Environ Resource Econ (2008) 39:199–221
DOI 10.1007/s10640-007-9105-8

Simulating the spatial distribution of population and emissions to 2100

Malcolm O. Asadoorian

Received: 9 January 2006 / Accepted: 3 March 2007 / Published online: 12 April 2007

Abstract Urbanization and economic development have important implications for many environmental processes including global climate change. Although there is evidence that urbanization depends endogenously on economic variables, long-term forecasts of the spatial distribution of population are often made exogenously and independent of economic conditions. It is common for research concerning long-run projections of global environmental change to use population density as the primary means to spatially distribute emissions projections. However, researchers typically utilize year 1990 cross-sectional population data to distribute their emissions projections for both the short- and long-term, without projecting any changes in population density. Thus, a beta distribution for individual countries/regions is estimated to describe the geographical distribution of population using a one-degree-by-one-degree latitude–longitude global population data set. Cross-sectional country/regional data are then used to estimate an empirical relationship between parameters of the beta distribution and macroeconomic variables as they vary among countries/regions. This conditional beta distribution allows the simulation of a changing distribution of population, including the growth of urban areas, driven by economic forecasts until the year 2100.

Keywords Computable general equilibrium model · Emissions distribution · Population distribution · Spatial econometric · Urbanization

1 Introduction

Urbanization and economic development have many important implications, particularly effects on global environmental change. In general, concentrated urban development leads to higher concentrations of air pollutants. Although there exists evidence that urbanization depends on economic variables, long-term forecasts of the spatial distribution of population

M. O. Asadoorian
Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology,
1 Amherst Street, Building E40-407, Cambridge, MA 02139-4307, USA
e-mail: malcolma@mit.edu

are often made exogenously and independent of economic growth assumptions (Henderson 2003). This paper seeks to fill this gap by developing a model of urbanization that is used to project the spatial distribution of population as driven by long-term economic forecasts.

Urban air pollution is now recognized to be a global problem due to the long-range transport of pollution. Moreover, urban air pollution and climate are closely connected due to shared generating processes (e.g., combustion) for emissions of the driving gases and aerosols. They are also connected because the atmospheric life-cycles of common air pollutants such as carbon monoxide (CO), nitrogen oxides (NO_x), and volatile organic carbons (VOCs), and of the climatically important methane gas (CH_4) and sulfate aerosols, all involve the fast photochemistry of the hydroxyl free radical (OH) (Prinn et al. 2005).

It is common for research concerning long-run projections of global environmental change to use population density as the primary means to spatially distribute emissions projections. For example, the Dutch National Institute for Public Health and the Environment's (RIVM's) Emissions Database for Global Atmospheric Research (EDGAR) utilize population density as the means to distribute emissions projections for non-point sources (Olivier et al. 2002). In addition, the MIT Integrated Global System Model's (IGSM's) coupled atmospheric chemistry and climate model includes an urban air pollution sub-model that utilizes emissions projections distributed by population density (Mayer et al. 2000). However, given that an adequate time-series of data is not readily available, these groups and others typically utilize cross-sectional population data (e.g., year 1990) to distribute emissions projections for both the short- and long-term, without projecting any changes in population density. Ideally, one would prefer to acquire spatial data on economic determinants of urban growth and urbanization in order to distribute emissions projections at relatively low-spatial scales. However, relatively speaking, because data on population is more readily available, it is frequently used as a proxy for all economic activity.

Modeling regional climate change, including the effects of aerosols and other relatively short-lived substances, is a next major step in climate change research (IPCC 2001). For these projections, the spatial distribution of emissions within countries is of great importance (IPCC 2000). The growing need for spatially explicit emissions forecasts that depend on where major population centers are located makes it critical to model the spatial distribution of population, and *dynamically* simulate it, driven by forecasts of economic development with long-term time horizons. Such a model can be used to predict the emergence of new urban areas and the growth in existing ones.

This paper develops a model to simulate a changing distribution of population, including the growth of urban areas and distribution of emissions projections. The model is constructed to be driven by long-term economic forecasts, specifically from MIT's Emissions Prediction and Policy Analysis (EPPA) Model, a computable general equilibrium (CGE) model of the world economy, and applied to examine possible future levels of NO_x in the absence of environmental policies to control them; the population model is not designed ad hoc so that it can be driven by other economic models besides EPPA. This model is intended to be used to distribute projected emissions of pollutants that are actually emitted and for which population is an appropriate and justifiable basis to distribute them. Relevant pollutants include: sulfur dioxide (SO_2), NO_X, CO, non-methane volatile organic compounds (NMVOCs), black carbon (BC), and organic carbon (OC).

The main sections of the article proceed as follows: Background Literature and Issues, Empirical Models and Data, Empirical Results and Analysis, and Conclusions.

2 Background literature and issues

One of the enduring observations of urbanization is that of Zipf (1949), which has come to be known as Zipf's law. Essentially, it applies to the distribution of cities by size. The approach is empirical in nature and involves ranking all the cities in a country or region and then regressing the natural logarithm of the rank on the natural logarithm of population. The basic observation of Zipf's law is: "when we draw log-rank against log-size, we get a straight line, with a slope, which we shall call ζ, that is very close to 1. In terms of the distribution, this means that the probability that the size of a city is greater than some S is proportional to $1/S$: $\mathrm{P(Size} > \mathrm{S)} = \alpha/\mathrm{S}^{\zeta}$, with $\zeta \cong 1$" (Gabaix 1999, p. 740). Gabaix (1999) provides a comprehensive review of this literature and demonstrates that Zipf's law for the distribution of city sizes is robust, indicating reasons for this universal relationship.

Recent theoretical literature has been largely characterized by micro-theoretic geographical economic models (Krugman 1991; Krugman and Venables 1995; Fujita et al. 1999; Brakman et al. 2001; Forslid et al. 2003), generally known as "core-periphery models" that examine, under what conditions in a two-region country, industrialization (i.e., urbanization) is spread across, or concentrated in, a single region. However, as Henderson (2004) points out, "…core-periphery models have limited implications for urbanization per se, since in many versions including Krugman's (1991) initial paper, the agricultural population is fixed…Urban models are focused on the city formation process, where the urban sector is composed of numerous cities, endogenous in number and size" (pp. 33–34). More recent work has indeed focused on such urban models, identifying underlying economic variables that explain growth of particular urban areas. For example, Henderson and Wang (2004) develop an endogenous growth model and then empirically estimate the determinants of growth in a number of cities by economic factors such as: urban and rural wages, costs of commuting, levels of technology, levels of education, and urban and rural employment rates.

For the task here, three main issues arise when developing a model to simulate future urban development for the entire world over a relatively long period. First, the definition of an "urban area" (e.g., city) is not uniform across countries. For example, the United States Census Bureau defines an urban area as a population settlement that has a population density of at least 1,000 people per square mile with a population size of at least 2,500 people (US Census Bureau 2000). In contrast, Albania defines an urban area as "towns and other industrial centres with more than 400 inhabitants"; Chile defines an urban area as "Populated centers with definite urban characteristics, such as certain public and municipal services" (United Nations 2001b). Because of this variability, a list of "urban areas" as self-defined by custom in different countries leads to tremendous inconsistency among them.

Second, a fixed list of urban areas throughout the world obtained from, say, national databases would provide no basis for adding to the list as populations grow and economies change over time. Using pre-specified urban areas treats them as exogenous, when it is more informative to allow urban areas and urbanization to be modeled as endogenously determined by changes in economic and demographic variables.

Third, in order to estimate a model and use it to generate forecasts, it is necessary to identify economic determinants of urban growth and urbanization for which historical data is readily available and which can be forecast (Henderson 2003; Henderson and Wang 2004). An estimated model that appears to explain urbanization extremely well is of little use for forecasting purposes if one has no forecasts of the future evolution of the explanatory variables.

Table 1 EPPA Version 4.0—regional aggregation

Regions/countries	Description
USA	United States of America
CAN	Canada
MEX	Mexico
JPN	Japan
ANZ	Australia and New Zealand
EUR	European Union (EU) and European Fair Trade Association (EFTA) (Iceland, Liechtenstein, Norway, Switzerland)
EET	Eastern Europe (Czech Republic, Slovakia, Poland, Hungary, Romania, Bulgaria, Slovenia)
FSU	Former Soviet Union Countries
ASI	South Korea, Malaysia, Philippines, Singapore, Taiwan, Thailand
CHN	China and Hong Kong
IND	India
IDZ	Indonesia
AFR	Africa
MES	Middle East (excluding FSU countries and Turkey)
LAM	Latin America (does not include Mexico)
ROW	Rest of the World (remainder of Asia and Turkey)

Ordering is not intended to reflect any ranking of regions

3 Empirical models and data

3.1 Lorenz curve approach

In developing an empirical model of urbanization and urban growth, the three aforementioned issues must be addressed: the variability of the definition of urban areas, the exogenous treatment of urban areas, and the "appropriate" economic variables that determine urban growth and urbanization. The approach outlined in this section does not define urban areas a priori, but instead models the spatial distribution of population. In doing so, population density can then be used consistently across countries/regions to define "urban areas" and allow them to be endogenously, rather than exogenously, determined.

A possible approach to the first two issues indicated above is to estimate a Lorenz curve for the spatial distribution of population. The Lorenz curve is commonly used to represent and analyze the size distribution of income and wealth; the curve relates the cumulative proportion of income units to the cumulative proportion of income received when the units are arranged in ascending order of their income (Kakwani and Podder 1976).

