

II THE RESEARCH DESIGN AND THE SAMPLE OF CITIES

1. Introduction

In the broader context of the study of urbanization, this study focuses on the *spatial consequences* of urbanization, rather than on the demographic dimensions of the growth of urban populations. It seeks to describe, measure and explain the patterns and dynamics of the urban use of land, and it seeks to do so in a global comparative framework by focusing on a sample of cities of different sizes in all world regions.

The first phase of the study—and the one that is reported on here—makes four important contributions to the present study of urbanization and urban expansion:

- a. **The global sample of cities:** The study introduces a new and improved list of 3,943 cities and metropolitan areas with populations in excess of 100,000—the *universe* of cities—and identifies a global, stratified sample of 120 cities from this universe of cities. This sample is of sufficient size, so as to derive global estimates and global norms, as well as to engage in rigorous econometric modeling than can explain variations—as well as change over time—in the urban extent of cities;
- b. **The rigorous classification of remote-sensing data:** The study uses an innovative and cost-effective methodology for classifying built-up and non-built up pixels in *Landsat* satellite images of all the cities in the global sample—approximately a decade apart—that permits the accurate and detailed measurement of the built-up area of cities and its change over time;
- c. **The construction of metrics:** The study defines, constructs and tests a number of descriptive metrics of urban extent and expansion that correspond to our intuitive grasp of these phenomena, making it possible—for the first time—to estimate urban extent and expansion for the universe of cities by geographic regions, income classes and city size groups;
- d. **The derivation of explanatory models:** The study introduces a number of hypotheses and tests a number of econometric models associated with these hypotheses that explain the variation in urban extent and expansion in the sample of cities, using available geographic, demographic and economic data.

The second phase of the study—now already under way, supported by a grant from the US National Science Foundation (NSF)—improves on the first phase by making two additional contributions:

- a. **Improving the classification of remote-sensing data:** Engaging local consultants in each city in the sample to provide ground-truth checks in selected locations, so as to test and improve the initial classification of images;
- b. **Improving the explanatory models:** Using data from census and municipal offices, real estate agencies, and visits to informal settlements to improve the

explanations of variations in urban extent and expansion among cities, as well as to test several hypotheses regarding the effects of urban extent and expansion on quality of life and on key dimensions of poverty in the sample of cities.

The third phase of the study—now already under way as well and supported by the Japanese Trust Fund to the World Bank—examines the alternatives available to cities in making effective preparations for absorbing their future population growth in the coming decades. The study will focus on three cities soon to be chosen—one in China, one in India and one in Sub-Saharan Africa—engaging a consulting firm to collect data, exploring alternatives with groups of local stakeholders, and preparing policy prescriptions for managing urban expansion in these cities in the years to come. The consultants will prepare handbooks so that other cities will also be able to prepare for urban expansion.

The following sections in this chapter will provide greater detail on the universe of cities and sample of cities developed for the study. Subsequent chapters will focus on the classification of *Landsat* images in the sampled cities into built-up and non-built-up areas; on the development of metrics for measuring urban extent and expansion; and on the development and testing of econometric models that explain urban extent and expansion.

2. The rationale for sampling

Our most recent investigation found a total of 3,943 distinct metropolitan areas that had populations in excess of 100,000 in the year 2000. These metropolitan areas had an estimated population of 2.1 billion, and they constitute the ‘universe’ of cities for the purpose of this study. The actual number of cities in this category may be of the order of 3,500–4,500, and it is indeed both a surprise and a pity that a complete and reliable list of metropolitan areas (with their corresponding latitude/longitude location) does not exist at the present time.

The few global comparative studies of cities undertaken in the past²⁰ have had to rely on haphazard collections of cities for which data were available, usually in different and non-compatible forms. Needless to say, these were generally either cities in developed countries, or large capital cities in developing countries that are already part of the global network of world cities—Buenos Aires, Mexico City, Shanghai, Seoul, or Bangkok to take typical examples. Smaller and lesser-known cities are rarely, if ever, included in global comparative studies. Possible biases are only to be expected in drawing conclusions about the universe of cities from such non-random collections of cities. Even basic questions such as whether, on the whole, cities are becoming more or

²⁰ See, for example, Newman, P.W.G. and J. Kenworthy, 1989, *Cities and Automobile Dependence—An International sourcebook*, Aldershot, UK: Gower Publishing; Angel, Shlomo, 2000, *Housing Policy Matters: A Global analysis*, New York: Oxford University Press; or Malpezzi Bertaud, 2002, *op.cit.*

less compact over time can still not be answered without referring to a properly drawn sample of cities while using identical definitions and procedures for collecting and aggregating data.²¹

This chapter describes how our universe of cities was originally obtained at the outset of the study, how the original—as well as the final—sample of 120 cities was selected, and how the final sample of cities corresponds to the new universe of cities that emerged from our investigation.

