Probabilistic Projections of Urbanization for All Countries

Leontine Alkema, Patrick Gerland and Thomas Buettner *

Extended abstract, February 24, 2011

1 INTRODUCTION

The UN Population Division produces estimates and projections of the proportion urban in every country, which are revised every two years and published in the World Urbanization Prospects (United Nations, Department of Economic and Social Affairs, Population Division 2010). The UN estimates and projections are of interest to a range of scientists and policy makers worldwide, because they are the *only* source for estimates and projections of urbanization for all countries. Projections are crucial, as urbanization interacts with/influences/affects a variety of global trends like energy consumption, climate change, future outbreaks of epidemics (EIA 2009; Gubler 2002; World Bank 2009).

The UN projections of the proportion urban are based on the urban-rural growth differential in a country: the growth rate of the ratio of the urban to rural population. The last observed urban-rural growth differential in each country is projected forward to converge to a "global norm", based on observed urbanization trends in countries in the past. Two issues arise in the UN projections. Several researchers have pointed out that the UN projections are biased upwards; the UN projections from the past have significantly overestimated the rate of urbanization (National Research Council 2003; Cohen 2004; Bocquier 2005; Montgomery 2008). Secondly, the UN produces deterministic urbanization projections; an uncertainty assessment of the range of possible future outcomes based on the country's current situation and past trend is lacking. Bocquier (2005) proposed an alternative model for projecting the proportion urban in each country, but this model does not include an uncertainty assessment of future outcomes either.

---

*Leontine Alkema, Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546; Email: alkema@nus.edu.sg. Patrick Gerland, Population Estimates and Projections Section, United Nations Population Division, New York, NY 10017; Email: gerland@un.org. Thomas Buettner, United Nations Population Division, New York, NY 10017; Email: buettner@un.org. The project described was partially supported by the Global Asia Institute, National University of Singapore. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the United Nations. Its contents have not been formally edited and cleared by the United Nations. The authors are grateful to Kirill Andreev, Deborah Balk, Alex Cook, Danan Gu, Gerhard Heilig, Gavin Jones, Mark Montgomery, Nan Li, Adrian Roellin and Thomas Spoorenberg for helpful discussions and insightful comments.

Our objective is the development of probabilistic projection methodology for the proportion urban in every country; to construct country-specific projections that include an uncertainty assessment. We propose to use a Bayesian hierarchical model (Lindley and Smith 1972; Gelman et al. 2004) to estimate the parameters in the projection model, such that projections are based on the observed relation in the country of interest, as well as on the overall regional or global experience.

The extended abstract is organized as follows. The next section describes the data on urbanization. The following section describes the current UN projection methodology, and the model details for the Bayesian projection model. The paper concludes with preliminary results of the Bayesian projection model, and a discussion of possible model improvements.

2 DATA

As of March 2008, the UN Population Division has collected over 2200 empirical estimates of the proportion urban for 227 countries or areas. After evaluating the various estimates and making adjustments for definitional changes and urban/rural reclassifications, a subset of 1711 data points spanning from 1800 to 2007 were used to produce national estimates and projections of the proportion urban from 1950 to 2050 as part of the 2007 revision of the World Urbanization Prospects (WUP) (United Nations, Department of Economic and Social Affairs, Population Division 2007).

The primary data sources used in this paper are censuses and official estimates from the WUP 2007 data set. Within the last decade, UN analysts have rated the reliability of these observations. Generally, well-documented census or official estimates (especially within the last two decades) have been defined as "Complete/High quality" observations, except for a few cases in which the data quality was questionable. About two-thirds of the data set (mostly before 1990) has not be rated yet, due to the lack of meta-information allowing a proper assessment of data reliability. Table 1 gives the break-down of the observations by source and completeness/data quality.

In this analysis, all observations that were categorized as "Complete/ high quality" as well as "Unknown" are included. We excluded (i) countries which are 100% urban, (ii) countries with only one observation and (iii) observations before 1950. Observation years are rounded (only two countries have two observations within the same year, we take the average of these outcomes). The resulting data set contains 1311 observations in 209 countries. Figure 1 gives an overview of the number of observations for each country.

