world development report 2009

Reshaping Economic Geography

BACKGROUND PAPER

AGGLOMERATION INDEX: TOWARDS A NEW MEASURE OF URBAN CONCENTRATION

HIROTSUGU UCHIDA

Department of Environmental and Natural Resource Economics
University of Rhode Island

and

ANDREW NELSON

Institute for Environment and Sustainability
Global Environment Monitoring Unit
European Commission—DG Joint Research Centre

Current version: February 15, 2008

Agglomeration Index: Towards a New Measure of Urban Concentration*

Hirotsugu Uchida
Research Assistant Professor
Department of Environmental and Natural Resource Economics
University of Rhode Island
1 Greenhouse Rd, Kingston, RI 02879, U.S.A.
and
Consultant
World Development Report 2009
World Bank
1818 H St NW, Washington, D.C., 20433, U.S.A.
uchida@uri.edu

Andrew Nelson
Institute for Environment and Sustainability
Global Environment Monitoring Unit
European Commission—DG Joint Research Centre
TP 440, Via Fermi 2749, I-21027, Ispra(VA), Italy
andrew.nelson@jrc.it

Draft as of February 15, 2008
Do not cite without permissions from the authors

* This paper was prepared as a background paper for World Development Report 2009. The authors would like to thank Uwe Deichman, Timothy Thomas, Ravi Kanbur, Indermit Gill, and David Satterthwaite, participants in the conference "Beyond the Tipping Point: Development in an Urbanizing World" held in London on October 18-20, 2007 for their comments. The usual disclaimer applies.

Agglomeration Index: Towards a New Measure of Urban Concentration

Abstract

A common challenge in analyzing urbanization is the data. The United Nations compiles information on urbanization (urban population and its share of total national population) that is reported by various countries but there is no standardized definition of "urban," resulting in inconsistencies. This situation is particularly troublesome if one wishes to conduct a cross-country analysis or determine the aggregate urbanization status of the regions (such as Asia or Latin America) and the world. This paper proposes an alternative measure of urban concentration that we call an agglomeration index. It is based on three factors: population density, the population of a "large" urban center, and travel time to that large urban center. The main objective in constructing this new measure is to provide a globally consistent definition of settlement concentration to conduct cross-country comparative and aggregated analyses in the same way that the $1 per day poverty line is used in poverty issues-related studies. As an accessible measure of economic density, the agglomeration index lends itself to the study of concepts such as agglomeration rents in urban areas, the "thickness" of a market, and the travel distance to such a market with many workers and consumers. With anticipated advances in remote sensing technology and geo-coded data analysis tools, the agglomeration index can be further refined to address some of the caveats currently associated with it.

1. Introduction

One common challenge in analyzing urbanization is the data. The United Nations (UN) compiles information on urbanization that is reported by various countries and makes the data readily available. But one must understand the nature of the data provided. Cohen (2004) summarizes the issues stemming from the limitations in the UN data:

> "Although invaluable to those interested in studying urban change, the data in the UN report are somewhat deceptive in their apparent completeness and there is a great deal of misunderstanding and misreporting by nonspecialists about what these data mean and how they should be interpreted. … Most end-users cite the UN data as if it is absolute truth rather than treating them as simply indicative of general broad trends. There is a general underappreciation of the fact that the UN is forced to rely on member countries' existing definitions of what constitutes an urban or a rural area. Not only do these definitions differ widely by country, in many places the traditional urban/rural dichotomy is becoming increasingly inadequate" (p.24–25)

For example, the share of India's population that resided in urban areas in 1991 would be 39% instead of the official figure of 26% if 113 million inhabitants of 13,376 villages with populations of 5,000 or more were classified as urban. The share would be even higher if the Swedish definition of urban (settlements with more than 200 inhabitants) were applied. The notion that South Asia is densely populated but that a relatively small proportion of the population lives in urban areas may not be as paradoxical as it sounds. In Mexico, the urban population in 2000 was 74.4% when settlements of 2,500 or more were defined as urban. If that threshold was changed to 15,000 or more (Nigeria and Syria, for example, have cut-offs of 20,000), the urban share of the population would drop to 67%. In Mauritius in 2000, about a quarter of the population resided in settlements of between 5,000 and 20,000 inhabitants, some of which were district capitals but were not classified as urban. If they were reclassified, the urban percentage in 2000 would have been more than two-thirds instead of less than half (42.7%). A country's definition also can change over time, adding yet another layer of confusion. In China, for example, the urban share in 1999 could have been 24%, 31%, or 73% depending on the official definition of urban population used (Satterthwaite 2007). These examples should not be interpreted as asserting that one definition is correct and another incorrect. There is a problem, however, and it is one that is particularly critical when making a cross-country analysis.

Researchers in the field are aware of the issues inherent in the UN data. At the heart of the problem is how to measure urban concentration in a consistent and systematic way. One option used in the literature is the Hirschman-Herfindahl index of concentration, which is constructed by squaring the share of population apportioned to each city in the national urban population and summing those squares (e.g., Wheaton and Shishido 1981). Another option is to use the Pareto parameter, which indicates how quickly city size declines as one moves from the largest to the smallest in the size distribution (e.g., Rosen and Resnick 1980). Others use primacy, which is measured by the share of population contained in largest city—or metro area—in a national urban population (e.g., Ades and Glaeser 1995). Since the UN's World Urbanization Prospects report includes data on the population of settlements with more than 750,000 inhabitants, data on primacy are available for many countries across time. The availability of data and its relation to Zipf's law in a distribution and ranking of cities based on size have prompted researchers to use primacy as a measure of urban concentration when the analysis calls for wide and long panel data as when studying the relationship between urbanization and economic growth (e.g., Henderson 2000, 2003).

Note, however, that these three measures of concentration still depend heavily on (a) how a city (or metropolitan area) is defined and (b) how an urban area is defined. Henderson (2005) points out several issues regarding systematic and consistent definitions of a city or metropolitan area both across countries and over time. Satterthwaite (2007) presents examples of how a metropolitan area can be assigned markedly different population sizes depending on how the area is defined. The population of Mexico City, for example, ranges from 1.7 million for the "central city" to 19.4 million for the "megalopolis of central Mexico" (p.13–14). Data on national urban populations are also necessary to calculate percentage for the Hirschman-Herfindahl index or primacy. Cohen (2004) outlines the difficulties associated with obtaining and interpreting urban population data.

