9)
ln(GDP per Capita)	9.6349 **	5.4900 **	5.1640 **
	(0.5223)	(1.0124)	(0.9815)
GDP Growth	-0.08388	0.1042	0.02686
	(0.1099)	(0.2051)	(0.1722)
Population Growth	-0.3935	1.1307	4.1689 **
	(0.4847)	(0.8498)	(0.7691)
Gov. Expenditure on Education	-0.2820	0.6574	0.05727
	(0.3089)	(0.6118)	(0.5151)
Primary School Enrollment	0.1878 **	0.2850 **	-0.0007565
	(0.03819)	(0.09611)	(0.09762)
Rural Access to Electricity		0.09305 **	0.1270 **
		(0.04366)	(0.03791)
ln(Fertilizer Consumption)		2.6805 **	2.3148 **
		(0.7717)	(0.6417)
ln(Cereal Yield)		-0.3611	0.1468
		(1.5193)	(1.2813)
Agricultural Land			0.3308 **
			(0.03676)
Agricultural Value			-0.70006 **
			(0.1544)
Country Specific F.E.	YES	YES	YES
Countries	44	44	44
Adjusted R-Squared	0.6548	0.6737	0.7751

In OLS (7) the results are consistent with past regressions. In OLS (8), we may note the strong significant effect of Fertilizer consumption, such that a 1% point increase in Fertilizer usage leads to a 2.7% point increase in urbanization. This can once again be interpreted by understanding how it affects agricultural labor practices. Put simply, increased fertilizer usage can be seen as an improvement in the agricultural factors of production. As such, an improvement in productivity allows for the same amount to be farmed, utilizing less labor, ceteris paribus. Therefore, increased fertilizer usage makes at least a part of the agricultural labor sector become obsolete, pushing these agents to look for employment elsewhere, most likely in an urban location.

It can also be noticed in OLS (9) that the coefficient on Agricultural Land is positively significant, such that a 1% point increase in Agricultural Land leads to a 0.33% point increase in urbanization. This result can be dissected to further understand the reason for its occurrence. From the raw data, it can be seen that most countries have a relatively stable agricultural land, i.e. not subject to much change, most likely due to the fact that the best farming areas in a country have long been discovered and used. As such, further increases in agricultural land are most likely due to the opening of new farming areas which were previously unable to be farmed because of technological constraints. Therefore, the new agricultural land being utilized is relatively more industrialized compared to other

areas of the country. From this, we may understand that the addition of this new farming land allows for increased food production in the country with less average agricultural labor per area. As such, it contributes to making agricultural labor in other parts of the country obsolete which leads to an increase in urbanization through the previously explained mechanism.

4.3 IV Regressions

Moving on, in order to improve the accuracy of the study, I conduct a 2SLS regression with lagged variables instrumenting for the potential endogenous regressors. This refers to GDP per Capita, GDP Growth, Government Expenditure on Education and Primary School Enrollment. The other regressors are assumed not to be endogenous due to theoretical considerations of causal inference. In table 8 one may see first the implementation of the lagged variable in the complete data OLS regression (holding all other variables the same), followed by the p-value of the Wu-Hausman Endogeneity Test. The selected optimal lag is highlighted in yellow, prioritizing the exogeneity of a variable.

Table 9 – IV Optimal Lag Determination

	OLS Implementation (OLS (3))				**IV Test**	
	Coefficient	Std. Error	t	P>|t|	2SLS Wu-Hausman Endogeneity Test (p=	Exogenous?
ln(GDP per Capita) (4 Period Lag)	2.4443	0.5746	4.25	0.000	0.4252	YES **
ln(GDP per Capita) (5 Period Lag)	2.5723	0.5703	4.51	0.000	0.1345	YES **
ln(GDP per Capita) (6 Period Lag)	2.6196	0.5570	4.70	0.000	0.0492	NO
GDP Growth (4 Period Lag)	-0.2215	0.09874	-2.24	0.025	0.0279	NO
GDP Growth (5 Period Lag)	-0.1534	0.1018	-1.51	0.132	0.1461	YES **
GDP Growth (6 Period Lag)	-0.2945	0.1028	-2.87	0.004	0.0046	NO
Expenditure on Education (4 Period Lag)	0.7850	0.2882	2.72	0.007	0.9803	YES **
Expenditure on Education (5 Period Lag)	0.6526	0.2909	2.24	0.025	0.7327	YES **
Expenditure on Education (6 Period Lag)	0.5213	0.2952	1.77	0.078	0.0464	NO
Primary School Enrollment (5 Period Lag)	0.1626	0.05077	3.20	0.001	0.000	NO
Primary School Enrollment (10 Period Lag)	0.2128	0.05597	3.80	0.000	0.0001	NO
Primary School Enrollment (12 Period Lag)	0.1751	0.05275	3.32	0.001	0.0001	NO
Primary School Enrollment (15 Period Lag)	0.2027	0.05090	3.98	0.000	0.0002	NO

