Econometric Modeling of City Population Growth in Developing Countries

A Dissertation Presented

by

Donghwan Kim

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Economics

Stony Brook University

May 2011

Stony Brook University

The Graduate School

Donghwan Kim

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Mark R. Montgomery - Dissertation Advisor
Professor of Economics

Warren Sanderson - Chairperson of Defense
Professor of Economics

Alexis Anagostopoulos
Professor of Economics

Nancy Mendell
Professor of Applied Mathematics and Statistics

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Econometric Modeling of City Population Growth in Developing Countries

by

Donghwan Kim

Doctor of Philosophy

in

Economics

Stony Brook University

2011

Urbanization process is one of great issues in our time. Cities are now home to more than half of the world's population. Especially, the towns and cities of developing countries are places where almost all the world population growth is occurring and thus preparation for the growth is required to meet a range of development needs. In this increasing urban era, it is of importance to monitor urban populations and environments, and understand how urban populations are changing in both the spatial and time dimensions for development and urban planning policies.

This dissertation considers econometric modeling of city population growth, aiming at estimating and forecasting rates of city population growth and size for almost all of the developing countries. Both classical and Bayesian spatial econometric models of city growth are used, which can provide us information on uncertainty and take into account of economic, demographic, geographic, and environmental factors promoting and hindering city population growth.

The main contributions of this dissertation are two-fold: (1) it develops Bayesian MCMC estimation and forecasting methods of a panel data model with spatially correlated errors and (2) it lays the foundation of a spatially-explicit cities database by linking two existing cities datasets, that of the United Nations Population Division (a panel dataset of city populations) and CIESIN's GRUMP dataset, housed at the Columbia University's Earth Institute, which is in geospatial format. This geospatial cities database provides us with a better understanding of patterns of urbanization.

By incorporating into the database additional city-level indicators using GIS and geospatial programming, this study analyzes how current urban populations are distributed by ecological environments and how the patterns evolve over time. This analysis documents urban settlements and their population sizes in the dryland ecozone and low-elevation coastal zone, and quantifies vulnerable populations to related climate-related hazards (e.g. storm surges, droughts). Also, the model estimation implies that growth of a city is affected not only by the city's characteristics but also by those of its neighboring cities.

As a baseline model, this study first develops fertility-based econometric models of city growth for developing countries. It finds that urban fertility rate has a significant positive impact on developing-country city growth rate, so re-confirms the important role of fertility in city growth of developing countries. The median future city growth rate is projected to decline as fertility rates continue their historical trend downward.

In summary, this dissertation deals with the population dimension of urbanization process, aiming at developing international-level estimation and forecasting methods of city growth, especially in developing countries. There are unresolved issues which should be addressed in the future. Also, simultaneous approach considering multi-dimensions of urbanization process should be studied for systematic analysis of complex urbanization process.

Table of Contents

List of Tables **viii**

List of Figures **xi**

1 Introduction **1**

2 Methodological Development **5**

2.1 Panel data spatial econometric model 5

2.1.1 Overview 5

2.1.2 Model specification 7

2.2 Review of classical approach 10

2.2.1 Generalized Method of Moments estimation 10

2.2.2 Best Linear Unbiased Prediction 13

2.3 Bayesian approach 16

2.3.1 Bayesian hierarchical modeling 16

2.3.2 Markov chain Monte Carlo estimator 18

2.A Overview of Markov chain Monte Carlo estimation 22

3 Baseline Models **27**

3.1 The UN and GRUMP cities data 28

3.1.1 The UN tabular data 28

3.1.2 The GRUMP geospatial data 31

3.2 Basic modeling of city growth 32

3.2.1 City population growth rates 35

3.2.2 Model specification 36

3.3 City growth forecasts 37

3.A Linking the UN data with the GRUMP data 49

3.B Implementing the Bayesian models 50

4 Spatial Econometric Models 54
4.1 Urban populations by ecological zones 55
4.1.1 Ecological zones 55
4.1.2 Analysis 58
4.2 Extended modeling of city growth 65
4.2.1 Econometric specifications 66
4.2.2 Spatial linkages among cities 69
4.3 Results 71
4.3.1 Random and fixed-effect city growth results 71
4.3.2 Estimation with spatial linkages 75
4.4 City growth forecasts 78
4.A Supplementary regression results 83
4.B Data sources 94
4.C Geo-processing algorithms 96

5 Conclusion 98
5.1 Concluding Remarks 98
5.2 Suggestions for Future Studies 100

Bibliography 103

List of Tables

3.1 Classical and Bayesian city growth regression models with national total fertility rates and child mortality rates as covariates, All cities in developing countries, 2005-2007. 38

3.2 Growth rates and changes in definition, Relative to the baseline category with city defined in terms of urban agglomeration at start and end of spell, Classical random-effects Models. 41

3.3 Classical city growth regression models with observed and projected urban total fertility and child mortality rates as covariates. 47

4.1 Distribution of the Asian urban population and land area in the LECZ and drylands, by population size ranges. Population in thousands (000s) and land area in square kilometers. (Size and area in 2000, estimated using GRUMP methods.) 59

4.2 Percentages of the Asian urban population and land area in the LECZ and drylands, by population size ranges. Population in thousands (000s) and land area in square kilometers. (Size and area in 2000, estimated using GRUMP methods.) 60

4.3 Distribution and percentages of the African urban population and land area in the LECZ and drylands, by population size ranges. Population in thousands (000s) and land area in square kilometers.(Size and area in 2000, estimated using GRUMP methods.) 61

4.4 Distribution and percentages of the South American urban population and land area in the LECZ and drylands, by population size ranges. Population in thousands (000s) and land area in square kilometers. (Size and area in 2000, estimated using GRUMP methods.) 62

4.5 City population density in persons per square kilometer, by ecozone and city population size ranges, all regions. Figures are for cities that intersect more than one administrative area; cities contained within a single administrative area are omitted. 64

4.6 Number of cities in inland water, LECZ, and dryland ecozones. Dryland consists of dry subhumid, semiarid, and arid; the last category includes hyper-arid. 66

4.7 Number and percentage of cities by LECZ and aridity. 66

4.8 Descriptive regressions of observed urban rates on the UN's estimated national rates. `Time` variable set to 1 for 1950. 68

4.9 Regressions with estimated urban vital rates, all UN cities (Z-statistics in parentheses). 72

4.10 Classical and Bayesian panel data city growth regression models with spatially correlated errors. Spatial weights are given by distance-based ones, $1/d_{ij}$, in which d_{ij} denotes the Haversine great-circle distance (in kilometers) between cities i and j. Spatial weights are row-standardized. 77