This paper proposes an alternative measure of urban concentration that we call the agglomeration index. It is based on three factors: population density, the size of the population in a "large" urban center, and travel time to that urban center. Each factor used in the index is based on the conceptual framework of agglomeration economies. The index does not define what is urban *per se*—it does not incorporate urban characteristics such as political status and the presence of particular services or activities. Instead, the index creates a globally definition of settlement concentration that could be used to conduct cross-country comparative analyses. The accessibility of this measure of economic lends itself easily to the study of concepts such as agglomeration rents in urban areas, the "thickness" of a market, and the travel distance to such a market with many workers and consumers.

A new measure of agglomeration does not suggest that the UN's data are flawed. The matter is analogous to measurements of global poverty and comparisons of poverty levels across countries. Each country has its own definition based on legitimate factors, but the varying definitions among countries make cross-country analysis and aggregation nearly impossible. A uniform definition—like the $1 or $2 oera day index used in evaluating poverty levels across countries—makes analysis possible. This study likewise uses both country-reported measures and the agglomeration index as appropriate to examine national and subnational trends.

In addition to the limitations of the UN's data, there are caveats inherent to the agglomeration index (described in detail later) related to the plausibility of key assumptions associated with travel time. Finally, but perhaps most importantly, a critical constraint is that the index is available for only one year (2000).[1]

In section 2 we explain the concepts underlying the agglomeration index and then address the data and the mechanics of calculating the index in section 3. Section 4 presents the results, compares them with the results from the UN data, and an analysis of the index's properties. Section 5 discusses the strengths, caveats, and promise of the agglomeration index.

2. Concepts behind the Agglomeration Index

[1] GRUMP (Global Rural-Urban Mapping Project) human settlement data is developed by Center for International Earth Science Information Network (CIESIN) at Columbia University. The data were gathered primarily from official statistical offices (census data) and secondarily from other web sources, or from specific individual databases when official statistical databases were not available. Based on the data available and applying UN growth rates, population was estimated for 1990, 1995, and 2000 (c.f. http://sedac.ciesin.columbia.edu/gpw/index.jsp).

The proposed agglomeration index does not focus on the conceptual definition of "urban;" rather, it focuses on the economic significance of urban areas. Residents, workers, and firms are typically concentrated—or agglomerated—in urban areas, which gives rise to the notion of agglomeration economies. Clearly, people and firms agglomerate because there are benefits from doing so. The literature cites several sources of agglomeration rent, such as the existence of "thick" markets (both consumer and labor markets), ease of access to these markets, and the resulting, so-called forward and backward linkages associated with large local markets (Marshall 1961; Baldwin et al. 2003).

What are the key indicators of these sources of agglomeration economies and rent? We argue that it can be summarized into the following three indicators: population size, population density, and travel time. More specifically, population size refers to that of large city(ies) that can be regarded as the focal point(s) of an urban area, and travel time is to the nearest large city. The idea is illustrated in figure 1. To have a thick market, there must be a certain mass of people—a large city. We assume that a large city— defined by applying a high population threshold to the GRUMP human settlement data[2]—has a dense population: a city with a population size of 100,000 or more and a population density of *less* than 150 people per square kilometer does not exist.

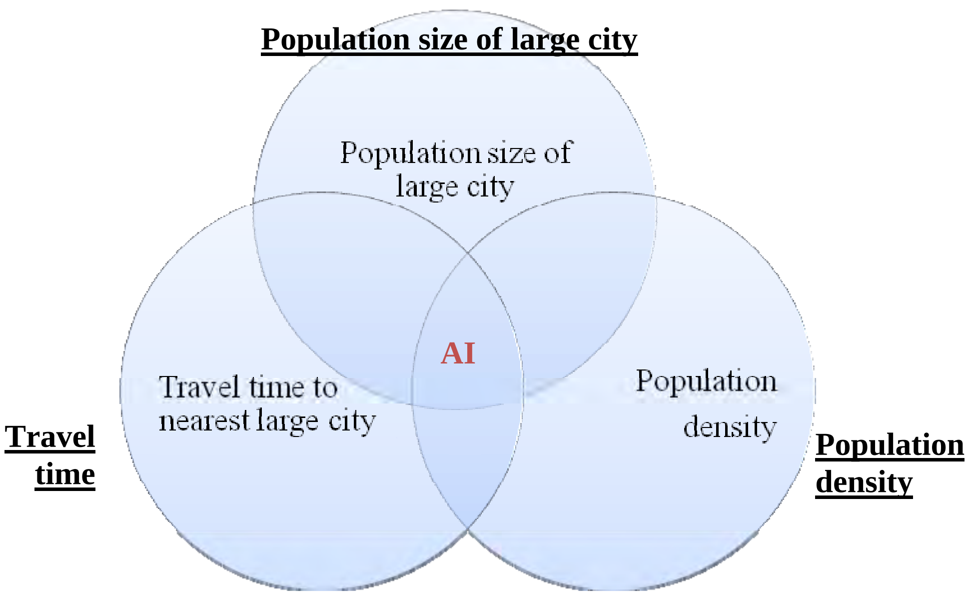

Figure 1. Key indicators constituting the agglomeration index (AI)

Consider some location outside of a large city's center. For that location to take part in agglomeration economies, it must, generally speaking, have both a relatively high population density, which is a proxy for market thickness, and be reasonably near the large city center, a proxy for market access and lower transportation costs. For example, even with a high population density, a small, isolated community such as a mining town is not likely to generate

[2] Here we rely on the GRUMP data to define what a city is, but definitions of urban vary from country to country and within countries the population of a single urban agglomeration can vary enormously depending on how it is defined, as described previously with Mexico City's example.

agglomeration economies. As depicted in figure 1, locations that satisfy all three indicators are included in calculating the agglomeration index.

The urban area thus can be identified and delineated by the combination of the three indicators. This is a significant step forward because measures of concentrations of settlement no longer depend on country-specific and sometimes ad hoc definitions of a city, an administrative boundary, or an urban area. Rather than treating urban and rural areas as dichotomous and discreet entities, we view the spatial transition as a gradient, a conceptual framework put forth by Chomitz, Buys, and Thomas (2005).