Table 1: Overview of data sources and quality of observations

Data source	Complete/ high quality	Incomplete/ questionable quality	Unknown
Census	377	8	778
Estimate	134	15	198
Register	3	0	3
Sample survey	5	1	5
Unknown	0	0	1

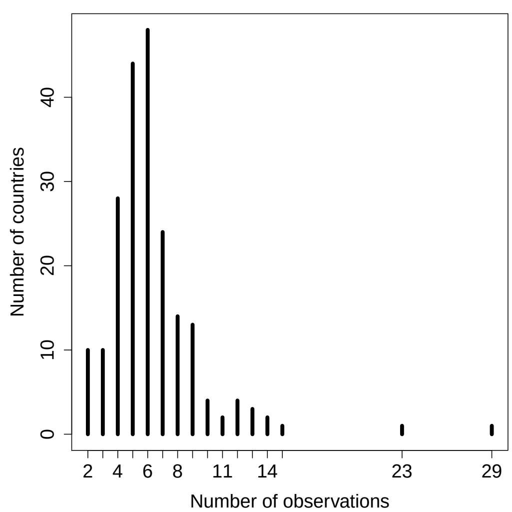


Figure 1: Overview of the number of observations for each country.

3 METHODS

3.1 UN methodology for projecting urbanization

In the UN projections, the process of urbanization is modeled with a logistic growth curve for the proportion urban, or equivalently, by assuming exponential growth of the ratio of the urban to rural population (United Nations, Department of Economic and Social Affairs, Population Division 2007). The UN projections of the proportion urban are based on the urban-rural growth differential in a country; the growth rate of the ratio of the urban to rural population, which is given by:

$$r_{c,t} = \left(log\left(\frac{p_{c,t}}{1-p_{c,t}}\right) - log\left(\frac{p_{c,t-1}}{1-p_{c,t-1}}\right)\right), \tag{1}$$

with $p_{c,t}$ the proportion urban in country c, year t. The growth differential $r_{c,t}$ is the difference in the proportion urban between year t and $t-1$ on the logit-transformed scale, and thus the "annual growth rate of ratio urban/rural populations":

$$\frac{U_{c,t}}{R_{c,t}} = \frac{U_{c,t-1}}{R_{c,t-1}} \exp\left(r_{c,t}\right),$$

with $U_{c,t}$ the urban and $R_{c,t}$ the rural population in country c, year t.

The last observed urban-rural growth differential in each country is projected forward to converge to a "global norm", usually in the next 20 years. The global norm is based on the observation that urban growth slows down at higher proportion urban, because of depletion of the pool of potential rural-urban migrants. The global norm gives the growth differential $r_{c,t+s}$ as a linear function of proportion urban $p_{c,t}$, and is illustrated in Figure 2 with the blue solid line. Its outcome is based on regressing observed differentials on their initial observed proportions urban. This exercise was carried out for the 1996 revision of the World Population Prospects (United Nations, Department of Economic and Social Affairs, Population Division 1998), and included 113 countries with more than one million inhabitants in 1995.

The projection approach is illustrated in Figure 2 for China and India. Observed urban-rural growth differentials are plotted against proportion urban for China in red and India in green. For both countries, the UN projections of their growth differentials are illustrated with the blue dotted line. For China, in which the current growth differential is above the global norm, its growth differential is projected to decrease linearly until it reaches the global norm, after which the differential is given by the solid blue line. For India, in which the growth differential is currently below the global norm, it is projected to increase. Once the growth differential hits the global norm, the future rates of urbanization are no longer country specific. An uncertainty assessment of the range of possible future outcomes based on the country's current situation and past trend is lacking.