It can be noted that the first 3 variables (GDP per Capita, GDP Growth, and Government Expenditure on Education) are all exogenous with a 5 period lag. For Primary School Enrollment however, none of the tested lags allow for exogeneity, as such, I have resolved to choose the most logical time lag. A period of 12 years is perhaps the best time lag for this variable due to its nature. Considering the fact that most children entre primary school at the age of 6 or 7, a 12 year lag period is the time taken for these children to reach the age of adulthood and be allowed to move out of their home, potentially making a rural to urban transition. In table 9, we may see the first implementation of the instrumental variable results in the complete data set (with no truncation).

Table 10 – IV Results for Complete Data

(Urbanization)	IV (1)	IV (2)	IV (3)
ln(GDP per Capita) (5 Period Lag)	7.7240 ** (0.3644)	5.0340 ** (0.5770)	2.1069** (0.7049)
GDP Growth (5 Period Lag)	-0.1272 (0.09651)	-0.05259 (0.1385)	-0.006422 (0.1317)
Population Growth	1.4681 ** (0.3407)	2.2108 ** (0.5146)	2.4817** (0.4975)
Gov. Expenditure on Education (5 Period Lag)	0.08754 (0.2567)	0.7034 * (0.3457)	0.8890** (0.3291)
Primary School Enrollment (12 Period Lag)	0.3752 ** (0.03207)	0.4209** (0.06116)	0.2513 ** (0.06233)
Rural Access to Electricity		0.1290 ** (0.03699)	0.09199 ** (0.03552)
ln(Fertilizer Consumption)		-0.7567 (0.6009)	-0.5315 (0.5758)
ln(Cereal Yield)		1.2805 (1.2404)	2.3273 * (1.1960)
Agricultural Land			0.03482 (0.02549)
Agricultural Value			-0.8897** (0.1228)
Country Specific F.E.	YES	YES	YES
Countries	122	122	122
Adjusted R-Squared	0.6697	0.6553	0.6897

We may see that most results are in line with the OLS regressions, with particularly the coefficient on GDP Growth remaining small and insignificant despite the instrumental variable approach. Both Government Expenditure on Education and Primary School Enrollment provide significant coefficient estimates with the added time lag, however their absolute values remain relatively small. Fertilizer Consumption is a negative and insignificant result, but cereal yield becomes positively significant at a 10% confidence level, having a strong impact of 2.3% point increase in urbanization for a

1% point increase in cereal yield. Next, looking at the truncated data set, we may see some notable changes in the coefficient values compared to the non-truncated IV regression.

Table 11 – IV Results for Truncated Data Set (10% +)

(Urbanization)	IV (4)	IV (5)	IV (6)
ln(GDP per Capita) (5 Period Lag)	9.1335 ** (0.4441)	6.3241** (0.6981)	3.9577 ** (0.8276)
GDP Growth (5 Period Lag)	-0.1881 * (0.1117)	-0.1547 (0.1591)	-0.04211 (0.1409)
Population Growth	1.4623 ** (0.3579)	2.4780 ** (0.5616)	2.9256 ** (0.5070)
Gov. Expenditure on Education (5 Period Lag)	-0.1462 (0.2752)	0.2587 (0.3931)	0.2738 (0.3471)
Primary School Enrollment (12 Period Lag)	0.2694 ** (0.03578)	0.3582 ** (0.06712)	0.1563 ** (0.06345)
Rural Access to Electricity		0.1072 ** (0.04224)	0.03849 (0.03790)
ln(Fertilizer Consumption)		-0.9779 (0.7015)	0.1972 (0.6374)
ln(Cereal Yield)		3.3865 ** (1.3930)	4.6746 ** (1.2463)
Agricultural Land			0.16412 ** (0.03174)
Agricultural Value			-0.9527 ** (0.1280)
Country Specific F.E.	YES	YES	YES
Countries	86	86	86
Adjusted R-Squared	0.6602	0.6691	0.7425

The coefficients of GDP per Capita and Population Growth further increase while remaining significant, indicating that their effects are more pronounced in countries experiencing higher changes in their levels of urbanization. Further, the coefficient of Cereal Yield has roughly doubled, and the coefficient on Agricultural Land has become positively significant, despite retaining a relatively small absolute value. Overall, the 10% truncation exacerbates the relative values of the coefficients for the ones retaining their significance, but also makes some variables lose their significance, indicating that perhaps their effect on urbanization is not as strong. As a side note, the latter point may also indicate that depending on the current level of urbanization, the chief determinants impacting this change may vary. In other words, at lower levels of urbanization, variables that improve the agricultural factors of production may have a larger impact, whereas at higher levels of urbanization, variables linked to human capital (such as education), begin playing a more important role. Next we may see the instrumental variable regressions for the truncated data set at a level of 20% or more change in urbanization.