Henderson and Wang (2004) suggest a similar ordering of area units of land from the least dense to the most dense to describe the spatial distribution of population. Just as the Lorenz curve is used to describe income inequality, a Lorenz curve for population distribution can describe how the population of a country/region is more or less equally distributed across the total land area.

This procedure represents the first step in my study. I order the area units by density to construct Lorenz curves and compute corresponding Gini coefficients for the 16 regions in MIT's EPPA Model using a 1990 $1^\circ \times 1^\circ$ latitude–longitude spatial population data set from the United Nations Environment Programme.[1] See Table 1 for a description of the EPPA regional aggregation and Fig. 1a, b for Lorenz Curves and Gini Coefficients. Since this is a

[1] Each grid cell in a $1^\circ \times 1^\circ$ space is an average area equal to $100\,\text{km}^2$, equivalent to $38.61\,\text{mi}^2$.

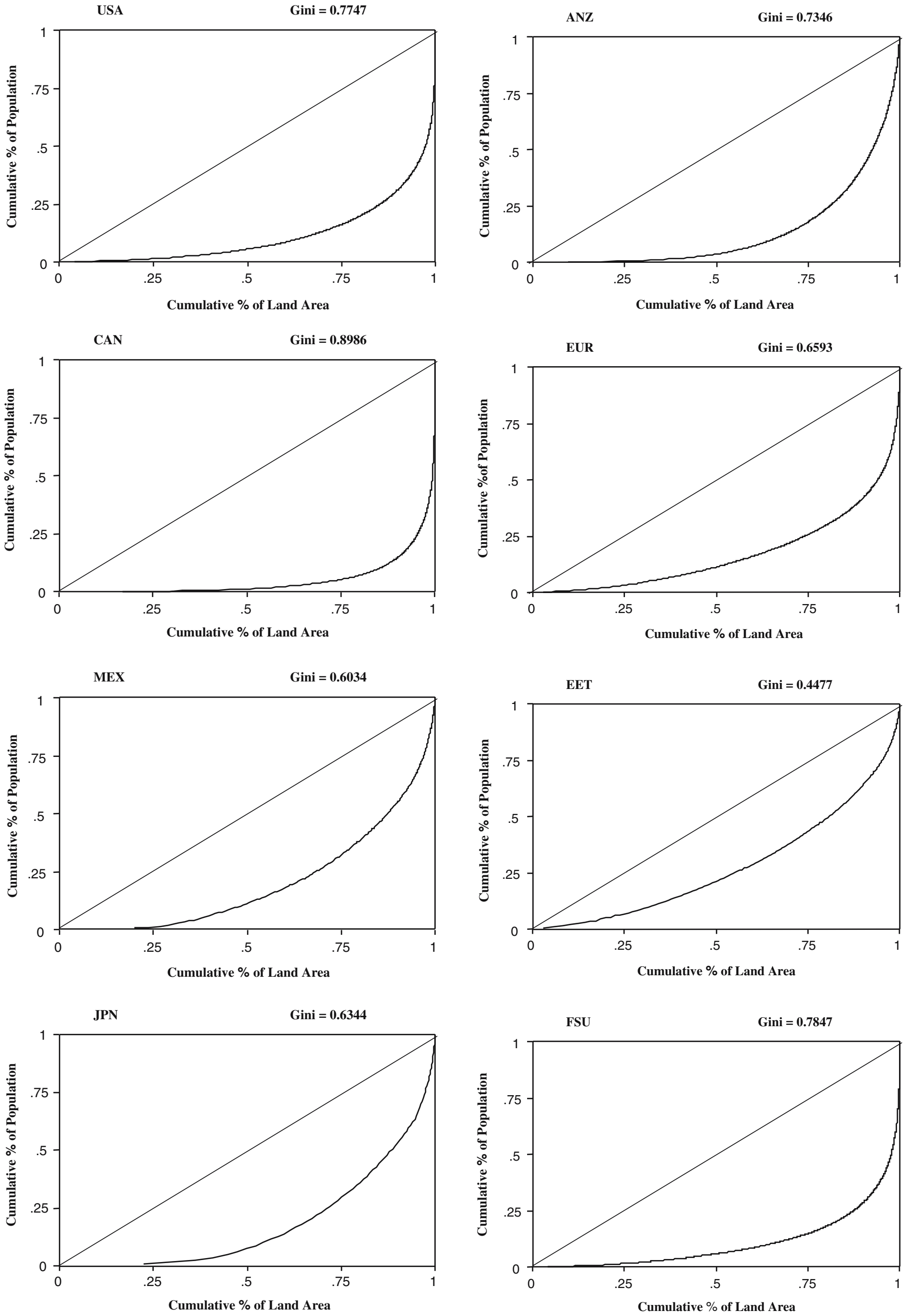

Fig. 1 **a** 1990 Lorenz curves and Gini coefficients for EPPA regions. **b** 1990 Lorenz curves and Gini coefficients for EPPA regions

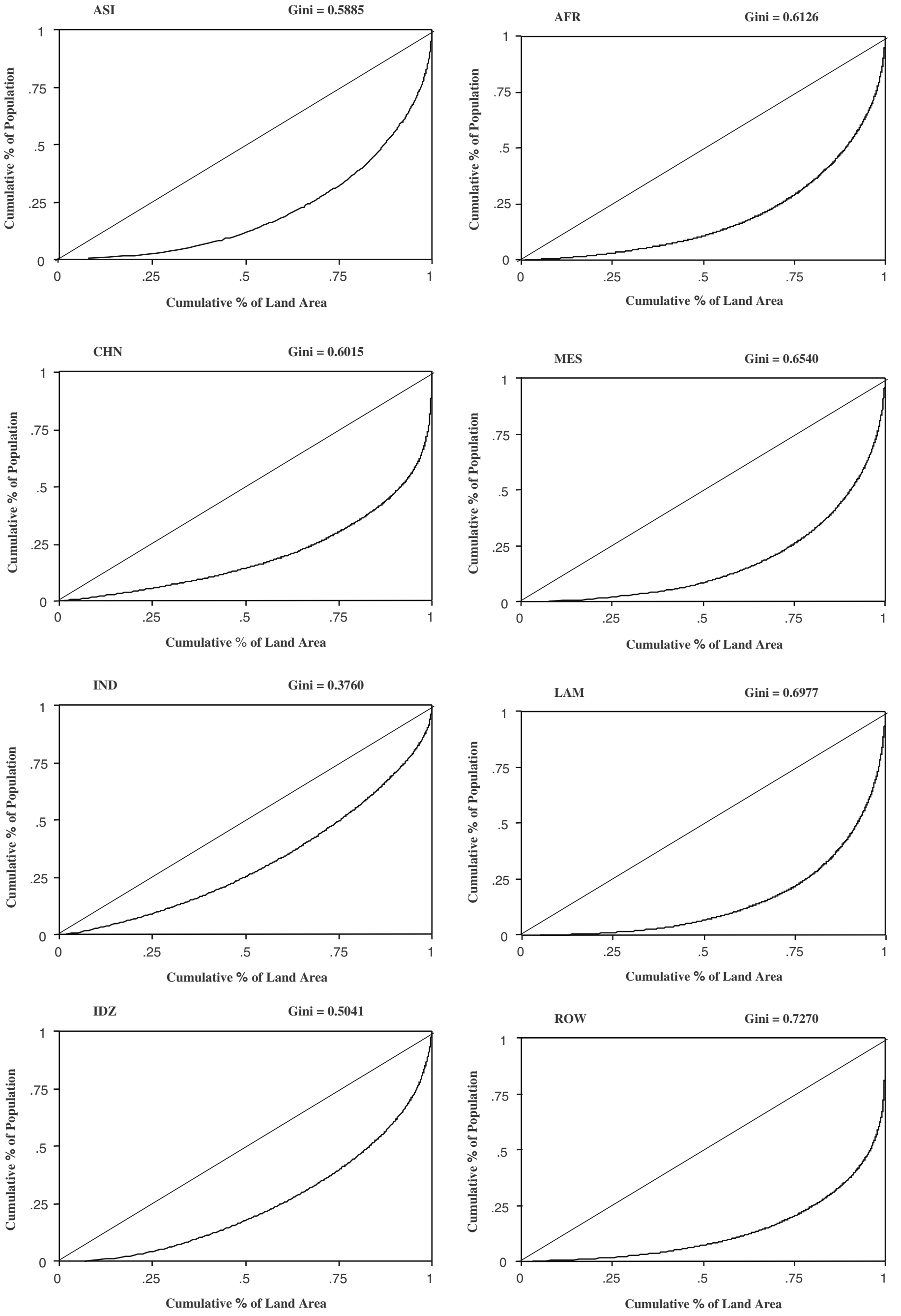

Fig. 1 continued

simple ordering, it is a non-parametric approach to describing the Lorenz curve for population and requires no a priori assumption about the functional form of the relationship.