As mentioned before in the Introduction, several researchers have pointed out that the UN projections are biased upwards; the UN projections from the past have significantly overestimated the rate of urbanization (National Research Council 2003; Cohen 2004; Bocquier 2005; Montgomery 2008). The observed bias in recent projections is illustrated in Figure 3. This figure shows the UN global norm (blue) together with local smoothers to illustrate levels and trends in growth differentials in different time

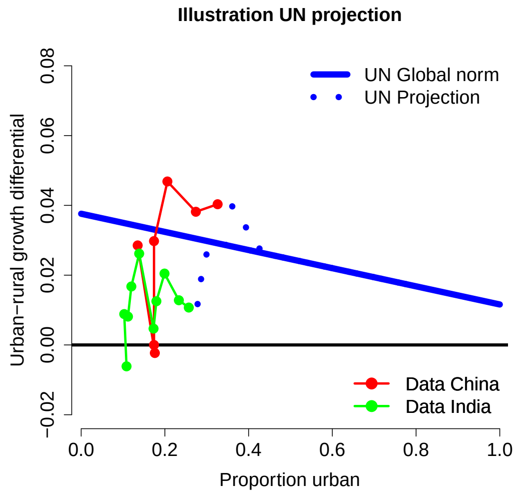


Figure 2: Urban-rural growth differential plotted against proportion urban for China (red) and India (green), with the UN projected growth differentials in blue (blue dotted line) and the UN global norm (blue solid line).

periods. The local smoothers are fitted to observed growth differentials from all countries, the results for the observations in 1950–1970 are shown in black, for 1970–1990 in red and for 1990–2010 in green (note that the outcomes of the local smoother are less informative at very low or high proportion urban). The smoothers show that on average, the growth differentials are smallest in the most recent period, and smaller than the global norm at most levels of urbanization during that period. Using the global norm for projecting urbanization in the past 20 years would have led to an overestimate of the proportion urban.

One possible explanation for the overestimate of growth differentials in recent years are that the UN projections are based on the assumption that all countries will eventually become 100% urban in the far future, which does not have to hold true, as illustrated in Figure 4 for Denmark and Germany. In both countries the proportion urban has been relatively stable for the last decades, but the UN projections indicate increasing levels of urbanization.

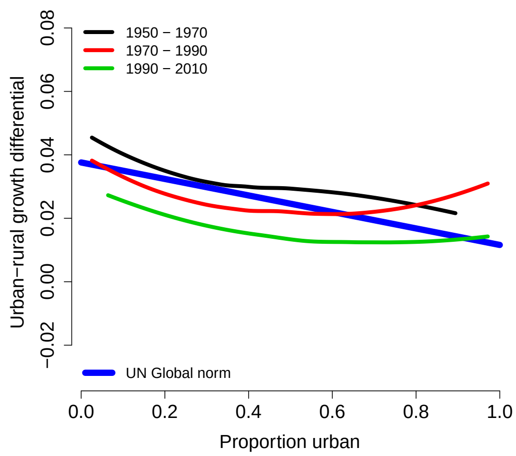
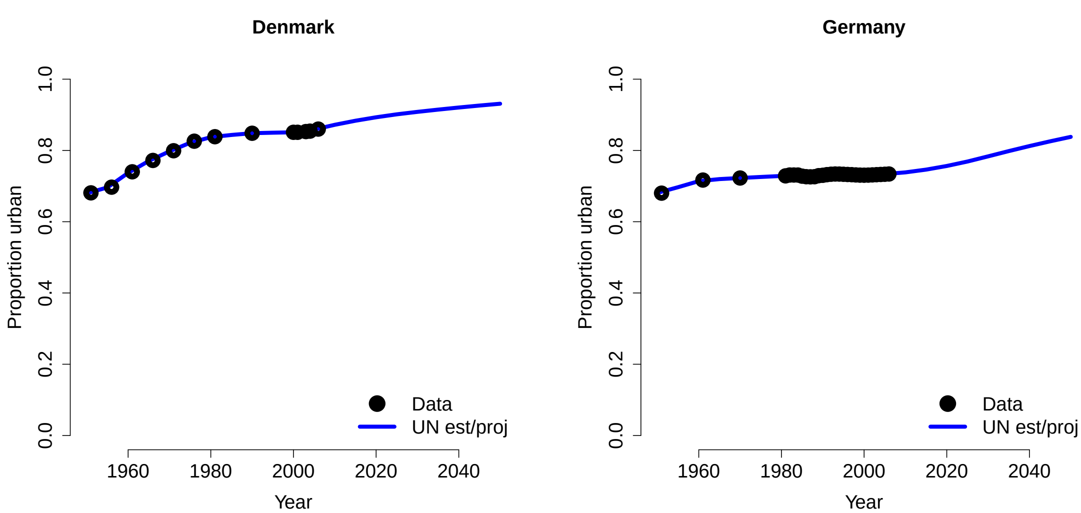