Table 12 – IV Results for Truncated Data Set (20% +)

(Urbanization)	IV (7)	IV (8)	IV (9)
ln(GDP per Capita) (5 Period Lag)	9.6900 ** (0.5536)	4.1601 ** (0.9365)	4.5365 ** (1.0513)
GDP Growth (5 Period Lag)	-0.07244 (0.1419)	-0.1352 (0.2354)	-0.2032 (0.2094)
Population Growth	1.3971** (0.5688)	0.6044 (0.8431)	3.144 ** (0.8582)
Gov. Expenditure on Education (5 Period Lag)	0.14304 (0.3801)	2.1075 ** (0.7382)	0.8853 (0.6641)
Primary School Enrollment (12 Period Lag)	0.1969 ** (0.03890)	0.08690 (0.08541)	0.04734 (0.07843)
Rural Access to Electricity		0.1341** (0.04400)	0.1258 ** (0.03960)
ln(Fertilizer Consumption)		2.1154 ** (0.8167)	2.2308 ** (0.7099)
ln(Cereal Yield)		3.2104 ** (1.6249)	2.3986 * (1.4447)
Agricultural Land			0.2905 ** (0.04653)
Agricultural Value			-0.5139 ** (0.1732)
Country Specific F.E.	YES	YES	YES
Countries	44	44	44
Adjusted R-Squared	0.6779	0.7444	0.8072

Looking at the last instrumental variable regression, we may see that Government Expenditure on Education and Primary School Enrollment both lose their significance. Perhaps the most interesting result here is the positive significance (with a relatively large coefficient) of both Fertilizer Consumption and Cereal Yield. Through the mechanisms explained previously, these seem to have a large impact on changes in urbanization, of roughly a 2.3% point increase for a 1% point increase in their respective values. Next, similar to the 20% truncated OLS regression, Agricultural Land attains a higher than previous positive significance of around 0.3.

4.4 Overview of Variable Effects

As a last point, I provide a summary of the range of coefficients that the variables take, in order to facilitate an easier discussion and answer the research question. The table refers only to values for the complete model specification, i.e. including all explanatory variables, that is OLS 3, 6, 9 and IV 3, 6, 9. In parenthesis we may see the data truncation applied to the specific regression, taking values of 0% +, i.e. the complete data set, as well as 10% + / 20% + urbanization change.

Table 13 – Range of Indicator Coefficients for Complete Model Specifications

Indicator	OLS Regressions		IV Regressions		Effect
	Min Value	Max Value	Min Value	Max Value	
ln(GDP per Capita)	2.7 ** (0% +)	5.2 ** (20% +)	2.1 ** (0% +)	4.5 ** (20% +)	Major, Positive, Significant
GDP Growth	0.03 (20% +)	0.25 ** (10% +)	-0.20 (20% +)	-0.006 (0% +)	Minor, Insignificant
Population Growth	2.2 ** (0% +)	4.2 ** (20% +)	2.5 ** (0% +)	3.1 ** (20% +)	Major, Positive, Significant
Gov. Expenditure on Education	0.06 (20% +)	1.52 ** (0% +)	0.3 (10% +)	0.9 ** (0% +)	Minor, Positive, Significant
Primary School Enrollment	0.0 (20% +)	0.2 ** (0% +)	0.05 (20% +)	0.25 ** (0% +)	Minor, Positive Significant
Rural Access to Electricity	0.06 * (10% +)	0.13 ** (20% +)	0.04 (10% +)	0.13 ** (20% +)	Minor, Positive, Significant
ln(Fertilizer Consumption)	-0.04 (0% +)	2.3 ** (20% +)	-0.53 (0% +)	2.2 ** (20% +)	Major, Positive, Significant (depends on truncation)
ln(Cereal Yield)	0.03 (0% +)	3.3 ** (10% +)	2.3 * (0% +)	4.7 ** (10% +)	Major, Positive, Significant (depends on truncation)
Agricultural land	0.04 * (0% +)	0.33 ** (20% +)	0.03 (0% +)	0.29 ** (20% +)	Minor, Positive, Significant (depends on truncation)
Agricultural Value	-0.99 ** (10% +)	-0.70 ** (20% +)	-0.95 ** (10% +)	-0.51 ** (20% +)	Major, Negative, Significant