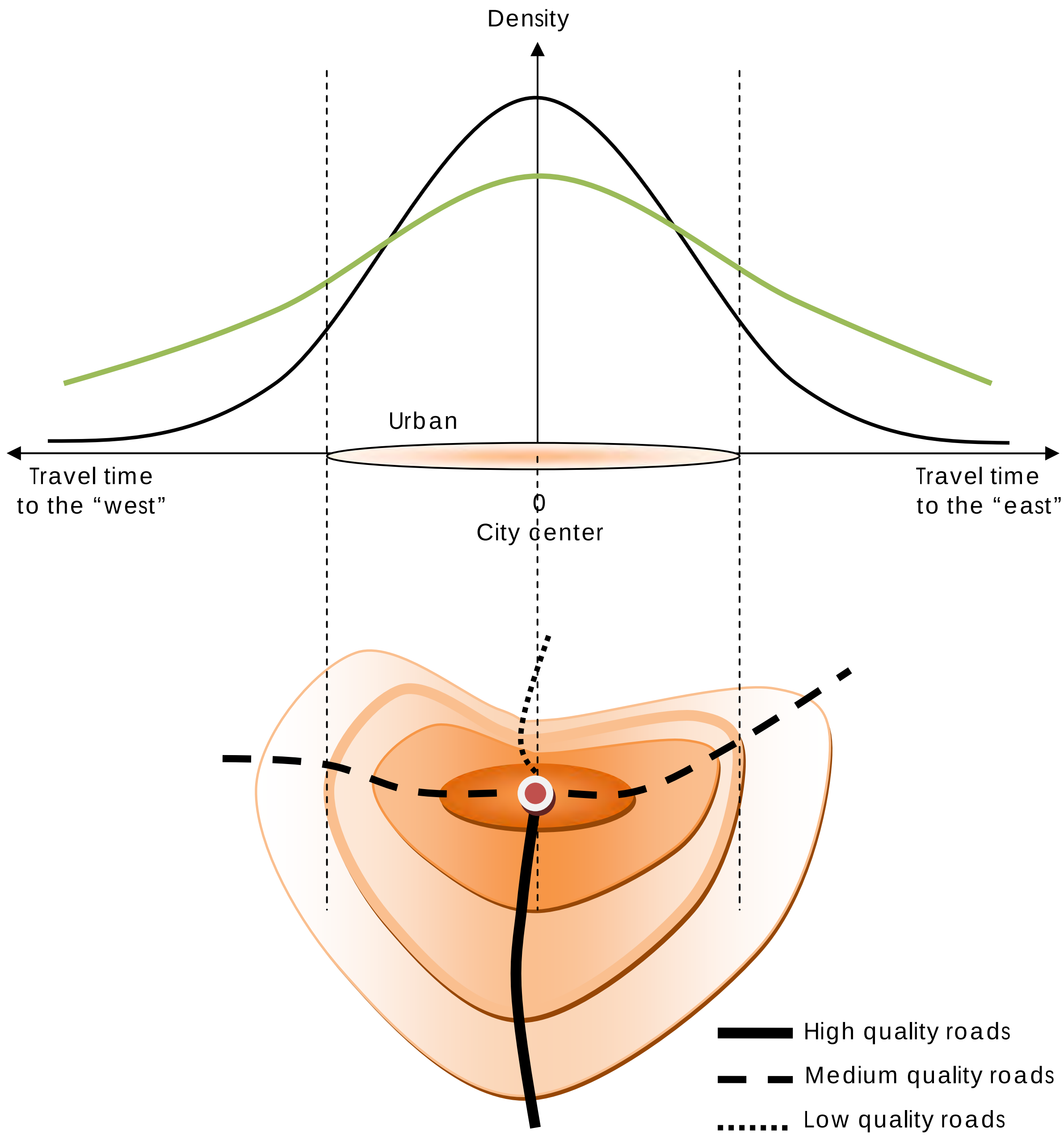

Figure 2 Concepts of the urban-rural gradient and travel time in the agglomeration index

The gradient framework is illustrated in figure 2. Assuming that density is the highest at the center of a large city[3] and that it gradually decreases as one moves away from the center, in calculating the agglomeration index, we defined some maximum travel time to the city center as the cut-off point (provided that the density criterion was met) of an urban area. Depending on the population density distribution, the density at the boundary varies; this is illustrated in the top panel of figure 2 by two distributions, one being a mean-preserving spread of the other.

The difference in spatial distributions of population density raises another important point—*why* are they different? It is easy to imagine a region with poor transportation infrastructure will result in a distribution that is concentrated near the center A flatter distribution of population density, then, may imply a better transportation infrastructure—accessibility over five miles of poor transportation infrastructure could be equivalent to ten miles over a good infrastructure system. In our agglomeration index, we used road networks as a proxy for transportation infrastructure. We assigned each road network to one of three categories of quality—low, medium, and high—and assigned an arbitrary traveling speed to each category. As a result, a physical distance identified by 60 minutes of travel time is further from the city center along a paved road than along a dirt road. The bottom panel of figure 2 presents the case in which concentric circle-like figures depict travel-time contours or isochrones. Note that further locations along the high-quality road are included in the urban area. For areas without roads we assumed the mode of transportation is either on foot or a ride on an animal, and incorporated information on land cover (grassland versus forest) and the slope of land in calculating the travel time.

3. Data and How the Agglomeration Index is Calculated

The procedure to calculate the agglomeration index can be summarized as follows:

1. Specify a threshold value to each of the three criteria: minimum population density, maximum travel time, and minimum population size that define large cities.
2. Locate the center of defined large cities from the GRUMP human settlements database.
3. Determine the border surrounding that large city center based on the maximum travel time. This boundary is computed from a cost-distance model that estimates travel time to the city over a *cost surface*. This surface has a spatial resolution of approximately 1km and is derived from GIS data on; (i) the transport network, (ii) off road surfaces derived from land cover data and (iii) slope and estimates of the average travel speeds for each permutation of these data (see table A.1 in the appendix).
4. Create (i) population and (ii) population density grids at 1km spatial resolution for year 2000. This is based on the average of two global gridded population data sources, GRUMP LandScan.[4]
5. Aggregate the population of all the grid cells that satisfy all three thresholds. The result is analogous to urban population. The proportion of this number to that country's total population is the agglomeration index.

[3] Generally true for the economically active percentage of the population
[4] LandScan was developed by Oak Ridge National Laboratory; http://www.ornl.gov/sci/landscan/.

A country's census data is the foundation of agglomeration index. The GRUMP population estimates are based on population data at the smallest available scale in the national census (state, province, county, or district), combined with web sources or from specific individual databases when official statistical databases were not available. GRUMP does not model population distribution within these administrative units. Conversely the LandScan population estimates are based on population estimates for larger units but the population within each unit is redistributed across the unit's grid cells based on likelihood coefficients, which are derived from other spatial data such as distance to roads, slope, and land cover. There are strengths and weaknesses in both of these population models, so we used both and took the computed the population per 1km pixel as an average of both sources.

Travel time to large cities is based on estimates of the time required to travel 1km over different road and off road surfaces. We constructed a cost surface from the data sources listed in table x and applied the estimated travel speeds in table y to each permutation of surfaces (i.e., major road on a moderate slope). The costsurface contains the time in minutes to cross each 1km cell. This surface and the location of the city centers are input to the costdistance model to determine the travel time to each city centre.