4.11 Ordinary least-squares and random-effect estimates by region, using estimated urban fertility and mortality rates. Models without controls for city size. 83

4.13 Regressions with national vital rates, all UN cities 85

4.14 Regressions with observed urban vital rates, cities with such information available 88

4.12 Fixed-effects city growth regression models, by region. Models without controls for city size. 91

4.15 Sources of data 95

List of Figures

3.1 UN cities database 2007 version (United Nations, 2008b): population records on some 3,000 cities in developing countries, 1950 - 2007 29

3.2 City population time-series for Cuiabá, Brazil 30

3.3 Administrative units and urban settlements around Cuiabá urbanized area in Brazil with population for each entity. Combined UN-GRUMP cities data. 33

3.4 Distributions of city population records and city growth rates, All cities in developing countries, 1950 - 2007 35

3.5 Posterior distributions of selected parameters, Bayesian random-effects models 40

3.6 Classical and Bayesian forecasts of city growth rates (in terms of urban agglomeration) with random-effects models in Table 3.1, along with the UN's forecasts of national total fertility rates. 2000–2045. Median growth rates are shown along with 25th and 75th percentiles (left axis) and median national total fertility rates are shown (right axis). 43

3.7 Classical forecasts of city growth (in terms of urban agglomeration) based on random-effects models in Table 3.3 with projected urban total fertility rates and child mortality rates. 2005–2045. 48

4.1 Visualization of drylands geospatial data along with national boundaries . 56

4.2 Bayesian posterior distributions of the random-effects model 76

4.3 Forecasts of city growth rates conditional on UN projections of fertility and mortality 79
4.4 Forecasts of city growth rates conditional on UN projections of fertility and mortality, Cities over 750,000 and above where UN forecasts are available. 80
4.5 Year 2000 estimates and year 2050 forecasts of city size in India, conditional on UN projections of fertility and mortality. Limited to cities in UN cities database that crosses UN's threshold of 100,000 persons and above. 82
4.6 Forecasts of city growth rates by region, conditional on UN projections of fertility and mortality 92
4.7 Forecasts of city growth rates by LECZ and aridity. 93

Chapter 1

Introduction

As the urban populations of poor countries continue to grow, these countries will come under increasing pressure to re-think their development strategies and set priorities with both rural and urban interests in mind. Ideally, the demographic research community would assist in priority-setting by providing countries and international aid agencies with informative city population estimates and scientifically credible forecasts of the pace and distribution of future growth. Although the urban transition has been in the making for decades, much remains to be done if demographers are to supply planners and policy-makers with useful guidance (UNFPA, 2007; Montgomery, 2008).

In this dissertation, I am concerned with estimating and forecasting city population growth in developing countries. Methodologically, this study uses both classical and Bayesian panel data econometric modeling of city growth. It includes spatial econometric modeling which considers a correlation among the growth rates disturbance of the cities that are linked within a spatial network. More importantly, this study draws upon a newly-assembled, comprehensive, spatially-explicit database with demographic, ecological and geographic indicators for almost all the developing countries. The database was constructed by: (1) combining the United Nations Population Division's panel dataset of city populations and GRUMP (Global Rural-Urban Mapping Project)'s geospatial population data from CIESIN (Center for International Earth Science Information Network) at Columbia University's Earth Institute; (2) incorporating geospatial ecological data on drylands, inland water systems, the low-elevation coastal zones; and (3) linking to those data total fertility rates and child mortality rates at both the national and urban levels.

The main contributions of the thesis are two-folds: (1) methodological development and its application to estimating and forecasting developing-country city population growth and (2) geo-referencing of the United Nations Population Division's cities database (United Nations, 2008b). Methodologically, it develops Bayesian MCMC estimation and forecasting methods of a panel data model with spatially correlated errors. And, as an infrastructure for international-level urbanization research, this study geographically references cites in the database of the United Nations Population Division for the first time in its history. So it lays the foundation for a database for more extensive global urban research. Also, this study shows how GIS and geospatial data and analysis, as new forms of data and technology, can be used to give new perspective on patterns on urbanization.

Definitions used in the study

Developing Countries This study follows the definition of developing countries used in the United Nations System. United Nations Statistical Division (1999) states that "The designations "develope" and "developing" are intended for statistical convenience and do not necessarily express a judgment about the stage reached by a particular country or area in the development process." Also, the current explicit definition [1] is that "In common practice, Japan in Asia, Canada and the United States in northern America, Australia and New Zealand in Oceania, and Europe are considered "developed" regions or areas." The cities database of the United Nations Population Division (United Nations, 2008b), from which city population data in this study drawn, divides countries by the practical criteria.

[1] http://unstats.un.org/unsd/methods/m49/m49regin.htm#ftnc

City and City Boundary This study follows the definitions of both city and city boundary presented in the UN Population Division's cities database (United Nations, 2008b). As will be mentioned in Section 3.1, city populations in the cities database, which was used for World Urbanziation Prospects 2007, are recorded with one of the statistical concepts. These are City Proper, Urban Agglomeration, and Metropolitan Region. The details are also discussed in Section 3.1.

The structure of the dissertation

The thesis consists of as follows: In Chapter 2, I investigate recently-developed spatial econometric modeling and develops a Bayesian Markov chain Monte Carlo estimator and predictor of a panel data spatial econometric model when panel data are irregularly-spaced in time dimension like the UN's cities database. Spatial econometrics is a rather recent methodological development in econometrics and has been rapidly growing as useful statistical tools for analysis of spatial effect including spillovers, spatial interactions, social network effects, and peer effects in economic, demographic, political and social studies.

In Chapter 3, I first develop fertility-based econometric models of developing-country city growth and forecast future city growth rates based on the models. The forecasting method proposed has advantages: (1) it can take account of demographic, socioeconomic, and environmental factors affecting city growth, and (2) it is a probabilistic forecasting method and thus has information on uncertainty.

In Chapter 4, I analyze how urban populations are distributed by some ecological environments, and how the patterns will evolve over time. Recent developments of methodologies and geospatial technologies make it possible to construct cities data on demographic, ecological and geographic indicators in a spatial framework, to analyze the spatial distribution of city population and growth by ecological environments, and to estimate panel data regression models of city growth rates which take into account ecological variables and a correlation among the growth rate disturbances of the cities that are linked within a spatial network.

In the last chapter, I conclude the study with policy implications and suggestions for future studies.

Chapter 2

Methodological Development

This chapter is wholly devoted to the discussion of methodology with which city growth rate is modeled in the later chapters. The reason why we need this methodology is provided in 4.2. Section 2.1 investigates recently-developed spatial econometric model specification, Section 2.2 reviews classical estimation and forecast methods of the specification for programming purpose, and Section 2.3 develops its Bayesian counterpart.