3. The initial universe of cities

Two universes of cities were identified at the time of our initial sample selection in 2003. The first was the matrix of city data prepared by Vernon Henderson at Brown University, as part of a World Bank research project entitled “Successful cities: Determinants of City Growth Rates.” This matrix provided information on the urban population in the period 1950–2000 for 2,719 metropolitan areas that had populations in excess of 100,000 in the year 2000. The second was a list of 4,574 metropolitan areas prepared by the United Nations Human Settlements Programme (UN Habitat). This second list also focused on cities that had populations in excess of 100,000 in the year 2000. Both lists provided exact latitude/longitude locations for most cities.

The difference between the two lists was largely due to: the inclusion in the UN Habitat list of a larger number of smaller-size cities; a more complete listing of Chinese cities; the inclusion of more countries; and more double-counting. Most large cities were included in both lists: the average size of the cities in the UN list not included in Henderson’s list was 250,000 and the median was 150,000. Of the cities not included that were larger than 500,000 almost half were in China. The UN Habitat list also incorporated some countries—*e.g.* Algeria, Libya and North Korea—not included in Henderson’s list. And finally, the UN Habitat list included more cities that were part of larger metropolitan areas—*e.g.* Giza, which is part of greater Cairo; Quezon City, which is part of Metro Manila; and St. Paul, which is part of the metropolitan area of Minneapolis–St. Paul.

The research team created a new universe of cities in mid-2005 by combining the UN Habitat list of 4,574 cities with an updated list of 2,884 cities provided by Henderson into a new, comprehensive list of 3,943 cities. In this new list, most double counting was eliminated, cities for which an exact location could not be found were eliminated, and cities that were estimated to be part of larger metropolitan areas were eliminated as well. An explanation of the procedure used to derive this new universe of cities appears in Section 7 of this chapter.

²¹ Even Burchfield et al, 2004, *op.cit.*, seeking to compare ‘urban sprawl’ in two time periods in the United States—where data are plentiful and systematically collected—had to use different sources of data for the two time periods that are not strictly comparable.

4. Metropolitan areas as individual data units

It is important to note at the outset that the focus of this study is on *metropolitan areas* rather than on city administrative jurisdictions, and that the satellite images, the population data, and the derived measures of urban extent and expansion all pertain to metropolitan areas.²² Indeed, both the Henderson and the UN Habitat lists sought to focus on metropolitan areas, rather than on individual city jurisdictions, so as to avoid double-counting of cities that were part of larger metropolitan agglomerations. This is by no means a simple and well-defined task. It is often difficult to determine how far a metropolitan area extends or—in the cases of the U.S. Eastern Seaboard or Japan’s Kanto plain, for example—where one ends and another begins. In other cases—say, in Yulin, China, for example—it is difficult to tell where the city ends and the rural area begins as they gradually dissolve into each other. In addition, given the paucity of travel data, one cannot rely on commuting patterns to determine the outer limits of functional metropolitan areas. The lists of metropolitan areas, therefore, can only be taken as provisional lists of loosely defined but unique urban places, where initial attempts have been made to agglomerate all contiguous urban jurisdictions into single metropolitan areas. The lists themselves are thus to be considered work in progress, as we shall see more clearly in Section 7 below.

5. Sample size

UN Habitat selected a sample of 355 cities from its universe of 4,574 cities, drawing approximately 40 cities in each of nine world regions. In addition, it selected a small sub-sample of 35 cities from this larger sample. The method of sampling used by UN Habitat involved selecting approximately 40 cities at random in each region of the universe of cities, so that the probability of being selected was proportional to the population in each city. Larger cities therefore had much higher probabilities of being selected than smaller ones.²³ Considering that future funding for collecting global city data on a regular and sustained basis is likely to be rather limited, the study team decided to use the two UN Habitat samples as the basis for creating a new sample. This should increase the probability that, in the coming years, the data collected for the present study could be supplemented and updated by panel data to be collected by UN Habitat in its larger sample.

The study team considered the size of the sample necessary to derive global norms and estimates as well as to model global urban extent and expansion. While there was no rigorous analytical procedure employed in deciding on the exact sample size of 120, it was determined at the outset that 35 cities would be too few and 355 would be too

²² In the following discussion, however, the terms ‘city’ and ‘metropolitan area’ will be used interchangeably.

²³ In practice, random selection involved selection “with replacement”. Each city selected was returned into the regional sub-universe and could be selected again until 40 cities were selected in each region. In this manner, the bias created by the removal of selected cities from the universe was avoided.

many. It was also determined that a sample of 120 cities would be adequate for deriving statistically-significant results for the universe of cities as a whole, provided it was a stratified sample. In a stratified sample, each city in the sample represents a group of cities in the universe and is given a weight that is proportional to the share of the population of this group in the total population of the universe. The weight given to each city in the sample is then used in calculating global measures of urban extent and expansion, as well as in the statistical modeling of these measures.

