Urbanization and Development: Is Latin America and the Caribbean Different from the Rest of the World?[1]

Preliminary draft

Mark Roberts, Brian Blankespoor, Chandan Deuskar and Benjamin Stewart

The World Bank
1818 H Street, NW
Washington, D.C.
20433, USA

This version: 31st October, 2016

Abstract: Two long-established stylized facts in the urban and development economics literatures are that: (*a*) a country's level of economic development is strongly positively correlated with its level of urbanization; and (*b*) a country's level of urbanization is strongly negatively correlated with the size of its agricultural sector. However, countries in the Latin America and Caribbean region appear to depart significantly from the rest of the world with regards to these two basic relationships. In particular, while Latin American countries appear to be significantly more urbanized than predicted based on these global relationships, Caribbean countries appear significantly less urbanized. Analysis involving cross-country comparisons of urbanization levels are, however, undermined by systematic measurement errors arising from differences in how countries define their urban areas. In this paper, we re-examine whether LAC countries differ from the rest of the world when it comes to the basic stylized facts of urbanization, development and structural transformation. To do this, we make use of two alternative methodologies for the consistent definition of urban areas across countries – the Agglomeration Index methodology and a methodology based on the identification of dense spatially contiguous clusters of population. Both of these methodologies rely on globally gridded population data sets as input. There exist several such data sets and so the paper also assesses the robustness of its findings to the choice of input population layer.

Keywords: Latin America and Caribbean, urbanization, development, measurement error
JEL codes: O18, O47, C21

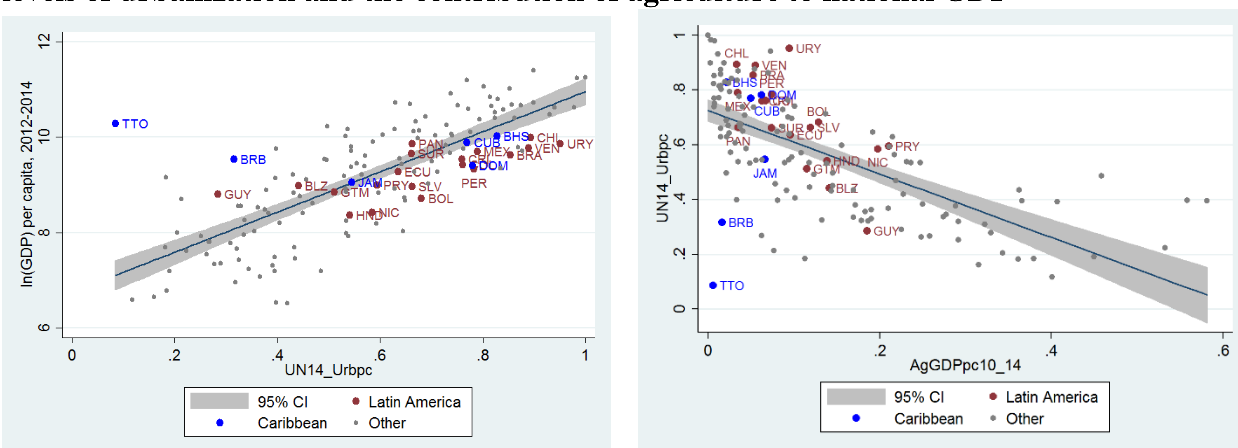
---

[1] This paper has been prepared as input into the LCR "Cities and Productivity" Report (P158837), which is a joint product of the World Bank's LCR Chief Economist's Office and its Social, Urban, Rural and Resilience Global Practice. The authors are extremely grateful to Katie McWilliams, Siobhan Murray and Jane Park for the inputs they have provided into the preparation of the paper. Corresponding author: Mark Roberts (mroberts1@worldbank.org).

1. Introduction

Urbanization, economic development and structural transformation have long been viewed as inextricably linked processes. In particular, in models of urbanization and economic growth, technological progress leads to an aggregate re-allocation of labor from agricultural to non-agricultural activities, which, in turn, drives both urbanization and economic development (see, *inter alia*, Henderson and Wang, 2005; Michaels, Rauch and Redding, 2012; Henderson, Storeygard and Roberts, 2013). Consistent with these models, cross-country data exhibit two key stylized facts: ***(a)*** the existence of a strong positive correlation between a country's level of development and its level of urbanization; and ***(b)*** the existence of a strong negative correlation between a country's level of urbanization and the relative importance of agriculture to its economy (see, *inter alia*, Davis and Henderson, 2003; World Bank, 2008; Ellis and Roberts, 2015).

Figure 1: Relationship between: (a) levels of economic development and urbanization; and (b) levels of urbanization and the contribution of agriculture to national GDP



The existence of these two key stylized facts is illustrated in Figure 1. In this figure, a country's level of urbanization is measured by the share of its population living in areas that its national government officially classifies as urban.[2, 3] However, from this figure, it is also clear that countries in the Latin American and Caribbean (LAC) region depart from the established global relationships that link urbanization, economic development and structural transformation in important ways. Thus, while more urbanized LAC countries also generally exhibit higher levels of development, Latin American countries tend to have significantly lower levels of GDP per capita than one would predict based on their levels of urbanization. At the same time, they also appear to be "over-urbanized" given the relative sizes of their agricultural sectors as measured by the share of agriculture in GDP. Conversely, for the Caribbean countries of Barbados, Trinidad and Tobago, and, to a much lesser extent, Jamaica, the opposite holds true. Hence, all three of these countries have significantly higher levels of GDP per capita than one would predict based on their urbanization

[2] In common with most studies, this data is taken from the United Nations' World Urbanization Prospects (2014 revision) database (https://esa.un.org/unpd/wup/). This database also provides the source of urban population data for the World Bank's World Development Indicators (WDI) (http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS).
[3] The Data Annex provides full details of the data used in this paper.

levels at the same time as they appear to be "under-urbanized" relative to other countries given the (small) sizes of their agricultural sectors.[4]

The fact that Latin American countries appear to be "over-urbanized" is consistent with a policy narrative that is emerging regarding the productivity performance of the region's cities. According to this narrative, these cities are failing to realize their full potential in terms of fueling national economic growth due to policy failures that have exacerbated congestion forces, thereby undermining their productivity performance (McKinsey Global Institute, 2011a, b).[5] Such apparent "over-urbanization" is also consistent with theories of urban bias, whereby policy distortions which skew the allocation of resources away from agriculture and towards manufacturing, spur higher urbanization than might have otherwise occurred (see, e.g., Davis and Henderson, 2003, and references therein). Particularly important in this context may be the trade policy distortions that were widespread across Latin America during the import substituting industrialization era of the 1960s and 1970s, which may continue to have residual impacts on levels of urbanization in the region today. Such policies have been linked in the literature to excessive urban primacy across Latin American countries with consequent adverse consequences for productivity and growth (Ads and Glaeser, 1995; Krugman and Elizondo, 1996).[6]

An alternative hypothesis to urban bias that may explain Latin America's "over-urbanization", not to mention the "under-urbanization" of certain Caribbean countries, is simply that this is an illusion of the data which is attributable to systematic biases in the measurement of levels of urbanization. In particular, in common with the vast majority of published research on urbanization and development, the stylized facts depicted in Figure 1, are based on data which depend on national definitions of urban and areas. However, as is well-known, these definitions vary widely from country-to-country. Furthermore, these variations have a strong regional component. Hence, while South Asian countries have generally tighter criteria for designating areas as urban, countries in LAC tend to have more relaxed criteria (World Bank, 2008; Ellis and Roberts, 2015). Such systematic biases in the data give rise to the possibility of "urban myths", which are artefacts of the data rather than a reflection of on-the-ground reality (Satterthwaite, 2007).

Given the above, this paper re-examines the basic empirical relationships between urbanization, development and structural transformation depicted in Figure 1 with a view to addressing the question of whether LAC countries depart significantly from these relationships. Specifically, the paper re-estimates these relationships using two alternative methods for the consistent definition of urban areas across countries that have been proposed in the literature. The first method – the Agglomeration Index (AI) – was originally proposed by Uchida and Nelson (2008) and delineates urban areas on the basis of population densities and travel times to large urban centers.[7] Meanwhile,

[4] The existence of these significant differences between LAC and non-LAC countries is confirmed by the regression results presented in Section 4.2 (see columns [1] – [3] of Table 2).

[5] These congestion forces arise from a number of sources including congestion of not only local transport infrastructure, but also of other forms of urban infrastructure and of land and housing markets.

[6] Whether or not policy distortions that favor urban over rural populations stimulate a higher level of urbanization will depend on how they affect the margin of choice for the representative rural-urban migrant. Hence, policies which, for example, re-distribute agricultural surpluses to urban elites are unlikely to affect a country's overall level of urbanization (Henderson *et al*., 2013).

[7] Uchida and Nelson built on earlier work by Chomitz, Buys and Thomas (2005) which identified population density and remoteness from large metropolitan areas as important dimensions along which to classify an area's degree of

the second method, which, for brevity, we shall refer to as the "cluster method", has been developed by Dijkstra and Poelman (2014) and relies on the identification of dense spatially contiguous clusters of population. Having re-estimated the relationships, the paper identifies whether LAC countries continue to exhibit significant differences of the sort portrayed in Figure 1. In doing so, we are able to assess the extent to which Latin America's apparent "over-urbanization" qualifies as one of Satterthwaite's urban myths.

In addition to re-examining the key basic stylized facts of urbanization, development and structural transformation, this paper also makes two additional contributions to the literature. First, by implementing both the AI and cluster methods to defining urban areas, the paper provides the first systematic comparison of results between these two methods using a global data set. Secondly, because both of these methods rely on globally gridded population data sets for their implementation and there exist several alternative such data sets, the paper also provides the first systematic investigation as to how the choice of input data affects the results obtained from these methods. In particular, the paper analyzes how the use of three main gridded population data sets influence the results obtained from these methods. These data sets are the *LandScan 2012* global gridded population data set that is produced by Oak Ridge National Laboratory, the *WorldPop* family of gridded population data sets that are produced by the University of Southampton, UK, and the *GHSL Global Population Grid* (GHS-Pop) that has recently been generated by the European Commission's Joint Research Center.[8, 9]

The structure of the remainder of the paper is as follows. Section 2 discusses how official definitions of urban areas in LAC countries compare to those in non-LAC countries and introduces both the AI and cluster methods for the consistent measurement of urbanization. Section 3 describes the data used to implement both the AI and cluster methods, focusing, in particular, on the input gridded population data. Section 4 presents the results from re-assessing LAC's position relative to the basic stylized facts of urbanization, development and structural transformation when urbanization is consistently measured. Finally, Section 5 concludes.

rurality (and, therefore, implicitly, its degree of urbanization). Chomitz *et al.* found that many small settlements in LAC that are officially classified as urban are actually embedded in an agriculturally based countryside.

[8] At the time of writing, GHS-Pop remains a pre-public release product available only as a beta test version.

[9] Other global gridded population data sets include CIESIN's GRUMP (Global Rural-Urban Mapping Project) Population Count Grid (http://sedac.ciesin.columbia.edu/data/set/grump-v1-population-count) and its Gridded Population of the World version 4 (GPW v 4) (http://beta.sedac.ciesin.columbia.edu/data/collection/gpw-v4) data sets. Various institutions and authors have also created gridded population data sets for individual countries or regions. These include a European population grid for 2006 produced by the Austrian Institute of Technology (http://www.efgs.info/data/) which covers EU28 and EFTA countries; and population grids for Haiti, Pakistan and Rwanda produced by the US Census Bureau as part of its international *Demobase* project (http://www.census.gov/population/international/data/mapping/demobase.html) .