Recall, that $0 \leq \text{Gini} \leq 1$, where a value of 0 indicates perfect equality and a value of 1 indicates perfect inequality. In the spatial context here, as a country's Gini value approaches one, greater inequality implies a greater degree of concentration of the population in a relatively small land area within a region (and vice-versa). From the diagrams in Fig. 1b, the most fundamental observation that can be made is that regions with greater land area, as indicated by the number of grid cells that compose the region, have higher Gini coefficients; for example, regions as USA, Canada, and the Former Soviet Union have greater land area (i.e., more grid cells) than Japan, Eastern Europe, and Indonesia. A more formal Spearman (1904) rank correlation test between the Gini coefficients for each region and the number of grid cells indicates a positive correlation significant at the one percent level. This raises the question, "why does this significant positive correlation exist?" It is not simply land area indicated by the number of grid cells but, most importantly, the arable proportion of the total land area; it is arable land that captures the notion of possible "spatial spread" within a country/region. Deserts and tundra are not areas for potential habitats, except in the USA where, for example, air conditioning makes desert areas habitable. Therefore, arable land is used as an index of habitable land.

Using 1990 data from the World Bank (2004), the two regions with the lowest Gini coefficients, Eastern Europe and India, respectively, have the largest percentages of arable land of total land area; in contrast, the region with the highest Gini coefficient, namely Canada, has the second-lowest percentage of arable land (second only to the Middle East). Granted, this is correlation and not causality. However, it indicates that the Lorenz curve only serves to describe the degree of inequality of population, but does not directly shed light on the economic determinants of this inequality. Thus, it is critical to estimate a functional distribution of population using a flexible functional form for each region and control for economic determinants in the process (e.g., arable land, measure of economic growth).

It is also possible to describe the Lorenz curve as a formal functional relationship, requiring the selection of some underlying distribution function. Some common distributions include: equal, exponential, shifted exponential, general uniform, and Pareto (1897) distributions (Gastwirth 1972). For example, Majumder and Chakravarty (1990) model the probability distribution of income, comparing the empirical performances of various distributions for USA income data, including the Pareto (1897), Lognormal, Gamma, Singh and Maddala (1976), Dagum (1977), and McDonald's (1984) Generalized Beta distribution. More recent work expands the realm of distribution functions with a distinct focus on more flexible forms, specifically the beta distribution, as described in Ortega et al. (1991) and Boccanfuso et al. (2003). For my purposes, the advantage of moving from a non-parametric description of the Lorenz curve to using an explicit functional relationship is that it allows me to then estimate the parameters as a function of economic variables. In other words, the distribution is then "conditional" on the value of economic variables, shifting the distribution over time as economic variables change.

3.2 Beta distribution approach

I adopt the approach of Nelson and Preckel (1989), utilizing the conditional beta distribution to model the probability distribution of population density individually for each of the 16 EPPA regions. Nelson and Preckel (1989) demonstrated the broad application of the conditional beta distribution by using it to model the probability distribution of agricultural output, estimating a stochastic production function and allowing the shape-parameters of the distribution of output to be functions of economic variables (i.e., conditioned on economic variables).

As Nelson and Preckel (1989) point out, "…distributions can be significantly skewed either to the right or to the left. The beta distribution has such flexibility. In addition, the beta distribution is well known and mathematically tractable. All of the moments of the distribution exist and are simple functions that are ratios of polynomials in the parameters of the distribution" (p. 371). Given a beta distributed random variable, $X \sim \text{Beta}(\alpha, \beta)$ with $0 \leq X \leq 1$, $\alpha > 0$, and $\beta > 0$, the probability density function of an (unconditional) beta random variable can be expressed as:

$$f(X) = \frac{X^{\alpha-1}(1-X)^{\beta-1}}{\int_0^1 X^{\alpha-1}(1-X)^{\beta-1}\mathrm{d}X}$$

The distribution can be conditioned on a vector of determinants, $\mathbf{Z}$, by expressing the shape-parameters α and β as functions of $\mathbf{Z}$, namely $\alpha(\mathbf{Z})$ and $\beta(\mathbf{Z})$.

The third issue indicated previously is concerned with identifying the "appropriate" economic determinants of urban growth and urbanization to condition our beta shape-parameters. Recall, determinants found to be important by Henderson and Wang (2004) include such economic factors as: urban and rural wages, costs of commuting, levels of technology, levels of education, and urban and rural employment rates. However, it is relatively difficult to generate long-term projections for these variables. Moreover, they are not variables typically predicted within a global CGE framework. Yet, over the long-term and across the wide range of economic conditions among countries, one can reasonably expect variation in these shape-parameters to be functions of broad economic measures such as gross domestic product (GDP) per capita and national population per unit of arable land area, most of which are predictions of the long-term EPPA Model.

3.3 Data sources

The 1990 $1° \times 1°$ latitude–longitude spatial population data set from the United Nations Environment Programme is utilized to construct a distribution of population density for the 16 EPPA regions in the world. With the goal of using long-term economic forecasts from EPPA to make projections of urbanization and the spatial pattern of pollutant emissions, the urbanization model is therefore estimated for these specific EPPA regional groups. The EPPA Version 4.0 divides the world into its 16 economic regions each with a number of economic sectors and input factors and produces projections, including emissions of both greenhouse gases and major criteria air pollutants. Most importantly for this study, GDP is endogenously projected by EPPA and depends on population, labor force, and labor productivity growth as they interact with other endogenous variables including prices and resource availabilities (Paltsev et al. 2005). For base-line model data, year 1997 data from Purdue University's Global Trade, Assistance, and Production (GTAP) Version 5 database is the primary source used by EPPA (Dimaranan and McDougall 2002). In addition, it utilizes the United Nations (2001a) national population projections. These EPPA model projections of GDP and national population per unit of arable land are utilized as the primary economic determinants of urban growth and urbanization. Longer-term environmental change may likely have implications for the amount and distribution of arable land. Although the simulations presented here treat arable land as exogenous, the model offers the flexibility to treat changes in arable land as variable given projections of such changes.

3.4 Conditional beta distribution model & data

Maximum likelihood estimation is first employed to fit the distribution of population density to the two-parameter (unconditional) beta distribution for each of the EPPA regions individually. Maximum likelihood estimation of the beta distribution produces consistent, asymptotically normal and efficient estimates of α and β, subject to the condition that both parameters be greater than one in order for the beta distribution to be unimodal; this condition was fulfilled for all EPPA regions.

Empirical implementation of the conditional beta model requires that a functional forms for $\alpha(\mathbf{Z})$ and $\beta(\mathbf{Z})$ be selected. As Nelson and Preckel (1989) point out, the functions $\alpha(\mathbf{Z})$ and $\beta(\mathbf{Z})$ must be consistent with the regularity conditions for maximum likelihood estimation. In addition, "…arguments for simplicity and parsimony might justify linear or log-linear functions" (p. 372). After some experimentation, log-linear specifications were selected for both the α and β functions. Thus, the functions $\alpha(\mathbf{Z})$ and $\beta(\mathbf{Z})$ can be expressed generally as:

$$\mathrm{LN}(\alpha)_i = \gamma \mathrm{LN}(\mathbf{Z})_i + \varepsilon_\alpha, \tag{1}$$

$$\mathrm{LN}(\beta)_i = \xi \mathrm{LN}(\mathbf{Z})_i + \varepsilon_\beta, \tag{2}$$

where i indicates the ith EPPA region, γ and ξ are vectors of parameters, and ε_α and ε_β are stochastic error terms with the usual properties.

Year 1997 data for $\mathbf{Z}$ variables are obtained from the most recent edition of the World Bank's (2004) World Development Indicators. Formally, I estimate the following equations:

$$\mathrm{LN}(\alpha)_i = \gamma_0 + \gamma_1 \mathrm{LN}(\mathrm{GDPPC})_i + \gamma_2 \mathrm{LN}(\mathrm{POPDEN})_i + \varepsilon_\alpha, \tag{3}$$

$$\mathrm{LN}(\beta)_i = \xi_0 + \xi_1 \mathrm{LN}(\mathrm{GDPPC})_i + \xi_2 \mathrm{LN}(\mathrm{POPDEN})_i + \varepsilon_\beta, \tag{4}$$

where GDPPC is the real GDP per capita measured in Purchasing Power Parity (PPP); POPDEN is the national population per unit of arable land.

As estimated, the α and β parameters affect the normalized shape of the beta distribution. In addition, I also make the maximum density grid cell a function of both GDPPC and POPDEN, econometrically estimate the model, and test the significance on the maximum population density for each region (i.e., the grid cell with the highest population); this shifts the distribution in absolute terms. Specifically, the following additional equation is empirically estimated:

$$\mathrm{LN}(\mathrm{MAX})_i = \xi_0 + \xi_1 \mathrm{LN}(\mathrm{GDPPC})_i + \xi_2 \mathrm{LN}(\mathrm{POPDEN})_i + \varepsilon_\beta, \tag{5}$$

where MAX is the maximum population density for each EPPA region.

The correct exchange rate to use in the context of long-term projections has been an issue of some recent contention. It is generally recognized that market exchange rates (MER) provide an unreliable basis for making cross-country comparisons of income. Thus, it is necessary to estimate relationships such as those above using PPP conversion factors (e.g., McKibbin et al. 2004). Note that the EPPA model is solved in MER because it must deal with international trade, and trade occurs at MER. For purposes of forecasting urbanization using my estimated model, I apply fixed PPP conversion factors for 1997 (the EPPA base year) to convert EPPA projections of MER-based GDP to PPP.