6. Sample stratification

Three important characteristics were used to define the strata in our stratified sample of 120 cities: (a) the world region in which the city is located; (b) city size; and (c) the level of economic development of the country in which the city is located, measured by Gross National Income (GNI) per capita. The universe of cities was divided into nine regions, into four size categories, and into four per-capita income groups:

Table II-1: Comparison of the Study Sample with the Universe of Cities, by Region

Region	Urban Pop. in 2000	Cities in 2000	Sample Population		Sample Cities	
			Population	%	Number	%
East Asia & the Pacific	410,903,331	550	57,194,979	13.9%	16	2.9%
Europe	319,222,933	764	45,147,989	14.1%	16	2.1%
Latin America & the Caribbean	288,937,443	547	70,402,342	24.4%	16	2.9%
Northern Africa	53,744,935	125	22,517,636	41.9%	8	6.4%
Other Developed Countries	367,040,756	534	77,841,364	21.2%	16	3.0%
South & Central Asia	332,207,361	641	70,900,333	21.3%	16	2.5%
Southeast Asia	110,279,412	260	36,507,583	33.1%	12	4.6%
Sub-Saharan Africa	145,840,985	335	16,733,386	11.5%	12	3.6%
Western Asia	92,142,320	187	18,360,012	19.9%	8	4.3%
Total	2,120,319,475	3,943	415,605,624	19.6%	120	3.0%

a. Geographic regions: UN Habitat used a breakdown of countries into nine regions to draw its sample of 355 cities, and it is this regional classification that was used for constructing our study sample.²⁴ The nine regions are: (1) Europe—including both Western and Eastern Europe, as well as the Russian Federation; (2)

²⁴ Unfortunately, there is no agreed-upon classification of countries into regions. International organizations, such as the UN and the World Bank, typically classify countries into regions, and these classifications tend to change over time. The UN Habitat sample, as noted earlier, was selected from a 9-region classification. The World Bank currently divides *developing* countries into six regions: East Asia and the Pacific, Europe and Central Asia, Latin America and the Caribbean, the Middle East and North Africa, Sub-Saharan Africa, and South Asia. The United Nations now divides all countries into 20 world macro-regions, but has five regional commissions in developing countries: Asia and the Pacific, Western Asia, Europe, Africa, and Latin America and the Caribbean.

East Asia and the Pacific—comprising China, the two Koreas, Mongolia and the Pacific islands; (3) Latin America and the Caribbean; (4) Northern Africa; (5) Other Developed Countries—comprising the United States, Canada, Japan, Australia and New Zealand; (6) South and Central Asia, including Iran; (7) Southeast Asia; (8) Sub-Saharan Africa; and (9) Western Asia, including Turkey. A minimum of eight cities was selected from each of these nine regions. Five of these nine regions have approximately 15–20% each of the global urban population. Sixteen cities were selected from each of these five regions. Two of the regions have 5–8% each of the global urban population, and twelve cities were selected from each one of them. A comparison of the universe of cities and the sample of cities appears in table II-1.

b. City size categories: The smaller universe of cities provided by Henderson was used to divide cities into four size strata.²⁵ This universe was divided into four classes so that the total urban population in each size class was approximately equal. The total population in the Henderson universe of 2,719 metropolitan areas was 1.815 billion. This population was divided into 4, so that each size category contained approximately 454 million people. This resulted in the following size categories:

1. Size class 1: cities with populations between 100,000 and 528,000 (1,982 cities);
2. Size class 2: cities with populations between 528,000 and 1,490,000 (498 cities);
3. Size class 3: cities with populations between 1,490,000 and 4,180,000 (190 cities); and
4. Size class 4: cities with populations in excess of 4,180,000 million (49 cities).

To the extent possible, the cities in each of the nine regions were sampled so that there was to be an equal number of cities in each size category. For example, in Latin America and the Caribbean, a total of sixteen cities were selected for the sample, four cities in each size category. The sample of 120 cities therefore contained approximately 30 cities in each size category. As a result, although the resulting final sample contains only 120 cities (3% of the total number of cities), it contains 415 million people (20% of the world's urban population). Because urban land consumption is closely related to the urban population, the share of the built-up area examined and analyzed in the sample cities amount to approximately one-fifth of the built-up area in urban use in the global universe of cities.

Table II-2 compares the final universe of cities and the sample in terms of population size categories. Three characteristics of the table merit special attention: first, the size categories in the universe are no longer equal in the new universe of cities because so many small cities were added to the original Henderson universe. Second, in some regions there were not enough cities in the largest size category, and so cities in the second-largest size category were selected instead. Third, while the

²⁵ At the time the sample was drawn, the UN Habitat universe of cities was not available to the study team. Only the UN Habitat sample of 355 cities and the sub-sample of 35 cities were available.

number of cities in each size category in the sample is still approximately the same, only 1.3% of the population and 0.9% of the cities in the smallest size category are in the sample, compared to 57% of the population and 48% of the cities in the largest size category. This necessarily means that in the assignment of weights to the cities in the sample, the smaller cities will be assigned much heavier weights than the larger ones.