2.1 Panel data spatial econometric model

2.1.1 Overview

Spatial econometrics is a rather recent methodological development in econometrics and has been rapidly growing as useful statistical tools for analysis of spatial dependence and spatial effect including externalities, spillovers, spatial interactions, social network effects, and peer effects in economic, demographic, political and social studies (See, for its review, Anselin, 1999, 2007). As will be discussed, however, spatial econometric modeling is methodologically and computationally challenging, especially with large-scale geospatial datasets since it requires specifying multidirectional relations among spatial units.

This chapter discusses both classical and Bayesian estimation and prediction of panel data panel data models with spatially correlated errors when panel data are irregularly-spaced in time dimension like international-level city population data of the United Nations Population Division (United Nations, 2008b). This chapter derives Bayesian Markov chain Monte Carlo estimator and predictor of the model using Tierney (1994)'s Metropolis-within-Gibbs algorithm. For methodological comparison with Bayesian inference and programming purpose, classical estimation and prediction methods of the model using the generalized method of moments (GMM) estimator and Goldberger (1962)'s best linear unbiased predictor (BLUP) are reviewed.

Spatial correlation is at the heart of spatial econometrics. Historically, Sir Ronald A. Fisher recognized the issue in statistics in 1930s but he did not address statistical modeling for spatial correlation directly. Cliff and Ord (1969) laid the foundation of a type of spatial autocorrelation model which is widely used in econometrics and later Anselin (1988) extended it. As Voss et al. (2006) argue, when there is reason to suspect that spatial error correlation exist, regression models that do not take it into account will likely be biased in terms of coefficient standard errors, thus contaminating inference and causing forecast error variances to be calculated incorrectly. There have been studies on spatial econometric models both for cross-sectional (Ord, 1975; Anselin, 1988; Kelejian and Prucha, 1999: among others) and for panel data (Anselin, 1988; Elhorst, 2003; Baltagi, Egger, and Pfaffermayr, 2006, 2007; Kapoor, Kelejian, and Prucha, 2007: among others).

Econometric analysis of panel data, cross-sectional time-series data, dates back to the seminar work by Balestra and Nerlove (1966). Since then, there has been substantial growth in the number of panel data studies, be it methodological or empirical, due to its advantages, availability of panel data (Hsiao, 2007). Frequently, the panel data set under study is incomplete in the sense that some observations are missing randomly in the time dimension. The missing observations problem is discussed both in time-series models (Wansbeek and Kapteyn, 1985; Shively, 1993: among others) and in panel data models (Wansbeek and Kapteyn, 1989; Baltagi and Chang, 1994; Baltagi and Wu, 1999; Mckenzie, 2001: among others).

2.1.2 Model specification

Suppose data having two dimensions: one is the time dimension denoted by index t $(t = 1, \cdots, T)$ and the other is the cross-sectional unit having spatial dimension like cities or administrative units. Assume there are a total of N spatial units in the panel dataset and each spatial unit can be numbered with integer numbers from 1 to N. However, for each time t, not all the N spatial units are available due to unbalancedness of data; the set of spatial units available vary across time in terms of both the number and its composition. Researchers frequently face this kind of unbalanced panel datasets.

As such, the model I consider is the unbalanced version of the panel data with spatial correlation both in individual effects and remainder error term (Baltagi, Egger, and Pfaffermayr, 2007). In the model, the dependent variable of interest, $\mathbf{y}_{i,t}$, is assumed to be generated by

$$\mathbf{y}_{i,t} = \mathbf{X}_{i,t}\beta + \varepsilon_{i,t} \tag{2.1}$$

for individual i $(i = 1(t), \cdots, N(t))$ [1] and time t $(t = 1, \cdots, T)$ with the spatial dependence in the error term $\varepsilon_{i,t}$ specified by

$$\varepsilon_{i,t} = \rho \sum_{j \neq i}^{N(t)} w_{i,j}\varepsilon_{j,t} + u_{i,t} \tag{2.2}$$

and with two error components in $u_{i,t}$ written as

$$u_{i,t} = \mu_i + \nu_{i,t} \tag{2.3}$$

in which μ_i is the individual effect and $\nu_{i,t}$ is the remainder disturbance. This model is usually named the KKP model due to Kapoor, Kelejian, and Prucha (2007) or called

[1] For the balanced version, $i = 1, 2, \cdots, N$.

SAR-RE (spatial autoregressive random-effects model) by Baltagi, Bresson, and Pirotte (2009).

It is worth noting some properties of the model specification: The model specified above is an unbalanced model in the sense that, given time t, only $N(t)$ observations are available among N individual spatial units in total in which $N(t) <= N$ and $1 <= 1(t) < 2(t) < \cdots, N(t) <= N$. That is, the index i goes from $1(t)$ to $N(t)$. Note that, in the standard balanced model, $i = 1, \cdots, N$ for all t.

The model is a spatial model by incorporating, in the model, the spatial relationship across individual spatial units, which is specified as equation (2.2). In the Ord (1975) type of the spatial autoregressive model, the disturbance term of spatial unit i, $\varepsilon_{i,t}$, is related to that of other spatial units, say, k through the term ρw_{ik} with spatial weight w_{ik} and spatial correlation coefficient ρ. When ρ is zero, the model is standard panel data model. The random disturbance $v_{i,t}$ is assumed to be normally distributed with mean 0 and variance σ_v^2. In addition, the random-effects panel data model assumes that μ_i is also a random variable but the fixed-effects model assumes the non-random and fixed μ_i.