Figure 3: UN global norm (blue) together with local smoothers to illustrate levels and trends in growth differentials in different time periods. The local smoothers are fitted to observed growth differentials from all countries in 1950–1970 (black), 1970–1990 (red) and 1990–2010 (green).

Figure 4: Proportion urban in Denmark and Germany, with UN estimates and projection (blue).



3.2 Bayesian projection model

The objective in this paper is to extend the current UN projection model such that the urbanization projections are country-specific, include an uncertainty assessment and are well-calibrated. Our model is an extension to the current UN model using a slightly modified version of $r_{c,t}$, to take into account that not all countries necessarily converge to a 100% urban population.

Modified growth differential: The modified growth differential is defined as:

$$d_{c,t} = \left(log\left(\frac{p_{c,t}}{p^{(c)}_{max} - p_{c,t}}\right) - log\left(\frac{p_{c,t-1}}{p^{(c)}_{max} - p_{c,t-1}},\right)\right),$$

where $p^{(c)}_{max}$ is the maximum proportion urban for a country of interest, its "urbanization asymptote". The modified differential $d_{c,t}$ can be seen as the urban-rural growth differential within proportion $p^{(c)}_{max}$ of the population, thus leaving out proportion $(1 - p^{(c)}_{max})$ of the population that will never become urban. Using $d_{c,t}$, urbanization projections are given by:

$$\frac{U_{c,t}}{p^{(c)}_{max}N_{c,t} - U_{c,t}} = \frac{U_{c,t-1}}{p^{(c)}_{max}N_{c,t-1} - U_{c,t-1}} \exp(d_{c,t}),$$

with $N_{c,t}$ the total population in country c, year t.

We assume that $d_{c,t}$ fluctuates around its country-specific average growth differential, denoted by ω_c, and model this with an AR(1) time-series model:

$$d_{c,t} - \omega_c = \rho_c(d_{c,t-1} - \omega_c) + \varepsilon_{c,t}, \quad (2)$$
$$\varepsilon_{c,t} \sim N(0, \sigma^2), \quad (3)$$

with autoregressive parameter $0 < \rho_c < 1$, such that the expected $d_{c,t}$ depends on its previous outcome, and how far that outcome was from ω_c.

Bayesian hierarchical model: Estimating the country-specific parameters ρ_c, ω_c and $p^{(c)}_{max}$ presents a challenge because of the limited number of observations for each country (see Figure 1). We use a Bayesian hierarchical model (Lindley and Smith 1972; Gelman et al. 2004) to estimate these parameters in each country, such that the estimates are based on the observations in the country of interest, as well as on the overall regional or global experience. A hierarchical approach to projecting demographic outcomes for a number of countries is a natural way to exchange information between countries while constructing country-specific probabilistic projections. The fewer the number of observations in the country of interest, the more its projection is driven by the experience of other countries, while in countries with many observations the projection will be driven more by its own history.

In the Bayesian hierarchical modeling approach for ω_c we assume that for all countries, ω_c is drawn from a probability distribution that represents the range of outcomes of the average annual growth differential across all countries. For ω_c in a specific country, its probability distribution based on the world-level experience is then updated using Bayes' theorem with the observed trend in the country, which results in the posterior distribution for ω_c. The resulting estimates (draws from the posterior distribution) can be viewed as weighted averages of a "world pattern" and information from the country data. The hierarchical distribution for ω_c is given by:

$$\log(\omega_c) \sim N(\omega^*, \sigma^2_\omega),$$

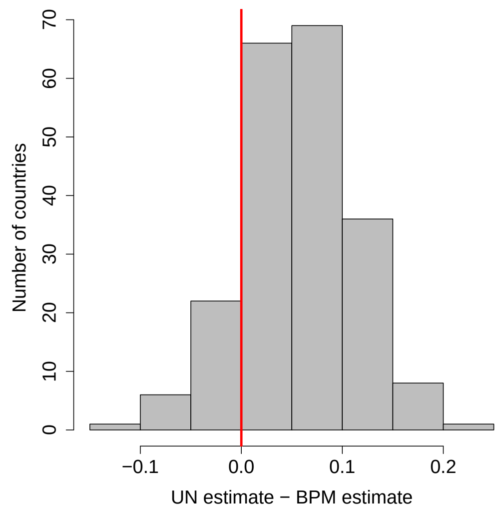
with hierarchical mean and variance ω^* and σ^2_ω on the log-transformed scale, to restrict ω_c to positive outcomes. Spread out prior distributions are assigned to ω^* and σ^2_ω (the full model is given in the Appendix). A similar approach is used for the autoregressive parameter ρ_c and the asymptote $p^{(c)}_{max}$.

Estimation: All model parameters are estimated in a Bayesian framework. Spread out prior distributions are assigned to the additional model parameters. A Markov Chain Monte Carlo (MCMC) algorithm is used to get samples of the posterior distributions of the parameters (Gelfand and Smith 1990). The MCMC sampling algorithm was implemented using Winbugs software (Lunn et al. 2000). The result is a set of future urbanization trajectories for each country.

4 PRELIMINARY RESULTS

The UN projections (World Urbanization Prospects 2007) of the proportion urban in 2050 tend to be higher than the preliminary projections by the Bayesian urban projection model, as illustrated in Figure 5, which show the differences between UN and Bayesian projection model (BPM) projections for the proportion urban in 2050 in all countries.

Figure 5: Histogram of the differences between UN and Bayesian projection model (BPM) projections for the proportion urban in 2050 in all countries.



Preliminary results for a subset of countries are given in Figures 6 and 7. The selected countries differ with respect to past trends and levels of urbanization, as well as the number of observations.

In Figure 6 for China, the median projection of the Bayesian projection model (BPM) shows a decrease in the pace of urbanization and suggests that China will be approximately 50% urban by 2050. In India, the pace of urbanization so far has been very slow, which results in a continuation of a slow pace of urbanization in the projections, and a low asymptote around 60%, which is significantly lower than in China. Because of the low levels of urbanization in Angola and Madagascar, and because the last observation years are well before 2010, there is a lot of uncertainty in current and future outcomes of the proportion urban in both countries. The UN and BPM projections are similar for Madagascar, the UN projections are within the 50% projection intervals of the BPM. In Egypt, the level of urbanization has not changed much since 1960, and the projections suggest that the level will increase only slightly.

Figure 7 shows the results for Denmark, Germany, Belize and Columbia. In Denmark and Germany, the proportion urban has not changed much in the last decades, this is reflected in the projection intervals which are every narrow and just above the current value. For Belize, the level of urbanization has decreased slightly, followed by a slight increase in the most recent observation period. The BPM projects a slight further increase. In Columbia, the pace of the urbanization process has decreased recently; the projections suggest a further leveling off.

Figure 6: Projection intervals for selected countries; The median projection is are shown in red and the UN projections (WUP 2007) are shown in blue. The 50% projection intervals are represented by the dark grey area, the 95% projection intervals by the lighter grey area. The asymptote $p_{max}^{(c)}$ and its 95% confidence interval are show in dark green.