2. Measuring Urbanization

2.1. *Official definitions of urban areas in LAC versus non-LAC countries*[10]

Most published research on urbanization and development relies on data which utilizes national definitions to measure cross-country differences in levels of urbanization, and it is this same data that underpins the stylized facts depicted in Figure 1. This data is usually taken from the United Nations' World Urbanization Prospects (WUP) database or, equivalently, the World Bank's World Development Indicators (WDI) database (see also footnote 2). As part of its database, the UN provides supporting documentation on how each of the 232 included countries defines its urban areas. Table 1 shows that 133 of these countries use one or more of four basic types of criteria to define their urban areas.[11] By far the most commonly used criterion is a minimum population size threshold (103 countries). The remaining three criteria are less frequently used. Hence, 30 countries include references to the availability of infrastructure and services (e.g. piped water or schools) or other "urban" characteristics in their definitions; 16 countries to the structure of the local economy, mostly referring to the presence or predominance of non-agricultural activity, and 11 countries to a minimum population density threshold. The remaining 99 countries do not make use of any explicitly stated criteria. Rather, they either simply list their urban areas by name, or state a designation of administrative units that constitutes cities.

Compared to the overall sample of WUP countries, a smaller share of the 40 LAC countries covered by the WUP database use a specific criterion or criteria to define urban areas (i.e. 42 percent of countries versus 57 percent). LAC countries are also less likely to employ a minimum population threshold in their definitions of urban than the sample overall (30 percent versus 44 percent). These differences between the LAC countries and the overall sample are mainly driven by the Caribbean countries. Conversely, LAC countries are more likely to include reference to the provision of infrastructure or services in their definition (i.e. 20 percent versus 13 percent). In this case, the difference between LAC and the global sample is driven by the Latin American countries.[12]

The use of more relaxed definitions of urban areas in LAC versus non-LAC countries is most evident from the minimum population threshold criterion, where this is used. Hence, among the countries that use this criterion, the mean threshold is 1,958 people for LAC as compared to 4,799 people globally. This mean difference is not, furthermore, driven by the small island nations of the Caribbean – the mean population thresholds for the Latin American and Caribbean countries are almost identical. The mean difference is also statistically significant at the 10 percent level in a simple one-sided two-sample t test where the alternative hypothesis is that the mean difference is negative (i.e. the mean for LAC is less than that for non-LAC countries). Likewise, a Mann-Whitney U test – which may be more appropriate given the large difference in size between the LAC and non-LAC samples and the absence of normality – rejects the null hypothesis that the LAC and non-

[10] This sub-section draws on a previous conference paper by two of the authors (Deuskar and Stewart, 2016)

[11] For all countries (except Austria) urban criteria fall into at least one of these four broadly defined categories. However, every country's definition is slightly different, and may include particularities not fully reflected among these four. For example, the definition for Honduras (see Annex 2), was counted in the 'population size' and 'urban services or characteristics' categories, but also has elements that do not fit neatly in either (e.g. "*communication by land (road or train) or regular air or maritime service*"). Austria's definition is unique in that, at least according to the definition listed in the WUP, it is based on commuting patterns into an urban core, and does not fall into any of these categories.

[12] A complete list of official definitions of urban used by LAC countries is provided in Annex 2 (see Table A2).

LAC samples come from the same underlying population. Hence, on average, LAC countries include smaller settlements in their urban population figures than do non-LAC countries. This may at least partially explain the apparent "over-urbanization" that we observe in Latin America.[13]

Table 1: Types of criteria used by countries in their urban definitions, by region

	Total no. countries	*No specific criteria*		*Population Size*			*Pop Density*		*Urban services or characteristics*		*Economic activity*	
		n	**%**	**n**	**%**	**Mean value**	**n**	**%**	**n**	**%**	**n**	**%**
World	**232**	**99**	**43%**	**103**	**44%**	**4,799**	**11**	**5%**	**30**	**13%**	**16**	**7%**
LAC	**40**	**23**	**58%**	**12**	**30%**	**1,958**	**0**	**0%**	**8**	**20%**	**0**	**0%**
Non-island	20	8	40%	9	45%	1,938	0	0%	7	35%	0	0%
Caribbean islands	20	15	75%	4	20%	2,000	0	0%	1	5%	0	0%
Rest of the world	**169**	**59**	**35%**	**87**	**51%**	**5,226**	**7**	**4%**	**21**	**12%**	**15**	**9%**
EAP	36	17	47%	15	42%	5,607	3	8%	6	17%	1	3%
SSA	48	13	27%	31	65%	5,694	0	0%	7	15%	8	17%
SAR	7	4	57%	3	43%	5,125	1	14%	2	29%	0	0%
ECA	54	18	33%	22	41%	4,395	1	2%	3	6%	5	9%
MNA	21	6	29%	14	67%	5,607	0	0%	3	14%	1	5%
N. America	3	1	33%	2	67%	1,750	2	67%	0	0%	0	0%

Notes: Criteria are not mutually exclusive, i.e. countries with multiple criteria are listed in more than one of these columns (Source: authors' analysis of data from UN, 2014)

2.2. Approaches to the consistent measurement of urbanization

A growing recognition of the problems posed for cross-country analyzes of urbanization and development by the use of national definitions of urban areas has led to several proposed methodologies for the consistent measurement of urbanization across countries.[14] In this paper, we focus on two of these methodologies – the Agglomeration Index (AI) that was originally developed by Uchida and Nelson (2008), and Dijkstra and Poelman's (2014) cluster method. The AI methodology has already featured in several Flagship World Bank publications (World Bank, 2008; World Bank and IMF, 2013; and Ellis and Roberts, 2015). By contrast, the cluster method, which was originally developed for application to the European Union, is more recent and its use at the global level is just beginning to be explored. The two methodologies differ both in terms of their underlying conceptual approaches to defining urban areas and their data requirements. The decision to focus on these two methodologies is based on the fact that they can be feasibly implemented to produce estimates of urban population shares for a large global cross-section of countries. This is a basic requirement for the (re-)examination of the basic stylized facts of urbanization, development

[13] The fact that LAC countries tend to count smaller settlements as 'urban' may be related to the history of human settlement in the region. Unlike, for example, Asia and Europe, which experienced more organic urban growth, most of Latin America's cities were purposefully and 'abruptly' founded by European colonizers (Morse, 1962; World Bank, 2009). Perhaps due to these origins, Latin American countries might consider these settlements to be 'urban' regardless of how small or sparsely populated they are, whereas an Asian village that has expanded and densified over time might still be considered 'rural'. However, this is a speculative hypothesis which would need further investigation to substantiate.

[14] The proposal of new methodologies has also been aided by the increasing availability of high-resolution global gridded population data sets and data sets on built-up area derived from satellite imagery.

and structural transformation, and the comparison of LAC with non-LAC countries. Below we provide a more detailed overview of the two methodologies.

2.2.1. The Agglomeration Index

From a labor market perspective, individual urban areas should ideally be conceptualized and defined as functional urban areas (FUAs), where a functional urban area captures the spatial extent over which a city or town's labor market may be considered integrated. Approaches to delineating FUAs typically begin with an urban core, defined by size and/or density criteria, and then identify a 'commuting shed', i.e. a surrounding area from which workers commute to the core. [15] A good example of an algorithm to identify FUAs is that used in the construction of the OECD's Metropolitan Areas database.[16] Hence, this algorithm defines FUAs by first identifying densely inhabited urban cores using gridded population data, and then adding non-contiguous interconnected urban cores, as well as other surrounding municipalities, based on commuting patterns (OECD, 2012). More recently, Duranton (2015) has proposed on algorithm to identify FUAs which dispenses with the need to pre-define an urban core, and instead relies solely on a minimum commuting threshold. Duranton tests this algorithm on data for Colombia, obtaining good results.

Unfortunately, the commuting data necessary to implement either the OECD's or Duranton's algorithm for identifying FUAs is not available at a global scale. In this context, the AI can be seen as a methodology which adopts a similar labor market perspective to defining urban areas, but which instead relies on estimated travel times in combination with data on population density to allow for a consistent method of measuring urbanization levels that can be feasibly implemented for a broad cross-section of countries. More specifically, the AI identifies as "urban, agglomerated, or dense" those areas that meet a certain population density threshold and which are within a given travel time radius of a "sizeable" settlement, where "sizeable" settlements are identified on the basis of a population threshold. As with the OECD algorithm for identifying FUAs, the AI typically relies on gridded population data as input.

To see more formally how the AI measures a country's level of urbanization, let i index a cell in a population grid for that country, and p_i and d_i the population and population density respectively associated with that cell. Furthermore, let $j = 1, \dots, J$ index a settlement point within the country that has an associated population of p_j, where $p_j \geq \bar{P}$ and $\bar{P}$ denotes the threshold population level that defines a "sizeable" settlement. Given this, associated with each of these J settlement points there will be an agglomeration A_j and grid cell i will be classified as belonging to agglomeration j *iff* $d_i \geq \bar{D}$ and $t_{ij} \leq \bar{T}$, where $\bar{D}$ is a threshold population density level, $t_{i,j}$ is the estimated travel time from grid cell i to the settlement point j, and $\bar{T}$ is a threshold travel time level. Based on this, a country's level of urbanization, as measured by the share of its population living in urban areas, is given by:

---

[15] The use of commuting data to identify metropolitan areas in the United States dates back to the 1960s (Berry, Goheen and Goldstein, 1969). This approach is also similar to the delineation of Labour Market Areas in Europe, or Travel to Work Areas in the UK, with the main difference being that these other exercises aim to exhaustively divide up an entire country or region into distinct areas based on commuting patterns, rather than identifying urban areas specifically (UK Office for National Statistics, 2015; Coombes, Casado-Diaz, Martinez-Bernabeu and Carausu, 2012).

[16] http://www.oecd.org/gov/regional-policy/regionalstatisticsandindicators.htm.

$$U = \left(\frac{\sum_{j=1}^{J} \sum_{i \in j} p_i}{\sum p_i}\right) \qquad [1]$$

From this, it is clear that there are three key thresholds to which values need to be assigned in the implementation of the AI methodology: namely, $\bar{P}, \bar{D},$ and $\bar{T}$. Following Uchida and Nelson (2008) and World Bank (2008), it has become standard to set $\bar{P} = 50{,}000, \bar{D} = 150$ people per km^2, and $\bar{T} = 60$ minutes. In other words, an area is defined as urban if it has a population density of at least 150 people per km^2 and falls within a 60 minute travel time radius of a settlement which itself has a population of at least 50,000. A country's level of urbanization is then given by the sum of the population living within all such areas.

2.2.2. The cluster method

In contrast to the AI, the cluster method adopts a more spatio-demographic approach to the consistent cross-country measurement of urbanization. In doing so, it follows a long tradition of using population size and density thresholds to define urban areas. As noted above, over a hundred countries use population size to define urban areas, while 11 use population density. The US Census Bureau began using population size, population density, and contiguity criteria to standardize the definition of metropolitan areas as far back as the early 20th century (Berry, Goheen and Goldstein, 1969).