Firstly, GDP per Capita can be seen as a strong determinant of urbanization across all regressions. While it may seem that its effect is endogenous, the fact that it retains a high level of positive significance even under the instrumental variable approach shows that controlling for endogeneity issues does not affect its importance. This is a straightforward result in line with theoretical expectations: increased wealth across the population accelerates the rural to urban transition by reducing the costs of migration and making urban centers more attractive. Assuming that the largest GDP per Capita increases occur in cities, this factor also contributes to increasing wages and thus attracting international migrants which are most likely to settle in an urban location.

On the other hand, GDP Growth is consistently insignificant and displaying a relatively large variation in its coefficient values depending on model specification and truncation. This does not necessarily come as a surprise given the fact that each country is subject to its own macroeconomic factors (conductive of booms and recessions), while the general trend of urbanization continues to grow

despite these varying quantities. Overall, we may discard GDP Growth as a strong explanatory variable in terms of its effect on urbanization.

Population Growth maintains a strongly positive and significant effect across all regressions. This indicates that the largest growth in population occurs in cities thus contributing a relatively large amount to changes in urbanization.

Government Expenditure on Education provides a generally positive and significant estimate, albeit subject to large variation depending on the model specification and truncation. We may generally see that narrowing the data set to the countries experiencing the largest changes in urbanization lowers both the absolute value and significance of the coefficient. This may be explained by the following process: firstly, looking at the countries with the highest urbanization changes generally means focusing on less developed economies (because developed economies already have high urbanization levels and thus do not experience much further change). Within these countries, the link between Government Expenditure in any matter and the resolved changes in that matter is severely weakened by forces such as corruption or lower productivity. As such, some money is essentially lost in the process, decreasing the real effect of any expenditure program.

Primary School Enrollment has a high level of significance for the non-truncated data set, but has a lower absolute value as well as significance when truncating the data. This may occur because even though gaining an education in a developing nation is undoubtedly valuable, the ability of agents to make the rural to urban transition might be lessened by the capacity of underdeveloped cities. As such, despite the positive impact of gaining an education, it cannot be fully utilized unless the cities themselves develop in order to accommodate growth. Overall, there is a lower effect of education metrics as data is truncated. This indicates that the effect of education on urbanization is less pronounced in developing nations, a rather surprising result.

Rural Access to Electricity maintains its significance across most regressions, but has a very small absolute value effect. While this in theory is a good indicator for the level of industrialization of a country in its rural environment, we may understand that gaining access to electricity does not necessarily have a strong impact on the improvement of agricultural factors of production, perhaps due to the fact that many agricultural practices still do not make full use of the industrial technology available, especially in developing nations.

Fertilizer Consumption and Cereal Yield both see an increase in their significance as well as absolute value of their coefficients as the data is truncated. This provides a rather straightforward

interpretation: when referring to developing nations, increased productivity in the agricultural sector lowers the amount of human labor needed, thus incentivizing at least a part of the agents to pursue an urban transition due to lowered job opportunity in the rural environment. The reason this effect becomes more pronounced through truncation (when looking at developing nations), is because developed nations already have a higher amount of Fertilizer Consumption and Cereal Yield due to their established agricultural practices already being more productive.

Agricultural Land sees an increase in both its significance and value of its coefficient when the data is truncated. While the effect on urbanization remains small (around 0.3% increase for a 1% increase in agricultural land), it is once again indicative of the importance of improved agricultural productivity in facilitating an urban transition. As previously mentioned, increases in agricultural land are most likely associated with improvements in productivity for the reason that new farming land that is being used is presumably more industrialized. This occurs because a country was most likely already making use of its best farming areas (naturally farming the best areas first), as such, the usage of new farming areas is accompanied by improved technology to farm this area.

Lastly, Agricultural Value is a good estimator for the relative importance of the agricultural sector to other sectors. As agricultural value is diminished, this indicates that labor (or higher productivity) is being focused in other sectors of the economy, thus increasing the value of these sectors relative to the agricultural one. As such, due to a restructuring of labor from the rural, agricultural areas to urban ones, this variable can be seen as main driver of urbanization. In Table 14, we may see a summary of how the coefficients change due to truncation of the data. The truncation essentially allows us to focus on countries that experience larger changes in urbanization, i.e. observing a larger change in the development level of a country. As such, this is a useful way of showing how the relative impact of the determinants changes based on the current level of development.