Table. II-2: Comparison of the Study Sample with the Universe of Cities, by Size Class

City Size Category	Urban Pop. in 2000	Cities in 2000	Sample Population		Sample Cities	
			Population	%	Number	%
100,000 - 528,000	650,874,692	3,131	8,308,191	1.3%	29	0.9%
528,000 - 1,490,000	496,583,987	560	30,400,467	6.1%	31	5.5%
1,490,000 - 4,180,000	468,804,459	197	87,925,743	18.8%	33	16.8%
More than 4,180,000	504,056,338	55	288,971,224	57.3%	27	48.2%
Total	2,120,319,475	3,943	415,605,624	19.6%	120	3.0%

c. Per capita income categories: The World Bank's World Development Report provides a regular breakdown of countries into four annual Gross National Income (GNI) per capita categories. The 2003 World Development Report was initially used to obtain the classification of the universe of cities into four 2001 per-capita income groupings.²⁶ This initial classification was later changed to reflect annual Gross National Product (GNP) per capita in Purchasing Power Parities (PPP), using World Bank data for 1995. This resulted in the following annual GNP per-capita categories:

1. GNP per-capita category 1: cities in countries with annual GNP per-capita measured in PPP of less than \$3,000;
2. GNP per-capita category 2: cities in countries with annual GNP per-capita measured in PPP between \$3,000 and \$5,200;
3. GNP per-capita category 3: cities in countries with annual GNP per-capita measured in PPP between \$5,200 and \$17,000; and
4. GNP per-capita category 4: cities in countries with annual GNP per-capita measured in PPP of \$17,000 or higher.

Table II-3 compares the final universe of cities and the sample in terms of annual GNP per capita categories. As can be seen from the table, the share of cities in each category in the sample is of the order of 3% of the cities in the universe in all GNP per capita categories.

²⁶ World Bank, *World Development Report-2003: Sustainable Development in a Dynamic World*, Washington DC: The World Bank, 243.

Table II-3: Comparison of the Study Sample with the Universe of Cities, by GNP per Capita Class

1995 GNP Per Capita in PPP	Urban Pop. in 2000	Cities in 2000	Sample Population		Sample Cities	
			Population	%	Number	%
Less than \$3,000	537,574,166	1,075	92,568,021	17.2%	32	3.0%
\$3,000 - \$5,200	518,840,787	855	85,044,633	16.4%	25	2.9%
\$5,200 - \$17,000	516,674,573	1,082	124,057,217	24.0%	35	3.2%
More than \$17,000	547,229,950	931	113,935,753	20.8%	28	3.0%
Total	2,120,319,475	3,943	415,605,624	19.6%	120	3.0%

7. Sample selection

It must be noted here that the research team did not have access to the UN Habitat universe of 4,574 cities when constructing the global sample of 120 cities, but rather only to the sample of 355 cities drawn by UN Habitat from this universe. The only available universe of cities was that prepared by Henderson. As noted earlier, this universe of 2,761 cities was stratified into the nine geographical regions, then further stratified into the four size categories, and then further stratified into the four income categories. This stratification resulted in a total of 144 cells, of which 60 cells were found to be non-empty. The cities in the UN Habitat sample of 355 were then allocated among these 60 cells. After ensuring that as many cities in the UN Habitat sub-sample of 35 cities were included in our initial sample selection, other cities from the UN sample of 355 were selected at random from each non-empty cell.

This procedure resulted in an initial sample of 120 cities. There then followed a prolonged period of replacing individual cities by other cities from the UN Habitat sample in case essential data were not found. To be included in the final sample, three conditions had to be met:

- a. The country in which the city was located had to have conducted and published two population censuses during the years 1985–2002²⁷;
- b. Statistical information on the country in which the city was located had to be collected by the World Bank's World Development Indicators (WDI); and
- c. Cloud-free *Landsat* images of the city had to be available for two time periods, each one within not more than three years of the time of each national census.

The first constraint eliminated cities in some 20 countries. Most countries were in the midst of political strife: Afghanistan, Angola, the Congo Democratic Republic, Burma, Cambodia, Lebanon, the Palestinian Territories, Colombia and Haiti. Eliminating these countries introduces a bias in the sample: it is largely restricted to cities in peaceful countries and says little or nothing about urban expansion in the midst of civil or international conflict, or in failed states. The only large country for which

²⁷ Data on the national censuses can be found in U.S. Census, "Census Dates for Countries and Areas of the World: 1945–2014", available on line at www.census.gov/ipc/www/cendates.

census data were not available was Pakistan. The second constraint eliminated all small countries and most small island countries, as the World Bank does not regularly collect demographic or economic information about them. It also eliminated Cuba, Libya and North Korea, three countries that are presently not members of the Bank. Eliminating these countries introduces another bias in the sample: it eliminates cities in the remaining centrally-planned economies. The third constraint eliminated more than about a dozen cities from the original sample, and there is some bias introduced by this constraint too. The *Landsat* acquisition plan favors the United States, and there is less frequent coverage of some parts of the world. Also, there is a bias against those parts of the world commonly in cloud cover (equatorial, tropical, and sub-tropical areas) for which there are fewer cloud-free scenes available. These latter biases were largely overcome by insisting on picking the required number of cities in all the nine regions.