Table 2 Estimation results

Equation Number:	(3)	(4)	(5)
Number of Obs:	16	16	16
Dep. Var:	LN(α)	LN(β)	LN(MAX)
Intercept	3.676 (4.73)*	14.980 (6.05)*	15.519 (16.32)*
LN(GDPPC)	−0.259 (3.12)*	−0.946 (3.37)*	0.010 (1.91)*
LN(POPDEN)	0.023 (1.94)*	−0.486 (2.27)*	0.438 (4.11)*
R^2:	0.430	0.377	0.473

Absolute t-statistics in parentheses
* = Statistically significant at the 10% level

4 Empirical results and analysis

4.1 Regression results

Table 2 reports the regression results for Eqs. 3–5 using White's (1980) robust variance estimator to correct for heteroscedasticity. In general, goodness-of-fit based on the R^2 measure is relatively good for all models given the cross-sectional nature and small sample size of the data. In terms of estimated coefficients, we find that all are consistently significant at the 10% level. Most importantly, however, it is the predicted impact of the independent variables on both shape-parameters, α and β, as well as the maximum population density, which will collectively determine the shape of the overall probability distribution.

4.2 Projected beta distributions

For present purposes, the output of EPPA solved without any emissions-reducing policy provisions (i.e., "business-as-usual"), is used to generate projections of GDP from year 2005 to 2100, in 5 year time-steps, for each of the 16 economic regions. These projections are then applied to the estimated Eqs. 3–5, respectively, to generate the predicted shape-parameters, maximum population density, and corresponding beta distributions for each EPPA region.

For ease-of-exposition, projected conditional beta distributions for all sixteen EPPA regions are generated at 20 year time-steps from years 2020 to 2100, as well as the estimated conditional beta distribution for base-year 1997.[2] A sample of three regions is presented in Figs. 2 through 4. These projected distributions are generated only to illustrate the impact of changing α and β shape-parameters and maximum population density on the general shape of the beta distribution. It is important to note that these projected distributions truncate the tails as they asymptotically approach the horizontal axis. Thus, the right end-point of the horizontal axis does not indicate the maximum population density value for each region.

From Figs. 2 to 4, the most fundamental observation that can be made is that, as GDP and national population increase over time, we project that the distribution is shifting rightward to more population-dense grids for each of the regions. Moreover, the shape of the distribution is becoming less right-skewed and more normalized, with a relatively larger spread. Specifically, for China in Fig. 3, we predict that ninety percent (90%) of the distribution consists of grids ranging between 0.07 and 0.3 million persons per 100 km^2 in year 1997, with the largest cities located to the right of this interval. In contrast, we predict that by year 2100, ninety percent (90%) of this distribution will consist of grids ranging between 0.4 and 3.7

[2] The amount of arable land is held constant for all calculations given the lack of availability of future projections for the time period until year 2100.

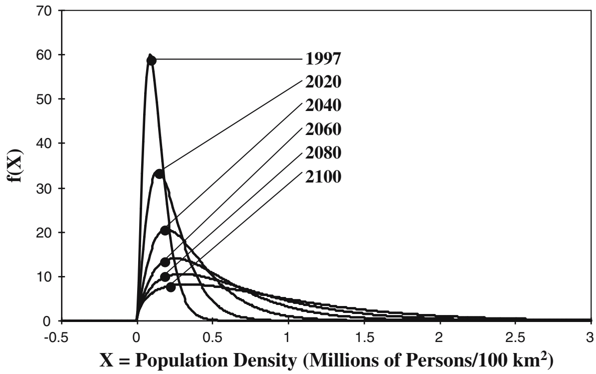

Fig. 2 United States (USA)—projected probability distribution of population density to year 2100

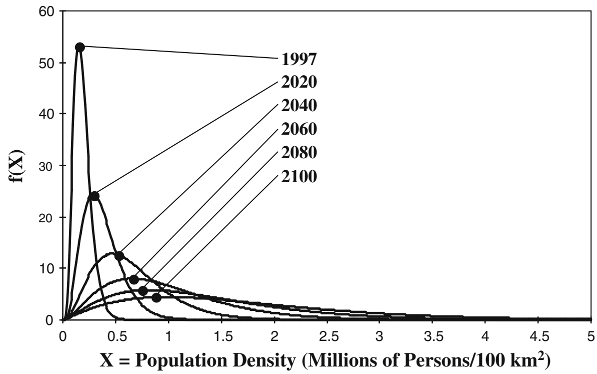

Fig. 3 China (CHN)—projected probability distribution of population density to year 2100

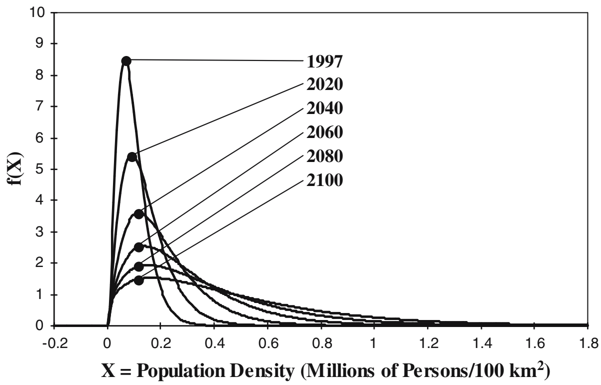

Fig. 4 Europe (EUR)—projected probability distribution of population density to year 2100

million persons per 100 km^2; for China, the grid cell with the maximum population density is predicted to increase from 16.4 million in 1997 to 17.5 million in 2100. Essentially, this implies that relative growth is expected to yield a change in the distribution of population from, what can be considered, more rural to urban areas with urban sprawl. That is, the area with the most-dense population becomes slightly more populous, but many areas become moderately to extremely densely populated, relatively speaking. Table 3 provides a summary of the projected conditional beta distributions for all 16 EPPA regions.

Table 3 Summary of projected beta distributions 1997 vs. 2100

	Density of the grid cells that contain 90% of the population		Maximum population density	Total number of grid cells
	Lower limit	Upper limit		
USA – 1997	33,535	281,724	7,320,575	1028
USA – 2100	110,091	2,066,114	8,766,257	1028
CAN – 1997	13,431	107,840	5,072,912	685
CAN – 2100	44,660	878,157	5,902,739	685
MEX – 1997	60,237	252,102	10,799,286	227
MEX –2100	231,501	2,386,186	13,933,452	227
JPN – 1997	517,415	3,439,821	26,103,307	77
JPN – 2100	1,195,778	14,863,422	26,133,053	77
ANZ – 1997	10,724	74,867	4,271,523	632
ANZ – 2100	34,915	655,432	5,204,611	632
EUR – 1997	27,596	201,060	7,037,624	744
EUR – 2100	56,380	1,159,443	6,360,667	744
EET – 1997	36,204	98,605	7,227,918	199
EET – 2100	62,333	609,020	6,216,168	199
FSU – 1997	23,032	123,375	8,651,612	2451
FSU – 2100	70,376	949,400	8,030,357	2451
ASI – 1997	29,928	58,485	6,159,596	208
ASI – 2100	63,273	377,545	7,866,321	208
CHN – 1997	70,085	293,984	16,413,928	653
CHN – 2100	352,285	3,676,095	17,496,956	653
IND – 1997	161,968	261,312	12,971,459	329
IND – 2100	354,239	1,944,539	16,907,118	329
IDZ – 1997	118,980	374,572	17,228,131	292
IDZ – 2100	457,873	3,311,639	22,153,368	292
AFR – 1997	528,554	2,796,153	3,451,149	2233
AFR – 2100	653,573	5,232,690	5,128,681	2233
MES – 1997	34,331	136,184	9,494,958	490
MES – 2100	186,080	1,411,696	15,469,112	490
LAM – 1997	30,946	168,194	9,738,322	1716
LAM – 2100	161,411	1,661,996	12,849,604	1716
ROW – 1997	194,818	977,044	23,217,873	893
ROW – 2100	1,107,162	8,671,352	33,454,239	893

4.3 Robustness tests

Given these results, it is critical to isolate and explain which of the three factors of the beta distribution, namely α, β, or MAX, has the most significant impact on its shape. From the regression results in Table 2, it is evident that the estimated β Eq. 3 yields the largest estimated coefficients in terms of magnitude, particularly for LN(GDPPC). Although United Nations (2001a) national population projections approach a zero rate of growth for each region by 2100, EPPA projects consistent increases in GDPPC for each region. In short, it is the relatively large estimated coefficient for LN(GDPPC) combined with increasing GDPPC which are primarily driving the relatively large decreases in our predicted β's until year 2100 and, ultimately, driving the shift in the population distribution to more population-dense grids.

Because of the significant impact of the predicted β's on the shape of our beta distributions, I test the robustness of the estimated Eq. 3. Following Kennedy (2003), I first test for "influential observations" by eliminating a single observation (i.e., a single region) from the estimation, repeating this process for each of the 16 regions (p. 373). Next, I run a series of regressions with observation-specific dummies for each of the regions, testing whether each observation is an outlier (p. 379). Results of both tests support the robustness of the estimated Eq. 3 as reported in Table 2.

4.4 Projected spatial distribution of population

As noted earlier, I do not assume or impose a fixed definition of an urban area. Figures 5 and 6 represent global Geographical Information System (GIS) population density maps for both actual 1997 and projected changes of 1997 vs. 2100, respectively. A graduated scale is employed for each map in order to illustrate the fact that, depending on how one defines an urban area, one may draw different conclusions as to the rate of urban growth and urbanization. In other words, the threshold or "cut-off" imposed on the graduated scale will affect the projected number of urban areas.