More specifically, the cluster method classifies cells in a population grid according to density, and then groups them into "urban clusters." Again using the notation p_i to denote the population of a grid cell and d_i its population density, a grid cell i is classified as belonging to an urban cluster j if $d_i \geq \bar{D}_{UC}$ and $(\sum_{i \in j} p_i) \geq \bar{P}_{UC}$, where $\bar{D}_{UC}$ denotes a population density threshold and $\bar{P}_{UC}$ a population threshold. In addition, the cell must be spatially contiguous with at least one other cell within the same cluster. In other words, a spatially contiguous group of grid cells will be classified as constituting an urban cluster if each of these cells satisfies a certain minimum population density threshold and the aggregate population of these cells also exceeds some specified threshold value. Given this, a country's level of urbanization can then be defined as the share of the overall national population living in urban clusters. Dijkstra and Poelman (2014) set values of $\bar{D}_{UC} = 300$ people per km^2 and $\bar{P}_{UC} = 5000$. Within urban clusters, they also define a sub-set of "high density clusters". These high density clusters are defined similarly to urban clusters, but instead use threshold values of 1,500 people per km^2 for population density and 50,000 people for overall population.

2.2.3. Comparison between the two approaches

While it is not the intention of this paper to draw any conclusion as to which of the above two methods is to be preferred, some remarks on their similarities and differences are nevertheless in order. Hence, it will be noted that both crucially rely on a source of gridded population data for their implementation. Furthermore, while the AI may well be considered preferable to the cluster method to the extent that it more closely approximates an attempt to identify FUAs, it suffers the disadvantage of being more data-intensive. Thus, in addition to gridded population data, the AI also requires travel time data and a settlement point layer as inputs.

There is also a slight contradiction in how the AI approaches the definition of urban areas insofar as it implicitly relies on national definitions of urban in identifying the "sizeable" settlements around which it defines agglomerations. This is because the population data for the settlement points is invariably derived from national sources.

Finally, while, according to the cluster method, grid cells belonging to a given urban area must be spatially contiguous, the AI imposes no such restriction. As a corollary, development within a given agglomeration area in the AI approach can be discontinuous in the sense that urban cells can be interspersed with non-urban cells. Such a pattern of development may be consistent with the existence of topographical constraints (e.g. a river) that divide a city into different parts or with patterns of leapfrog development.[17]

3. Data

As noted above, both the AI and cluster methods rely on gridded population data for their implementation. In addition, the AI also requires a settlement point layer and travel time data.

3.1 Gridded population data

In their original AI analysis, Uchida and Nelson (2008) constructed estimates of urban population shares for a global cross-section of countries for the year 2000 using two global gridded population data sources – CIESIN's *GRUMP* (Global Rural-Urban Mapping Project) v 1 and Oak Ridge National Laboratories' *LandScan XXXX* population grids. Uchida and Nelson's estimates were then subsequently updated to the year 2010 as part of the World Bank and IMF's Global Monitoring Report 2013 on *Rural-Urban Dynamics and the Millennium Development Goals* (World Bank and IMF, 2013). In this update, the same two global gridded population data sources were used except that the gridded population data instead now referred to the year 2010. In the case of GRUMP, the 2010 population data is based on the projection of data from earlier years.[18, 19] Meanwhile, for the application of the cluster method, Dijkstra and Poelman (2014) limited their attention to the EU-27 countries, and used a mixture of population grids for the year 2006 based on different methods.[20]

[17] It is easy to imagine a methodology which combines different aspects of the AI and cluster methods. This methodology could, for example, follow the cluster approach of identifying high-density clusters and then define agglomerations around these clusters using criteria of travel time and population density *à la* the AI. Adopting this combined methodology would overcome the AI's weakness of needing an *a priori* defined layer of settlement points and associated populations.

[18] In updating the AI, the World Bank and IMF used the same settlement point and travel time layers as in Uchida and Nelson's (2008) original work, both of which referred to the year 2000. In this sense, the update performed was only partial.

[19] Ellis and Roberts (2015) also make use of the same 2010 AI estimates.

[20] For Croatia, Denmark, Sweden, Finland, Austria, the Netherlands, Slovenia, Switzerland and Norway they used grids that were constructed based on "bottom-up" methods from geo-coded address locations and population register data. By contrast, for the remaining EU countries they used a population disaggregation grid created by the European Commission's Joint Research Center (JRC) based on administrative area population and CORINE land cover data. Finally, they used LandScan data for French overseas regions, as well as Madeira and Açores in Portugal (see Dijkstra and Poelman, 2014, pp 7-8 for more details).

Subsequent to this, Deuskar and Stewart (2016) have also applied the cluster method to a further 53 countries using *WorldPop* gridded population data.[21]

In this paper, we use updated estimates of the national share of population living in urban areas based on the application of both the AI and cluster methods. We generate these estimates using three different population grids – *LandScan 2012*, *GHS-Pop* and *WorldPop*. All three of these grids have extensive cross-country coverage and are derived using "top-down" methods which involve the dasymetric distribution of population data for sub-national administrative units across grid cells. Where the three methods differ is in the exact weights they use to distribute the population within administrative areas. Although the exact methods are propriety, *LandScan* derives the weights using a modelling process that relies on "*primary geospatial input or ancillary datasets, including land cover, roads, slope, urban areas, village locations, and high resolution imagery analysis; all of which are key indicators of population distribution*." *LandScan* also applies *ex post* adjustments to the weights where obvious errors are identified through a manual verification and modification process.[22] *WorldPop* uses a similar range of input data to derive weights through a "random forest" modelling approach (Stevens, Gaughan, Linard and Tatem, 2015). Finally, *GHS-Pop* assigns the population of a given administrative unit to the built-up area within that unit. The built-up area data is taken from JRC's Global Human Settlement Layer (GHSL), which is itself derived from satellite imagery (Freire, MacManus, Pesaresi, Doxsey-Whitfield and Mills, 2016). For all three population grids, we use data at a spatial resolution of approximately 1 km.[23] While the *LandScan* and *GHS-Pop* data is available for the entire globe, the *WorldPop* data is limited to 53 countries, including 27 LAC countries. Another key difference is that *LandScan* provides estimates of ambient population (i.e. average population over 24 hours), while both *GHS-Pop* and *WorldPop* provide estimates of residential population.

3.1. Travel time and settlement point data

The travel time data used in the original AI analysis was based on "*… estimates of the time required to travel 1 km over different road and off-road surfaces…"* derived from a cost surface that was constructed from a variety of GIS data layers (Uchida and Nelson, 2008, p 6). These layers included data on road and rail networks, navigable rivers and water bodies, travel delays for crossing international borders, roughness of terrain and foot based travel for off-road and paths.[24] The AI estimates generated for this paper are based on an updated version of this same cost surface layer that is taken from Berg, Blankespoor, Li and Selod (2016). This updated layer is derived from more recent (i.e. *circa* 2010 versus *circa* 2000) data on roads, railroads, and land cover.

In contrast to the population and travel time data, we use the same settlement point layer as Uchida and Nelson (2008) – i.e. GRUMP Settlement Points v 1.[25] Because this data relates to the year 2000,

[21] Although *WorldPop* data sets are available for more countries, Deuskar and Stewart limit themselves to countries for which the data has been derived using "random forest" modelling methods. These methods are thought to be superior to earlier methods that were used to generate the population grids for other countries.
[22] See http://web.ornl.gov/sci/landscan/landscan_documentation.shtml for more information.
[23] Both *WorldPop* and *GHS-Pop* data are available at a higher spatial resolution (100 m in the case of *WorldPop*). For this paper, however, the data was aggregated up to a resolution of 1 km to allow for comparability with the *LandScan* data.
[24] See Appendix Table A.1 in Uchida and Nelson (2008) for more details.
[25] http://sedac.ciesin.columbia.edu/data/set/grump-v1-settlement-points.

it means that the updated AI estimates of national urban population shares generated as part of this paper are likely to be biased downwards. This will be the case, in particular, to the extent that there are settlements that were below $\bar{P}$ (i.e. the population threshold used to define a "sizeable" settlement) in 2000, but whose populations have since grown to exceed this threshold. The new agglomerations that will have been created as a result of this will be missed by our updated AI estimates.[26]

4. Results

In implementing the AI and cluster methods we use the same parameter values as Uchida and Nelson (2008) and Dijkstra and Poelman (2014) respectively. Hence, for the AI, we assume $\bar{P} = 50{,}000, \bar{D} = 150$ people per km^2, and $\bar{T} = 60$ minutes, while, for the cluster method, we assume $\bar{D}_{UC} = 300$ people per km^2 and $\bar{P}_{UC} = 5000$. For these parameter values, we generate estimates of the share of each country's population that lives in urban areas using each of the three gridded population data sets described above – i.e. *LandScan 2012*, *GHS-Pop* and *WorldPop*. We also estimate national urban population shares using the cluster method based solely on the share of population living in high density clusters, for which $\bar{D}$ = 1,500 people per km^2 and $\bar{P}$ = 50,000 people. Hence, overall, we generate a total of nine different sets of "consistent" estimates of national urban population shares.[27] Below, we briefly assess how, compared to the WUP data, these estimates change our impression of LAC's level of urbanization *vis-à-vis* other regions of the world. Using the AI and cluster urban population share estimates, we then re-examine the basic empirical relationships between urbanization, development and structural transformation that were discussed in Section 1 with a view to addressing the question of whether LAC countries depart significantly from these relationships. Because the *WorldPop* data is only available for a much more limited sample of countries, we focus on the results obtained using the *LandScan 2012* and *GHS-Pop* data. In doing so, we concentrate on a global sample of 146 countries – including 25 LAC countries – for which we have complete data.[28]

4.1. Re-assessing LAC's level of urbanization relative to the rest of the world

Figure 2 shows levels of urbanization for both the world and major regions, including LAC, as estimated using both the WUP data and the different permutations of methods (AI and cluster) and input population data sets (*LandScan* and *GHSPop*).[29] The WUP data shows LAC to be far more urbanized than the world overall. Hence, while, based on national definitions of urban areas, the share of the population living in urban areas in LAC in 2014 is estimated to be 80 percent, for the world as a whole it is only estimated to be 53 percent. Furthermore, LAC also appears to be the

[26] Performing a complete update of the AI remains an obvious area for future research.