4. Results

Comparison with the UN data

Figure 3 presents a comparison between the agglomeration index and the UN's urbanization figures. The index is based on a minimum threshold of 150 people per square kilometer for density, 60 minutes of maximum for travel time, and a minimum of 50,000 inhabitants as defining a large city. Each country's agglomeration index and UN urban population share are given in a table A.2 in the appendix. Here, we present the results by world-regions as defined by the World Bank: Sub-Sahara Africa (SSA), Middle East and North Africa (MENA), South Asia (SAS), East Asia and the Pacific (EAP), Europe and Central Asia (ECA), and Latin America and the Caribbean (LAC). OECD and non-OECD high-income countries are excluded from these regional groups; the latter is called Other High Income Economies (OHIE).

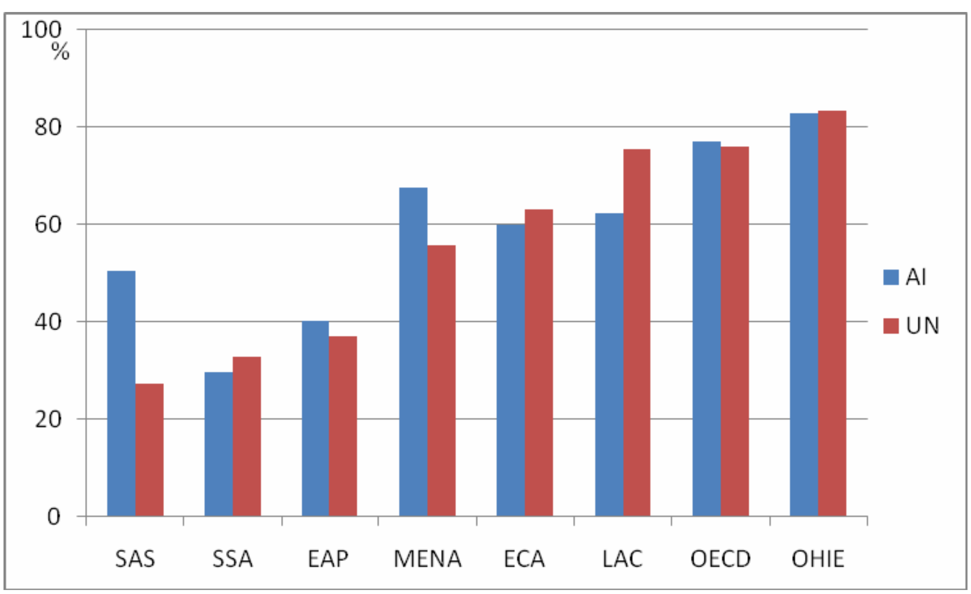

Figure 3. Agglomeration index and UN urban population share for year 2000

Source: United Nations (2006).

The results of the two calculations are qualitatively consistent given the issues associated with the UN's data. For example, the agglomeration index for South Asia (SAS) is much higher than the UN result (50.4% versus 27.2%). This suggests that the notion of south Asia as being densely populated but having a low urban population share may not be as paradoxical as it sounds. A similar pattern—where the agglomeration index is higher than the UN figure—is observed for the East Asia and the Pacific and for the Middle East and North Africa. The converse is true, however, for regions such as Latin America and the Caribbean (LAC) and Sub-Saharan Africa. The LAC region is of particular interest as this region has often been cited as having the most urbanized region among the developing countries. According to the UN figure, urban areas accounted for 75.4% of the region's population in 2000; the agglomeration index puts the urban share at 62.4%. Thus, the LAC region is probably still heavily urbanized but not as much as commonly thought. In fact, the LAC region's agglomeration index is lower than that of MENA's (67.5%) and about the same as Europe and central Asia's (ECA's) (59.8%). Lastly, one can observe that the agglomeration index and the UN figures are virtually identical for high-income countries (OHIE and OECD).

Sensitivity of the results to thresholds

The danger of using the agglomeration index to infer the state of urbanization in the world in place of more conventional urban share is that the results can be manipulated to fit any conclusion by changing the threshold/criteria combination. It is very important, therefore, to have good justification for the combination chosen and to interpret the results as *conditional* on that choice. This cautionary note raises an interesting question, however: To what extent agglomeration index sensitive to different combinations of thresholds?

We considered three levels of threshold for each indicator and calculated an agglomeration index for every combination. Those levels are:

- Population density: 150, 300, and 500 people per square kilometer;
- Travel time: 30, 60, and 90 minutes;
- Large city's population: 50,000, 100,000, and 500,000 inhabitants.

The base combination was 150 people per square kilometer, travel time of 60 minutes, and a minimum of 50,000 inhabitants to qualify as a large. The minimum density of 150 people per square kilometer is what OECD used for its statistics (Chomitz, Buys, and Thomas 2005) and is equivalent to placing each person approximately 81.6 meters apart. We viewed this threshold as rather sparse, especially for certain parts of the world, which led us to try thresholds of 300 and 500 people per square kilometer (corresponding to placing each person about 57.7 meters and 44.7 meters apart, respectively). For the large city's population size, inclusion of a minimum population of 20,000 was suggested. However, it became apparent that the GRUMP settlement data does not have sufficient coverage at this low population level in many countries.

There are 27 different combinations in total; however we present only combinations in which a single factor was increased or decreased from the base case and attempt to deduce trends. Figure 4 shows the results of changing the population density threshold while holding the other two criteria constant at 60 minutes (travel time) and 50,000 (population size). As expected, with a more stringent threshold (i.e., a higher population density cut-off level) the agglomeration index falls, though not by a large amount. Consequently, the relative magnitudes of the index for all of the regions with respect to the UN figures are qualitatively the same as the base case. The agglomeration indexes are higher for the SAS and MENA regions and lower for the SSA, LAC, and ECA regions. Remarkably, the results are fairly robust against the changes in population density threshold.