The common way to work with the spatial model is to sort data by grouping time first and then spatial units to account for the spatial correlation among spatial units (Anselin, 1988; Elhorst, 2003). Thus, given time t $(t = 1, \cdots, T)$, the equation (2.2) is written as

$$\varepsilon_t = \rho W_t \varepsilon_t + \mathbf{u}_t \tag{2.4}$$

with $\varepsilon_t = (\varepsilon_{1(t),t}, \cdots, \varepsilon_{N(t),t})$ and the spatial weight matrix of dimension $N(t) \times N(t)$

$$\mathbf{W}_t = \begin{bmatrix} 0 & w_{1(t),2(t)} & \cdots & w_{1(t),N(t)} \\ w_{2(t),1(t)} & 0 & \cdots & w_{2(t),N(t)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N(t),1(t)} & w_{N(t),2(t)} & \cdots & 0 \end{bmatrix}$$

and the equation (2.3) as

$$\mathbf{u}_t = \mathbf{D}_t \mu + v_t \tag{2.5}$$

with N column vector $\mu = (\mu_1, \mu_2, \cdots, \mu_N)'$ and the $N(t) \times N$ matrix $\mathbf{D}_t$ (termed *extraction matrix*) which is obtained by deleting from the identity matrix of order $N(t)$ those rows that correspond to the missing observations (Wansbeek and Kapteyn, 1985: 470). Stacking all observations first by time and then by spatial unit, the model is written as

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

$$\varepsilon = \rho \mathbf{W} \varepsilon + \mathbf{u} \tag{2.6}$$

$$\mathbf{u} = \mathbf{D}\mu + \nu \tag{2.7}$$

in which $\mathbf{Y} = (\mathbf{Y}_1', \cdots, \mathbf{Y}_T')'$, $\mathbf{X} = (\mathbf{X}_1', \cdots, \mathbf{X}_T')'$. $W = diag(W_1, \cdots, W_T)$ is a block-diagonal $n \times n$ matrix with block W_t, where $n = \sum_{t=1}^{T} N_t$, the total number of observations. $\mathbf{D} = (\mathbf{D}_1', \mathbf{D}_2', \cdots, \mathbf{D}_T')'$ is a $n \times N$ matrix.

Choice of spatial weights

Specification of spatial autoregressive models is completed with the choice of spatial weights. Unlike that of time series models, the choice is not obvious (Ord, 1975). Anselin (1988) gives a general guideline about the weights saying that the weight matrix should bear a direct relation to a theoretical conceptualization of the structure of dependence.

Despite the guideline, the choice of a spatial weights matrix specification is not clear-cut, mostly is done an ad hoc manner, and seems to be governed primarily by convenience and convention (Griffith, 1996). In fact, there are few studies on specification of spatial weights except a few mentioned below. It is case-by-case, depending on the subject of interest. The "traditional" approach, serving as at least a starting point, lies on geographical location of the observations using a *contiguity* matrix (See Anselin, 1988: for details), distance-based matrix, or k nearest neighbors.

In economics and social science studies, the notion of distance can be extended to the more general sense. In a study about interdependence of government expenditure decisions in US states, Case et al. (1993) uses income-based weighting (i.e. income differentials) as a measure of distance between states. In a study of some commodity prices in 64 countries, Aten (1996) uses trade-based interaction measures (i.e. the volume of trade between countries measured by their exports and imports). More formally, Leenders (2002) discusses how theories under study can be incorporated in the specification of W in a social network analysis context.

Finally, Griffith (1996) provides some guidance on the specification of the spatial weights matrix based on theorems on the effect of mis-specification of weights matrix on estimators. Among the five rules-of-thumbs he states, the first and last rules are as follows:

- RULE-OF-THUMB 1: It is better to posit some reasonable geographic weights matrix specification than to assume all entries are zero (the independent observations situation of conventional statistics), the extreme case of under-specification.
- RULE-OF-THUMB 5: In general, it is better to employ a somewhat under-specified than a somewhat over-specified geographic weights matrix.

2.2 Review of classical approach

This section reviews classical estimation and forecast methods of the model. The objective is to compare it with Bayesian inference and to derive the estimation and forecasting procedure in same mathematical notations for programming purpose. The procedure is programmed in Fortran 95.

2.2.1 Generalized Method of Moments estimation

Assume, like the standard assumption of random-effects model, that the individual-specific effects $\mu \sim N(\mathbf{0}, \sigma_\mu^2 \mathbf{I}_N)$, the remainder error term $\varepsilon \sim N(\mathbf{0}, \sigma_v^2 \mathbf{I}_n)$, and both are independent of each other. The classical estimation of the model is approached by the conventional generalized least squares. Assuming that $(\mathbf{I}_n - \rho \mathbf{W})^{-1}$ exists, the variance-covariance matrix of the model error term ε is

$$E(\varepsilon\varepsilon') \equiv \sigma_v^2\Omega = \sigma_v^2\left[(\mathbf{I}_n - \rho\mathbf{W})^{-1}\Omega_{\mathbf{u}}[(I_n - \rho W)^{-1}]'\right]$$

with E$(\mathbf{u}'\mathbf{u}) = \sigma_v^2\Omega_{\mathbf{u}} = \sigma_\mu^2\mathbf{D}'\mathbf{D} + \sigma_v^2\mathbf{I}_n$. When ρ, σ_μ^2, σ_v^2 are known, the GLS estimator for β is

$$\widehat{\beta} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Y}$$

which is estimated from OLS with a two-step transformation of all variables (Kapoor, Kelejian, and Prucha, 2007; Baltagi, Egger, and Pfaffermayr, 2007) usinga Spatial Cochrane-Orcutt type transformation, say, $\mathbf{z}* = (I_n - \rho\mathbf{W})\mathbf{z}$ first and then GLS-transformation on each elements of $\mathbf{z}*$; $z_{i,t}^{**} = z_{i,t}^{*} - \theta_i\bar{z}_{i,\cdot}^{*}$ in which $\theta_i = 1 - \sqrt{\sigma_v^2/(T_i\sigma_\mu^2 + \sigma_v^2)}$ and $\bar{z}_{i,\cdot}^{*}$ is its average of $z_{i,t}^{*}$ over time and T_i is the number of observations of individual i. The calculation of $\Omega_{\mathbf{u}}^{-1}$ in this unbalanced version with two symmetric idempotent matrices $\mathbf{P} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ and $\mathbf{Q} = \mathbf{I}_n - \mathbf{P}$ is of analogy to the Magnus (1982: p.242) method applied in the balanced version.

The generalized method of moments (GMM) or Generalized Moments (GM) estimation method of the spatial correlation coefficient ρ, along with σ_μ^2 and σ_v^2 in this model, is developed by Kelejian and Prucha (1999) for the cross-section model first, by Kapoor, Kelejian, and Prucha (2007) for the panel data model and by Baltagi, Egger, and Pfaffermayr (2007) for unbalanced version of panel data model. The key of the method is to derive moment conditions involving the model disturbance, $\mathbf{u}$ of equation (2.7). Let $\bar{\mathbf{u}} = \mathbf{W}\mathbf{u}$. Note the following facts: $\mathbf{QD} = \mathbf{0}$, trace$(W) = \mathbf{0}$, trace$(P) = N$, and $\mathbf{PD} = \mathbf{D}$. Using the above facts along with the property of trace operators, the following six moments conditions are drawn:

$$\begin{bmatrix} \mathrm{E}(\mathbf{u}'\mathbf{Q}\mathbf{u}) \\ \mathrm{E}(\bar{\mathbf{u}}'\mathbf{Q}\bar{\mathbf{u}}) \\ \mathrm{E}(\mathbf{u}'\mathbf{Q}\bar{\mathbf{u}}) \\ \mathrm{E}(\mathbf{u}'\mathbf{P}\mathbf{u}) \\ \mathrm{E}(\bar{\mathbf{u}}'\mathbf{P}\bar{\mathbf{u}}) \\ \mathrm{E}(\mathbf{u}'\mathbf{P}\bar{\mathbf{u}}) \end{bmatrix} = \begin{bmatrix} \sigma_v^2(n-N) \\ \sigma_\mu^2\,\mathrm{trace}(\mathbf{W}'\mathbf{QWDD}') + \sigma_v^2\mathbf{W}'\mathbf{QW} \\ 0 \\ n\sigma_\mu^2 + N\sigma_v^2 \\ \sigma_\mu^2\,\mathrm{trace}(\mathbf{W}'\mathbf{PWDD}') + \sigma_v^2\mathbf{W}'\mathbf{PW} \\ 0 \end{bmatrix} \tag{2.8}$$

Let $\bar{\varepsilon} = \mathbf{W}\varepsilon$, $\bar{\bar{\varepsilon}} = \mathbf{W}\bar{\varepsilon}$. From the model relating $\mathbf{u}$ to ε in equation (2.6), we know that

$$\mathbf{u} = \varepsilon - \rho\bar{\varepsilon} \;\text{ and }\; \bar{\mathbf{u}} = \bar{\varepsilon} - \rho\bar{\bar{\varepsilon}} \tag{2.9}$$

Substituting the above two equations into equation (2.8) and rearranging give the following population moment conditions:

$$\begin{bmatrix} 2\mathrm{E}(\varepsilon'\mathbf{Q}\bar{\varepsilon}) & -\mathrm{E}(\bar{\varepsilon}'\mathbf{Q}\bar{\varepsilon}) & n-N & 0 \\ 2\mathrm{E}(\bar{\bar{\varepsilon}}'\mathbf{Q}\bar{\varepsilon}) & -\mathrm{E}(\bar{\bar{\varepsilon}}'\mathbf{Q}\bar{\bar{\varepsilon}}) & g_{23} & g_{24} \\ \mathrm{E}(\varepsilon'\mathbf{Q}\bar{\bar{\varepsilon}} + \bar{\varepsilon}'\mathbf{Q}\bar{\varepsilon}) & -\mathrm{E}(\bar{\bar{\varepsilon}}'\mathbf{Q}\bar{\varepsilon}) & 0 & 0 \\ 2\mathrm{E}(\varepsilon'\mathbf{Q}_1\bar{\varepsilon}) & -\mathrm{E}(\bar{\varepsilon}'\mathbf{P}\bar{\varepsilon}) & N & 1 \\ 2\mathrm{E}(\bar{\bar{\varepsilon}}'\mathbf{Q}_1\bar{\varepsilon}) & -\mathrm{E}(\bar{\bar{\varepsilon}}'\mathbf{P}\bar{\bar{\varepsilon}}) & g_{53} & g_{54} \\ \mathrm{E}(\varepsilon'\mathbf{P}\bar{\bar{\varepsilon}} + \bar{\varepsilon}'\mathbf{P}\bar{\varepsilon}) & -\mathrm{E}(\bar{\bar{\varepsilon}}'\mathbf{P}\bar{\varepsilon}) & 0 & 0 \end{bmatrix} \begin{bmatrix} \rho \\ \rho^2 \\ \sigma_v^2 \\ \sigma_\mu^2 \end{bmatrix} - \begin{bmatrix} \mathrm{E}(\varepsilon'\mathbf{Q}\varepsilon) \\ \mathrm{E}(\bar{\varepsilon}'\mathbf{Q}\bar{\varepsilon}) \\ \mathrm{E}(\varepsilon'\mathbf{Q}\bar{\varepsilon}) \\ \mathrm{E}(\varepsilon'\mathbf{P}\varepsilon) \\ \mathrm{E}(\bar{\varepsilon}'\mathbf{P}\bar{\varepsilon}) \\ \mathrm{E}(\varepsilon'\mathbf{P}\bar{\varepsilon}) \end{bmatrix} = \mathbf{0} \tag{2.10}$$

with

$$g_{23} = \mathrm{trace}(\mathbf{W}'\mathbf{QW}), \;\; g_{24} = \mathrm{trace}(\mathbf{W}'\mathbf{QWDD}')$$
$$g_{53} = \mathrm{trace}(\mathbf{W}'\mathbf{PW}), \;\; g_{54} = \mathrm{trace}(\mathbf{W}'\mathbf{PWDD}')$$

The sample analogue to the above equation is obtained by replacing ε for the OLS residuals $\mathbf{e} = Y - X\tilde{\beta}$ where $\tilde{\beta}$ is the OLS estimator. The sample analogue can be written as $\mathbf{G}\delta - \mathbf{g} = \xi$ with $\delta = (\rho, \rho^2, \sigma_v^2, \sigma_\mu^2)'$ and the error vector ξ. The GM estimator is defined as

$$\hat{\delta} = \mathrm{argmin}\;\; (\mathbf{G}\delta - \mathbf{g})'(\mathbf{G}\delta - \mathbf{g})$$

The GM estimator can be obtained from the non-linear regression of $-\mathbf{g}$ on $-\mathbf{G}$ with non-linear optimization algorithms. Since we have three coefficients to be estimated, the first four equations of $\mathbf{G}$ and $\mathbf{g}$ is sufficient to estimate the coefficients (Egger et al., 2005; Baltagi et al., 2007). The asymptotic properties of the GM estimator and other details are available in Kelejian and Prucha (1999) and Kapoor, Kelejian, and Prucha (2007).

2.2.2 Best Linear Unbiased Prediction

I here discuss prediction of future dependent variables based on the model and data in hand. Our discussion about prediction begins with Goldberger (1962)'s best linear unbiased predictor (BLUP) of the generalized linear regression model, written as

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon,\ \mathrm{E}(\varepsilon) = \mathbf{0},\ \mathrm{E}(\varepsilon\varepsilon') = \Sigma$$

in which y is a $T \times 1$ vector of dependent variables. At a future time $T+s$, the actual realization of the dependent variable y_{T+s} given the regressor X_{T+s} will be

$$y_{T+s} = X_{T+s}\beta + \varepsilon_{T+s}$$

with the prediction disturbance ε_{T+s}. Assume that $\mathrm{E}\,\varepsilon_{T+s} = 0$, $\mathrm{E}\,\varepsilon_{T+s}^2 = \sigma_{T+s}^2$ and $\mathrm{E}\,\varepsilon_{T+s}\varepsilon = \mathbf{w}$ in which $\mathbf{w}$ is the $T \times 1$ vector of covariances of the prediction disturbances with the vector of sample disturbances.