To evaluate the performance of the model, a "back-casting" (i.e., "project backwards" to compare projections against actual data) exercise was performed. The principal challenge in such an exercise is the availability of data. Time-series gridded data for the spatial population distribution is generally not available, as only in recent years has this geographic representation of data become of interest.[3]

As noted previously, for the future projections, I rely on reference GDP projections of the EPPA model (Paltsev et al. 2005) as well as United Nations (2001a) population projections. Table 4 presents the United Nations (2001a) 21st century national population projections. As seen from Table 4, the United Nations (2001a) forecasts zero national population growth by year 2100. Recall, since GDP in EPPA is projected at MER in real terms, I converted GDP per capita in MER to PPP terms using the 1997 (the EPPA base year) MER-PPP conversion factors. Table 5 presents the MER-PPP conversion factors, year 2000 GDP, and the 21st century reference GDP per capita growth rate projections from EPPA recently updated from Paltsev et al. (2005). For purposes of comparison, Table 5 also includes historical GDP per capita growth rates constructed for each of 16 EPPA regions using data from Maddison (2001). In general, Paltsev et al. (2005) conclude that "the GDP growth rates projected for different regions are within the range of historical growth rates as computed by Maddison (2001)…" (p. 50).

Based on the Maddison (2001) data, many regions realized relatively rapid growth in 1950–1973 period, and slower growth in 1973–2001 period. The main exceptions are China, India, and Indonesia, where the experience is reversed with growth rates accelerating in 1973–2001 period relative to 1950–1973 period. The EPPA future projections are "near the middle" of the experience for these two periods over the next 25–50 years, with a general slowing of growth over time. Notably, the rates of growth in many developing country regions are not particularly more rapid than in developed countries, reflecting the historical relationship. Thus, this is not a scenario where developing countries catch-up to developed countries. However, consistent per capita economic growth implies that all regions are relatively better-off than they are today, at least as measured by per capita market income.

[3] Validation activities are ongoing as the quantity and quality of spatially explicit population data becomes increasingly available. To date, back-casting exercises have been undertaken for US and China using county-level data. Specifically, US population distribution was validated using US Census Bureau (2000) county-level data for each of the 50 states from years 1990 to 2004; China's population distribution was back-cast using 1997 county-level data for each of its 30 provinces (excluding Tibet) from the University of Michigan's China Data Center (2006). A goodness-of-fit calculation between actual and projected county-level population yielded a value of 85% (on average) for US and 72% for China, respectively.

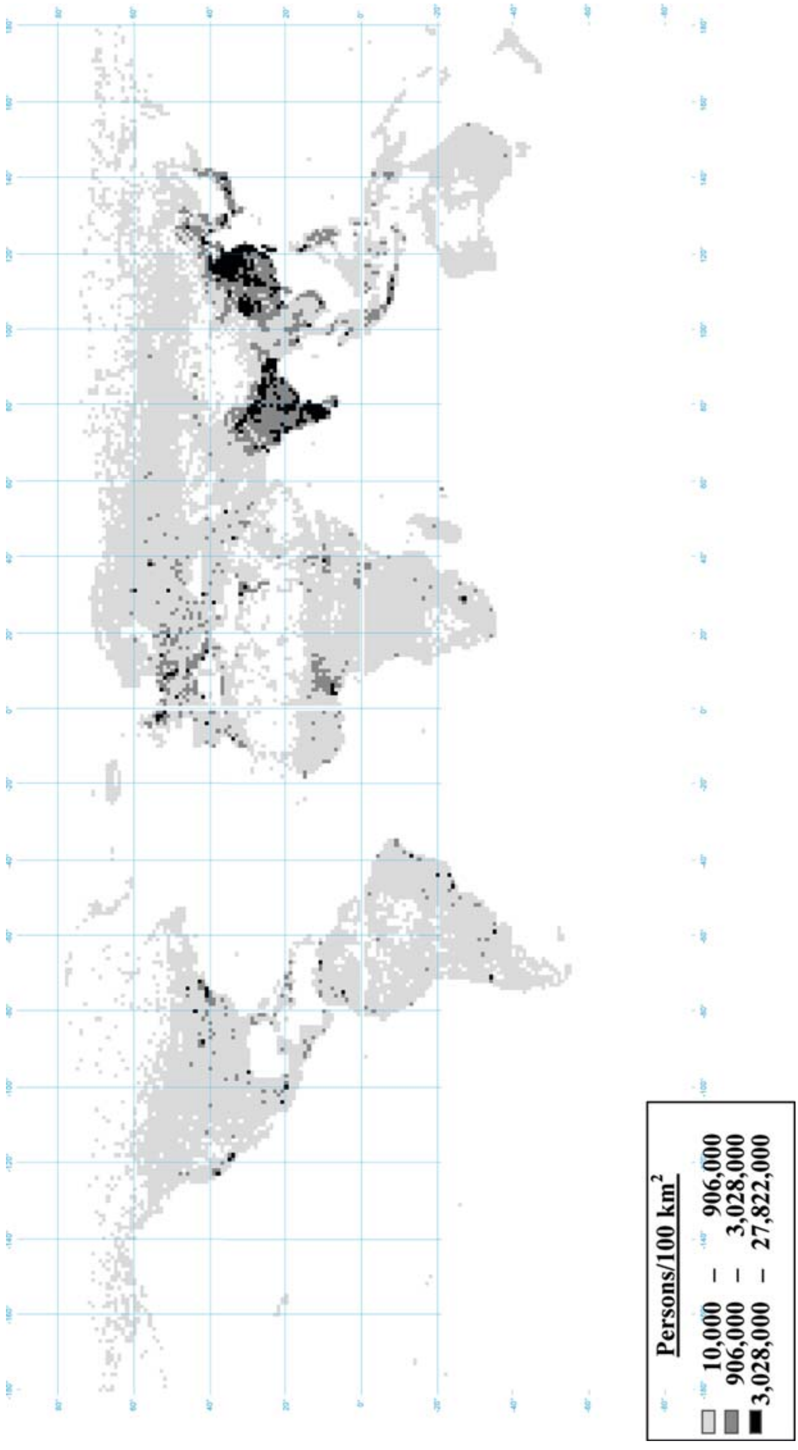

Fig. 5 1997 actual population density

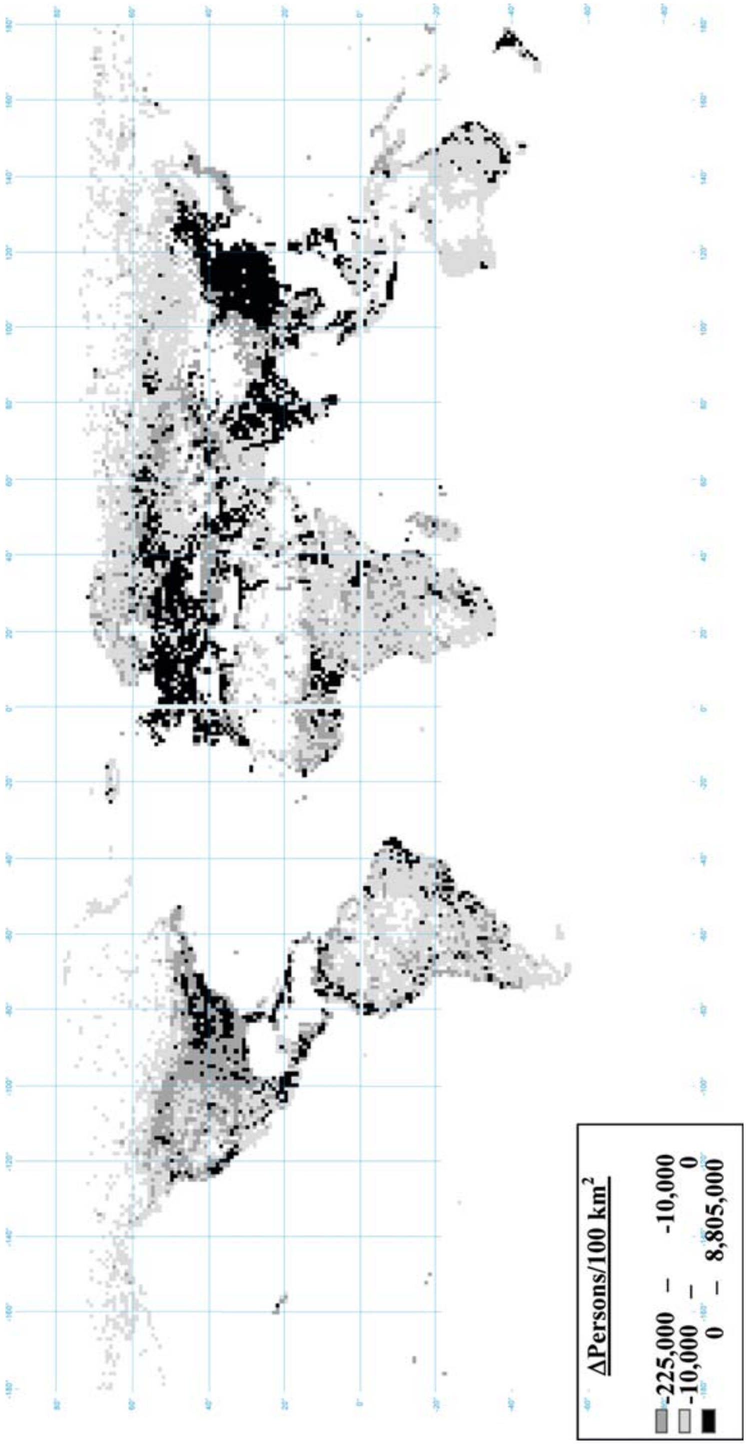

Fig. 6 Change (Δ) in population density—1997 vs. 2100

Table 4 United Nations population forecast (in millions)