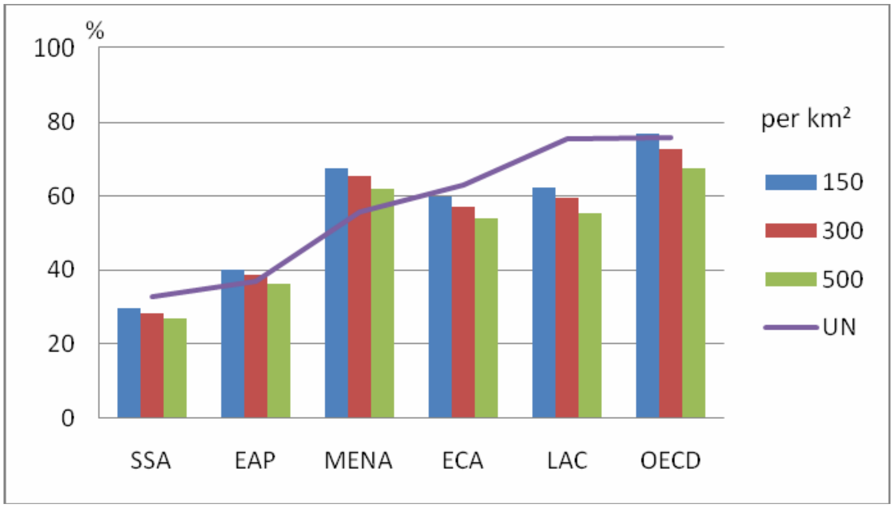

Figure 4. Sensitivity to indicators: minimum population density

Figure 5 shows how the indexes change as the travel time threshold is changed while holding the other two indicators constant at base-case levels. Unlike the case of the population density criterion, changes in travel time result in large changes in the agglomeration index. The difference is particularly significant in the SAS and EAP regions—with a moderate increase in the threshold (from 60 to 90 minutes), the agglomeration index for the SAS region jumps to 60.8%, more than double the UN figure. The results suggest that the choice of the travel time threshold can easily alter the conclusions about a country's urbanization state, a cautionary note when using the index.

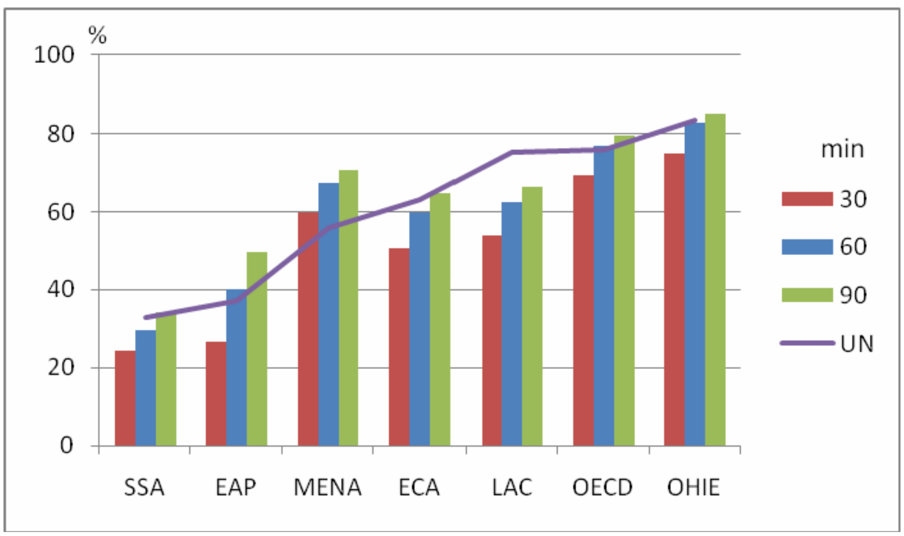

Figure 5. Sensitivity to indicators: maximum travel time

Lastly, figure 6 shows the results when the minimum population size for large cities is changed. This case also generates a substantial change in the values of the resulting agglomeration indexes and the magnitude of the change is not affected by whether the region is developing or developed. The drop is particularly steep when going from 100,000 to 500,000 inhabitants for the large city threshold. The figure suggests that a threshold of 500,000 may not be suitable for many regions because it is too restrictive, driving the agglomeration indexes too low both in absolute terms and relative to the UN figures.

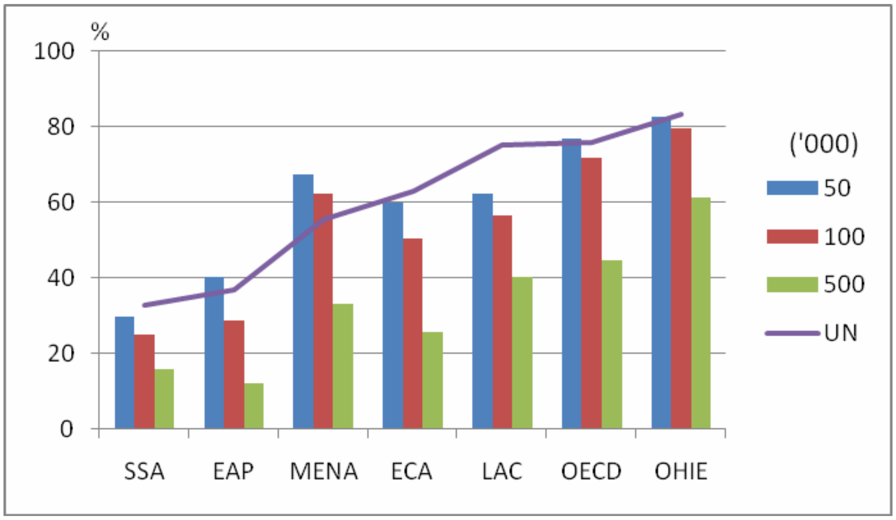

Figure 6. Sensitivity to indicators: minimum population size of large cities

One last note regarding the choice of combinations of thresholds concerns feasibility. The current method estimates travel time from the center of the nearest large city. However, large cities vary in the degree to which they are dispersed geographically based both on their populations and on their density. So, while 30 minutes away from a densely populated city of 50,000 may be genuinely outside of that city's boundary (however defined), this is less likely for a city of 500,000. The bottom line is that threshold combinations such as a city having a minimum population of 500,000 and a travel time of 30 minutes or less may simply be implausible.

5. Discussions: Strength, Caveats, and Promise

Provided that the conceptual framework and the variables used to calculate the agglomeration index are plausible, the main advantage of this measure is the consistent measurement across the countries that enables direct comparisons among them. While commonly used estimation techniques fundamentally rely on census data, the agglomeration index is not influenced by country-specific definitions of what is urban. The index, however, is not isolated from ad hoc definitions of city boundaries since it uses city population as a criterion by which to define and locate large cities. However, unlike the primacy measure, the agglomeration index uses this information only to identify the point at which the travel time is measured; in other words, population figures are not used in the calculations of the index. Therefore, the accuracy of population counts is far less critical in the index than in the primacy measure used by Henderson (2003) and others.

The agglomeration index is designed solely to quantify the degree of settlement concentration. Nonetheless, it can make a significant contribution to the debate regarding urbanization. At the heart of discussion about the environmental footprint of a city, congestion problems, and provision of public service infrastructures in densely packed areas is overconcentration of settlement However, the urbanization issues may be much less severe if, instead of a few large cities growing ever bigger, there are numerous small cities sprouting in what was a sparsely populated areas. Since both types of cities will be defined as urban, urbanization data alone cannot differentiate the two. In our agglomeration index, this is possible: population density captures the concentration and the population size of the nearest largest city distinguishes large cities growing bigger from many small cities emerging. Furthermore, the impacts of concentrated settlements will be greater if the population distribution is skewed toward a single point, such as the city center. This characteristic is captured by travel time. In sum, as Satterthwaite (2007) points out, some of the assertions regarding how much a country or region has urbanized based on the UN data can be misleading given their ad hoc definitions of urban. The agglomeration index can potentially resolve this data issue both in terms of the conceptual framework just described and through use of a uniform definition. As a result, it may lay to rest some myths about urbanization in various regions of the world.

