Levels of urbanization in the world's countries:
alternative estimates

Leontine Alkema, Gavin W. Jones, Cynthia U. R. Lai*
March 31, 2012

Abstract

BACKGROUND Estimates of urbanization levels for all countries, as published in the United Nations World Urbanization Prospects, are not comparable across countries because they are based on national definitions of urbanization, which vary greatly across countries.

OBJECTIVE To construct alternative estimates of urbanization for all countries for 1990 and 2000 which are comparable across countries and can be used to point out countries with definitional problems or interesting urban characteristics.

METHODS We assume that the UN estimates represent the true proportion urban in a country, with measurement errors caused by definitional issues. To construct alternative estimates, we regress this proportion urban against variables that are associated with urbanization, as well as a categorical variable that summarizes the urban definition.

RESULTS Among the 181 countries included in the analysis, 21 per cent have a restrictive definition, and 31 per cent a generous definition of urbanization. Among the set of candidate predictor variables, GDP per capita, percent of employment in the agriculture sector and population density were selected as important predictors for urbanization. In 2000, the difference between the alternative estimate and the UN estimate was more than 10 per cent for 67 countries. Examples of countries with large differences include Trinidad and Tobago, Guyana, Sri Lanka and Thailand.

CONCLUSIONS Comparison of urbanization estimates across countries should be treated with care, given the wide range of definitions used. To produce more informative estimates of urbanization, several countries would need to clarify or re-think their definition of urban.

Introduction

The United Nations Population Division produces estimates of urbanization levels of all the world's countries, updated every two years (the latest being *World Urbanization Prospects: The 2009 Revision*). These are the only detailed set of estimates of the world's urban population, and as such, are widely used. The Population Division accepts the data as supplied by countries, based on the definitions of urban as adopted in each country. These definitions (which are listed in the United Nations publication) vary widely. To list some examples of definitions used:

- Afghanistan: Sixty-six localities and provincial centers.
- Bulgaria: Towns, that is, localities legally established as urban.
- Cambodia: Municipalities of Phnom Penh, Bokor and Kep and 13 additional urban centers.
- Lesotho: District headquarters and other settlements with rapid population growth and with facilities that tend to encourage people to engage in non- agricultural economic activities.
- Morocco: Respectively 250 and 370 localities in the 1982 and 1994 censuses.

This diversity in definitions of urbanization clearly creates problems of comparability. The 1996 Habitat Report (UN 1996) urges caution in using such data. "The proportion of the world's population currently living in urban centres is best considered not as a precise percentage (i.e. 45.2 per cent in 1995) but as being between 40 and 55 per cent, depending on the criteria used to define what is "an urban centre" (UN, 1996: 14)."

It is the aim of this paper to find more comparable estimates of the level of urbanization of the world's countries. But before pursuing this issue further, we need to pay attention to a more fundamental challenge. There is an increasingly prevalent argument that urban-rural distinctions are losing their meaning. There are two versions of this argument. One is that in countries where the proportion urban is 90 per cent or more, there is not much point in focusing on urban-rural distinctions, because so few people live in rural areas. The other is that, even in countries where the proportion urban is less than half, the blurring of the urban-rural distinction lowers the importance of estimating urban populations. Not only in the developed countries, but in less developed ones as well, complexities in settlement patterns, for example the emergence of extended metropolitan regions, makes it increasingly difficult to define town and countryside (see, e.g. Jones, 1997). For example, throughout most of Southeast Asia, 'urban' facilities have permeated rural areas to an astonishing degree over the past 40 years, as the forces of modernization "impinge on formerly isolated, inward-looking, self-sufficient and agriculturally-based communities" (Rigg, 1997: 157).