Region	2000	2025	2050	2075	2100
USA	283.4	347.1	397.4	395.6	392.9
CAN	30.8	36.7	40.4	40.2	39.9
MEX	98.9	130.2	146.7	155.3	159.0
JPN	127.1	123.8	109.2	117.0	119.5
ANZ	22.9	27.8	30.9	32.4	32.9
EUR	389.6	384.8	352.0	307.5	288.5
EET	96.6	89.6	78.2	68.2	63.9
FSU	290.9	274.1	248.4	235.8	229.8
ASI	211.4	272.9	304.9	326.5	333.6
CHN	1,282.0	1,479.5	1,471.7	1,379.9	1,334.3
IND	1,008.9	1,351.8	1,572.1	1,633.5	1,643.3
IDZ	212.8	274.1	312.7	334.9	342.2
AFR	792.9	1,357.2	1,999.4	2,344.7	2,499.8
MES	174.1	298.3	430.3	460.8	470.8
LAM	419.3	563.7	657.8	696.5	713.4
ROW	615.1	925.1	1,170.1	1,247.9	1,273.1
World	6,056.7	7,936.7	9,322.2	9,776.7	9,936.9

From Fig. 6 we predict that, for a relatively large number of grid cells, population density will decrease by the year 2100. This result is primarily driven by the fact that United Nations national population projections from 1997 to 2100 exhibit slowing rates of growth, with the United States and Canada being two of the major regions achieving zero growth by 2100 (United Nations 2001a). The first interval indicates grids with the largest predicted decreases. Although relatively difficult to see on the map, the cell with the largest predicted decrease of 225,000 persons is Lucknow, the modern capital of Uttar Pradesh, representing the most populous state in India. However, this predicted decrease is accompanied by predicted increases as large as 195,000 persons in grid areas surrounding Lucknow such as the southern state of Madhya Pradesh.

The second largest predicted decrease of 134,000 persons is located in the largely rural Nara and Mie prefectures of Japan. Most interestingly, the immediately neighboring grid to the west of Nara and Mie is the Osaka/Kyoto/Kobe portions of the Kansai region, which is predicted to have the largest global increase in population of 8,805,000 persons from 1997 to 2100. Moreover, the immediately neighboring grid to the north of Nara and Mie is the Chubu region, which includes Nagoya with a predicted increase of 387,000 persons by 2100; Osaka and Nagoya represent the second and third largest cities in Japan, respectively. It is also important to note that the second and fifth largest predicted increases in population globally are in the largest city of Tokyo and Fukuoka located in the Kita/North-Kushu regions, respectively. Although United Nations (2001a) national population projections indicate that, by the year 2100, Japan will achieve an approximately zero national population growth rate, this analysis predicts that there will be a relatively large change in the distribution of its population.

In addition to Japan, the region of Mexico illustrates a similar pattern. More specifically, the model predicts decreases in population for Monterrey, the third largest city of Mexico, as well as for the relatively mountainous area of Sierra Madre Occidental in Durango city. This is accompanied by the third largest predicted global increase in population for Mexico City by year 2100. Like Japan, United Nations (2001a) national population projections indicate that by year 2100, Mexico will reach an approximately zero national population growth rate and, apparently, also experience a relatively large change in the distribution of its population.

Table 5 GDP and per capita growth, historical and projected

Region	Real purchasing power index and base GDP, 1987 billion US$		Average Annual Per Capita GDP Growth Rates (%)								
	MER-PPP Conversion Rates, 1997	2000 GDP, MER	Historical Maddison (2001)				EPPA projected updated from Paltsev et al. (2005)				
			1870–1913	1913–1950	1950–1973	1973–2001	2000–2025	2025–2050	2050–2075	2075–2100	2000–2100
USA	1.00	9,072	1.9	1.8	2.6	2.0	2.3	2.0	1.7	1.3	1.8
CAN	1.21	727	2.4	1.6	2.7	1.9	2.5	2.2	1.7	1.3	1.9
MEX	1.51	453	2.1	0.9	3.1	1.6	1.7	1.9	1.7	1.7	1.7
JPN	0.69	4,415	1.5	0.9	8.1	2.1	2.9	3.4	1.6	1.2	2.3
ANZ	1.27	524	1.3	1.1	2.3	1.5	2.7	2.4	1.7	1.3	2.0
EUR	1.16	9,176	1.3	0.8	4.1	1.9	2.8	2.9	2.4	1.6	2.4
EET	2.60	331	1.4	0.6	3.8	0.7	3.6	3.7	3.1	2.3	3.1
FSU	4.08	644	1.1	1.8	3.4	−1.0	4.3	3.6	2.4	1.9	3.0
ASI	1.94	1,261	0.7	0.1	3.5	2.4	2.4	2.6	2.1	1.8	2.2
CHN	4.46	1,232	0.1	−0.6	2.9	5.3	4.2	3.7	2.8	2.0	3.2
IND	5.37	487	0.5	−0.2	1.4	3.0	2.9	2.1	2.3	2.1	2.3
IDZ	3.99	222	0.7	1.3	2.4	3.2	2.3	2.6	2.1	1.9	2.2
AFR	3.10	610	0.6	0.9	2	0.2	1.4	0.7	1.6	1.7	1.4
MES	1.81	572	0.8	0.9	0.9	0.7	1.1	0.8	1.7	1.6	1.3
LAM	1.91	1,645	1.8	1.4	2.6	0.9	2.0	2.9	2.3	1.8	2.3
ROW	3.72	680	2.3	1.4	2	1.0	1.3	1.1	2.1	2.2	1.7
World	–	32,053	1.3	0.9	2.9	1.4	2.0	2.1	1.8	1.5	1.8

In Fig. 6, an interval with limits ranging from $-10,000$ to 0 is utilized to highlight the grids in which there is largely no change in population density. From the map, regions such as the Former Soviet Union, Latin America, Africa, as well as Australia and New Zealand have a relatively large proportion of grid cells in which there are no predicted changes in population, 1997 vs. 2100.

EPPA consistently projects increases in GDPPC for each region and the United Nations (2001a) projects of zero national population growth by year 2100 for most regions. Thus, the amount of arable land across regions, an index of habitable land, appears to largely account for the differences in predicted changes of the distribution across regions. More specifically, India possesses the largest percentage of arable land among the regions, 55%, and may be reason for the possible change in distribution outside of major existing cities into more rural areas. In contrast, Japan and Mexico have relatively small proportions of arable land, both approximately 13%, and may explain the predicted increases in population mostly for existing cities and predicted decreases in rural areas.

The United Nations (2001b) projects that, "Over the next 15 years, the number of mega-cities in the more developed regions will remain unchanged as will that in the least developed countries, but five additional mega-cities are expected to emerge in the less developed regions" (p. 75). Although these are relative short-term projections, in general, the predicted trend is the development of mostly urban agglomerations with less than ten million inhabitants (i.e., "small cities") as opposed to "mega-cities" that exceed this population (United Nations 2001b). From the results here, the amount of arable land in conjunction with relative growth in both income per capita and national population collectively predict a region-specific pattern of urban growth and urbanization.

Interestingly, the spatial pattern empirically projected here appears to be consistent with the theoretical conclusions of recent geographical economic models (Krugman 1991; Krugman and Venables 1995; Fujita et al. 1999; Brakman et al. 2001; Forslid and Ottaviano 2003). For example, Brakman et al. (2001) conduct a simulation exercise which illustrates that these core-periphery models, in general, predict increasing agglomeration followed subsequently by decreasing agglomeration as transaction and other interaction costs decline. The problem these models have yet to reconcile is the timing of the transition to the decreasing agglomeration phase. Although the model developed in this paper does not explicitly represent the theoretical variables driving this behavior, it is interesting that with the relative simple formulation based on aggregate economic, population, and land data, empirical projections consistent with these theoretical conclusions are obtained. The empirical projections provide some support that the 21st century may possibly be characterized by the second, decreasing agglomeration phase. In further work, providing data is available, it would be interesting to explore more closely the relationship between the theoretical literature and the empirical formulation of the model estimated here to better identify those forces tending toward increasing and decreasing agglomeration. As currently formulated, the tendency is derived from the relative value of conditioned terms of the beta distribution for multiple variables and is not traceable to a single variable.