There are several notable caveats associated with the current version of the agglomeration index. First is data availability related to information for population derived from censuses and road networks, which is the most important factor in the costsurface. Fortunately, the improvements in available population data over the last 10 to 15 years, in terms of spatial resolution, census availability and sub national growth estimates have been immense. This trend is continuing, especially in developing countries. The problems associated with data on road networks are multi-layered. First, the information generally is old. We used data from VMAP0 (also known as the Digital Chart of the World, or DCW), which has not been updated since 1992. With more roads and highways being built, especially in emerging economies, having to rely on data more than a decade old presents problems. Second, we would like to see more detailed information about the quality of roads. Our current version of the index used three levels of quality but it could be refined with more comprehensive data. Again, fortunately, this issue is being addressed on several fronts.[5]

On the conceptual front, the burning issue is how many people actually travel the roads—in other words, how many people in developing countries drive? We have witnessed newly built, multiple-lane highways in developing countries that are effectively empty because few people own cars. Thus, the existence of a road does not necessarily translate into people living on the outskirts of a city having easy access to the center. Related to this issue is how to incorporate the case in which the majority of travel is by public transportation. Large cities in developed countries (Tokyo and London, for example) have well-developed mass transit system that many commuters and residents use, and these systems may spread quickly to cities in developing countries—especially in light of global warming and climate change. Thus, some measures of accessibility other than roads need to be developed.

Despite these caveats, developments on the horizon, including more extensive remote sensing technologies and techniques for using and interpreting the resulting data, promise to resolve many of the limitations With an ever increasing number of accessible satellite images of high resolution—think of free services such as Google Maps and Google Earth, updating of road network data is within reach. Contiguous areas also can be determined directly using satellite images. Of course, manually determining these characteristics from the images will likely consume a large amount of manpower, which may limit the affordability of such endeavor. As a

[5] , An excellent summary of the situation and possible approaches to improving the roads data can be found on the CIESIN website http://www.ciesin.org/confluence/display/roads/.

result, population models such as the GRUMP and LandScan are still needed. However, satellite images and other remote sensing technologies can reduce the reliance on census data for calculating the index.

References

Ades, F., and E. L. Glaeser. 1995. Trade and circuses: explaining urban giants. *Quarterly Journal of Economics* 110: 195-227.

Baldwin, R., R. Forslid , P. Martin, G. Ottaviano and F. Robert-Nicoud. 2003. *Economic Geography and Public Policy*. Princeton, NJ: Princeton University Press.

Chomitz, K., P. Buys, and T. S. Thomas. 2005. *Quantifying the rural-urban gradient in Latin America and the Caribbean*. World Bank Policy Research Working Paper 3634. Washington, D.C.: World Bank.

Cohen, B. 2004. Urban growth in developing countries: a review of current trends and a caution regarding existing forecasts. *World Development* 32(1): 23-51.

Henderson, J. V. 2000. *The effects of urban concentration on economic growth*. NBER Working Paper 7503. Cambridge, MA: National Bureau of Economic Research.

----------. 2003. The urbanization process and economic growth: the so-what question. *Journal of Economic Growth* 8: 47-71.

----------. 2005. Urbanization and growth. In Aghion, P., and S. Durlauf (eds). Handbook of Economic Growth Volume 1. Amsterdam: North Holland.

Marshall, A. 1961. Principles of Economics. London: Macmillan (9th edition).

Rosen, K., and M. Resnick. 1980. The size distribution of cities: an examination of the Pareto law and primacy. *Journal of Urban Economics* 8: 165-186.

Satterthwaite, D. 2007. *The Transition to a Predominantly Urban World and Its Underpinnings*. London: International Institute for Environment and Development.

United Nations. 2006. *World Urbanization Prospects: The 2005 Revision*. New York: United Nations, Department of Economic and Social Affairs, Population Division.

Wheaton, W., and H. Shishido. 1981. Urban concentration, agglomeration economies, and the level of economic development. *Economic Development and Cultural Change* 30: 17-30.

Appendix

Table A.1 List of datasets used to calculate the costsurface

Travel time is estimated based on the combination of several GIS data layers that are merged into a costsurface layer which represents the time required to cross each pixel. The higher the cost of a pixel, the more time required to cross it.

The friction layer is composed from the following global GIS data layers.

Data layer	Proxy for	Source
VMAP0 Road layer	Major and minor road network	http://earth-info.nga.mil/publications/vmap0.html
VMAP0 Rail layer	Railway network	http://earth-info.nga.mil/publications/vmap0.html
WDBII Rivers	Navigable rivers	http://www.ngdc.noaa.gov/seg/cdroms/ged_iib/datasets/b14/mw.htm
VMAP0 Country borders	Travel delay for crossing international borders	http://earth-info.nga.mil/publications/vmap0.html
GLWD water bodies	Navigable water bodies	http://www.wwfus.org/science/data/globallakes.cfm
SRTM elevation	Inaccessible areas of very high elevation	http://www2.jpl.nasa.gov/srtm/
SRTM slope	Slope factor to reduce travel speed	http://www2.jpl.nasa.gov/srtm/
GLC2000 global land cover map	Foot/animal based travel for off road & paths	http://www-gem.jrc.it/glc2000/defaultglc2000.htm