The final sample that emerged from this procedure is shown in table 4 below. 117 cities in this sample are from the UN Habitat sample of 355 cities, and 22 are from its sub-sample of 35 cities. Three cities that were not in the UN Habitat sample were added to our sample—Fukuoka, San Salvador and Moscow—because no appropriate replacement cities for their particular cells were found in the UN Habitat sample. Figure II.2 graphically depicts the locations of the 120 cities, by UN Habitat Region, categorized by population and incomes classes.

8. The new universe of cities

As mentioned earlier in this chapter, a new universe of cities was created in mid-2005 by combining the original UN Habitat list of 4,578 cities with an updated list of 2,884 cities provided by Henderson into a new, common list of 3,943 cities.

The two lists were first compared to identify metropolitan areas that appeared in both lists. The two lists used different naming conventions, sometimes citing the name in the local language and sometimes the international name (*e.g.* München as against Munich, Germany), and sometimes spelling the same name differently. Cities appeared more than once in the same list with different spellings or different names. The latitudes and longitudes of all cities were then compared to check where cities were, in fact, identical. Many missing latitudes and longitudes were then obtained from other lists of cities that have now become available:

1. The *NASA World Wind* (worldwind.arc.nasa.gov) has a Place Finder that, given a city name, finds all places with that name and their latitudes and longitudes;
2. The *Tageo* website (www.tageo.com) has a list of some 3,850 cities with populations of 100,000 or more (no date given), their population and their latitudes and longitudes; and
3. The *Maxmind GeoIP City Database* (www.maxmind.com) has a list of 2,760 cities with population in excess of 100,000.

After identifying as many latitudes and longitudes for the cities in the Henderson and UN Habitat lists, some 70 cities for which no location was found or for which multiple locations were found were eliminated from the new universe.

As it turned out, there were serious discrepancies between the two lists regarding cities in the U.S. and the U.K. The UN Habitat list had 339 metropolitan areas in the U.S. and 219 in the U.K., while the Henderson list had 208 in the U.S. and 50 in the U.K. According to the latest censuses in the two countries, there were 260 metropolitan areas in the U.S. and 73 in the U.K. with populations in excess of 100,000 in the year 2000. In the new universe of cities, the lists of metropolitan areas from the recent censuses in both countries replaced the U.S. and U.K. cities in the Henderson and UN Habitat lists.

Finally, an attempt was made to try to rid the new universe of cities from cities that were parts of larger metropolitan areas. The procedure that was adopted was by no means perfectly accurate. For every city, the geographical distance to twenty nearest neighbors was computed.²⁸ Cities that were within the orbit of larger cities were then eliminated. The radii of the orbits of cities were computed as a function of their population: 30 kms for cities of 10 million or more; 20 kms for cities of 4.7 million or more; 10 kms for cities of 1.2 million or more; 5 kms for cities of 300,000 or more; and 3 kms for cities of 100,000 or more.²⁹ This procedure made it possible to eliminate several hundred cities from the combined UN and Henderson list that were either in the orbit of larger cities or were identical cities with different names.

Applying all these procedures resulted in a new universe of 3,943 cities. This universe is by no means complete. It does require more work, but it appears to be a considerable improvement on the other available universes of cities at the present time.

9. The provisional sample of 90 cities and the assignment of weights

The global sample of cities assembled for this study contains 120 cities. Satellite images and population data were obtained for all 120 cities for two time periods approximately a decade apart, and all these images were classified into built-up and non built-up pixels. The classification is now complete, as are the corresponding population estimates. However, for this draft report it was only possible to obtain summary measures—*e.g.* built-up area totals, densities, and annual changes in built-up areas and densities—for 90 cities in the sample.

²⁸ Nearest neighbors were identified first by sorting the list of cities by latitude and then by longitude and choosing five nearest cities with greater latitudes and the five with smaller latitudes; and second by sorting the list of cities by longitude and then by latitude and choosing five nearest cities with greater longitudes and the five with smaller longitudes.

²⁹ A circular city of 30-km radius will have an area of some 2,800 km². Assuming an average density of 7,500 persons per km², such a city will house a population of some 20 million people. Assuming that only half the circle will be built-up, such a city will house some 10 million people. The orbit of a city of 10 million was thus taken to be 30 km. Similar calculations were made for other city sizes.

As mentioned earlier, the new universe of cities contained 3,943 cities, some 50% more than those found in the Henderson sample from which the original sample was drawn. In the new universe, there are altogether 90 non-empty strata out of a total of 144 strata (9 regions x 4 income classes x 4 city size categories = 144). The 90-city sample contained cities in 50 strata out of these 90 strata. In order to use the 90-city sample data in modeling and calculations, strata for which there was no representative city in the sample had to be combined with strata that did have such representatives. This resulted in 50 *merged* strata that now contained all non-empty cells in the new universe of cities. The merged strata are shown in figure II-1.