Table 14 – Effect of Data Truncation on Coefficients

Indicator	Coefficient Change due to Truncation	Impact on Urbanization
GDP per Capita	Increase	High
GDP Growth	No significant change	Low
Population Growth	Increase	High
Gov. Expenditure on Education	Decrease	Medium
Primary School Enrollment	Decrease	Low
Rural Access to Electricity	Increase	Low
Fertilizer Consumption	Increase	High
Cereal Yield	Increase	High
Agricultural Land	Increase	Low
Agricultural Value	Decrease (in absolute value)	Medium

Overall, we may notice that the effects of GDP per Capita, Population Growth, Rural Access to Electricity, Fertilizer Consumption, Cereal Yield, and Agricultural Land all see an increase in their coefficient value as the data is truncated. This indicates that the importance of these factors in promoting urbanization is more pronounced in developing nations rather than developed ones. Conversely, the coefficients of Gov. Expenditure on Education and Primary School Enrollment see a decrease in their coefficient values as the data is truncated. This may show that the link between education and urbanization is not as strong in developing nations for a number of reasons. Most importantly perhaps, it has to do with the nature of the cities in developing nations. Due to the fact that the cities themselves are not as developed, there is a lack of job opportunities even if the potentially migrating workforce becomes more educated. As such, the relative impact on urbanization of getting an education in a developing nation is diminished compared to a developed nation.

Lastly, we may recall from the Variance Inflation Factor table (Table 4) that the coefficient of GDP per Capita, Agricultural Value and Rural Access Electricity, might all be artificially exacerbated. Even if that may be the case, given the high absolute values and significance of GDP per Capita and Agricultural Value, it can be concluded that despite the presence of this effect, their impact on urbanization remains noticeable. With reference to Rural Electricity Access, its absolute value effect is small (and not always significant), as such this can be seen as a low-impact determinant of urbanization.

5. Conclusions

5.1 Summary of Key Results

Perhaps the most useful starting point for this section is to refer back to the research question:

What are the key macroeconomic factors responsible for the increase in urbanization over the last 50 years across the world? How does the relative importance of these factors change based on the level of urbanization within a country?

In order to answer the first part of the question, I refer to the complete data set regressions, after which the second part is answered by observing how the coefficients change under different truncation conditions. Overall, even though the reliability of the regressions is questionable at times, what is perhaps more valid is the change in the value of the coefficients based on truncation. Truncating the data essentially allows seeing the differences in effects based on the level of development of a nation. That is not to say that the level of urbanization is a direct indicator of the level of development: as some papers have shown, there is in fact a best degree of urbanization based on country-specific factors, i.e. some countries might be optimally developed at 70% urbanization, while others at 90%. What *is* a good

indicator of development is the *change* in the level of urbanization: as a country increases its level of urbanization, we may understand it becomes more developed, getting closer to their optimal level. With this in mind, I formulate the following conclusion points: point (1) to answer the first part of the question, points (2) and (3) to answer the second part.

1. ***GDP per Capita**, **Population Growth**, and **Agricultural Value Added** are the main drivers of urbanization under both the OLS and IV approach. These retain a high absolute value and significance across all regressions. These results are in line with theoretical expectations.*
2. *When truncating the data, i.e. focusing on countries experiencing larger changes in their respective urbanization levels, **Cereal Yield** and **Fertilizer Consumption** are shown to be strong drivers of urbanization under both OLS and IV approaches. **Agricultural Land** and **Rural Electricity Access** also have a significant impact, but remain small in absolute value. From this, we may understand that increased industrialization (improvements in the agricultural factors of production) have a larger impact on urbanization in less developed countries.*
3. *When truncating the data, we may notice that education metrics have a lower impact on urbanization in less developed countries. **Government Expenditure Education** and **Primary School Enrollment** both decrease in their absolute value and significance as the data is truncated. This is indicates a lesser connection between improvements in human capital and changes in urbanization in less developed countries, most likely due to fewer formal job opportunities in their respective urban locations.*

5.2 Policy Recommendations

Given the main conclusions formulated above, several pertinent policy recommendations can be made, particularly in regards to the pace of urban development. Overall, while it naturally varies on a case by case basis, the main problem currently experienced by developing nations can be boiled down to an inability to properly convert the labor force from an agricultural, rural-based one to an industrial/service, urban dwelling one. This issue is exacerbated from multiple angles:

1. Improved agricultural factors of production make at least part of the rural population obsolete, forcing agents to search for new job opportunities.
2. Improved human capital via education promotes a rural to urban transition, but this is constrained by the number and type of job vacancies in cities: too few good jobs.
3. Increased wealth in developing countries allows for the expansion of urban centers, but this most likely happens in a non-sustainable fashion, with too little money invested in the infrastructure (physical and otherwise), necessary to allow cities to scale in a manner conductive of future growth.