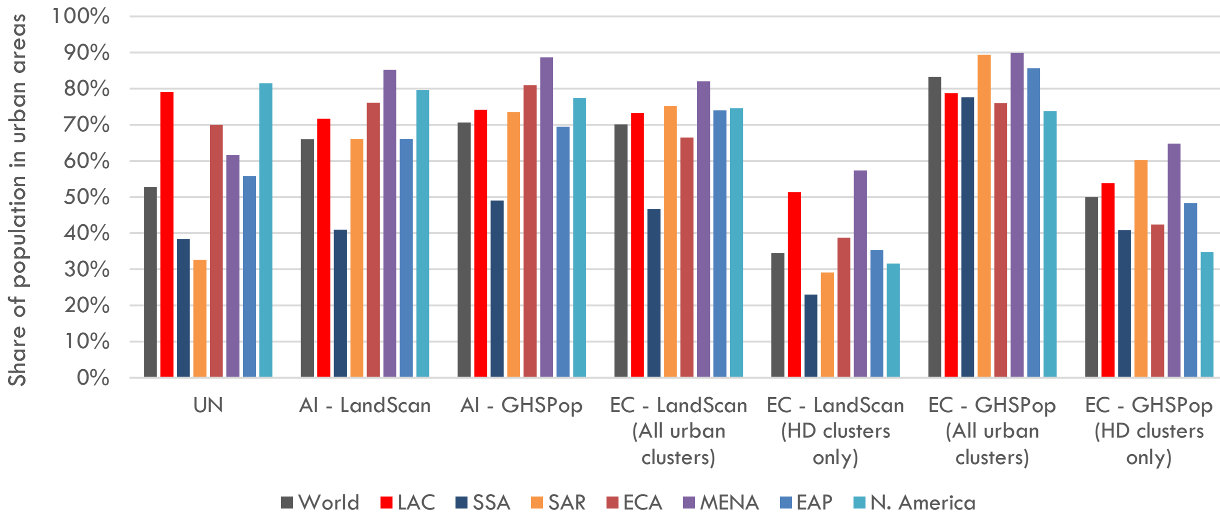
[27] Annex 3 examines the extent to which the regional and global maps of urban areas generated as a side-product of the analysis correlate with each other. It shows that there is a generally high level of agreement between the maps generated using the AI and cluster methods and different gridded population data sets. Interestingly, the main differences are driven less by the method adopted (AI versus cluster) and more by the choice of population layer.

[28] In addition to these 146 countries, estimates of urban population shares were generated for a further 65 countries using either the AI and/or cluster methods. The urban population share estimates for these countries are reported in Table A4.2 in Annex 4.

[29] Figure 2 has been constructed using the 146 country sample for which we have complete data. However, it also looks very similar when instead constructed using the data for all countries for which we are able to generate estimates of urbanization levels using the AI and cluster methods.

most urbanized region in the world after North America. However, for the various consistent sets of estimates of urban population shares, the picture is very different. In particular, LAC's level of urbanization appears much closer to that for the world overall. They also show LAC to be less urbanized than the Middle East and North Africa region in all six measures, less urbanized than South Asia in three measures, less urbanized than Europe and Central Asia in two measures, and East Asia and the Pacific in one measure. In four of the six measures, LAC appears as the fourth most urbanized among seven regions.

Figure 2: Estimated levels of urbanization using different methods and data sets



Note: EC refers to the cluster method for the consistent measurement of urbanization. "All urban clusters" refers to the share of population living in all urban clusters, while "HD clusters" refers to the share of population living just in high-density clusters.

The conclusion to be drawn from the above is that, when measured consistently, LAC appears far less urban relative to other regions than we are generally led to believe. This finding is consistent with the use of more relaxed standards for classifying areas as urban in LAC versus non-LAC countries (see Section 2). It also confirms findings reported by Uchida and Nelson (2008) and the World Bank (2008) using the original version of the AI.

4.2. Re-assessing LAC's position in relation to the basic stylized facts

Tables 2 and 3 show results from two basic sets of regressions. In Table 2, the (natural log) of a country's GDP per capita level is regressed on its level of urbanization. Meanwhile, in Table 3, a country's level of urbanization is regressed on the share of its national GDP that is generated by agriculture. In both tables, columns (1) – (3) report results obtained when using the WUP data, which is based on national definitions, to measure levels of urbanization, while columns (4) – (6) report the corresponding results when instead using the AI with *LandScan 2012* population data as input.

Table 2: Regression results for relationship between ln(GDP per capita) and urbanization

	WUP data			AI – *LandScan 2012*		
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**
Urban pop. share	4.215***	4.338***	4.254***	4.050***	4.102***	4.077***
	[14.70]	[15.21]	[14.64]	[12.36]	[12.20]	[12.27]
LA_dummy		-0.386**			-0.049	
		[-2.090]			[-0.237]	
C_dummy		0.595*			-0.267	
		[1.934]			[-0.753]	
LAC_dummy			-0.143			-0.099
			[-0.853]			[-0.537]
Constant	6.742***	6.697***	6.744***	6.685***	6.670***	6.685***
	[38.22]	[38.67]	[38.20]	[31.48]	[31.01]	[31.40]
n	146	146	146	146	146	146
R^2	0.600	0.623	0.602	0.515	0.517	0.516
$\overline{R^2}$	0.597	0.616	0.597	0.511	0.507	0.509

Dependent variable: ln(GDP per capita);
Notes: t-statistics in brackets. *** Significant at 1 % level; ** at 5 % level; * at 10 % level

Table 3: Regression results for relationship between urbanization and the share of national GDP derived from agriculture

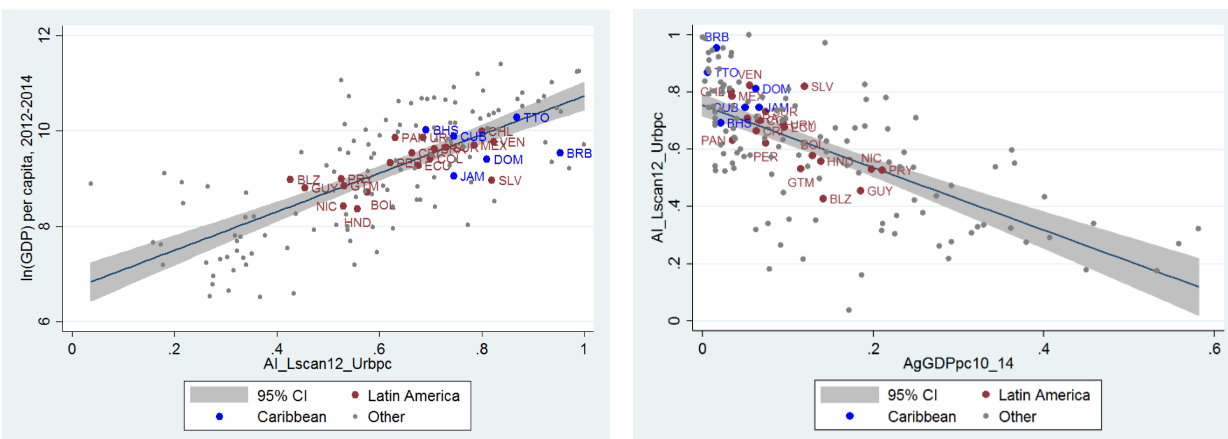
	WUP data			AI – *LandScan 2012*		
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**
Agric. share	-1.159***	-1.171***	-1.145***	-1.092***	-1.067***	-1.079***
	[-10.72]	[-10.87]	[-10.44]	[-10.29]	[-9.890]	[-10.03]
LA_dummy		0.0779*			0.007	
		[1.965]			[0.164]	
C_dummy		-0.124*			0.096	
		[-1.831]			[1.419]	
LAC_dummy			0.0308			0.027
			[0.848]			[0.769]
Constant	0.725***	0.722***	0.718***	0.754***	0.746***	0.748***
	[37.37]	[34.73]	[33.90]	[39.61]	[35.87]	[35.97]
n	146	146	146	146	146	146
R^2	0.444	0.473	0.446	0.424	0.432	0.426
$\overline{R^2}$	0.440	0.462	0.439	0.420	0.420	0.418

Dependent variable: urban share of national population
Notes: t-statistics in brackets. *** Significant at 1 % level; ** at 5 % level; * at 10 % level

As can be seen, the tables confirm the results for the WUP data that were depicted visually in Figure 1. Hence, a country's level of development is significantly positively correlated with its level of urbanization, while there is a significant negative relationship between a country's urbanization level and the relative importance of agriculture to its economy. Furthermore, Latin American (LA) and Caribbean (C) countries depart significantly from these relationships, but in opposite directions. This is evident in the opposite signs on the dummy variables for Latin American and Caribbean countries in column (2) in both tables. On average, Latin American countries have significantly lower levels of GDP per capita than one would predict based on their levels of urbanization. Consistent with this, they also appear "over-urbanized" given the contribution that agriculture makes to GDP for countries in the region. By contrast, Caribbean countries appear "under-urbanized." However, this result is driven by two extreme outliers – Barbados and Trinidad and Tobago.[30]

When we turn to the results using the AI based on the *LandScan 2012* population data, however, the estimated co-efficients on the Latin American and Caribbean dummies are much smaller in magnitude and statistically insignificant (see, in particular, column (5) in Tables 2 and 3). As a consequence, Latin American countries, as a group, no longer appear "over-urbanized." Likewise, the Caribbean countries no longer appear "under-urbanized." This can also be seen in Figure 3, which shows the scatterplots that correspond to the regressions in column (4) of Tables 2 and 3. In particular, comparing Figure 3 with Figure 1, one can see that the Latin American countries are no longer systematically positioned below the fitted line, but, rather, are more evenly distributed around it. Meanwhile, for the Caribbean countries, Barbados and Trinidad and Tobago no longer have the appearance of outliers.

Figure 3: Relationship between: (a) levels of economic development and urbanization; and (b) levels of urbanization and the contribution of agriculture to national GDP. Level of urbanization measured using the Agglomeration Index with LandScan 2012 population data



Qualitatively similar results to those reported above for Latin America are also found using the AI based on both the *GHS-Pop* and *WorldPop* data – i.e. there is no longer any evidence of "over-urbanization" (see Annex 5 for full results). For the Caribbean countries, on the other hand, there is some continued evidence of "under-urbanization" in the relationship between GDP per capita and urbanization using the *WorldPop* data, but the sample size in this regression is only small (i.e. 40

[30] This is confirmed when the Caribbean dummy in the column [2] regressions in Tables 3 and 4 is replaced by two separate dummies – one for Barbados and Trinidad and Tobago, and the other for the remaining Caribbean countries. In particular, for these regressions, the Barbados and Trinidad and Tobago dummy is statistically significant with the expected signs, while the dummy for the other Caribbean countries is insignificant.

countries). Likewise, with one exception, the cluster method based on the *LandScan* and *GHS-Pop* data gives broadly similar results (see Annex 6, Table A6.1 – A6.4). The exception is when we restrict our attention to high-density clusters and use the *LandScan* data, in which case the Latin American dummy remains significantly position (indicating "over-urbanization") in the regression of urbanization on the share of the agricultural sector in a country's GDP. Finally, for the cluster method using the *WorldPop* data, Caribbean countries continue to have significantly higher levels of GDP per capita than predicted given their levels of urbanization (see Annex 6, Table 6.5). However, again, the sample size for this regression is restricted to only 40 countries. As an additional observation, it will be noted that the basic stylized relationships between urbanization, development and structural transformation break down when the cluster method is used in conjunction with either the *GHS-Pop* or *WorldPop* data. Insofar as the basic stylized facts are more stable across different input population data sets, this suggests that the AI results are more robust than those for the cluster method. Given that the AI also relies on additional sources of data – in particular, travel time estimates – this should perhaps come as no surprise.

5. Conclusion

When levels of urbanization are measured based on national definitions of urban areas, countries in the LAC region are seen to depart significantly from the basic stylized facts of urbanization, development and structural transformation. Most notably, Latin American countries appear to be "over-urbanized" relative to their levels of development and the importance of the agricultural sector to their economies. However, when levels of urbanization are instead measured using a consistent set of criteria across countries, this apparent "over-urbanization" is revealed to be an "urban myth." This finding tends to hold irrespective of whether the Agglomeration Index or a cluster-based methodology is adopted as the approach for consistently defining urban areas across countries. It, furthermore, tends to hold irrespective of the choice of gridded population data used as input into these two methodologies.