Table II-4: The Global Sample of 120 Cities

No	City				Country		
	Name	Population in 2000	Size Class	Rank	Name	GNP/cap. in PPP (\$)	Income Class
Eastern Asia							
1	Shanghai	12,900,000	4	1	China	3,547	2
2	Beijing	10,800,000	4	2	China	3,547	2
3	Seoul	9,887,779	4	1	Republic of Korea	13,958	3
4	Hong Kong	6,927,000	4	4	China	3,547	2
5	Guangzhou	3,893,000	3	9	China	3,547	2
6	Pusan	3,830,000	3	2	Republic of Korea	13,958	3
7	Zhengzhou	2,070,000	3	23	China	3,547	2
8	Yulin	1,558,000	3	46	China	3,547	2
9	Yiyang	1,343,000	2	67	China	3,547	2
10	Leshan	1,137,000	2	88	China	3,547	2
11	Ulan Bator	738,000	2	1	Mongolia	1,491	1
12	Changzhi	593,500	2	185	China	3,547	2
13	Anqing	566,100	2	196	China	3,547	2
14	Ansan	549,900	2	15	Republic of Korea	13,958	3
15	Chinju	287,100	1	24	China	13,958	3
16	Chonan	114,600	1	47	Republic of Korea	13,958	3
Europe							
1	Paris	9,624,000	4	1	France	23,225	4
2	Moscow	9,321,000	4	1	Russian Fed.	6,644	3
3	London	8,219,226	4	1	United Kingdom	22,652	4
4	Milan	4,251,000	4	1	Italy	22,875	4
5	Madrid	4,072,000	3	1	Spain	18,314	4
6	Warsaw	2,269,000	3	1	Poland	9,114	3
7	Vienna	2,070,000	3	1	Austria	25,694	4
8	Budapest	1,825,000	3	1	Hungary	11,301	3
9	Thessaloniki	789,000	2	2	Greece	15,280	3
10	Palermo	684,300	2	7	Italy	22,875	4
11	Sheffield	640,048	2	9	United Kingdom	22,652	4
12	Astrakhan	486,100	1	36	Russian Fed.	6,644	3
13	Leipzig	446,491	1	19	Germany	23,913	4
14	Le Mans	194,825	1	34	France	23,225	4
15	Castellon	144,500	1	40	Spain	18,314	4
16	Oktyabrsky	111,500	1	147	Russian Fed.	6,644	3
Latin America and the Caribbean							
1	Mexico City	18,100,000	4	1	Mexico	8,182	3
2	Sao Paulo	17,800,000	4	1	Brazil	6,781	3
3	Buenos Aires	12,600,000	4	1	Argentina	11,131	3
4	Santiago	5,538,000	4	1	Chile	8,412	3
5	Guadalajara	3,908,000	3	2	Mexico	8,182	3

Table II-4: The Global Sample of 120 Cities (continued)

No	City				Country		
	Name	Population in 2000	Size Class	Rank	Name	GNP/cap. in PPP (\$)	Income Class
Latin America and the Caribbean (continued)							
6	Guatemala City	3,242,000	3	1	Guatemala	3,633	2
7	Caracas	3,153,000	3	1	Venezuela	5,174	2
8	San Salvador	1,408,000	2	1	El Salvador	4,307	2
9	Montevideo	1,236,000	2	1	Uruguay	8,130	3
10	Tijuana	1,167,000	2	7	Mexico	8,182	3
11	Kingston	912,500	2	1	Jamaica	3,370	2
12	Ribeirão Preto	502,333	2	23	Brazil	6,781	3
13	Valledupar	274,300	1	16	Colombia	5,618	3
14	Guarujá	269,104	1	70	Brazil	6,781	3
15	Ilhéus	161,898	1	85	Brazil	6,781	3
16	Jequié	130,207	1	102	Brazil	6,781	3
Northern Africa							
1	Cairo	10,600,000	4	1	Egypt	3,253	2
2	Alexandria	4,113,000	3	2	Egypt	3,253	2
3	Casablanca	3,541,000	3	1	Morocco	3,195	2
4	Algiers	2,760,740	3	1	Algeria	4,979	2
5	Marrakech	736,500	2	4	Morocco	3,195	2
6	Port Sudan	384,100	4	2	Sudan	1,512	1
7	Aswan	219,017	4	15	Egypt	3,253	2
8	Tébessa	163,279	4	13	Algeria	4,979	2
Other Developed Countries							
1	Tokyo	26,400,000	4	1	Japan	23,828	4
2	Los Angeles	16,373,645	4	2	United States	31,338	4
3	Chicago	9,157,540	4	3	United States	31,338	4
4	Philadelphia	6,188,463	4	6	United States	31,338	4
5	Houston	4,669,571	4	10	United States	31,338	4
6	Sydney	3,664,000	3	1	Australia	24,013	4
7	Minneapolis	2,968,806	3	14	United States	31,338	4
8	Pittsburgh	2,358,695	3	21	United States	31,338	4
9	Cincinnati	1,979,202	3	23	United States	31,338	4
10	Fukuoka	1,341,470	2	10	Japan	23,828	4
11	Tacoma	700,820	2	62	United States	31,338	4
12	Springfield	591,932	2	72	United States	31,338	4
13	Modesto	446,997	1	94	United States	31,338	4
14	St. Catharines	389,600	1	14	Canada	25,456	4
15	Victoria	317,506	1	16	Canada	25,456	4
16	Akashi	293,117	1	60	Japan	23,828	4