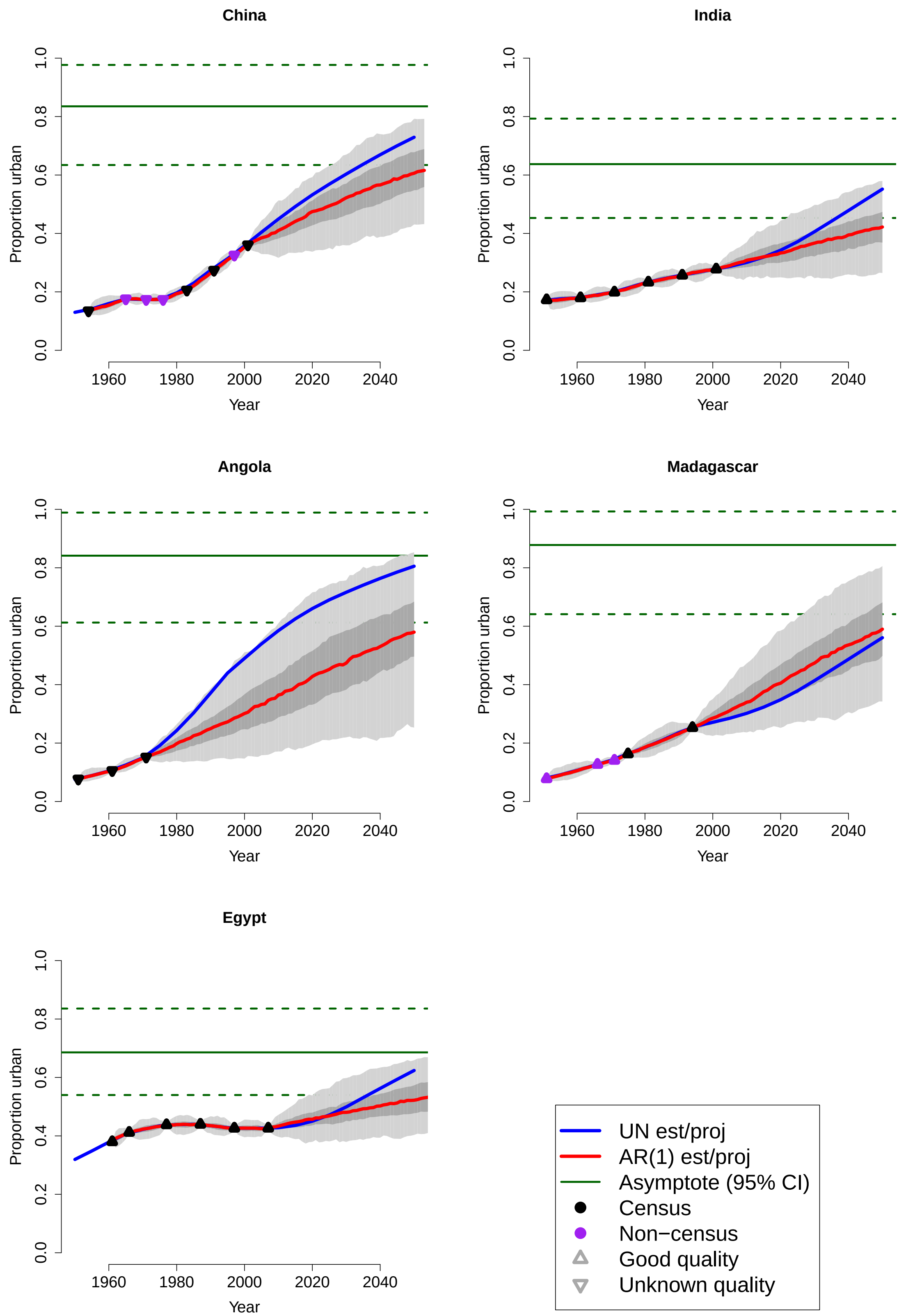