References

Ades, A. F., & Glaeser, E. L. (1995). Trade and Circuses: Explaining Urban Giants. *The Quarterly Journal of Economics, 110*(1), 195-227.

Berg, Blankespoor, Li, & Selod. (2016), unpublished manuscript under preparation.

Berry, B. J., Goheen, P. G., & Goldstein, H. (1969). *Metropolitan area definition: A re-evaluation of concept and statistical practice.* Washington: U.S. Bureau of the Census.

Chomitz, K. M., Buys, P., & Thomas, T. S. (2005). *Quantifying the rural-urban gradient in Latin America and the Caribbean.* World Bank, Development Research Group, Infrastructure and Environment Team. Washington, DC: World Bank.

Cohen J. A. (1960). Coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, *20*, 37-46.

Coombes, M., Casado-Díaz, J. M., Martínez-Bernabeu, L., & Carausu, F. (2012). *Study on comparable Labour Market Areas - Final research report.*

Davis, J., & Henderson, J.V. (2003). Evidence on the Political Economy of the Urbanization Process. *Journal of Urban Economics, 53*(1), 98-125.

Deuskar, C., & Stewart, B. (2016, March 15). Measuring Global Urbanization using a Standard Definition of Urban Areas: Analysis of Preliminary Results. Washington, DC: presented at World Bank Land & Poverty Conference, March 2016.

Dijkstra, L., & Poelman, H. (2014). *A harmonised definition of cities and rural areas: the new degree of urbanization.* Regional Working Paper, European Commission, Directorate-General for Regional and Urban Policy.

Duranton, G. (2015). A Proposal to Delineate Metropolitan Areas in Colombia. *Desarrollo y Sociedad* (75), 223-264. Retrieved from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-35842015000100006&lng=en&tlng=en. 10.13043/DYS.75.6.

Ellis, P., & Roberts, M. (2015). *Leveraging Urbanization in South Asia : Managing Spatial Transformation for Prosperity and Livability.* World Bank. Washington, DC: World Bank. Retrieved from https://openknowledge.worldbank.org/handle/10986/22549 License: CC BY 3.0 I

Foody, G.M. (2004). Thematic map comparison. *Photogrammetric Engineering & Remote Sensing*, 70(5), 627-633.

Freire, S., MacManus, K., Pesaresi, M., Doxsey-Whitfield, E., & Mills, J. (2016). Development of new open and free multi-temporal global population grids at 250 m resolution. *proceedings of 19th AGILE Conference on Geographic Information Science* . Helsinki.

Henderson, J. V., & Wang, H. G. (2005). Aspects of the rural-urban transformation of countries. *Journal of Economic Geography, 5*(1), 23-42.

Henderson, J.V., Storeygard, A., & Roberts, M. (2013). *Is Urbanization in Sub-Saharan Africa Different?* (World Bank Policy Research Working Paper No. 6481). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2280643.

Krugman, P. & Elizondo, R. L. (1996). Trade policy and the third world metropolis. *Journal of Development Economics, 49*(1), 137-150.

McHugh, M.L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276-282.

McKinsey Global Institute (2011a). *Urban World: Mapping the Economic Power of Cities*. McKinsey Global Institute.

McKinsey Global Institute (2011b). *Building Globally Competitive Cities: The Key to Latin American Growth*. McKinsey Global Institute.

Michaels, G., Rauch, F., & Redding, S. J. (2012). Urbanization and structural transformation. *The Quarterly Journal of Economics, 127*(2), 535-586.

Monserud, R.A. & Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, 62(4), 275-293.

Morse, R. M. (1962). Some Characteristics of Latin American Urban History. *The American Historical Review, 67*(2), 317-338.

Organisation for Economic Co-operation and Development. (2012). *Redefining "urban": A new way to measure metropolitan areas.* Paris: OECD.

Satterthwaite, D. (2007). *The transition to a predominantly urban world and its underpinnings.* London: International Institute for Environment and Development (IIED).

Stevens, F., Gaughan, A., Linard, C., & Tatem, A. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS One, 10*(2).

Uchida, H., & Nelson, A. (2008). *Agglomeration Index: Towards a New Measure of Urban Concentration.*

UK Office for National Statistics. (2015). *Overview of 2011 Travel to Work Areas.* Retrieved from http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/other/travel-to-work-areas/index.html

United Nations, Department of Economic and Social Affairs, Population Division. (2014). World Urbanization Prospects: The 2014 Revision. (CD-ROM Edition).

World Bank (2008). *World Development Report, 2009: Reshaping economic geography.* Washington, DC: World Bank.

World Bank (2009). *Reshaping Economic Geography in Latin America and the Caribbean: A Companion Volume to the 2009 World Development Report*. Washington, DC: World Bank.

World Bank & IMF (2013). *Global Monitoring Report 2013: Rural-Urban Dynamics and the Millennium Development Goals.* Washington, DC: World Bank.

World Bank (2015). *East Asia's Changing Urban Landscape : Measuring a Decade of Spatial Growth.* Washington, DC: World Bank. Retrieved from https://openknowledge.worldbank.org/handle/10986/21159

Annex 1: Definitions of samples and variables, and data sources

Definition of samples

Countries were excluded from the main analysis if values of relevant variables were missing. These variables, which are further explained below, were: the urban population share, GDP per capita, agricultural share of GDP, and national population density. For measures of urbanization constructed using the *GHSPop* and *LandScan 2012* data, this resulted in a sample of 146 countries. For measures involving *WorldPop* data, the resultant sample consisted of 44 countries.

Definition of variables and data sources

Urban share of population (World Urbanization Prospects - WUP)
The urban share of population based on official national definitions of urban areas is the 2014 figure from the World Development Indicators (SP.URB.TOTL.IN.ZS, last updated September 2015). These figures are themselves based on those reported in the United Nations' World Urbanization Prospects database (https://esa.un.org/unpd/wup/)

GDP per capita
GDP per capita is measured in constant 2011 international dollars using Purchasing Power Parity exchange rates, taken from the World Development Indicators (NY.GDP.PCAP.PP.KD, last updated in September 2015). The value for each country is calculated as the mean over all non-missing observations of the 2012, 2013, and 2014 values.

Agricultural share of GDP
Agricultural share of GDP refers to the value added by forestry, hunting, fishing, cultivation of crops, and livestock production, as a percentage of the country's GDP, taken from the World Development Indicators (NV.AGR.TOTL.ZS, last updated in November 2015). The value for each country is calculated as the mean over all non-missing observations of the 2010-2014 values.

National population density
The national population density figures used are the 2014 values from the World Development Indicators (EN.POP.DNST, last updated in December 2015). It refers to "midyear population divided by land area in square kilometers".

Annex 2: Official definitions of urban areas used in LAC countries

Country	Urban definition
Antigua and Barbuda	St. John's (capital).
Argentina	Localities with 2,000 inhabitants or more.
Aruba	Oranjestad (capital) and Sant Nicolas.
Bahamas	For the 1980, 1990 and 2000 censuses, sum of the cities.
Barbados	Bridgetown (capital).
Belize	Belize City and all towns.
Bolivia	Localities with 2,000 inhabitants or more.
Brazil	Administrative centres of 'municipios' and districts, including suburban zones.
Cayman Islands	Entire population.
Chile	Populated centres with defined urban characteristics, such as certain public and municipal services.
Colombia	Administrative headquarters ('población cabecera') with 2,000 inhabitants or more.
Costa Rica	Administrative centres of cantons, including adjacent areas with clear urban characteristics such as streets, urban services and electricity.
Cuba	Places with 2,000 inhabitants or more, and places with fewer inhabitants but with paved streets, street lighting, piped water, sewage, a medical centre and educational facilities.
Curaçao	Willemstad (capital).
Dominica	Cities and villages with 1,000 inhabitants or more.
Dominican Republic	Administrative centres of communes and municipal districts.
Ecuador	Capitals of provinces and cantons.
El Salvador	For 1971, areas where authorities of the municipality reside, as determined by those authorities. For 2007, the head of the municipality, where the primary civil, religious and military authorities reside, and those areas having a continuous cluster of at least 500 dwellings, with street lighting service, basic education schools, regular transportation service, paved or cobbled streets, and telephone services.
Grenada	Parishes of St. George's Town (capital) and St. George.
Guatemala	The 'municipio' of Guatemala Department and officially recognized centres of other departments and municipalities. The urban population for 1981 is officially adjusted to include the urbanized suburbs bordering the 'municipio' of Guatemala in a way consistent with the previous census.
Guyana	City of Georgetown (capital), and four other towns.
Haiti	Administrative centres of communes.
Honduras	Populated centres with 2,000 inhabitants or more that also meet the following criteria: piped water service; communication by land (road or train) or regular air or maritime service; complete primary school (six

Country	Urban definition
	grades); postal service or telegraph; and at least one of the following: electrical light, sewer system, or a health centre.
Jamaica	Kingston metropolitan area and selected main towns.
Mexico	Localities with 2,500 inhabitants or more.
Nicaragua	Department, region and municipality headquarters, and population centers of 1,000 inhabitants or more, with some features such as: streets, electricity service, commercial and / or industrial establishments, etc.
Panama	Localities with 1,500 inhabitants or more, with all or most of the following urban characteristics: electricity, water-supply and sewerage systems, paved roads and access to commercial establishments, secondary schools and social and recreational centres. Some places with most of the mentioned features were defined as urban, even if they did not fulfilled the population requirement.
Paraguay	Administrative centres of the official districts.
Peru	Populated centres with 100 dwellings or more grouped contiguously and administrative centres of districts.
Puerto Rico	Densely settled territory that meets minimum population density requirements and with 2,500 inhabitants or more. A change in the definition for the 2000 census from place-based to density-based affects the comparability of estimates before and after this date.
Saint Kitts & Nevis	Basseterre (capital) and Charlestown.
Saint Lucia	No official definition available. In the present publication, the urban agglomeration of the city of Castries, its suburban and three towns (Gros Islet, Soufrière, and Vieux Fort).
Saint Vincent & Grenadines	No official definition available.
Sint Maarten (Dutch part)	Entire population.
Suriname	The district of Paramaribo (capital) and Wanica district.
Trinidad & Tobago	Port-of-Spain (capital), Arima borough and San Fernando town.
Turks & Caicos Islands	The islands of Grand Turk and Providenciales.
US Virgin Islands	For the 2000 and 2010 censuses, densely settled territory that meets minimum population density requirements and with 2,500 inhabitants or more. For the 1950 and 1960 censuses, the proportion urban was adjusted for consistency with the new definition.
Uruguay	Cities officially designated as such.
Venezuela	Places with 2,500 inhabitants or more.

Source: United Nations, Department of Economic and Social Affairs, Population Division (2014). World Urbanization Prospects: The 2014 Revision, CD-ROM Edition.

Annex 3: Comparison of urban extent maps

In this annex we present a comparison of the maps of urban areas generated as a side-product of the analysis of this paper. In particular, we perform pairwise comparisons of the maps generated using both the Agglomeration Index and cluster methods based on the three different gridded population data sets – *LandScan 2012*, *GHSPop* and *WorldPop*. We perform the comparisons both at the global level and for LAC as a region. To perform the comparisons, we make use of Cohen's kappa statistic (Cohen, 1960), which assesses the degree of agreement between two maps taking into account the degree of agreement which is expected to occur by chance (Monserud and Leemans, 1992; Foody, 2004).