Table II-4: The Global Sample of 120 Cities (continued)

No	City				Country		
	Name	Population in 2000	Size Class	Rank	Name	GNP/cap. in PPP (\$)	Income Class
South and Central Asia							
1	Mumbai	18,100,000	4	1	India	2,220	1
2	Kolkata	12,900,000	4	2	India	2,220	1
3	Dhaka	12,300,000	4	1	Bangladesh	1,427	1
4	Teheran	7,225,000	4	1	Iran	5,460	3
5	Hyderabad	6,842,000	4	4	India	2,220	1
6	Pune	3,489,000	3	9	India	2,220	1
7	Kanpur	2,450,000	3	11	India	2,220	1
8	Jaipur	2,145,000	3	13	India	2,220	1
9	Coimbatore	1,292,000	2	23	India	2,220	1
10	Vijayawada	1,237,000	2	28	India	2,220	1
11	Rajshahi	1,016,000	2	4	Bangladesh	1,427	1
12	Ahvaz	997,000	2	7	Iran	5,460	3
13	Shimkent	360,100	1	4	Kazakhstan	4,215	2
14	Jalna	244,523	1	158	India	2,220	1
15	Gorgan	188,710	1	33	Iran	5,460	3
16	Saidpur	114,000	1	25	Bangladesh	1,427	1
Southeast Asia							
1	Metro Manila	10,900,000	4	1	Philippines	3,668	2
2	Bangkok	7,281,000	4	1	Thailand	5,846	3
3	Ho Chi Minh City	4,615,000	4	1	Vietnam	1,854	1
4	Singapore	3,567,000	3	1	Singapore	21,832	4
5	Bandung	3,409,000	3	2	Indonesia	2,807	1
6	Medan	1,879,000	3	4	Indonesia	2,807	1
7	Palembang	1,422,000	2	5	Indonesia	2,807	1
8	Kuala Lumpur	1,378,000	2	5	Malaysia	8,217	3
9	Cebu	718,821	2	10	Philippines	3,668	2
10	Ipoh	566,211	2	2	Malaysia	8,217	3
11	Bacolod	429,076	1	7	Philippines	3,668	2
12	Songkhla	342,475	1	2	Thailand	5,846	3
Sub-Saharan Africa							
1	Addis Ababa	2,639,000	3	1	Ethiopia	648	1
2	Johannesburg	2,335,000	3	2	South Africa	8,667	3
3	Accra	1,976,000	3	1	Ghana	1,804	1
4	Harare	1,752,000	3	1	Zimbabwe	2,372	1
5	Ibadan	1,731,000	3	3	Nigeria	808	1
6	Pretoria	1,508,000	3	4	South Africa	8,667	3
7	Kampala	1,212,000	2	1	Uganda	1,164	1
8	Bamako	1,131,000	2	1	Mali	683	1
9	Ouagadougou	1,130,000	2	1	Burkina Faso	931	1

Table II-4: The Global Sample of 120 Cities (continued)

No	City				Country		
	Name	Population in 2000	Size Class	Rank	Name	GNP/cap. in PPP (\$)	Income Class
Sub-Saharan Africa (continued)							
10	Ndola	568,600	2	3	Zambia	715	1
11	Banjul	399,386	1	1	Gambia	1,542	1
12	Kigali	351,400	1	1	Rwanda	1,019	1
Western Asia							
1	Istanbul	9,451,000	4	1	Turkey	5,731	3
2	Tel Aviv-Jaffa	2,181,000	3	1	Israel	18,895	4
3	Baku	1,936,000	3	1	Azerbaijan	2,358	1
4	Sana'a	1,653,300	3	1	Yemen	760	1
5	Yerevan	1,406,765	2	1	Armenia	2,222	1
6	Kuwait City	1,190,000	2	1	Kuwait	14,471	3
7	Malatya	437,000	1	14	Turkey	5,731	3
8	Zugdidi	104,947	1	6	Georgia	1,722	1
	Total	415,605,624					

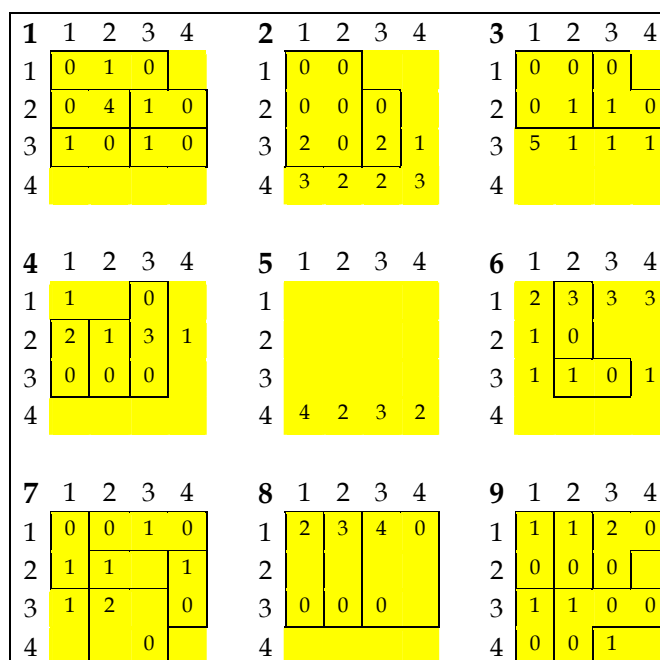


Figure II-1: Merged Strata in the Universe of Cities

Each one of the nine yellow boxes in figure II-1 represents a region, with the vertical dimension representing income classes and the horizontal dimension representing city