Let $p = c'\mathbf{y}$ be any linear predictor of y_{T+s} with c being a $T \times 1$ vector of constant. The best linear unbiased predictor (BLUP) of y_{T+s} is the predictor p such that

$$\min_{p}\ \sigma_p^2 = E(p - y_{T+s})^2$$
$$\text{s.t. } \mathrm{E}(p - y_{T+s}) = 0$$

in which σ_p^2 is the prediction variance. Goldberger (1962) shows that the best linear unbiased predictor is

$$\widehat{p} = \mathbf{X}_{T+s}\widehat{\beta} + w'\Sigma^{-1}e \tag{2.11}$$

with the best linear unbiased estimator $\widehat{\beta} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y}$ and its sample residuals $\mathbf{e}$. The second term $w'\Sigma^{-1}e$ utilizes a priori knowledge of the interdependence of disturbances along with the sample residuals (which are estimates of the sample disturbances) to estimate the prediction disturbance ε_{T+s} (Goldberger, 1962: p.371). The prediction variance can be shown to be

$$\begin{aligned}\sigma^2_{\widehat{p}} =&\sigma^2_{T+s} + \mathbf{X}'_{T+s}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}_{T+s} - 2\mathbf{X}_{T+s}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{w} - \\ &\mathbf{w}'[\Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1})\mathbf{X}'\Sigma^{-1}]\mathbf{w}\end{aligned}$$

The approach can be applied to various types of regression models. There have been studies on prediction with the random-effects panel data model. Taub (1979) derived the best linear unbiased predictor for the one-way error component model, which is same as our model when ρ is zero (i.e. no spatial correlation). The forecast of the future value of $\widehat{y}_{i,T+s}$ is, when ρ is zero,

$$\widehat{y}_{i,T+s} = \mathbf{X}_{T+s}\widehat{\beta} + \frac{T_i\sigma^2_\mu}{T_i\sigma^2_\mu + \sigma^2_\nu}(\overline{\mathbf{y}}_{i,\cdot} - \overline{\mathbf{X}}_{i,\cdot}\widehat{\beta}) \tag{2.12}$$

with the GLS estimator $\widehat{\beta}$ in which $\overline{\mathbf{y}}_{i,\cdot}$ is the average of $\mathbf{y}_{i,t}$ over time for individual i. Lee and Griffiths (1979), Judge et al. (1985: p.524), and Lee (2006: p.149) also discuss the second term of the right-hand side of equation (2.12). For other specifications of panel data models, the prediction of a one-way error component model with serial correlation is studied by Baltagi and Li (1992) and the random coefficient model by Lee and Griffiths (1979). Recently, Baltagi and Li (2004) and Baltagi et al. (2009) considered the problem of prediction in a panel data regression model with spatial correlation.

Let us consider our model when ρ is not zero. In this case, it is sufficient to derive the BLUP with the model's balanced version. The $N \times 1$ disturbance vector at any time t, $\varepsilon_t = (\varepsilon_{1,t}, \cdots, \varepsilon_{N,t})$, in terms of its component $\mu = (\mu_1, \cdots, \mu_N)'$ and ν_t is expressed as $\varepsilon_t = \mathbf{B}_N^{-1}\mu + \mathbf{B}_N^{-1}\nu_t$ in which $\mathbf{B}_N = \mathbf{I}_N - \rho\mathbf{W}_N$ with assumption that its inverse exists.

Define $\iota_T = (1, \cdots, 1)^{'}$ be a $T \times 1$ column vector of ones. Let us first look at the covariance between $NT \times 1$ error vector $\varepsilon = (\varepsilon_1^{'}, \cdots, \varepsilon_T^{'})^{'}$ and $N \times 1$ error vector ε_{T+s} at any s time ahead. The covariance matrix between ε and ε_{T+s} is expressed as

$$\begin{aligned} \mathrm{E}(\varepsilon_{T+s}\varepsilon') &= E(\mathbf{B}_N^{-1}\mu + \mathbf{B}_N^{-1}\nu_{T+s})[(\iota_T \otimes \mathbf{B}_N^{-1})\mu + (\mathbf{I}_T \otimes \mathbf{B}_N^{-1})\nu]^{'} \\ &= \sigma_\mu^2(\iota_T^{'} \otimes (\mathbf{B}_N'\mathbf{B}_N)^{-1}) \end{aligned} \tag{2.13}$$

in which $\otimes$ denotes the Kronecker product.[2] The second equality comes from our assumptions about error terms using the operations and properties of the Kronecker product (See Abadir and Magnus, 2005: p.273-281). The $N \times NT$ covariance matrix of ε and ε_{T+s} depends on the structure of the matrix $\mathbf{B}_N = \mathbf{I}_N - \rho\mathbf{W}_N$ and thus the spatial weight matrix $\mathbf{W}$.

Let us move onto the covariance matrix of ε. Define $\mathbf{P} = \iota_T(\iota_T^{'}\iota_T)^{-1}\iota_T^{'}$ and $\mathbf{Q} = \mathbf{I}_T - \mathbf{P}$. Both are symmetric and idempotent. The covariance matrix of ε is expressed as

$$\begin{aligned} \mathrm{E}(\varepsilon\varepsilon') \equiv \Sigma &= \sigma_\mu^2(\iota_T\iota_T^{'} \otimes (\mathbf{B}_N'\mathbf{B}_N)^{-1}) + \sigma_\nu^2(\mathbf{I}_T \otimes (\mathbf{B}_N'\mathbf{B}_N)^{-1}) \\ &= (T\sigma_\mu^2 + \sigma_\nu^2)(\mathbf{P} \otimes (\mathbf{B}_N'\mathbf{B}_N)^{-1}) + \sigma_\nu^2(\mathbf{Q} \otimes (\mathbf{B}_N'\mathbf{B}_N)^{-1}) \end{aligned}$$

The second equality originated from the lemma of Baltagi (1980: p.1548) and Magnus (1982: p.242), widely used in analysis of random-effects panel data models. The inverse of the covariance matrix is, by the lemma,

$$\Sigma^{-1} = \frac{1}{\sigma_\nu^2}\left[\frac{\sigma_\nu^2}{T\sigma_\mu^2 + \sigma_\nu^2}(\mathbf{P} \otimes \mathbf{B}_N'\mathbf{B}_N) + (\mathbf{Q} \otimes \mathbf{B}_N'\mathbf{B}_N)\right] \tag{2.14}$$

Substituting (2.13) and (2.14) into equation (2.11) gives the best linear unbiased predictor of the model. Note the fact that $\iota_T^{'}\mathbf{P} = \iota_T$ and $\iota_T^{'}\mathbf{Q} = \mathbf{0}$. The BLUP of the vector of future dependent variables $\widehat{\mathbf{y}}_{T+s}$ at s time ahead is

$$\widehat{\mathbf{y}}_{T+s} = \mathbf{X}_{T+s}\widehat{\beta} + \frac{\sigma_\mu^2}{T\sigma_\mu^2 + \sigma_\nu^2}(\iota_T^{'} \otimes \mathbf{I}_N)(\mathbf{y} - \mathbf{X}\widehat{\beta})$$

[2] An early reference on the Kronecker product is (MacDuffee, 1933: p.81-84). In the book, it is named as *right direct product* with the symbol combining × and ·.