Table A.2 Agglomeration index and UN urban population share comparison

Economy	Income	World region	Agg. Index (%) (a)	Agg. Index (%) (b)	UN Urban share (%)
Afghanistan	LIC	SAS	26.0	23.8	21.3
Albania	LMC	ECA	50.8	28.8	41.8
Algeria	LMC	MENA	56.9	49.8	59.8
American Samoa	UMC	EAP	-	-	87.9
Andorra	HIC	OHIE	-	-	92.4
Angola	LIC	SSA	26.7	26.5	50.0
Antigua and Barbuda	UMC	OHIE	-	-	37.7
Argentina	UMC	LAC	70.5	64.4	89.2
Armenia	LIC	ECA	71.8	68.0	65.1
Aruba (Neth.)	HIC	OHIE	-	-	46.7
Australia	HIC	OECD	75.2	69.9	87.2
Austria	HIC	OECD	58.1	54.0	65.8
Azerbaijan	LIC	ECA	48.7	32.1	50.9
Bahamas, The	HIC	OHIE	70.0	70.0	89.0
Bahrain	UMC	OHIE	99.3	99.3	94.6
Bangladesh	LIC	SAS	42.8	32.3	23.2
Barbados	UMC	OHIE	97.5	-	50.0
Belarus	LMC	ECA	59.8	53.0	70.0
Belgium	HIC	OECD	89.2	88.5	97.1
Belize	LMC	LAC	4.8	4.8	47.9
Benin	LIC	SSA	36.4	26.3	38.4
Bermuda	HIC	OHIE	-	-	100.0
Bhutan	LIC	SAS	7.0	-	9.6
Bolivia	LMC	LAC	54.7	51.8	61.8
Bosnia and Herzegovina	LMC	ECA	33.7	19.3	43.2
Botswana	UMC	SSA	23.8	14.3	53.2
Brazil	UMC	LAC	60.4	53.6	81.2
Brunei Darussalam	HIC	OHIE	71.7	-	71.2
Bulgaria	LMC	ECA	61.3	47.0	68.9
Burkina Faso	LIC	SSA	14.4	9.9	16.6
Burundi	LIC	SSA	17.8	7.6	8.6
Cambodia	LIC	EAP	24.8	19.1	16.9
Cameroon	LIC	SSA	41.2	34.4	50.0
Canada	HIC	OECD	71.0	67.0	79.4
Cape Verde	LMC	SSA	52.6	-	53.4
Cayman Islands	HIC	OHIE	-	-	100.0
Central African Republic	LIC	SSA	21.0	16.6	37.6
Chad	LIC	SSA	12.2	11.3	23.4
Channel Islands	HIC	OHIE	-	-	30.6
Chile	UMC	LAC	74.0	69.5	85.9
China	LMC	EAP	36.2	22.5	35.8
Colombia	LMC	LAC	64.8	59.9	71.2
Comoros	LIC	SSA	-	-	33.9
Congo, Dem. Rep.	LIC	SSA	24.9	23.0	29.8
Congo, Rep.	LIC	SSA	54.7	51.9	58.3
Costa Rica	LMC	LAC	55.4	54.5	59.0
Côte d'Ivoire	LIC	SSA	36.6	30.6	43.0
Croatia	UMC	ECA	34.2	21.8	55.6
Cuba	LMC	LAC	69.0	63.4	75.6
Cyprus	HIC	OHIE	66.6	-	68.7
Czech Republic	UMC	OECD	66.8	45.0	74.0
Denmark	HIC	OECD	58.1	50.2	85.1
Djibouti	LMC	MENA	50.4	50.4	83.4
Dominica	UMC	LAC	-	-	70.5
Dominican Republic	LMC	LAC	72.3	66.5	62.4
Ecuador	LMC	LAC	58.8	55.9	60.3
Egypt, Arab Rep.	LMC	MENA	92.6	90.2	42.5
El Salvador	LMC	LAC	72.3	68.1	58.4
Equatorial Guinea	LMC	SSA	25.2	-	38.8
Eritrea	LIC	SSA	20.9	19.0	17.8
Estonia	UMC	OHIE	49.9	44.4	69.3
Ethiopia	LIC	SSA	10.9	7.8	14.9
Faeroe Islands	HIC	OHIE	-	-	37.0
Fiji	LMC	EAP	21.3	21.3	48.3
Finland	HIC	OECD	52.0	40.0	61.1
France	HIC	OECD	71.4	66.2	75.8
French Guiana	N.A.	N.A.	40.1	-	-
French Polynesia	HIC	OHIE	-	-	52.5
Gabon	UMC	SSA	37.4	37.4	80.2
Gambia, The	LIC	SSA	46.3	-	-
Georgia	LIC	ECA	55.8	47.5	52.7
Germany	HIC	OECD	77.6	73.5	75.1
Ghana	LIC	SSA	36.6	30.2	44.0
Gibraltar	N.A.	N.A.	99.9	99.9	-
Greece	HIC	OECD	59.3	49.9	58.8
Greenland	HIC	OHIE	-	-	82.1
Grenada	UMC	LAC	-	-	30.4
Guadeloupe	N.A.	N.A.	76.5	-	-

Economy	Income	World region	Agg. Index (%) (a)	(b)	UN Urban share (%)
Guam	HIC	OHIE	-	-	93.5
Guatemala	LMC	LAC	34.7	33.8	45.1
Guinea	LIC	SSA	25.3	20.9	31.0
Guinea-Bissau	LIC	SSA	21.9	21.9	29.7
Guyana	LMC	LAC	37.4	37.4	28.6
Haiti	LIC	LAC	36.9	26.2	35.6
Honduras	LMC	LAC	41.4	34.0	44.4
Hong Kong, China	HIC	OHIE	99.8	99.8	100.0
Hungary	UMC	ECA	67.6	51.8	64.6
Iceland	HIC	OECD	65.4	65.4	92.5
India	LIC	SAS	51.9	42.9	27.7
Indonesia	LMC	EAP	57.5	51.1	42.0
Iran, Islamic Rep.	LMC	MENA	62.8	55.4	64.2
Iraq	LMC	MENA	72.2	71.4	67.8
Ireland	HIC	OECD	45.6	39.2	59.1
Isle of Man	HIC	OHIE	-	-	51.9
Israel	HIC	OHIE	87.4	81.2	91.4
Italy	HIC	OECD	77.0	68.5	67.2
Jamaica	LMC	LAC	69.0	51.2	51.8
Japan	HIC	OECD	92.9	90.6	65.2
Jordan	LMC	MENA	79.4	76.5	80.4
Kazakhstan	LMC	ECA	51.3	45.4	56.3
Kenya	LIC	SSA	27.2	21.9	19.7
Kiribati	LMC	EAP	-	-	43.3
Korea, Dem. Rep.	LIC	EAP	51.8	51.8	60.2
Korea, Rep.	UMC	OECD	89.6	88.1	79.6
Kuwait	HIC	OHIE	90.3	89.4	98.2
Kyrgyz Republic	LIC	ECA	34.0	25.8	35.4
Lao PDR	LIC	EAP	12.5	12.2	18.8
Latvia	LMC	ECA	50.7	47.0	68.1
Lebanon	UMC	MENA	86.8	75.7	86.0
Lesotho	LIC	SSA	18.6	10.6	17.8
Liberia	LIC	SSA	20.4	19.0	54.3
Libya	UMC	MENA	83.4	76.6	83.1
Liechtenstein	HIC	OHIE	4.7	-	15.2
Lithuania	LMC	ECA	54.6	53.4	67.0
Luxembourg	HIC	OECD	73.6	1.7	83.9
Macao, China	HIC	OHIE	100.0	100.0	100.0
Macedonia, FYR	LMC	ECA	60.8	32.7	64.9
Madagascar	LIC	SSA	17.8	14.2	26.0
Malawi	LIC	SSA	17.6	15.5	15.1