size classes. Empty boxes represent strata in which there was no city in the new universe of cities. Those marked with 0 represent strata that had cities in the universe, but no cities in the 90-city sample. Non-zero values represent the number of cities in the 90-city sample in that stratum. The outlines represent merged strata. If we represent each box with a 3-digit number (region, income group, size group), then the upper left square in figure II-2 indicates that one representative city in the 90-city sample represents three strata—111, 112, and 113; 4 cities represented strata 121 and 122, and so on.

Representative cities in both merged and non-merged boxes were assigned weights for each stratum they represented. For example, the city representing strata 111, 112, and 113 was assigned a weight corresponding to each stratum it represented—the weight being equal to the total population in the cities in the universe belonging to that stratum, divided by the population of the city representing the stratum. Similarly, if more than one city represented a stratum, each representative city was assigned the same weight—the weight being equal to the total population in the cities in the universe belonging to that stratum, divided by the total population of the cities in the 90-city sample representing the stratum.

The assignment of weights to each city in the 90-city sample made possible the calculation of several measures—both totals and averages—of urban extent and expansion that will be discussed in Chapter 4. They also made possible the construction of more accurate econometric models that explain the variation in urban extent and expansion in our universe of cities. Those will be presented and discussed in detail in Chapter V.

* * *

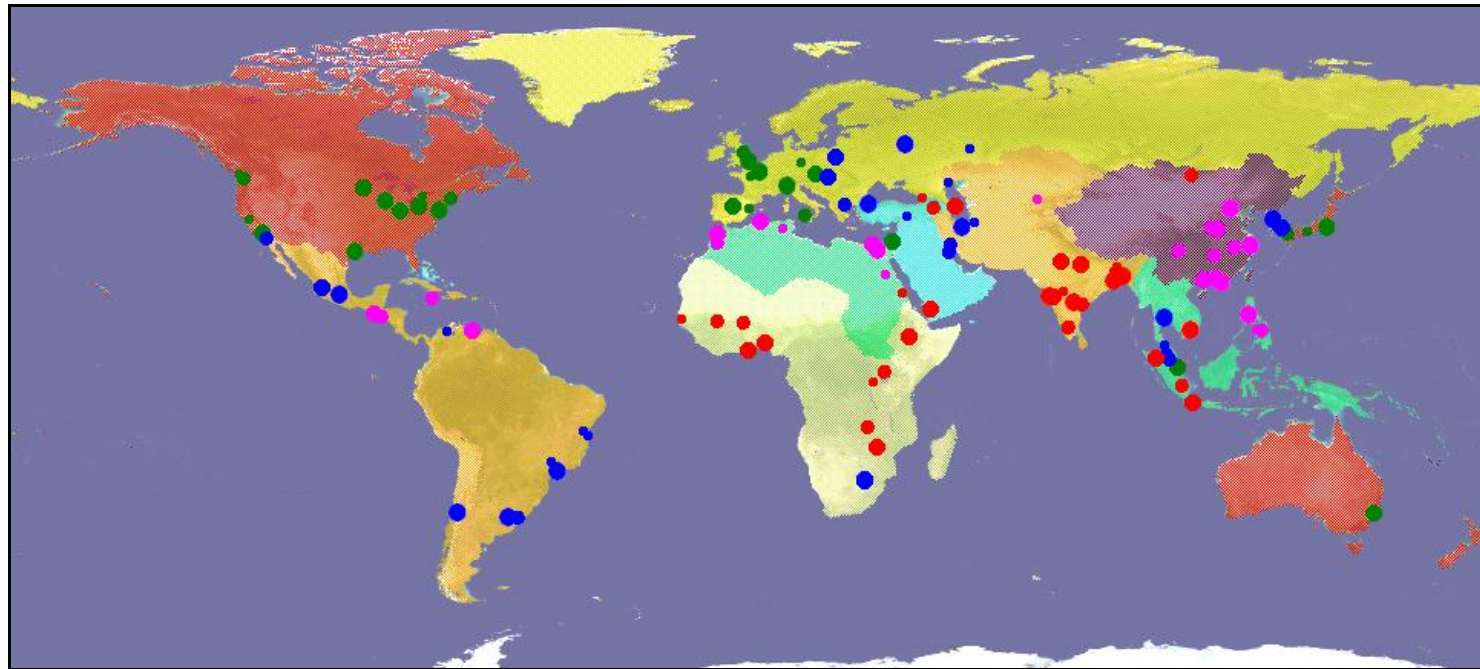


Figure II-2. The Global Sample of 120 Cities, by Regions, Population Size, and Income Classes