For each pairwise comparison of maps, Cohen's Kappa statistic is calculated as follows:

$$K = \frac{p_o - p_e}{1 - p_e}$$

Where p_0 is the relative observed agreement between the compared maps, and p_e is the expected agreement occurring by chance

Cohen's kappa has a theoretical range of -1 to +1, with 1 indicating perfect agreement. In his original paper, Cohen provided guidelines on the interpretation of the statistic. According to these guidelines, $K < 0$ indicates no agreement, $0 \leq K \leq 0.20$ slight agreement, $0.20 < K \leq 0.40$ fair agreement, $0.40 < K \leq 0.60$ moderate agreement, $0.60 < K \leq 0.80$ substantial agreement, and $K > 0.80$ almost perfect agreement. It is important to note, however, that the interpretation of the kappa statistic remains a matter of debate of the literature with alternative guidelines having been published (McHugh, 2012).

Results

A total of 6 pairwise comparisons were performed at the global level and 15 for the LAC region.[31] The number of comparisons for LAC is greater than for the globe because while *WorldPop* data is available for LAC, it is not available for the entire globe. Table A3.1 reports the estimated Kappa statistics from the pairwise comparisons. As will be seen, the statistics range from a minimum of 0.535 to a maximum of 0.761 with 7 of the pairwise comparisons for LAC falling in the range that Kappa categorized as representing "substantial" agreement and the remaining 8 in the range that he classified as showing "moderate" agreement. For the global statistics, 3 fall in the range of "substantial" agreement and the remaining 3 in the range of "moderate agreement." Levels of agreement tend to be slightly higher for LAC than for the globe overall. Interestingly, the highest Kappa statistics are obtained when comparing between the AI and cluster methods using the same input gridded population data set. Hence, the highest estimated Kappa statistics for both LAC and the globe are obtained by comparing the maps generated using the AI and cluster methods using *LandScan 2012* data as input. This suggests that the spatial form of the urban areas resulting from different urbanization measures is more sensitive to the input population distribution data than the urban definition method.

[31] Comparisons were not performed involving the cluster method using only high-density clusters.

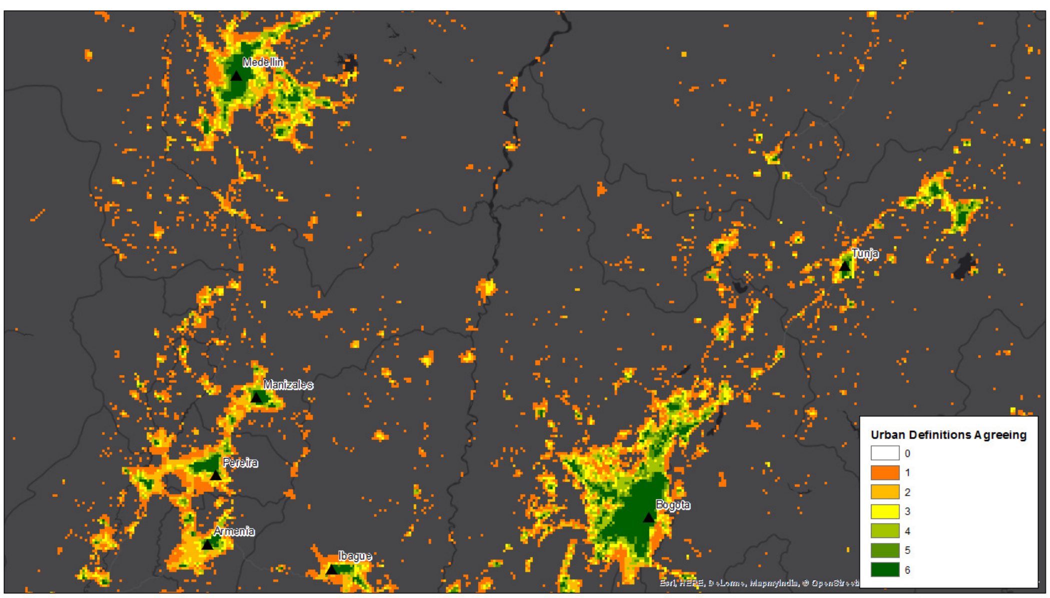
Table A3.1: Kappa statistics for pairwise map comparisons

Definition 1	Definition 2	LAC	Global
AI - LandScan	AI - GHSPop	0.700	0.646
AI – Landscan	AI – WorldPop	0.569	-
AI – Landscan	**CL – Landscan**	**0.761**	**0.707**
AI – Landscan	CL – GHS Pop	0.597	0.589
AI – Landscan	CL – WorldPop	0.569	-
AI – GHS Pop	AI – WorldPop	0.603	-
AI – GHS Pop	CL – Landscan	0.632	0.557
AI – GHS Pop	**CL – GHS Pop**	**0.712**	**0.692**
AI – GHS Pop	CL – WorldPop	0.594	-
AI – WorldPop	CL – Landscan	0.560	**-**
AI – WorldPop	CL – GHS Pop	0.563	-
AI – WorldPop	**CL – WorldPop**	**0.706**	-
CL – Landscan	CL – GHS Pop	0.612	0.591
CL – WorldPop	CL – GHS Pop	0.535	-
CL – WorldPop	CL – Landscan	0.548	**-**

Notes: AI = Agglomeration Index; CL = cluster method.

As a visualization of the degree of agreement between the different maps, Figure A3.1 shows the maps overlaid on one another for the Bogota area of Colombia with color coding according to the number of maps between which there is agreement. The figure shows that all maps agree on the classification of the core urban areas of major cities with the number of maps showing agreement becoming progressively less as one move towards the periphery of these cities. There is also less agreement for smaller settlements.

Figure A3.1: Visual comparisons of degree of agreement between different maps

z 

Annex 4: Estimates of national urban population shares using different methods

Table A4.1: 146 countries used in the regression analyses

Region	Country	GDP pc	Urban share (%)					
			AI			Cluster-based		
			LScan	*GHSPop*	*WPop*	*LScan*	*GHSPop*	*WPop*
ECA	Luxembourg	89,002	83.74	83.37		57.98	68.53	
EAP	Singapore	77,436	99.08	98.16		99.29	99.94	
MNA	Kuwait	75,746	98.64	98.78		96.69	97.30	
EAP	Brunei Darussalam	72,204	79.30	82.20		81.68	82.06	
ECA	Norway	63,654	52.56	53.06		58.53	58.58	
MNA	United Arab Emirates	60,903	93.68	97.57		92.59	97.06	
ECA	Switzerland	54,778	74.57	75.20		73.65	78.35	
NAC	United States	51,316	80.34	77.78		75.14	73.07	
MNA	Saudi Arabia	49,110	88.01	82.08		89.67	91.70	66.26
ECA	Ireland	45,316	54.06	57.87		52.38	55.32	
ECA	Netherlands	45,262	92.24	93.11		83.09	83.92	
ECA	Austria	44,060	70.76	69.11		60.84	59.45	
ECA	Sweden	43,580	68.31	72.08		60.34	69.59	
ECA	Germany	43,096	87.04	86.33		70.01	72.73	
ECA	Denmark	42,709	69.38	73.52		59.04	64.16	
NAC	Canada	42,298	72.92	74.75		74.39	75.83	
ECA	Belgium	40,726	91.08	91.08		75.51	80.25	
ECA	Iceland	40,707	63.87	63.90		73.92	74.38	
MNA	Oman	40,011	74.93	73.92		75.60	82.60	51.71
ECA	Finland	38,949	66.22	70.88		57.76	68.39	
ECA	United Kingdom	37,019	88.02	92.02		81.04	86.09	
EAP	Japan	35,412	94.64	95.60		87.89	92.16	
ECA	Italy	33,906	83.38	86.06		72.65	76.24	
EAP	New Zealand	33,083	78.26	77.31		73.43	73.43	
EAP	Korea, Rep.	32,738	95.46	96.70		90.44	93.42	
ECA	Spain	31,974	82.44	84.86		76.28	79.17	
ECA	Cyprus	30,488	81.40	83.05		70.23	71.74	
LAC	**Trinidad and Tobago**	**29,155**	**86.85**	**88.28**	**78.87**	**79.41**	**87.94**	**71.07**
ECA	Czech Republic	28,384	79.76	80.98		56.52	62.20	
ECA	Slovenia	27,703	61.43	61.30		47.35	53.00	

Region	Country	GDP pc	Urban share (%)					
			AI			Cluster-based		
			LScan	*GHSPop*	*WPop*	*LScan*	*GHSPop*	*WPop*
ECA	Portugal	25,895	73.98	74.75		79.44	70.33	
ECA	Slovak Republic	25,846	74.88	74.65		47.20	55.49	
ECA	Estonia	25,293	53.35	58.88		61.59	64.13	
ECA	Lithuania	24,717	64.12	71.20		53.57	78.47	
ECA	Greece	24,551	67.17	67.96		66.92	68.28	
ECA	Russian Federation	23,384	70.72	75.43		67.83	80.00	
ECA	Poland	23,279	81.33	84.45		61.39	65.60	
EAP	Malaysia	22,890	80.98	82.30		80.23	82.50	
ECA	Hungary	22,874	83.39	84.65		57.59	63.57	
LAC	**Bahamas, The**	**22,572**	**69.19**	**70.87**		**76.74**	**76.59**	
ECA	Kazakhstan	22,356	57.42	65.75		65.47	83.17	
LAC	**Chile**	**21,641**	**80.10**	**83.94**	**61.82**	**79.38**	**87.21**	**55.15**
ECA	Latvia	21,586	63.70	69.29		58.93	73.57	
ECA	Croatia	20,097	60.10	62.31		55.01	60.22	
LAC	**Cuba**	**19,706**	**74.56**	**89.22**	**58.62**	**59.44**	**82.52**	**44.44**
LAC	**Uruguay**	**19,226**	**68.59**	**74.06**	**64.54**	**79.44**	**86.85**	**61.43**
LAC	**Panama**	**18,973**	**63.07**	**62.48**	**52.70**	**65.55**	**68.86**	**42.95**
ECA	Turkey	18,484	73.49	84.44		63.43	81.95	
ECA	Romania	18,084	71.16	74.86		47.31	58.55	
SSA	Gabon	17,809	55.09	55.05		66.43	71.68	
LAC	**Venezuela**	**17,381**	**82.34**	**84.63**	**52.26**	**81.77**	**88.35**	**40.55**
SSA	Mauritius	17,176	92.52	93.84		86.40	90.57	
ECA	Belarus	17,114	74.00	85.00		67.97	82.30	
MNA	Lebanon	16,586	100.00	94.01		82.40	89.87	
ECA	Azerbaijan	16,397	57.94	63.48		59.42	79.14	
LAC	**Mexico**	**16,201**	**78.57**	**82.77**	**74.82**	**73.15**	**80.65**	**66.64**
ECA	Bulgaria	15,729	76.58	77.75		60.93	59.74	
LAC	**Suriname**	**15,577**	**73.13**	**73.82**	**57.34**	**65.75**	**65.96**	**45.47**
LAC	**Brazil**	**15,095**	**70.75**	**71.73**	**63.47**	**79.14**	**79.93**	**57.48**
SSA	Botswana	14,788	52.40	60.25		48.94	70.98	
ECA	Montenegro	14,075	44.79	49.34		60.70	75.48	
LAC	**Costa Rica**	**13,907**	**66.35**	**66.58**	**64.47**	**68.95**	**75.64**	**58.31**