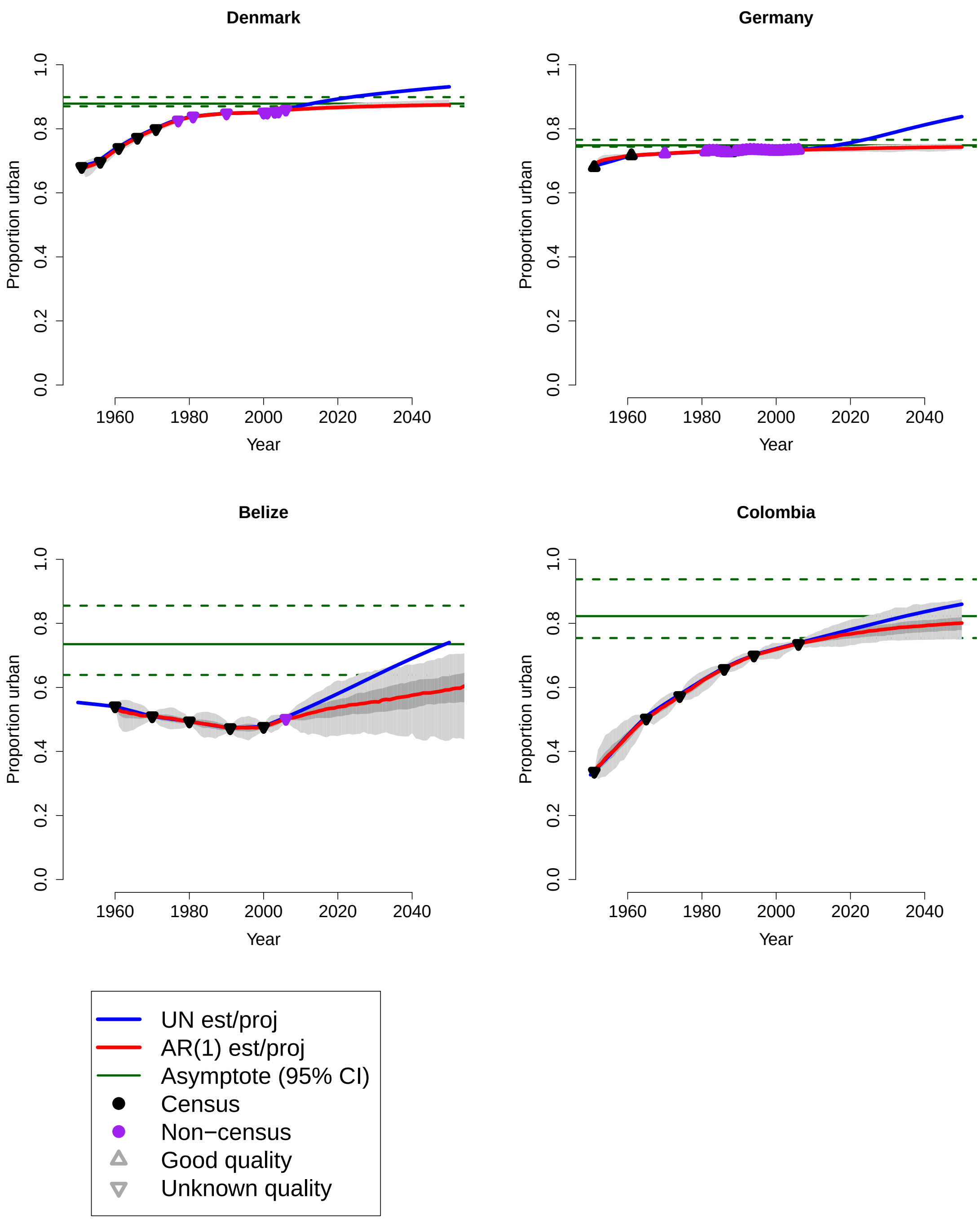