With this in mind, developing countries should carefully note the extent to which their urban centers can accommodate continually increasing populations, both due to migration and childbirth. With reference to the latter, as urban centers have the highest population density and countries continue to become more urbanized, the largest portion of population growth will occur in urban locations, placing a continually increasing strain on the limited resources of developing cities. Next, as agricultural productivity increases, a large part of the currently rural population will have to face obsolescence, and therefore search for new job opportunities, most likely in cities. As such, depending on the case, implementing the correct urban planning for a city should take priority to further industrializing the rural environment. In other words, resources should be focused first on the development of cities, and only second on the development of the rural environment. If this is not done, as parts of the rural population are made obsolete, but not given any alternative job opportunities, unemployment will increase or employment in the informal labor market will increase. Needless to say, both of these alternatives have a negative impact on further growth prospects, unemployment due to human capital atrophy and informal labor employment due to a lowering of the tax base and reduced future opportunities for the agent. Employment in the informal labor market is also associated with a reduction in living standards, furthering the negative impact on the city's populace.

Next, while education has long been upheld as one of the main drivers of development, its real effect on economic growth in any country is tied down to the opportunities available. Put otherwise, an education is undoubtedly useful, but the extent to which it makes a positive impact is constrained by what an agent can actually do with that education. As such, while improved human capital is a key factor to the development of a country, its effectiveness is based on how well the labor market can accommodate and capture it. In order to do this, developing countries should firstly prioritize the correct growth of their urban centers to ensure a higher capacity of human capital usage, rather than continually invest in human capital growth which cannot be properly utilized (at their current level of development).

While these comments essentially argue for relatively lower investment in the rural environment, also the part of a country most likely found in the worst living standards, this is a forward-looking approach which will allow for the sustainable development of the largest part of the population over the long-term. Sacrificing investment in the rural parts in order to ensure that the rest of the country will be able to accommodate the changes that this investment would bring is necessary in order to solve the problems of accelerated over-urbanization before they occur.

5.3 Further Improvements of the Study

Perhaps the biggest problem of the study is the potential endogeneity of regressors. GDP per Capita and GDP Growth are both show to be exogenous, but not at a very high level of significance.

Worse off, primary school enrollment maintains its endogeneity across all time lags. These issues put at least part of the results of the study into question, especially with regards to the direction of the effect on urbanization. In order to solve or improve this situation, one way to approach it would be to find better instrumental variables, however, that is not a simple matter, as the wealth of literature up to this date does not manage to come up with a convincing answer.

Another issue of the study is potential omitted variables. Needless to say, the complex process of urbanization is not determined solely due to the variables in the study. Rather, many other variables can help better pinpoint the channels through which urbanization occurs. However, the main limitation of introducing more explanatory variables is their lack of availability across multiple countries and time periods. Naturally, with any wide-scope study, there must be some sacrifices in order to discern the general common trend. This is nevertheless a limitation which can be tackled and improved in the future, as higher quality data on other variables becomes available.

5.4 Closing Words

In summation, we have seen that the process of urbanization is determined through many channels, varying in strength based on the specific level of development within a country. Generally, in less developed countries, urbanization is more strongly affected by improvements in the agricultural factors of production rather than improvements in human capital, most likely due to the labor market conditions a country is faced with at that stage of development. In order for sustainable development to be achieved, governments must be acutely aware of their specific situation, for example the level of rural industrialization, and adjust their urban planning policy with this in mind.

Overall, as the world continues to improve both its technology and human capital, more and more countries are going to be faced with increasing rural to urban migration. If this is not accounted for and managed, urban centers will be plagued by insufficient vacancies, an expanding informal labor sector, worsening average living standards, and other problems. All of these have the potential to deny cities their status as an engine of growth, and instead make them a drag on the country's resources, hindering long-term, sustainable growth. In 2018, 55% of the world's population was living in cities, as such, there is still much room left to decide whether the urbanization of the next large part of the populace will occur in a better fashion. In the next decades, urban planning will be a decisive point as to whether the currently less developed countries will make a smooth transition from agricultural based to modern industries. Ultimately, the successful incorporation of roughly half the world's population into the modern world will hinge on how governments and prefectures manage the flow of people from rural to urban environments.