Ideally, given the complexity of defining and delimiting urban areas, we need to be able to distinguish some sort of rural-urban continuum, rather than a binary urban-rural classification. However, for purposes of inter-country comparison spanning all the world's countries, data classified according to an urban-rural continuum remain no more than a dream. Though acknowledging the need for more graduated measures of the urban-rural continuum, we argue that it is unrealistic to hope for such measures covering almost all the world's countries, at least in the next decade or two. Happily, there is still some value in a simple urban-rural distinction,

because that does manage to capture important differentials in socio-economic variables. Urban-rural distinctions in fertility levels, occupational structure, levels of education, and accessibility to services are usually clear cut (Champion and Hugo, 204: 5). Thus it would clearly be of value to provide alternative estimates of levels of urbanization that are more comparable across countries.

Data and Methods

Our modeling approach to construct more comparable estimates of urbanization is based on the assumption that the UN estimates (national definitions of urban) represent "true" proportion urban (PU), with measurement errors caused by definitional issues. To construct alternative estimates which are more comparable between countries, we regress the proportion urban against variables expected to have a relationship with urbanization, as well as indicators for the urban definition used within the country. The estimates of proportion urban as given by the country covariates, based on a "neutral" definition are presented as our alternative estimates.

Data

We focused our analysis on all countries with a population of 200,000 or above in 2000 (181 countries), and analyzed their levels of urbanization in 1990 and 2000. We constructed a data set with covariates which are thought to be correlated with urbanization levels. The covariates are given by: employment in agriculture (percentage of total employment), total fertility rate, GDP per capita (current US$), under five mortality rate (deaths under age 5 per 1,000 live births), net migration rate (per 1,000 population), population growth rate, population density, total labour force participation rate (percentage of total population ages 15 and over), male labour force participation rate (percentage of total population ages 15 and over), gross secondary school enrollment ratios, desert countries with oil production. Detailed information on the covariates and the sources are listed in the Appendix (Table A1). The covariates were collected for 1990 and for 2000.

For a number of the covariates, the standard sources of data had gaps for many countries, and considerable time was taken in tracking down alternative sources of data. Of the three covariates finally used, one – the percentage of employment in agriculture – was incomplete for many countries in the World Bank's *World Development Indicators* and the database of the International Labour Organization (ILO). This was a crucial covariate, and considerable time was taken in searching for other sources for the missing data. Wherever possible, the data were filled in from national sources such as censuses, labour force surveys and reports of the planning commission.

We categorized countries based on their definition of urban into three categories: Restrictive, neutral or generous definition of urban. Objective rules were sought as far as possible. Thus, countries with minimum size of settlement of less than 5,000 inhabitants in their definition of settlement were placed in the Generous category; countries with minimum size of settlement

between 5,000 to less than 10,000 inhabitants in the Neutral category; and countries with minimum size of settlement of 10,000 and more inhabitants in the Restrictive category. However, for a substantial proportion of countries, no such clear cut-off is given. Sometimes, one or a few cities and towns are considered urban, but there is no indication of their size; other countries have more complex definitions; some have very vague definitions (examples given in the Introduction). The allocation of countries into the generous, neutral or restrictive categories was therefore a very time-consuming process, and necessarily somewhat arbitrary. In cases where there was no clear basis for assigning the country into one or other of the three categories, the country was assigned the "neutral" category.

Regression analysis

We carried out regression analyses (separately for the 1990 and 2000 data sets), in which we regressed logit-transformed proportion urban on the covariates which are candidate predictors of the level of urbanization.
In the analysis, we excluded countries which were more than 95 per cent urban: to avoid problems with the logit transform for these countries (as the logit-transformation of 100 per cent is undefined), and because this subset of countries includes "city states" like Singapore, Macao and Hong Kong, for which we consider the levels of urbanization to not be comparable to other countries.

We examined appropriate transformations of predictor variables in exploratory analysis, and used the BIC (Bayesian Information Criterion, Schwarz (1978), Raftery (1995)) to select an optimal model for both data sets (forcing the indicators for Generous and Restrictive definitions to be included into the model and excluding 12 countries with missing covariates).

Results

Classification of countries

Out of the 181 countries included in the analysis, about half of the countries were found to use either a restrictive or generous definition of urban populations; the definition used was restrictive for 38 countries (21 per cent of all countries) and generous for 57 countries (31 per cent) (see Table A1 in Appendix for a list of countries with restrictive/generous definitions).