Economy	Income	World region	Agg. Index (%) (a)	(b)	UN Urban share (%)
Malaysia	UMC	EAP	68.7	66.1	61.8
Maldives	LMC	SAS	14.5	-	27.6
Mali	LIC	SSA	19.0	14.1	27.9
Malta	UMC	OHIE	-	-	93.4
Marshall Islands	LMC	EAP	-	-	65.4
Mauritania	LIC	SSA	26.8	23.1	40.0
Mauritius	UMC	SSA	91.9	90.5	42.7
Mexico	UMC	LAC	66.7	61.0	74.7
Micronesia, Fed. Sts.	LMC	EAP	-	-	22.4
Moldova	LIC	ECA	48.2	44.9	46.1
Monaco	HIC	OHIE	87.4	87.4	100.0
Mongolia	LIC	EAP	34.4	31.3	56.6
Morocco	LMC	MENA	53.6	46.9	55.1
Mozambique	LIC	SSA	24.8	22.7	30.7
Myanmar	LIC	EAP	30.1	22.7	28.0
Namibia	LMC	SSA	12.9	12.9	32.4
Nepal	LIC	SAS	24.4	14.6	13.4
Netherlands	HIC	OECD	89.4	86.8	76.8
Netherlands Antilles	HIC	OHIE	87.0	-	69.3
New Zealand	HIC	OECD	66.0	55.3	85.7
Nicaragua	LIC	LAC	48.4	41.4	57.2
Niger	LIC	SSA	16.5	11.9	16.2
Nigeria	LIC	SSA	43.1	37.3	43.9
Northern Mariana Islands	UMC	EAP	-	-	92.9
Norway	HIC	OECD	51.8	38.0	76.1
Oman	UMC	MENA	72.1	58.1	71.6
Pakistan	LIC	SAS	56.5	49.0	33.1
Palau	UMC	EAP	-	-	68.4
Panama	UMC	LAC	52.9	47.3	65.8
Papua New Guinea	LMC	EAP	7.6	5.1	13.2
Paraguay	LMC	LAC	48.7	45.3	55.3
Peru	LMC	LAC	50.1	46.7	71.6
Philippines	LMC	EAP	56.4	50.9	58.5
Poland	UMC	ECA	63.5	55.1	61.7
Portugal	HIC	OECD	60.0	59.1	54.4
Puerto Rico	UMC	OHIE	87.7	80.2	94.7
Qatar	HIC	OHIE	87.0	87.0	95.0
Romania	LMC	ECA	61.2	49.9	54.6
Russian Federation	LMC	ECA	63.0	56.1	73.4
Rwanda	LIC	SSA	10.6	10.6	13.8
Samoa	LMC	EAP	-	-	22.0

Economy	Income	World region	Agg. Index (%) (a)	(b)	UN Urban share (%)
San Marino	HIC	OHIE	11.1	11.1	92.6
São Tomé and Príncipe	LIC	SSA	58.6	-	53.6
Saudi Arabia	UMC	OHIE	79.3	74.2	79.9
Senegal	LIC	SSA	47.9	44.1	40.6
Serbia and Montenegro		ECA	57.7	42.2	51.6
Seychelles	UMC	SSA	-	-	50.6
Sierra Leone	LIC	SSA	33.1	27.2	37.0
Singapore	HIC	OHIE	100.0	100.0	100.0
Slovak Republic	UMC	ECA	55.4	19.9	56.3
Slovenia	HIC	OHIE	42.3	41.2	50.7
Solomon Islands	LIC	EAP	8.3	-	15.8
Somalia	LIC	SSA	20.7	19.8	33.3
South Africa	UMC	SSA	49.4	45.3	56.9
Spain	HIC	OECD	75.3	71.4	76.3
Sri Lanka	LMC	SAS	44.1	33.7	15.7
St. Kitts and Nevis	UMC	LAC	-	-	32.5
St. Lucia	UMC	LAC	84.0	-	27.9
St. Vincent and the Grenadines	LMC	LAC	-	-	44.8
Sudan	LIC	SSA	30.7	26.4	36.1
Suriname	LMC	LAC	73.2	73.2	72.1
Swaziland	LMC	SSA	19.0	-	23.4
Sweden	HIC	OECD	53.8	41.0	84.0
Switzerland	HIC	OECD	72.8	64.9	73.1
Syrian Arab Republic	LMC	MENA	59.1	56.9	50.1
Taiwan, China	HIC	OHIE	84.4	82.7	-

Economy	Income	World region	Agg. Index (%) (a)	(b)	UN Urban share (%)
Tajikistan	LIC	ECA	36.2	23.3	25.9
Tanzania	LIC	SSA	25.8	20.5	22.3
Thailand	LMC	EAP	36.9	28.6	31.1
Togo	LIC	SSA	31.3	20.1	36.6
Tonga	LMC	EAP	-	-	23.0
Trinidad and Tobago	UMC	OHIE	88.0	87.1	10.8
Tunisia	LMC	MENA	51.9	39.0	63.4
Turkey	LMC	ECA	62.5	52.8	64.7
Turkmenistan	LIC	ECA	43.5	36.2	45.1
Uganda	LIC	SSA	25.0	14.3	12.1
Ukraine	LIC	ECA	61.7	52.8	67.1
United Arab Emirates	HIC	OHIE	75.1	70.5	77.4
United Kingdom	HIC	OECD	84.7	83.5	89.4
United States	HIC	OECD	72.3	65.9	79.1
Uruguay	UMC	LAC	62.2	51.4	91.4
Uzbekistan	LIC	ECA	58.8	49.9	37.3
Vanuatu	LMC	EAP	-	-	22.0
Venezuela, RB	UMC	LAC	78.5	72.9	91.1
Vietnam	LIC	EAP	49.5	41.2	24.3
Virgin Islands (U.S.)	HIC	OHIE	-	-	92.8
West Bank and Gaza	LMC	MENA	90.9	90.4	71.5
Yemen, Rep.	LIC	MENA	25.5	22.8	25.4
Zambia	LIC	SSA	30.5	27.8	34.8
Zimbabwe	LIC	SSA	33.2	30.0	33.8
World			**52.0**	**43.8**	**46.7**

Note: For agglomeration index, column (a) uses largest city size threshold of 50,000 or more, and column (b) uses the threshold of 100,000 or more.

Source: Authors' calculation (for agglomeration index); UN (2006).