6. Bibliography

1. Becker, M. Charles; *Urbanization and Rural-Urban Migration*, International Handbook of Development Economics, April 2007
2. Becker, M. Charles; Morrison, R. Andrew; *Observational equivalence in the modeling of African labor markets and rural-urban migration,* World Development 21 (4), 535-554, 1993
3. Becker, M. Charles; Grewe, Christopher; *Cohort specific rural-urban migration in Africa*, Journal of African Economies 5 (2), 228-270, 1996
4. Braimoh, K. Ademola; Onishi, Takashi; *Spatial determinants of urban land use change in Lagos, Nigeria*, Land Use Policy 24, 502-515, 2007
5. Brueckner, Markus; *Economic growth, size of the agricultural sector, and urbanization in Africa*, Journal of Urban Economics 71, 26-36, 2012
6. Chen, N.; Valente, P.; Zlotnick, H.; *What do we know about recent trends in urbanization*, Migration, Urbanization and Development: new directions and issues, New York United Nations Population Fund, pp. 59-88, 1998
7. Chen, Yingbiao; Chang, Kang-tsung; Han, Fuzhuang; Karacsonyi, David; Qian, Qinglan; *Investigating urbanization and its spatial determinants in the central districts of Guangzhou, China*, Habitat International 51, 59-69, 2016
8. Davis, C. James; Henderson, J. Vernon; *Evidence on the political economy of the urbanization process*, Journal of Urban Economics 53, 98-125, 2003
9. Everett, S. Lee; *A Theory of Migration*, Demography, Vol. 3, No. 1, pp. 47-57, 1966
10. Fan, Peilei; Chen, Jiquan; John, Ranjeet; *Urbanization and environmental change during the economic transition on the Mongolian Plateau: Hohhot and Ulaanbaatar*, Environmental Research 144, 96-112, 2016
11. Glaeser, Edward; Henderson, J. Vernon; *Urban Economics for the developing world: an introduction*, Journal of Urban Economics 98, 1-5, 2017
12. Harris, R. John; Todaro, P. Michael, *Migration, unemployment and development: a two-sector analysis*, American Economic Review 60 (1), 126-142, 1970
13. Henderson, J. Vernon; *The urbanization process and economic growth: the so-what question*, Journal of Economic Growth 8, 47-71, 2003

14. Hofmann, Anett; Wan, Guanghua; *Determinants of Urbanization*, Asian Development Bank working paper series, no. 355, July 2013
15. Lewis, W. Arthur; *Economic development with unlimited supplies of labor*, The Manchester School 22(2), 139-191, 1954
16. Lewis, W. Arthur; *The Theory of Economic Growth*, London: George Allen & Unwin, 1955
17. Michaels, Guy; Rauch, Ferdinand; Redding, J. Stephen; *Urbanization and Structural Transformation*, CEPR Discussion Paper No. DP7016, October 2008
18. Sabot, H. Richard; *Economic Development and Urban Migration: Tanzania, 1900-1971*, Oxford: Clarendon Press, 1979
19. Todaro, P. Michael; *A model of labor, migration, and urban development in less developed countries*, American Economic Review 59 (1), 138-148, 1969
20. You, Heyuan; Yang, Xuefeng; *Urban expansion in 30 megacities of China: categorizing the driving force profiles to inform the urbanization policy*, Land Use Policy 68, 531-551, 2017
21. Zhang, Wenjing; Wang, Mark Yaolin; *Spatial-temporal characteristics and determinants of land urbanization quality in China: Evidence from 285 prefecture-level cities*, Sustainable Cities and Society 38, 70-79, 2018

7. Appendix

Link to Complete Data Set (Google Drive Download)

https://drive.google.com/file/d/17IZXASf8cKkuAWZzlYf-CtkQv5fVYzd3/view?usp=sharing

Link to 10% + Urbanization Truncated Data Set (Google Drive Download)

https://drive.google.com/file/d/15Wg7islMoyC9h5zOLxRoAVTeQ4928J7x/view?usp=sharing

Link to 20% + Urbanization Truncated Data Set (Google Drive Download)

https://drive.google.com/file/d/1v1JG9qhMHQNA5HJO-8vq2YJG6xbxaZjC/view?usp=sharing