4.5 Projected spatial distribution of emissions

In order to demonstrate the application of this urbanization model to distributing emissions projections, Fig. 7 illustrates EPPA year 2100 projections for NO_x generated from non-agricultural (i.e., urban) sources, solved without any emissions-reducing policy provisions (i.e., "business-as-usual"), as compared to the base-line actual 1997 data distributed at the regional level. From a regional standpoint, the largest percentage increases in the Middle East,

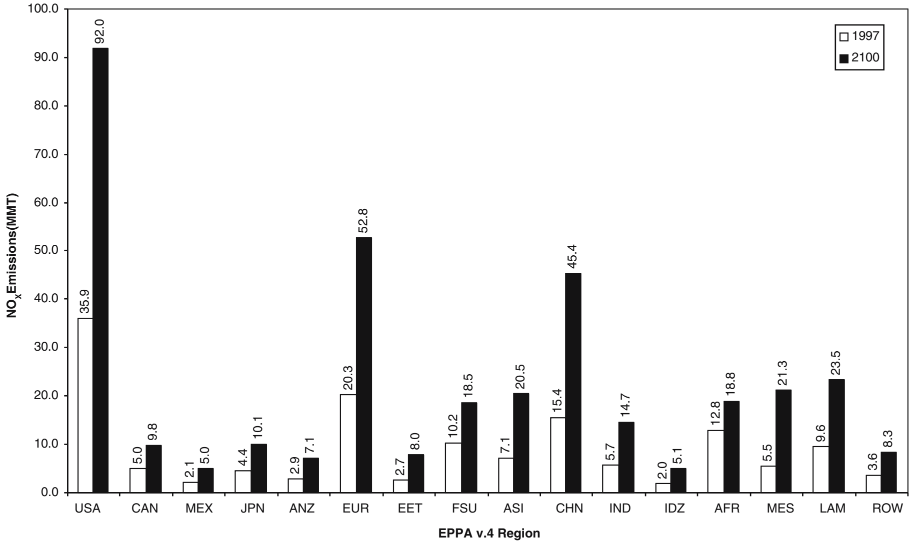

Fig. 7 EPPA regional—aggregate global distribution of NO_X emissions—1997 vs. 2100

Eastern Europe, and China; the smallest percentage decreases are in Canada, the Former Soviet Union, and Africa. Projected conditional beta distributions are utilized to calculate predicted probabilities for individual grid cells within each region. In this context, the predicted beta probability should be viewed as the *percentage* of NO_X emissions that are distributed to that grid cell. By following the common approach of utilizing a static cross-section of population density, it implies that the percentage of total emissions allocated to each grid cell is constant over time because of the fact that the population distribution is assumed constant. Therefore, even if increases in NO_X are projected to year 2100 for all regions as in Fig. 7, this means that the spatial pattern would remain unchanged with only increased projected emissions distributed to each grid. The greatest benefit of the model here is that the percentage of total emissions allocated to each grid cell *changes* over time according to changes in relative growth of income per capita and national population, as well as the (fixed) amount of arable land.

Figure 8 illustrates the actual 1997 base-line NO_X emissions, distributed within each region; Fig. 9 shows the predicted change in the spatial distribution of NO_X emissions for year 1997 versus year 2100. Although Fig. 7 indicates projected increases of NO_X for all EPPA regions as a whole, Fig. 9 clearly demonstrates decreases for some grid cells when emissions are distributed within each region; this is especially true for the Former Soviet Union, Eastern Europe, and India. In contrast, China, USA, and the European Union and European Fair Trade Association countries, generally show increases, though a relatively large degree of variability with respect to the magnitude of the change in emissions distributed across the entire spatial landscape.

5 Conclusions

In order to project the distribution of emissions for purposes of generating long-run projections of global environmental change, it is critical to dynamically model the spatial