Regression analysis

The model that included the predictors GDP per capita (on log-scale), proportion of employment in agriculture, and population density (also on the log-scale) was among the set of candidate models with lowest BICs, for the 1990 and 2000 data sets. Residual analysis did not suggest the presence of interaction terms or non-linear relations between any of the predictors and urbanization.

The first order model was chosen to construct alternative estimates for 1990 and 2000. We present the results for the model fit based on the 2000 data set for 175 countries with PU less than 95 per cent. The results for the 1990 data set are similar (the results for the 1990 data set and exploratory plots are given in the Appendix).

The estimates for the regression coefficients are given in Table 1. The proportion of total variance explained (R^2) is 69 per cent. Unsurprisingly, urbanization is negatively related to employment in agriculture, and positively related to GDP per capita. After controlling for GDP per capita and agricultural employment, the relation between urbanization levels and the population density within a country is negative; less densely populated countries tend to be more urbanized. A possible explanation for this finding is that in less densely populated countries, with similar levels of employment in agriculture and GDP per capita, the population tends to be more clustered which facilitates urbanization.

The estimates of the misclassification errors based on generous or restrictive definitions of urbanization (Table 2) are estimated on the logit-scale of proportion urban, and thus depend on the level of PU. This is illustrated in Figure 1, which shows the expected PU that would have been reported in a country with a generous (purple) or a restrictive (green) definition, as a function of its "true" PU based on our estimated regression coefficients for generous and restrictive definitions. For example, if a country that is 50 per cent urban would use a definition that is restrictive, its reported PU is expected to be 42 per cent (with 95 per cent confidence intervals ranging from 37 per cent to 49 per cent) compared to 56 per cent that would be reported by a country with a generous definition (with 95 per cent confidence interval ranging from 50 to 62 per cent). These estimated average differences based on the categorizations are smaller when PU is closer to 0 or 100 per cent.

Table 1: Estimated regression coefficients (data set year 2000).

Predictor	Coefficient	SE	p-value
Intercept	0.42	0.62	0.50
Definition: Restrictive	-0.29	0.13	0.02
Definition: Generous	0.25	0.12	0.03
Percentage Total Employment in Agriculture	-0.03	0.00	$< 10^{-8}$
log(GDP per capita)	0.13	0.06	0.05
log(pop. density)	-0.07	0.04	0.06

Figure 1: Illustration of the expected differences in reported PU between countries with different definitions (generous or restrictive). The purple line represents the expected PU reported in a country with a generous definition of PU, and the green line represents the expected PU reported in a country with a restrictive definition of PU, for "true" PU ranging from 0 to 100 per cent. Dashed lines represent the 95 per cent confidence intervals for reported PU.
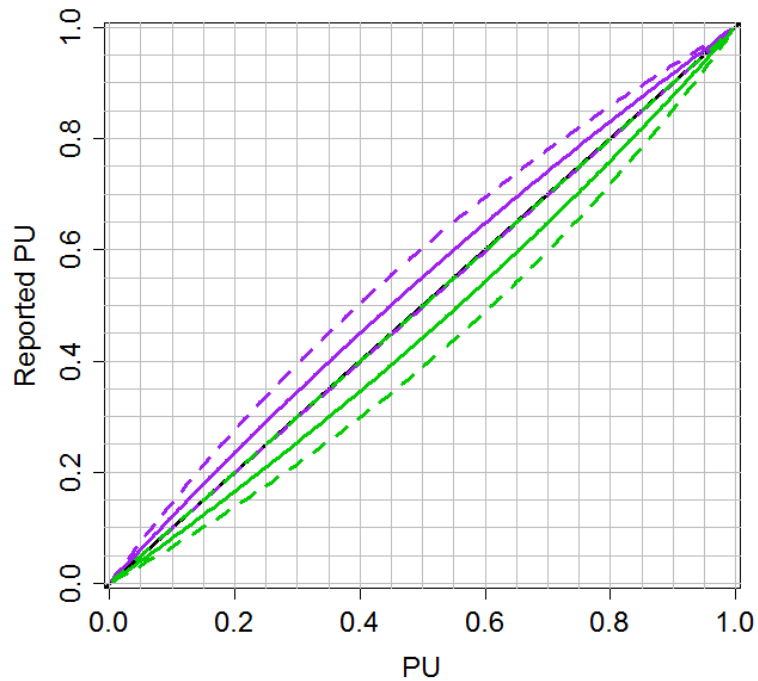
Overview of country estimates