When the observed data is unbalanced, the i'th element of $\widehat{\mathbf{y}}_{T+s}$ is same as that in equation (2.12). As shown in Baltagi et al. (2009: p.16), the result means that the predictor of the KKP model is the same as that of the RE model with no spatial correlation. While the predictor formula is the same, the GM procedure in the KKP model due to estimates of ρ yield different estimates β which in turn yield different residuals and hence different forecasts.

2.3 Bayesian approach

This section specifies the Bayesian approach of this model specification above, derives its Markov chain Monte Carlo (MCMC) estimator and predictor, and briefly discusses its computational issues. It is programmed in Fortran 95.

2.3.1 Bayesian hierarchical modeling

Bayesian modeling is comprised of two parts: model specification and specification of the model parameters' prior distributions. Recall our spatial econometric model for unbalanced panel data, in matrix notation, where observations are ordered by time t, $t = 1, \cdots, T$, first and then observed individual spatial unit i, $i = i(1), \cdots, N(t)$, among N individuals $(1 <= i(1) < \cdots, < N(t) <= N)$:

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}\beta + \varepsilon \\ \varepsilon &= \rho\mathbf{W}\varepsilon + \mathbf{u} \\ \mathbf{u} &= \mathbf{D}\mu + \nu\end{aligned}$$

in which spatial weight matrix $\mathbf{W} = diag(W_1, \cdots, W_T)$ is $n \times n$ block-diagonal matrix with blocks $\mathbf{W}_t$ of different dimensions over time due to unbalancedness. $\mu = (\mu_1, \cdots, \mu_N)'$ is the N column vector of individual effects and $\mathbf{D}$ is the *extract matrix* to account for the unbalancedness. Assuming that the inverse matrix of $(\mathbf{I} - \rho\mathbf{W})$ exists and the disturbance vector ε has independent normal distribution, the model is written as

$$\mathbf{y} = \mathbf{X}\beta + (\mathbf{I}_n - \rho\mathbf{W})^{-1}D\mu + (\mathbf{I}_n - \rho\mathbf{W})^{-1}\nu, \quad \nu \sim N(\mathbf{0}, \sigma_\mu^2\mathbf{I}_n)$$

Bayesian modeling is completed by specifying the prior distributions of the unknown model coefficients $\beta = (\beta_1, \cdots, \beta_k)$, $\mu = (\mu_1, \cdots, \mu_N)$, σ_μ^2, σ_ν^2 and ρ. We assume that the following prior distributions: [3]

$$\beta \sim N(\beta_0, \mathbf{M}_0^{-1}), \;\; \sigma_\nu^2 \sim iG(\nu_0/2, s_0/2), \;\; \mu_i \sim N(0, \sigma_\mu^2), \;\; \sigma_\mu^2 \sim iG(r_0/2, p_0/2) \quad (2.15)$$

and uniform prior distribution over (-1, 1) for ρ. The pair of the above priors for β and σ_ν^2 is called independent Normal-Gamma prior, a widely used conjugate prior in regression models.

From a Bayesian perspective, the latent individual effect u_i is considered as one of unknown coefficients. This idea comes from Tanner and Wong's (1987) idea of data augmentation and thus is used in, among others, Zeger and Karim (1991), Chib (1996), and Chib and Carlin (1999). In this way, we can generate samples of the latent individual effect. Note also the difference in estimation method of the individual-specific effect between classical and Bayesian approaches.

The priors for μ_i and σ_μ^2 (which is the parameter of distribution of μ) gives the nature of hierarchical modeling in Bayesian approach. This hierarchical structure corresponds to the usual random-effects model in classical approach. By assuming that model coefficients are random rather than fixed, Bayesian can give hierarchical structure for each of model coefficients though this model gives the structure only for the individual effect. This Bayesian capacity of hierarchical modeling looks attractive. See Banerjee et al. (2004)

[3] Here, the probability density function of a random variable Z which has the inverse gamma distribution is defined as

$$p(z; \nu, s) = \frac{1}{\Gamma(\nu)\beta^{-\nu}} z^{-\nu-1} e^{-\frac{s}{z}}, \quad 0 < z < \infty, \nu > 0, s > 0$$

in which $\Gamma(\nu)$ is the gamma function.

for hierarchical modeling for spatial data analysis and Koop et al. (2007) for exercises of Bayesian econometrics including hierarchial models.

2.3.2 Markov chain Monte Carlo estimator

Assuming that all the regression coefficient are random, Bayesian inference follows intuitive and coherent estimation procedure using the rules of probability. Bayesian estimation involves analysis of the distribution of the unknown model coefficients conditional on data, which is the formal expression of what we have learned from the data. In our model, by the laws of probability, the joint posterior distribution of the coefficients $\beta = (\beta_1, \cdots, \beta_k)$, $\mu = (\mu_1, \cdots, \mu_N)$, σ_μ^2, σ_v^2 and ρ is proportional to the likelihood function of the model and joint prior distribution of the coefficient:

$$p(\beta, \mu, \sigma_\mu^2, \sigma_v^2, \rho | \mathbf{Y}, \mathbf{X}) \propto p(\mathbf{Y}|\mathbf{X}, \beta, \mu, \sigma_\mu^2, \sigma_v^2, \rho) p(\beta, \mu, \sigma_\mu^2, \sigma_v^2, \rho)$$

which implies likelihood and prior determines posterior, our belief about the coefficient after observing data. The remaining part of the joint posterior distribution is just the normalizing constant of the joint posterior distribution.