Figure 7: Projection intervals for selected countries; The median projection is are shown in red and the UN projections (WUP 2007) are shown in blue. The 50% projection intervals are represented by the dark grey area, the 95% projection intervals by the lighter grey area. The asymptote $p^{(c)}_{max}$ and its 95% confidence interval are show in dark green.



5 DISCUSSION

We presented a model for probabilistic projections of urbanization, in which projections are based on a country's past trends as well as the global experience, and showed preliminary results. We plan to carry out extensive model checking, including in-sample and out-of-sample model validation.We also plan to investigate into model changes such as including demographic indicators into the time series model for the modified growth differential (e.g. the proportion of the population aged 15-39 (United Nations, Department of Economic and Social Affairs, Population Division 2009)).

It is well understood that national practises of defining what is an urban population differ significantly between countries, for an example see Jones (2004). In addition, it appears that countries are not likely to undertake efforts to harmonize national definitions in the interest of international comparability, despite long-standing requests by the research community. As the basis for monitoring the process of urbanization will remain divers, there is currently no alternative to relying on national definitions of urban and national estimates of urban population.

Is it possible to enhance the projection model by including (non-demographic) covariates? Kelley and Williamson (1984: 73) found that rapid rates of population growth are not, as often argued, the central influences driving the urbanization in developing countries, though they do foster rapid growth of urban populations. The most potent influences on urbanization in developing countries appear to have been the rate and unbalancedness of sectoral productivity advances, and relative prices. The problem in using such findings for the projection of urbanization is the unavailability of reliable projections of such variables, even for a few years, let alone decades. It is therefore not possible to use those and other covariates effectively for urban projections. This is well illustrated by comparing the projections from 1980 to 2000 performed by Kelly and Williamson with purely demographic projections (which they criticize), treating migration as exogenous. The actual rise in levels of urbanization in developing countries over the period 1980-2000 was around 11 percentage points (United Nations, 2007). This was close to, though a little lower than, the demographic projections of Ledent (1982) (14 per cent rise) and the United Nations (13 per cent rise). The Kelley and Williamson Baseline projection (which they favor) for the 1980-2000 period rises by a much higher 19 percentage points, although alternative scenarios vary widely between only 1 percentage point (assuming OPEC price rises continue) and 26 percentage points (in the stable projection in which relative prices of fuel and natural resource intermediates, manufactures and non-fuel primary products are all held constant over the 1980-2000 period). This comparison suggests that demographic projections with exogenous migration rates have predicted developing country urbanization fairly well over this period, and remain a good basis for urbanization projections.

References

Bocquier, P. (2005). World Urbanization Prospects: an alternative to the UN model of projection compatible with the mobility transition theory. *Demographic research 12(9).*

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.

Energy Information Administration, US Department of Energy (2009). *World Energy Outlook 2009*. Washington, D.C.

Gelfand, A. and A. F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis* (2nd ed.). Boca Raton, Fl.: Chapman & Hall/CRC.

Gubler, D. (2002). Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. *Trends in Microbiology 10. No.2*.

Jones, G. (2004). Urbanization trends in asia: the conceptual and definitional challenges. In Tony Champion and Graeme Hugo (Ed.), *New Forms of Urbanization.*

Kelley, A. and J. Williamson (1984). *What Drives Third World City Growth?* Princeton: Princeton University Press.

Ledent, J. (1982). Rural-urban migration, urbanization, and economic development. *Economic Development and Cultural Change 30(3)*, 507–538.

Lindley, D. V. and A. F. M. Smith (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B 34*, 1–41.

Lunn, D., A. Thomas, N. Best, and D. Spiegelhalter (2000). WinBUGS - A Bayesian modeling framework: Concepts, structure and extensibility. *Statistics and Computing 10(4)*, 325–337.

Montgomery, M. (2008). The urban transformation of the developing world. *Science 319, 761.*

Research Council, N. (2003). *Cities transformed - demographic change and its implications in the developing world*. Washington D.C.: The National Academies Press.

United Nations, Department of Economic and Social Affairs, Population Division (1998). *World Urbanization Prospects. The 1996 Revision, Estimates and Projections of Urban and Rural Populations and of Urban Agglomeration.* United Nations publication, New York. Sales No. E.98.XIII.6.

United Nations, Department of Economic and Social Affairs, Population Division (2007). *World Urbanization Prospects. The 2007 Revision*. United Nations publication, New York.

United Nations, Department of Economic and Social Affairs, Population Division (2009, forthcoming). *World Population Prospects. The 2008 Revision*. CD-ROM Edition - Extended Dataset in Excel and ASCII formats, United Nations publication, http://www.un.org/esa/population/publications/wpp2008/.

United Nations, Department of Economic and Social Affairs, Population Division (2010, forthcoming). *World Urbanization Prospects. The 2009 Revision*. United Nations publication, New York.

World Bank (2009). *World Development Report 2009: Reshaping Economic Geography*. Washington, D.C.