The alternative estimates for all 181 countries are shown in Figure A1 and Table A3 in the Appendix. Figure 2 summarizes differences between the alternative and UN estimates for all countries. For the majority of countries, the difference is less than 10 per cent (within grey area). For 37 per cent of all countries (67 countries) the difference is more than 10 per cent. The estimates for these countries are illustrated in Figure 3.

Figure 2: Alternative estimates of proportion urban plotted against UN estimates (2000). (The country codes are explained in the Appendix, Table A3).
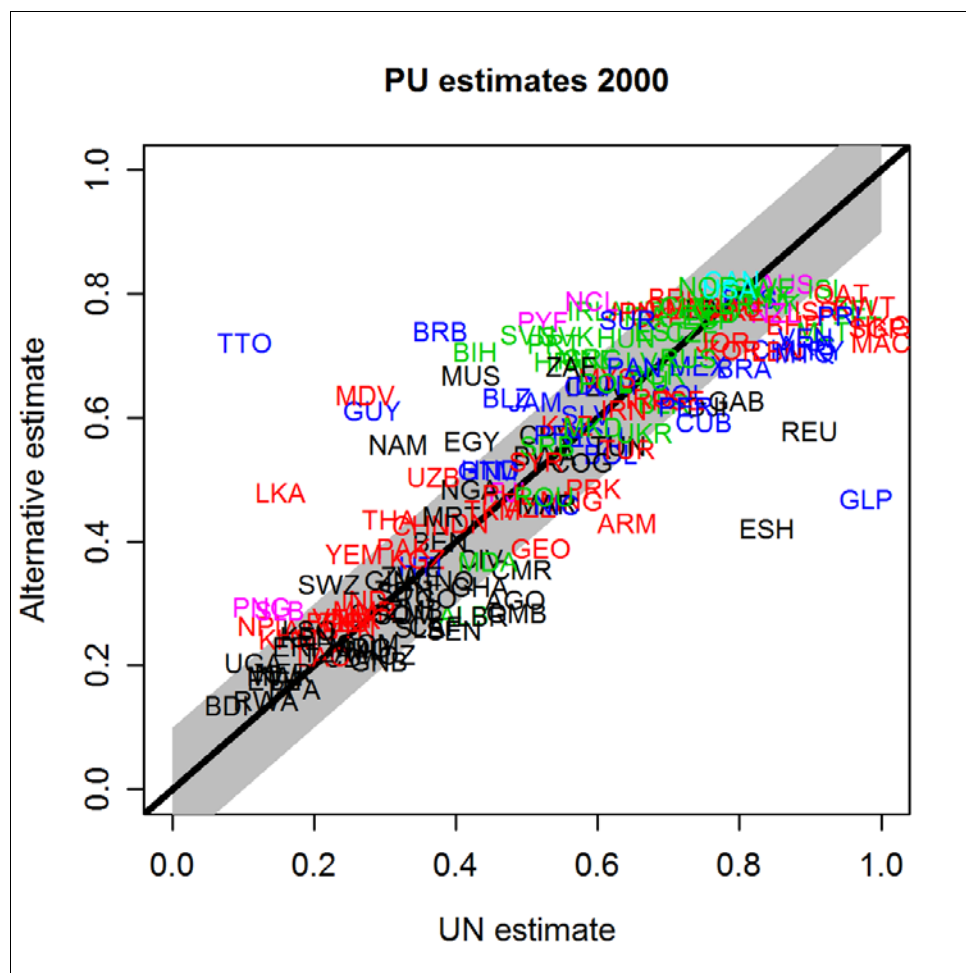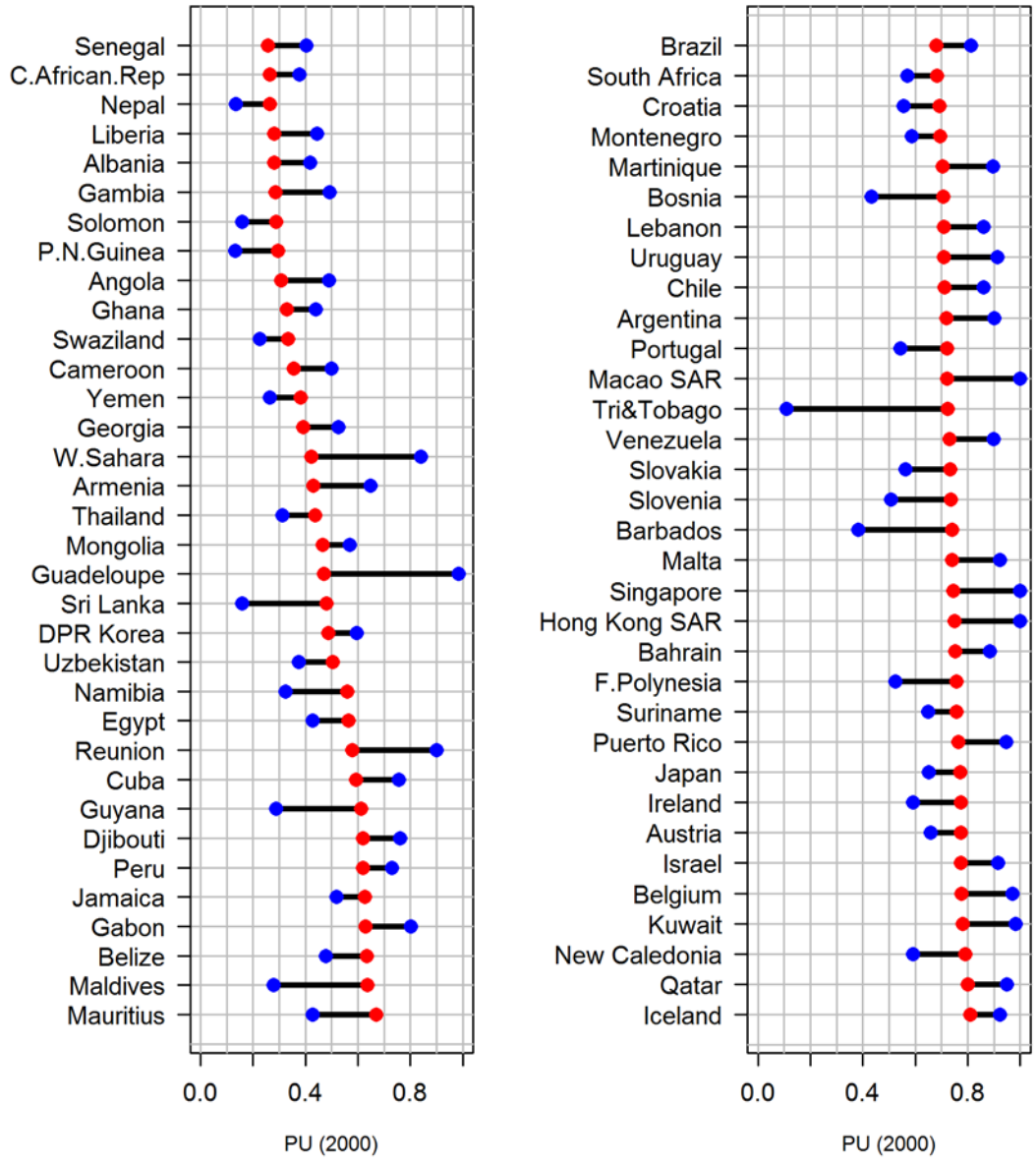
Figure 3: Countries with more than 10 per cent difference (absolute) between the alternative estimate of urbanization and the UN estimate for 2000 (alternative estimates in red, UN estimates in blue; countries are sorted by increasing alternative estimate).