Region	Country	GDP pc	Urban share (%)					
			AI			Cluster-based		
			LScan	*GHSPop*	*WPop*	*LScan*	*GHSPop*	*WPop*
LAC	**Barbados**	**13,755**	**95.36**	**97.40**		**83.51**	**84.55**	
EAP	Thailand	13,744	54.01	63.86	48.46	46.88	70.51	36.95
ECA	Turkmenistan	13,592	59.10	78.84		58.72	87.72	
MNA	Algeria	13,296	70.04	80.25	75.33	66.97	85.85	69.69
ECA	Serbia	12,704	77.70	79.94		57.60	65.74	
SSA	South Africa	12,425	58.46	65.83		58.89	77.43	
LAC	**Colombia**	**12,296**	**70.01**	**73.70**	**60.65**	**71.25**	**84.78**	**60.24**
LAC	**Dominican Republic**	**12,037**	**81.07**	**88.02**	**73.37**	**73.00**	**87.58**	**59.44**
ECA	Macedonia, FYR	11,905	75.01	78.57		68.10	76.25	
EAP	China	11,810	63.34	69.21		71.66	88.20	65.19
MNA	Jordan	11,414	93.75	87.81		86.82	91.96	
LAC	**Peru**	**11,204**	**62.18**	**64.84**	**40.45**	**63.51**	**77.88**	**40.24**
MNA	Tunisia	10,688	73.84	89.19		63.34	86.93	
EAP	Mongolia	10,654	49.75	54.75		63.37	70.21	
LAC	**Ecuador**	**10,567**	**67.66**	**74.58**	**61.72**	**65.68**	**82.39**	**56.46**
MNA	Egypt, Arab Rep.	10,054	97.20	97.53		96.32	98.08	
ECA	Albania	9,977	67.02	75.98		56.14	70.02	
EAP	Indonesia	9,664	64.29	66.72	64.45	81.69	90.33	71.55
SAR	Sri Lanka	9,442	57.86	56.89	54.83	80.03	80.41	76.21
ECA	Bosnia and Herzegovina	9,366	51.24	61.65		54.54	75.64	
SSA	Namibia	9,087	18.07	18.64	17.65	41.43	41.07	31.27
LAC	**Jamaica**	**8,564**	**74.63**	**74.11**	**67.15**	**74.74**	**81.30**	**59.52**
ECA	Ukraine	8,308	70.73	77.38		63.75	70.41	
LAC	**Paraguay**	**8,013**	**52.63**	**64.18**	**49.70**	**56.92**	**80.51**	**40.03**
LAC	**Belize**	**7,948**	**42.71**	**43.28**	**52.78**	**50.62**	**59.74**	**11.86**
LAC	**El Salvador**	**7,833**	**82.03**	**90.15**	**75.29**	**68.74**	**84.89**	**53.18**
EAP	Fiji	7,518	21.52	21.31		35.52	49.43	
ECA	Armenia	7,487	77.99	78.16		71.91	81.58	
SAR	Bhutan	7,262	3.70	13.34	0.91	26.60	31.55	11.55
LAC	**Guatemala**	**6,976**	**53.13**	**55.53**	**47.07**	**63.89**	**85.20**	**52.48**
ECA	Georgia	6,955	60.85	67.79		59.91	68.41	
MNA	Morocco	6,841	67.46	75.60	65.18	58.36	79.68	58.28

Region	Country	GDP pc	Urban share (%)					
			AI			Cluster-based		
			LScan	*GHSPop*	*WPop*	*LScan*	*GHSPop*	*WPop*
SSA	Angola	6,796	26.55	43.86		37.07	81.10	
LAC	**Guyana**	**6,631**	**45.47**	**48.05**	**43.00**	**59.08**	**64.86**	**54.06**
EAP	Philippines	6,363	61.65	60.60	59.80	77.69	79.47	75.23
SSA	Cabo Verde	6,163	57.23	31.69		66.85	89.54	
LAC	**Bolivia**	**6,067**	**57.74**	**58.64**	**30.02**	**61.27**	**76.42**	**33.88**
SSA	Swaziland	5,983	34.00	42.68		28.73	46.43	
SSA	Congo, Rep.	5,807	58.67	56.07		66.95	76.42	
SSA	Nigeria	5,466	55.46	65.12		60.30	86.96	
SAR	India	5,146	67.33	76.14		72.47	91.36	
EAP	Vietnam	5,136	68.08	73.10	69.35	74.45	92.06	75.60
ECA	Uzbekistan	5,009	82.57	92.46		76.00	91.11	
EAP	Lao PDR	4,792	21.80	25.19	16.70	26.17	76.17	12.52
LAC	**Nicaragua**	**4,538**	**53.00**	**53.86**	**41.00**	**52.75**	**76.35**	**44.16**
SAR	Pakistan	4,493	70.92	72.20		79.17	91.34	
ECA	Moldova	4,482	62.76	65.93		48.53	60.33	
LAC	**Honduras**	**4,299**	**55.75**	**57.82**	**40.26**	**52.08**	**54.85**	**36.46**
SSA	Sudan	3,844	43.72	47.20		40.98	88.41	
SSA	Ghana	3,798	53.62	66.28		45.63	78.85	
SSA	Zambia	3,616	35.56	45.56		35.55	77.35	
SSA	Mauritania	3,605	33.93	40.72	25.62	41.07	78.75	19.00
ECA	Kyrgyz Republic	3,053	51.63	56.95		61.27	87.16	
SSA	Sao Tome and Principe	3,046	77.31	78.42		73.47	80.32	
EAP	Cambodia	2,953	53.79	56.13		41.38	72.29	44.11
SSA	Cote d'Ivoire	2,936	58.80	60.99		50.47	61.78	
SAR	Bangladesh	2,846	62.76	70.44	65.51	97.50	97.20	90.19
SSA	Cameroon	2,748	53.85	67.61		48.77	81.80	
SSA	Kenya	2,745	47.34	46.39	51.89	58.59	73.29	72.55
ECA	Tajikistan	2,457	37.79	39.65		70.15	90.10	
SSA	Lesotho	2,442	31.87	29.42		39.56	58.20	
SSA	Tanzania	2,381	32.90	42.06	37.60	33.43	72.78	27.28
SAR	Nepal	2,182	32.32	36.49	31.96	68.90	72.90	61.51
SSA	Senegal	2,174	54.36	73.52	53.60	44.93	87.36	45.28

Region	Country	GDP pc	Urban share (%)					
			AI			Cluster-based		
			LScan	*GHSPop*	*WPop*	*LScan*	*GHSPop*	*WPop*
EAP	Timor-Leste	2,102	15.93	14.92		36.08	50.88	
SSA	Chad	2,016	17.47	25.95		25.88	85.55	
SAR	Afghanistan	1,877	42.17	43.91		50.79	81.56	
SSA	Sierra Leone	1,776	32.26	42.39		30.15	63.80	
SSA	Benin	1,733	59.63	69.86		54.73	82.77	
SSA	Zimbabwe	1,677	35.14	45.09		31.44	71.59	
SSA	Burkina Faso	1,562	30.44	48.32		20.40	73.97	
SSA	Rwanda	1,528	33.50	36.43	35.18	85.53	88.98	93.13
SSA	Mali	1,490	29.04	39.73	31.36	29.16	67.65	29.92
SSA	Madagascar	1,371	26.21	28.65		26.09	67.64	
SSA	Guinea-Bissau	1,331	34.03	42.73		31.28	72.61	
SSA	Ethiopia	1,329	17.75	22.73		37.79	87.87	
SSA	Togo	1,327	55.17	71.10		50.90	82.70	
SSA	Guinea	1,182	31.59	41.82		28.00	64.23	
SSA	Mozambique	1,040	27.65	35.51		28.03	80.81	
SSA	Niger	878	27.52	40.19		23.27	79.46	
SSA	Malawi	766	30.70	29.57	31.34	52.21	57.44	48.70
SSA	Burundi	726	43.29	47.92		76.87	86.97	
SSA	Central African Republic	684	27.02	28.80		33.81	65.83	
SSA	Congo, Dem. Rep.	675	36.73	42.64		49.68	73.99	

Note: Latin American & Caribbean (LAC) countries are shown in bold font.

Table A4.2: 65 countries excluded from the regression analyses (due to missing values)

Region	Country	GDP pc	Urban share (%)					
			AI			Cluster-based		
			LScan	*GHSPop*	*WPop*	*LScan*	*GHSPop*	*WPop*
MNA	Qatar	136,171	91.22	86.18		92.04	95.83	
EAP	Macao SAR, China	131,590		99.96		83.71	99.68	
EAP	Hong Kong SAR, China	51,518	99.59	99.83		97.45	99.53	
NAC	Bermuda	51,403					75.57	
EAP	Australia	42,860	76.19	78.35			81.10	
MNA	Bahrain	41,847	98.67	99.80		94.93	99.03	
ECA	France	37,248	81.18	81.70			63.57	

Region	Country	GDP pc	Urban share (%)					
			AI			Cluster-based		
			LScan	*GHSPop*	*WPop*	*LScan*	*GHSPop*	*WPop*
LAC	**Puerto Rico**	**33,699**	**90.15**	**90.35**		**82.86**	**83.89**	**77.29**
SSA	Equatorial Guinea	31,255	34.79	46.46		50.20	61.45	
MNA	Israel	30,875	94.03	90.63		91.33	92.97	
MNA	Malta	28,546				95.45	95.09	
SSA	Seychelles	24,509					63.22	
LAC	**St. Kitts and Nevis**	**20,759**					**62.16**	
LAC	**Antigua and Barbuda**	**20,576**					**74.71**	**63.46**
MNA	Libya	18,997	84.69	87.29		82.20	84.82	
MNA	Iran, Islamic Rep.	15,791	76.09	80.80		73.41	85.67	
MNA	Iraq	14,487	79.10	83.95		77.56	91.37	
EAP	Palau	13,587					40.23	
SAR	Maldives	13,130	37.91	76.50			60.79	
LAC	**Grenada**	**11,238**					**61.94**	
LAC	**Dominica**	**10,122**					**47.51**	
LAC	**St. Lucia**	**10,113**				**56.15**	**77.23**	
LAC	**St. Vincent & the Grenadines**	**10,105**					**66.67**	
ECA	Kosovo	8,601	77.50	87.14			75.89	
EAP	Tonga	4,966					73.22	
EAP	Marshall Islands	3,715					68.53	
MNA	Yemen, Rep.	3,636	35.49	36.37		60.08	85.23	
EAP	Tuvalu	3,508						
EAP	Micronesia, Fed. Sts.	3,352					21.24	
MNA	Djibouti	2,937	61.09	58.13		73.56	83.95	
EAP	Vanuatu	2,895					26.99	
EAP	Papua New Guinea	2,521	7.12	6.92		25.36	75.29	
SSA	South Sudan	2,087	14.43			14.72	89.88	
EAP	Solomon Islands	2,034	16.43	17.12		20.16	16.84	
EAP	Kiribati	1,694					53.52	
SSA	Uganda	1,672	43.57	48.66		59.20	84.77	46.41
LAC	**Haiti**	**1,622**	**51.88**	**57.86**	**56.44**	**60.79**	**88.78**	**63.64**
SSA	Gambia, The	1,567	51.34	59.48	55.57	48.64	76.58	51.85
SSA	Eritrea	1,468	35.22	29.11		39.89	86.75	