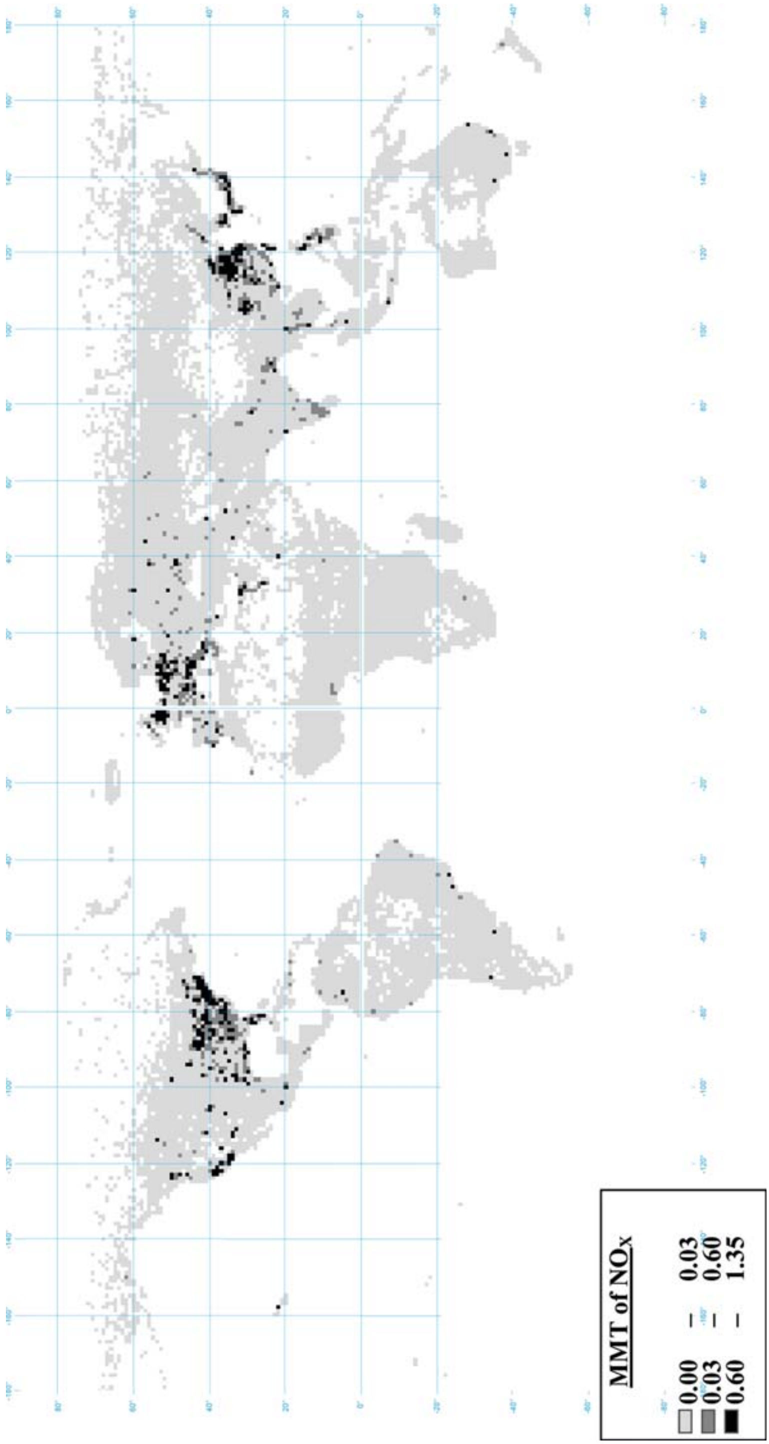

Fig. 8 1997 distribution of NO_X emissions

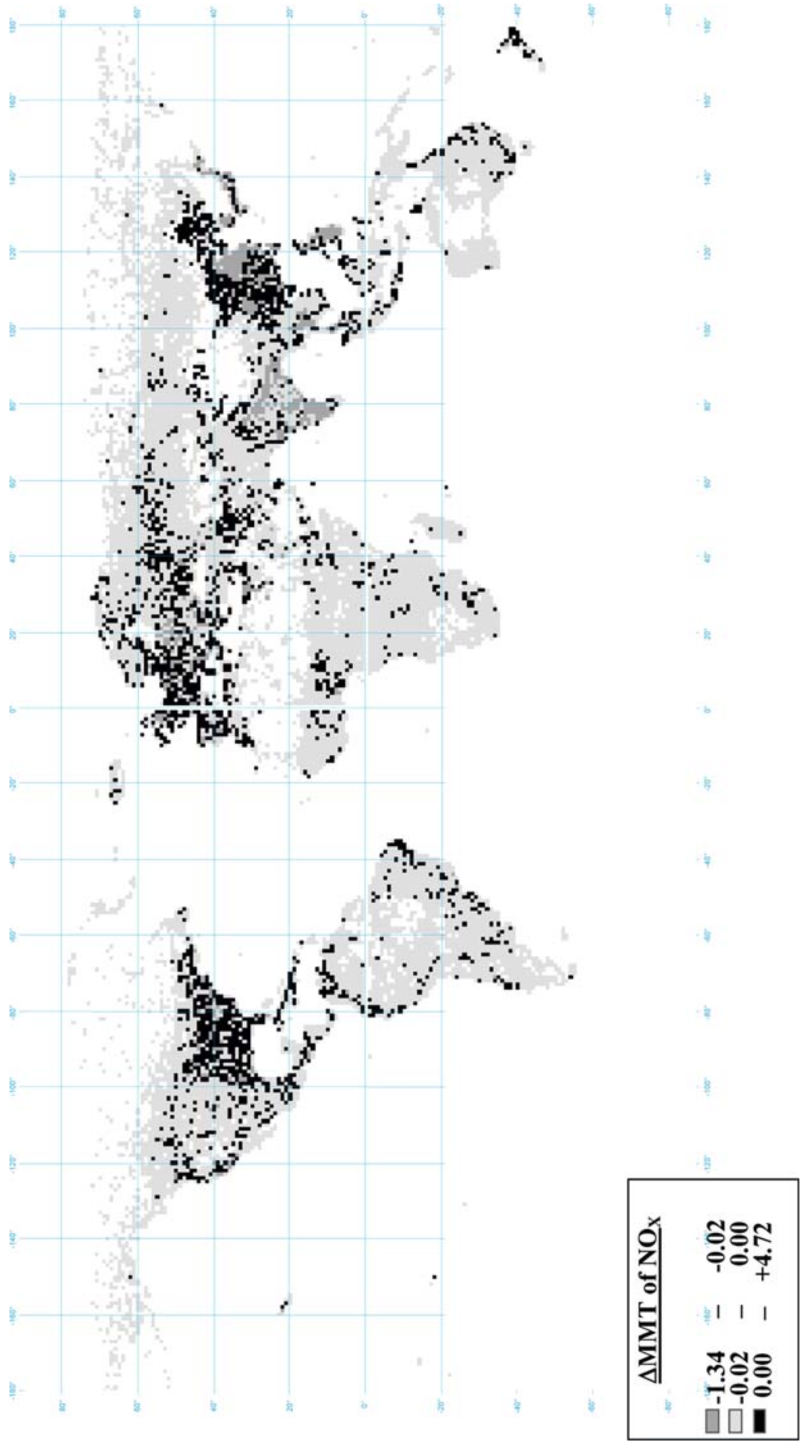

Fig. 9 Change (Δ) in distribution of NO_X emissions—1997 vs. 2100

distribution of population driven by forecasts of economic variables over the long-term. In this spirit, this paper developed an integrated approach that incorporates a CGE economic model and an estimated model of determinants of the spatial distribution of population.

Although the United Nations (2001b) predicts a general trend in the development of mostly urban agglomerations with less than 10 million inhabitants (i.e., "small cities") as opposed to "mega-cities" that exceed this population, the model in this paper demonstrates a more variable pattern of urban growth and urbanization across regions of the world. More specifically, the amount of arable land in conjunction with relative growth in both income per capita and national population collectively predict a region-specific pattern of urban growth and urbanization. Although the model developed in this paper does not have the same theoretical underpinnings as those in the relatively new economic geography literature, its predictions also provide some support that the 21st century may possibly be characterized by the second, decreasing agglomeration phase, a theoretical result that is typical of the core-periphery models.

Given that longer-term environmental change may likely have implications for the amount and distribution of arable land, future research priorities include linkage to a model that projects changes in biophysical conditions in order to generate projections on the amount and distribution of arable land to year 2100 along with EPPA projections. In addition, the application of this model to distribute emissions projections based on the projected distribution of population is essential. With a dynamically changing distribution of emissions, it is important to also examine the potential impact on resulting predictions of the concentrations of various pollutants utilizing the atmospheric chemistry component of the IGSM. Thus, the ability of this model to project urban growth and urbanization represents an important step to improve the spatial distribution of emissions projections in order to examine global environmental change.

Acknowledgments The author wishes to thank Richard S. Eckaus, Henry D. Jacoby, Sergey Paltsev, John E. Parsons, and John M. Reilly for their ongoing suggestions, support, and valuable input throughout this research. In addition, the constructive suggestions/comments provided by two anonymous Reviewers greatly improved this paper and is most appreciated. Lastly, the initial preliminary research and background for this paper is credited to Ms. Kira Matus.

References

Boccanfuso D, Decaluwe B, Savard L (2003) Poverty, income distribution and CGE modeling: does the functional form of distribution matter? (Working Paper 03–32) Centre interuniversitaire sur le risqué, les politiques economiques et l'emploi (CIRPE'E). Retrieved July 18, 2004, from http://www.wider.unu.edu/conference/conference-2003-2/conference%202003-2-papers/papers-pdf/Boccanfuso%20et%20al%20280403.pdf

Brakman S, Garretsen H, van Marrewijk C (2001) An introduction to geographical economics. Cambridge University Press, Cambridge, UK

China Data Center (2006) China county statistics. University of Michigan, Ann Arbor, MI. Retrieved on July 13, 2006 from, http://chinadataonline.org/

Dagum C (1977) A new model of personal income distribution: specification and estimation. Econ Appl 30:413–437

Dimaranan BV, McDougall RA (2002) Global trade, assistance, and production: the GTAP 5 data base. Center for global trade analysis. Purdue University, West Lafayette, IN

Forslid R, Gianmarco Ottaviano IP (2003) An analytically solvable core-periphery model. J Econ Geogr 3:229–240

Fujita M, Krugman P, Venables AJ (1999) The spatial economy. MIT, Cambridge, MA

Gabaix X (1999) Zipf's law for cities: an explanation. Q J Econ 113(3):739–767

Gastwirth JL (1972) The estimation of the Lorenz curve and Gini index. Rev Econ Stat 54(3):306–316

Henderson V (2003) Urbanization and economic development. Ann Econ Finance 4(2):275–341
Henderson V (2004) Urbanization and growth (Brown University Working Paper). Retrieved July 12, 2005, from http://www.econ.brown.edu/faculty/henderson/papers/Urbangrowhandbook261004.pdf
Henderson V, Wang HG (2004) Urbanization and city growth (Brown University Working Paper). Retrieved July 18, 2004, from http://www.econ.brown.edu/faculty/henderson/papers.html
IPCC (2000) Special report on emissions scenarios. In: Nakicenovic N et al. (eds) Cambridge University Press, New York
IPCC (2001) Climate change 2001: the scientific basis. In: Albritton DL et al. (eds) Cambridge University Press, New York
Kakwani NC, Podder N (1976) Efficient estimation of the Lorenz curve and associated inequality measures from grouped observations. Econometrica 44(1):137–148
Kennedy P (2003) A guide to econometrics, 5th edn. MIT Press, Cambridge, MA
Krugman PR (1991) Increasing returns and economic geography. J Polit Econ 99:483–499
Krugman PR, Venables AJ (1995) Globalization and the inequality of nations. Q J Econ 110(4):857–880
Maddison A (2001) The world economy: a millennial perspective. OECD, Paris
Majumder A, Chakravarty SR (1990) Distribution of personal income: development of a new model and its application to US income data. J Appl Econ 5:189–196
Mayer M, Wang C, Webster M, Prinn R (2000) Linking local air pollution to global chemistry and climate. J Geophys Res (Atmos) 105(D18):869–922
McDonald JB (1984) Some generalized functions for the size distribution of income. Econometrica 52:647–663
McKibbin WJ, Pearce D, Stegman A (2004) Long run projections for climate change scenarios (Brookings Discussion Papers in International Economics No. 160). The Brookings Institution, Washington, DC. Retrieved November 10, 2004, from http://www.brookings.edu/dybdocroot/views/papers/20040415_bdpie160.pdf
Nelson CH, Preckel PV (1989) The conditional beta distribution as a stochastic production function. Am J Agric Econ 71(2):370–378
Olivier J, Berdowski JM, Peters J, Bakker J, Visschedijk AJH, JanPieter JB (2002) Applications of EDGAR including a description of EDGAR 3.2: reference database with trend data 1970–1995 (Report No. 773301001) National Institute for Public Health and the Environment (RIVM), The Netherlands
Ortega P, Martin G, Fernandez A, Ladoux M, Garcia A (1991) A new functional form for estimating Lorenz curves. Rev Income Wealth 37(4):447–452
Paltsev S, Reilly JM, Jacoby HD, Eckaus RS, McFarland J, Sarofim M, Asadoorian M, Babiker M (2005) The MIT emissions prediction and policy analysis (EPPA) model: Version 4 (Report No. 125 August). MIT Joint Program on the Science and Policy of Global Change, Cambridge, MA. Retrieved September 8, 2005, from http://web.mit.edu/globalchange/www/MITJPSPGC_Rpt125.pdf
Pareto V (1987) Course d'economie politique, vol 2. F. Pichon, Paris
Prinn R, Reilly J, Sarofim M, Wang C, Felzer B (2005) Effects of air pollution control on climate (Report No. 118 January). MIT Joint Program on the Science and Policy of Global Change, Cambridge, MA. Retrieved February 12, 2005, from http://web.mit.edu/globalchange/www/MITJPSPGC_Rpt118.pdf
Singh SK, Maddala GS (1976) A function for the size distribution of incomes. Econometrica 23:963–970
Spearman C (1904) The proof and measurement of association between two things. Am J Psychol 15:72–101
United Nations (2001a) World population prospects: the 2000 revision. Department of Economic and Social Affairs—Population Division, New York, NY, Data in digital form
United Nations (2001b) World urbanization reports: the 2001 Revision. Department of Economic and Social Affairs—Population Division, New York, Retreived on July 12, 2004 from, http://www.un.org/esa/population/publications/wup2001/WUP2001report.htm
United Nations Environment Programme (1990) $1^\circ \times 1^\circ$ Global Population Density Map. Data in digital form
US Census Bureau (2000) United states census of population and housing 2000. US Government Printing Office, Washington, DC. Retrieved on July 20, 2004 from, http://www.census.gov/dmd/www/glossary.html
White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48:817–838
World Bank (2004) World evelopment indicators 2004. The World Bank Group, Washington, DC. Retrieved on June 30, 2004 from, http://devdata.worldbank.org/dataonline/
Zipf G (1949) Human behavior and the principle of last effort. Addison-Wesley Press, Cambridge, MA