Some of the differences in Figure 3 are extreme. In such cases there is a need to check the accuracy of the estimates for the explanatory variables GDP per capita, proportion of employment in agriculture, and population density. If there is no problem with these estimates, then there is a case for assuming that the official estimate of urbanization is idiosyncratic in some way, compared with that of the world's countries as a whole.

The most extreme case is that of Trinidad and Tobago (country code TTO), where the UN estimate of 11 per cent urban is much smaller than the alternative estimate of 72 per cent. Here, the official definition of urban is clearly the main reason. The very low percentage of employment in agriculture (7 per cent) is highly inconsistent with only 11 per cent of the population living in urban areas. The official definition of urbanization in Trinidad and Tobago is "Port of Spain (capital), Arima borough and San Fernando town". Other information on the urban population of Trinidad and Tobago indicates that these three localities greatly understate the true urban population of Trinidad and Tobago. The population of Port of Spain has been declining ever since 1960, as land shortage has developed in the capital and urban growth has spilled beyond its boundaries. At the same time, urban sprawl has focused on the East-West Corridor, stretching from Chaguramas, west of Port of Spain to Arima in the east. In 1980, this East-West corridor was estimated to contain 43 per cent of the national population (Fritz, 2008: 4); about one third of the population of Trinidad and Tobago currently lives in Port of Spain or its suburbs or within 16 km of them (*Encyclopedia of the Nations*).

Another extreme case is that of Guyana (country code GUY). This is a country with a population of only 751,000. Its urban definition is "City of Georgetown (capital) and four other towns". 41 per cent of the population lives in the Demerera-Mahaica region (centred on the city of Georgetown), with an area of 1,843 square kilometer, or 0.9 per cent of the country's land area. The population density in the Demerera-Mahaica region is 168 per square kilometer. These figures provide grounds for suspecting that the urban definition in Guyana, which results in only 28.7 per cent of the population being classified as urban, may exclude considerable "overspill" urbanization in the core urban region, though this needs to be further investigated. Another possible source of the discrepancy is the estimate of the percentage of employment in agriculture, which is 27.8 per cent according to the ILO "key indicators of labour market" database. Further checks confirm that the figure is roughly of the correct order. A detailed study on employment structure in Guyana using data from the 2002 census gave a figure of 22 per cent for employment in the agriculture and fishing sectors (Private Sector Commission of Guyana, 2007, Table 1). This strengthens our suspicion that the official urbanization estimate is too low.

It is not possible at this stage to double check every case in which our alternative estimate of urbanization differs widely from the UN estimate. However, the cases of Trinidad and Tobago and Guyana illustrate the kind of approach that can be used.

Two Asian case studies will also be briefly discussed here – Sri Lanka (LKA) and Thailand (THA). Sri Lanka is not only one of the most extreme outliers, but is a case where the recorded level of urbanization in 2000 was lower than it was 40 years before. Sri Lanka's recorded level of

urbanization rose gradually from 15.3 per cent in 1953 to 22.4 per cent in 1971. Thereafter, it fell slightly to 21.5 per cent in 1981, and more sharply to 14.6 per cent in 2001. (The UN estimate for 2000 was 15.83 per cent). Such a fall is inherently unlikely, in a country that is gradually industrializing, albeit slowly; the alternative estimate based on our model is 48 per cent. There are two main reasons for the low official estimate for 2000: (1) part of the country was not included in the later estimates, because of civil war; (2) the definition of urban areas was changed (see Abeykoon, 2005; De Silva, no date). The latter is almost certainly the main reason for the discrepancy; indeed, it is not certain that the exclusion of the Jaffna area from the estimates because of civil war would have lowered or raised the estimate of urbanization. Before 1981, areas identified as urban in Sri Lanka included municipal councils, urban and small town councils, and village councils (which were defined as rural). In 1981, small town councils and village councils were dissolved, and included in newly created Development Councils (*Pradesheya Sabbas*). Whether these Development Councils are urban or not cannot be determined, but in 1987 the definition of urban areas was changed to include only municipalities and urban councils. This meant that the populations of the former town councils, which had been considered urban, were no longer included in estimates of the urban population, and this served to lower the estimated urban population considerably.