Table 15 – Descriptive Statistics, 10%+ Truncated Data Set

Variable	Nr. Observations	Mean	Std. Dev.	Min // Max	Data Coverage
Urbanization	3,526	50.1	22.8	4.0 // 100	100%
Agricultural Land	3,389	41.8	20.6	2.5 // 90.6	96.1%
Agricultural Value	2,916	17.2	13.8	0.1 // 79.0	82.6%
Cereal Yield	3,389	2,346	1,612	132.3 // 21,865.5	96.1%
Enrollment (Primary Schooling)	1,788	84.2	17.9	12.6 // 100	50.7%
Electricity Access (Rural)	1,824	64.8	38.4	0 // 100	51.7%
Fertilizer	1,119	252.1	1,261.9	0 // 34,040	31.7%
GDP Growth	3,101	3.9	6.1	-62.0 // 123.1	87.9%
GDP per Capita	3,178	5409.5	10,398.3	94.6 // 102913.5	90.1%
Gov. Expenditure on Education	1,544	4.2	2.1	0 // 44.3	43.8%
Population Growth	3,522	1.9	1.6	-6.8 // 17.5	99.9%

Table 16 – Descriptive Statistics, 20%+ Truncated Data Set

Variable	Nr. Observations	Mean	Std. Dev.	Min // Max	Data Coverage
Urbanization	1,830	49.1	19.3	6.3 // 90.2	99.2%
Agricultural Land	1,797	41.8	18.8	3.4 // 90.6	97.4%
Agricultural Value	1,612	16.5	11.0	0.8 // 79	87.4%
Cereal Yield	1,797	2,164	1,589.3	132.3 // 13,956	97.4%
Enrollment (Primary Schooling)	999	82.5	18.1	12.6 // 100	54.1%
Electricity Access (Rural)	932	61.9	38.3	0 // 100	50.5%
Fertilizer	575	205.4	327	0.3 // 2,182.5	31.2%
GDP Growth	1,704	4.2	5.2	-42.4 // 49.4	92.4%
GDP per Capita	1,701	3821.2	6,453.7	128.1 // 57,644	92.2%
Gov. Expenditure on Education	806	4.1	1.6	0 // 10.7	43.7%
Population Growth	1,830	2.0	1.2	-1.9 // 7.9	99.2%

Table 17 – Specific Changes in Urbanization by Country (for Truncated Data Sets)

Ordered Position of Urbanization Change	Country	Change in Urbanization 1975-2015, % point
1	Botswana	55.271
2	Oman	43.137
3	China	38.1
4	Malaysia	36.561
5	Costa Rica	35.513
6	Indonesia	33.996
7	Korea, Rep.	33.601
8	Dominican Republic	32.837
9	Jordan	32.606
10	Turkey	32.025
11	Haiti	32.023
12	Bhutan	30.802
13	Mauritania	30.52
14	Algeria	30.518
15	El Salvador	28.149
16	Nigeria	28.058
17	Iran, Islamic Rep.	27.611
18	Cameroon	27.286
19	Bolivia	27.079
20	Netherlands	26.99
21	Belarus	26.567
22	Brazil	24.981
23	Saudi Arabia	24.831
24	Albania	24.722
25	Bangladesh	24.472
26	Ghana	24.037
27	Thailand	23.936
28	Mali	23.747
29	Mozambique	23.531
30	Namibia	23.249
31	Morocco	23.095
32	Honduras	23.055
33	Portugal	22.735
34	Paraguay	21.769

35	Burkina Faso	21.183
36	Lebanon	21.128
37	Ecuador	21.035
38	Tanzania	20.471
39	Tunisia	20.468
40	Yemen, Rep.	19.933
41	Cyprus	19.669
42	Mongolia	19.574
43	Colombia	19.458
44	Liberia	19.438
45	Somalia	17.747
46	Cambodia	17.711
47	Panama	17.658
48	Finland	17.394
49	Congo, Dem. Rep.	16.919
50	South Africa	16.719
51	Mexico	16.526
52	Libya	16.485
53	Bulgaria	16.428
54	Peru	15.896
55	Bosnia and Herzegovina	15.885
56	Japan	15.665
57	Guinea	15.626
58	Vietnam	15.027
59	Uganda	15.02
60	Nepal	13.727
61	Sierra Leone	13.1
62	Rwanda	13.006
63	Norway	12.904
64	Guatemala	12.846
65	Kenya	12.744
66	Cuba	12.672
67	Zimbabwe	12.512
68	Venezuela, RB	12.368
69	Croatia	12.291
70	Senegal	12.169
71	Uzbekistan	11.663
72	Uruguay	11.658
73	Lithuania	11.545
74	India	11.445
75	Slovenia	11.407

76	Afghanistan	11.162
77	Greece	11.106
78	Romania	11.056
79	Philippines	10.724
80	Ukraine	10.672
81	Kuwait	10.616
82	Argentina	10.534
83	Qatar	10.036
84	Spain	10.032
85	Ethiopia	9.968
86	Pakistan	9.685