Region	Country	GDP pc	Urban share (%)					
			AI			Cluster-based		
			LScan	*GHSPop*	*WPop*	*LScan*	*GHSPop*	*WPop*
SSA	Comoros	1,415				77.71	85.15	
SSA	Liberia	797	36.55	40.81		34.77	54.62	
LAC	**Aruba**						**93.60**	
ECA	Andorra						75.17	
LAC	**Argentina**		**74.08**	**81.52**	**73.68**	**75.90**	**89.57**	**62.38**
EAP	American Samoa						68.53	
LAC	**Curacao**		**94.08**				**89.07**	
LAC	**Cayman Islands**						**89.43**	
ECA	Faeroe Islands						29.86	
ECA	Greenland						38.77	
EAP	Guam					74.78	82.92	
ECA	Isle of Man						69.02	
ECA	Liechtenstein						82.13	
ECA	Monaco		100.00	100.00			99.27	
EAP	Myanmar		49.78	57.46		55.49	83.67	40.98
EAP	Northern Mariana Islands						76.02	
EAP	New Caledonia		54.81	68.04		45.45	64.61	
EAP	Korea, Dem. Rep.		62.63	64.77		69.49	87.34	
EAP	French Polynesia					60.56	41.34	
ECA	San Marino		3.57				87.98	
SSA	Somalia		23.26	27.54		33.30	74.43	
LAC	**Sint Maarten (Dutch part)**						**94.82**	
MNA	Syrian Arab Republic		75.46	85.35		72.26	79.78	
LAC	**Turks and Caicos Islands**						**61.27**	
EAP	Taiwan, China		96.88	97.74			95.47	
LAC	**Virgin Islands (U.S.)**						**75.46**	

Note: Latin American & Caribbean (LAC) countries are shown in bold font.

Annex 5: AI Regression Results for *GHS-Pop* and *WorldPop* Data

Table A5.1: Regression results for relationship between ln(GDP per capita) and urbanization

	AI – *GHS Pop*			AI – *World Pop*		
	[1]	[2]	[3]	[4]	[5]	[6]
Urban pop. share	3.923***	3.971***	3.938***	2.902***	2.040***	2.075***
	[10.67]	[10.51]	[10.57]	[4.187]	[3.118]	[3.348]
LA_dummy		0.001			0.809***	
		[0.006]			[3.859]	
C_dummy		-0.246			0.875**	
		[-0.644]			[2.374]	
LAC_dummy			-0.055			0.817***
			[-0.280]			[4.041]
Constant	6.601***	6.580***	6.601***	7.362***	7.362***	7.346***
	[26.23]	[25.73]	[26.14]	[19.38]	[21.91]	[22.90]
n	146	146	146	40	40	40
R^2	0.441	0.443	0.442	0.316	0.526	0.525
$\overline{R^2}$	0.438	0.431	0.434	0.298	0.486	0.500

Dependent variable: ln(GDP per capita)
Notes: t-statistics in brackets. *** Significant at 1 % level; ** at 5 % level; * at 10 % level

Table A5.2: Regression results for relationship between urbanization and the share of national GDP derived from agriculture

	AI – *GHS Pop*			AI – *World Pop*		
	[1]	[2]	[3]	[4]	[5]	[6]
Agric. share	-0.952***	-0.924***	-0.939***	-0.976***	-0.986***	-1.055***
	[-8.847]	[-8.471]	[-8.607]	[4.657]	[-3.596]	[-3.884]
LA_dummy		0.001			-0.035	
		[0.015]			[-0.632]	
C_dummy		0.112			0.064	
		[1.633]			[0.735]	
LAC_dummy			0.027			-0.025
			[0.735]			[-0.464]
Constant	0.777***	0.769***	0.771***	0.666***	0.677***	0.692***
	[40.27]	[36.56]	[36.58]	[17.94]	[10.12]	[10.39]
n	146	146	146	40	40	40
R^2	0.352	0.364	0.355	0.363	0.396	0.367
$\overline{R^2}$	0.348	0.351	0.346	0.347	0.346	0.333

Dependent variable: urban share of national population
Notes: t-statistics in brackets. *** Significant at 1 % level; ** at 5 % level; * at 10 % level

Annex 6: Cluster Method Regression Results

Table A6.1: Regression results for relationship between ln(GDP per capita) and urbanization

	Cluster – *LandScan (HDC only)*			Cluster – *LandScan*		
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**
Urban pop. share	3.604***	3.750***	3.761***	3.789***	3.807***	3.816***
	[7.040]	[6.996]	[7.032]	[8.262]	[8.091]	[8.167]
LA_dummy		-0.317			-0.100	
		[-1.210]			[-0.407]	
C_dummy		0.014			0.017	
		[0.031]			[0.041]	
LAC_dummy			-0.239			-0.072
			[-1.014]			[-0.330]
Constant	7.924***	7.914***	7.910***	6.842***	6.844***	6.838***
	[40.39]	[40.17]	[40.24]	[23.35]	[23.11]	[23.24]
n	146	146	146	146	146	146
R^2	0.256	0.264	0.261	0.322	0.322	0.322
$\overline{R^2}$	0.251	0.248	0.251	0.317	0.308	0.313

Dependent variable: ln(GDP per capita)
Notes: t-statistics in brackets. *** Significant at 1 % level; ** at 5 % level; * at 10 % level

Table A6.2: Regression results for relationship between urbanization and the share of national GDP derived from agriculture

	Cluster – *LandScan (HDC only)*			Cluster – *LandScan*		
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**
Agric. share	-0.629***	-0.584***	-0.582***	-0.740***	-0.714***	-0.719***
	[-6.439]	[-6.028]	[-6.059]	[-7.350]	[-6.973]	[-7.068]
LA_dummy		0.100***			0.036	
		[2.808]			[0.966]	
C_dummy		0.089			0.073	
		[1.465]			[1.133]	
LAC_dummy			0.098***			0.045
			[3.062]			[1.333]
Constant	0.427***	0.404***	0.404***	0.710***	0.698***	0.699***
	[24.35]	[21.64]	[21.75]	[39.27]	[35.38]	[35.58]
n	146	146	146	146	146	146
R^2	0.224	0.271	0.271	273	0.283	0.282
$\overline{R^2}$	0.218	0.256	0.261	0.268	0.268	0.272

Dependent variable: urban share of national population
Notes: t-statistics in brackets. *** Significant at 1 % level; ** at 5 % level; * at 10 % level

Table A6.3: Regression results for relationship between ln(GDP per capita) and urbanization

	Cluster – *GHS Pop (HDC only)*			Cluster – *GHS Pop*		
	[1]	**[2]**	**[3]**	**[1]**	**[2]**	**[3]**
Urban pop. share	1.727***	1.674***	1.684***	0.396	0.277	0.335
	[3.010]	[2.877]	[2.900]	[0.485]	[0.336]	[0.409]
LA_dummy		0.047			0.136	
		[0.164]			[0.461]	
C_dummy		0.440			0.547	
		[0.903]			[1.089]	
LAC_dummy			0.141			0.233
			[0.548]			[0.888]
Constant	8.437***	8.435***	8.431***	8.867***	8.917***	8.873***
	[32.29]	[32.12]	[32.17]	[14.07]	[14.05]	[14.07]
n	146	146	146	146	146	146
R^2	0.059	0.065	0.061	0.002	0.011	0.007
$\overline{R^2}$	0.053	0.045	0.048	-0.005	-0.010	-0.007

Dependent variable: ln(GDP per capita)

Notes: t-statistics in brackets. *** Significant at 1 % level; ** at 5 % level; * at 10 % level

Table A6.4: Regression results for relationship between urbanization and the share of national GDP derived from agriculture

	Cluster – *GHS Pop (HDC only)*			Cluster – *GHS Pop*		
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**
Agric. share	-0.275**	-0.252**	-0.253**	0.011	0.034	0.025
	[-2.532]	[-2.279]	[-2.303]	[0.141]	[0.411]	[0.303]
LA_dummy		0.046			0.012	
		[1.118]			[0.414]	
C_dummy		0.051			0.080	
		[0.740]			[1.551]	
LAC_dummy			0.047			0.028
			[1.289]			[1.042]
Constant	0.459***	0.448***	0.448***	0.761***	0.753***	0.755***
	[23.54]	[21.00]	[21.13]	[52.71]	[47.83]	[47.94]
n	146	146	146	146	146	146
R^2	0.043	0.054	0.054	0.000	0.017	0.008
$\overline{R^2}$	0.036	0.034	0.040	-0.007	-0.003	-0.006

Dependent variable: urban share of national population

Notes: t-statistics in brackets. *** Significant at 1 % level; ** at 5 % level; * at 10 % level

Table A6.5: Regression results for relationship between ln(GDP per capita) and urbanization

	Cluster – *World Pop (HDC only)*			Cluster – *World Pop*		
	[1]	[2]	[3]	[5]	[6]	[7]
Urban pop. share	2.831**	2.043**	2.032**	0.244	0.480	0.555
	[2.636]	[2.303]	[2.286]	[0.320]	[0.782]	[0.916]
LA_dummy		0.877***			0.999***	
		[4.031]			[4.359]	
C_dummy		1.259**			1.330***	
		[3.537]			[3.532]	
LAC_dummy			0.947***			1.062***
			[4.550]			[4.899]
Constant	8.149***	7.833***	7.835***	8.751***	8.044***	8.002***
	[25.94]	[26.58]	[26.67]	[20.43]	[21.43]	[21.56]
n	40	40	40	40	40	40
R^2	0.155	0.475	0.458	0.003	0.408	0.395
$\overline{R^2}$	0.132	0.431	0.429	-0.024	0.358	0.362

Dependent variable: ln(GDP per capita)
Notes: t-statistics in brackets. *** Significant at 1 % level; ** at 5 % level; * at 10 % level

Table A6.6: Regression results for relationship between urbanization and the share of national GDP derived from agriculture

	Cluster – *World Pop (HDC only)*			Cluster – *World Pop*		
	[1]	[2]	[3]	[5]	[6]	[7]
Agric. share	-0.542***	-0.684***	-0.655***	-0.241	-0.547	-0.588
	[-3.253]	[-3.105]	-3.055]	[-0.845]	[-1.480]	[-1.635]
LA_dummy		-0.033			-0.117	
		[-0.744]			[-1.587]	
C_dummy		-0.073			-0.059	
		[-1.049]			[-0.504]	
LAC_dummy			-0.036			-0.112
			[-0.843]			[-1.541]
Constant	0.338***	0.381***	0.374***	0.567***	0.671***	0.679***
	[11.44]	[7.084]	[7.125]	[11.23]	[7.439]	[7.717]
n	40	40	40	40	40	40
R^2	0.218	0.242	0.233	0.018	0.086	0.078
$\overline{R^2}$	0.197	0.179	0.191	-0.007	0.010	0.028

Dependent variable: urban share of national population
Notes: t-statistics in brackets. *** Significant at 1 % level; ** at 5 % level; * at 10 % level