Thailand is another country for which our estimate lies above the official estimate of urbanization. Actually, before 1999 Thailand's urbanization level were even more seriously understated, because only Municipalities were considered to the urban areas, while sanitary districts, some of which have populations as large as 20,000, were excluded from the urban population. In 1999, a major administrative reorganization took place, whereby all sanitary districts were upgraded to municipal status, raising the total number of municipalities from 131 to 1,081. Although this served to raise the estimate of Thailand's urbanization to 31 per cent, this still tended to underestimate Thailand's real level of urbanization, because the *tambon* (i.e. sub-district) administrative areas continue to be designated rural, although some of them, especially just outside municipalities, have distinctly urban characteristics (Jones, 2004: 115-117).

The four countries discussed – Trinidad and Tobago, Guyana, Sri Lanka and Thailand – all point to the value of our approach in flagging countries whose level of urbanization is, on the face of it, surprisingly low, and then investigating in more detail the plausibility of the official urbanization level. A similar approach is needed in the case of countries where the official urbanization level is surprisingly high.

With regard to a few other countries, Hong Kong, Macao and Singapore are (not surprisingly) each estimated to be 100 per cent urban, according to the national definitions of urban. For these countries, our alternative estimates are much lower (Hong Kong 75 per cent; Macao 72 per cent and Singapore 75 per cent), and not appropriate.

Belgium was one of the countries that was excluded during the variable selection and model fitting procedure; its UN estimate of PU is 97 per cent. The alternative estimate for Belgium is much lower than the UN estimate (78 per cent). The alternative estimate for Belgium makes it

correspond closely to other western European countries such as the Netherlands (UN estimate 76.8 per cent) and France (UN estimate 76.9 per cent), and thus improves the comparability across countries.

Discussion

The objective in this paper was to construct estimates of urbanization that are comparable across countries, and that can aid in pinpointing countries with either problematic definitions of urbanization, or countries with unusually high or low levels of urbanization, compared to the expected outcome based on the predictors. To construct such estimates, we classified countries according to their definition of urbanization, and used GDP per capita, percent of employment in the agriculture sector and population density as predictors for urbanization.

Out of the 181 countries included in the analysis, 21 per cent have a restrictive definition and 31 per cent a generous definition of urbanization. Among the set of candidate predictor variables, GDP per capita, percent of employment in the agriculture sector and population density were selected as important predictors for urbanization. In 2000, for 67 countries, the difference between the alternative estimate and the UN estimate was more than 10 per cent. Examples of countries with large differences include Trinidad and Tobago, Guyana, Sri Lanka and Thailand.

The underlying rationale for this project is the need for an alternative set of more comparable estimates of levels of urbanization of the world's countries, given the wide range of definitions and procedures underlying the official estimates of urban population in different countries. The derivation of alternative estimates turned out to require considerable time and effort, mainly because of the difficulty of preparing an internationally comparable set of values for the predictors. Population density figures were easily come by, GDP per capita readily available for most countries but problematic for some countries, and percent of employment in agriculture problematic for a larger number of countries. Some questions remain about the data for some countries, but on the whole we believe the estimates used are the best available.

Based on the linear regression model, confidence intervals and prediction intervals can be constructed. Both types of intervals include the uncertainty in the estimated coefficients of the predictors; prediction intervals also include the additional uncertainty that arises from the variability between countries in their levels of urbanization which is not explained by the model. We chose not to report either because we feel that appropriate uncertainty bounds are likely to be somewhere in between the confidence bounds and prediction bounds. Confidence bounds do not include the uncertainty associated with the "true" unexplained variability between countries; the fact that urbanization is a complex phenomenon which cannot be captured fully by predictors such as GDP, density or percentage working in agriculture alone. While prediction intervals do include that source of uncertainty, they also include the uncertainty due to the "measurement errors" in PU (the part that has not been captured in the definitional categorization). Instead of reporting inappropriate uncertainty bounds, we chose to

pinpoint to countries that need further investigation based on a "cut-off" rule of more than 10 per cent difference between the UN estimates and the alternative estimates (in line with the suggestion by UN habitat that the UN estimates could be off by around that margin of error, see Introduction). Additional analyses on the variability of the alternative estimates confirmed that the rule is robust to model specifications; additional model fits based on leaving out influential data points and using different subsets of predictors pinpointed to the same subset of countries for which the alternative estimates and the UN estimates differed by more than 10 per cent.

Comparable estimates of the level of urbanization – based on a binary division of settlements into rural and urban - are an important first step towards a more important goal – that of providing comparable estimates for each country of the distribution of their population across a rural-urban continuum. We are far from being able to reach this more important goal, but in the meantime, the first step is not unimportant. What this exercise has shown is that the wide differences in the procedures countries adopt for measuring their urban population makes a direct comparison of the results misleading in many cases. For example, definitions based on purely administrative criteria, such as considering all places designated as municipalities as urban areas, can lead to substantial underestimates of urban population growth if newly emerging towns are not accorded municipal status, and if boundaries of existing towns and cities are not adjusted on the basis of urban "overspill". By the same token, it is possible for urbanization to be overestimated when overly generous boundaries are used for areas designated as urban. Many countries need to clarify or re-think their definitions, taking into account the uses to which data according to urban-rural designations are put.

References

Abeykoon, A.T.P. L. (2005). Urbanization in Sri Lanka. In Ness, G. D. and Talwar, P. P. (eds*). Asian Urbanization in the New Millennium*. Singapore: Marshall Cavendish.

Champion, T. and Hugo, G. (2004). Introduction: Moving beyond the urban-rural dichotomy. In Champion, T. and Hugo, G. (eds). *New Forms of Urbanization: Beyond the Urban-Rural Dichotomy*. Aldershot. England: Ashgate.

De Silva and Indralal, W. no date. Dilemma in measuring urbanization in Sri Lanka and the Millennium Development Goals [unpublished paper]. University of Colombo.

Encyclopedia of the Nations – entry on Trinidad and Tobago.

Fritz, A. (2008). *Trincity millennium vision Trinidad and Tobago*. Paper presented at 44th ISOCARP Congress.

Jones, G. W. (1997). *The thoroughgoing urbanization of East and Southeast Asia. Asia Pacific Viewpoint*, 38(3): 237-50.

Key Indicators of the Labour Market, 6th Edition, International Labour Organization.

Private Sector Commission of Guyana Ltd. (2007). *Employment Structure in Guyana*. Technical Information Bulletin No. 5, Georgetown, Guyana.

Raftery, A. E. (1995). *Bayesian model selection in social research* (with discussion). Sociological Methodology 25, 111-196.

Rigg, J. (1997). *Southeast Asia: The Human Landscape of Modernization and Development*. London: Routledge.

Schwarz, G. E. (1978). *Estimating the dimension of a model*. Annals of Statistics 6 (2): 461–464.

United Nations Education, Scientific and Cultural Organization (UNESCO).

United Nations (1996). *An Urbanizing World: Global Report on Human Settlement.* Oxford: Oxford University Press.

United Nations Population Division (2009). *World Urbanization Prospects: the 2009 Revision*. New York: United Nations.

United Nations Population Division (2010). *World Population Prospects: the 2010 Revision*. New York: United Nations.

World Development Indicators (2010). The World Bank.