Define $\mathbf{Z} = \mathbf{Y} - \mathbf{B}^{-1}\mathbf{D}\mu$. From the joint posterior of the model, the kernel of the full conditional posterior distribution of β is,

$$p(\beta | \mu, \sigma_\mu^2, \sigma_v^2, \rho) \propto \exp\left\{ -\frac{1}{2}\left[(\mathbf{Z} - \mathbf{X}\beta)'(\sigma_v^{-2}\mathbf{B}'\mathbf{B})(\mathbf{Z} - \mathbf{X}\beta) + (\beta - \beta_0)' M_0 (\beta - \beta_0) \right] \right\}$$

Define $\mathbf{M}_1 = \mathbf{M}_0 + \sigma_v^{-2}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ with $\tilde{\mathbf{X}} = \mathbf{B}\mathbf{X} = (\mathbf{I}_n - \rho\mathbf{W})\mathbf{X}$ which is described as "spatially filtered" and $\beta_1 = \mathbf{M}_1^{-1}(\mathbf{M}_0\beta_0 + \sigma_v^{-2}\tilde{\mathbf{X}}'\tilde{\mathbf{Z}})$ in which $\tilde{\mathbf{Z}}$ is similarly defined as $\tilde{\mathbf{X}}$. Solving for β using completing the square, the full conditional for β is

$$\beta | \mu, \sigma_v^2, \sigma_u^2, \rho, \mathbf{Y}, \mathbf{X} \sim N(\beta_1, M_1^{-1}) \tag{2.16}$$

Define $\mathbf{Q} = \mathbf{Y} - \mathbf{X}\beta$. The kernel of the full conditional posterior for μ is

$$p(\mu, | \beta, \sigma_\mu^2, \sigma_v^2, \rho) \propto \exp\left\{ -\frac{1}{2}\left[\sigma_v^{-2}(\mathbf{BQ} - \mathbf{D}\mu)'(\mathbf{BQ} - \mathbf{D}\mu) + \sigma_\mu^{-2}\mu'\mu \right] \right\}$$

Define $\overline{\widetilde{\mathbf{Q}}}_{i,.}$ as the average of individual i's elements of matrix $\mathbf{Q}$ over time. The full conditional distribution of μ_i is

$$\mu_i|\beta,\sigma_\mu^2,\sigma_v^2,\rho,\mathbf{Y},\mathbf{X} \sim N\left(\frac{T_i\sigma_\mu^2}{T_i\sigma_\mu^2+\sigma_\varepsilon^2}\overline{\widetilde{\mathbf{Q}}}_{i,.},\frac{\sigma_\mu^2\sigma_v^2}{T_i\sigma_\mu^2+\sigma_v^2}\right) \tag{2.17}$$

The full conditional posterior of σ_μ^2 is

$$\sigma_\mu^2|\beta,\mu,\sigma_v^2,\rho,\mathbf{Y},\mathbf{X} \sim iG\left(\frac{N+r_0}{2},\frac{\mu'\mu+p_0}{2}\right) \tag{2.18}$$

The full conditional posterior of σ_v^2 is

$$\sigma_v^2|\beta,\mu,\sigma_u^2,\rho,\mathbf{Y},\mathbf{X} \sim iG\left(\frac{v_1}{2},\frac{S_1}{2}\right) \tag{2.19}$$

where $S_1 = (\mathbf{BY}-\mathbf{BX}\beta-\mathbf{D}\mu)'(\mathbf{BY}-\mathbf{BX}\beta-\mathbf{D}\mu)+S_0$ and $v_1 = n+v_0$.

The kernel of the full conditional posterior distribution of ρ is

$$p(\rho|\beta,\mu,\sigma_\mu^2,\sigma_v^2,\mathbf{Y},\mathbf{X}) \propto |\mathbf{B}|\exp\left[-\frac{1}{2\sigma_v^2}(\widetilde{\mathbf{Y}}-\widetilde{\mathbf{X}}\beta-\mathbf{D}\mu)'(\widetilde{\mathbf{Y}}-\widetilde{\mathbf{X}}\beta-\mathbf{D}\mu)\right] \tag{2.20}$$

Unlike the other parameters, the parameter ρ does not have a well-known distribution. Metropolis-Hastings algorithm can be used for sampling from the distribution.

The random-walk Metropolis-Hastings algorithm for ρ draws a candidate ρ^* from a candidate-generating function, here, a (truncated) normal distribution: at $i+1$-th iteration, draw ρ^* from

$$\rho^* \sim N(\rho_i,c^2)$$

in which ρ_i is the draw from previous i-th iteration and c is the tuning parameter which is used to adjust the acceptance rate of the MH algorithm. We then calculate the ratio $p(\rho^*)/p(\rho)$. If the ratio is greater than 1, the candidate is accepted (that is, $\rho_{i+1} = \rho^*$) and if the ratio is less than 1, the candidate is accepted with probability $p(\rho^*)/p(\rho)$; that is, take a uniform random number u between 0 and 1 and if $u < p(\rho^*)/p(\rho)$, accept the candidate and if $u > p(\rho^*)/p(\rho)$, do not accept the candidate (that is, $\rho_{i+1} = \rho_i$). The set of draws behaves like the draws from the the full conditional posterior distribution of ρ. The samples drawn through the Metropolis -Within-Gibbs are used to estimate the model using the usual Monte Carlo method.

Next, Bayesian predictor of the model is discussed. Bayesian approach to prediction is very straightforward compared to its classical counterpart. It uses rules of probability and the MCMC samples of the model parameters can be used to generate future values of the dependent variables.

Let $\theta = (\beta, \mu, \sigma_\mu^2, \sigma_v^2, \rho)$. Bayesian prediction of $\widetilde{\mathbf{Y}}_{t+s}$ at future time $t+s$ over N cities in the data is expressed as follows:

$$p(\widetilde{\mathbf{Y}}_{t+s}|\widetilde{\mathbf{X}}_{t+s}, \mathbf{Y}, \mathbf{X}) = \int_\Theta p(\widetilde{\mathbf{Y}}_{t+s}|\widetilde{\mathbf{X}}_{t+s}, \mathbf{Y}, \mathbf{X}, \theta) p(\theta|\widetilde{\mathbf{X}}_{t+s}, \mathbf{Y}, \mathbf{X}) d\theta$$

In this model, $p(\theta|\widetilde{\mathbf{X}}_{t+s}, \mathbf{Y}, \mathbf{X}) = p(\theta|\mathbf{Y}, \mathbf{X})$ and $p(\widetilde{\mathbf{Y}}_{t+s}|\widetilde{\mathbf{X}}_{t+s}, \mathbf{Y}, \mathbf{X}, \theta) = p(\widetilde{\mathbf{Y}}_{t+s}|\widetilde{\mathbf{X}}_{t+s}, \theta)$. We can simulate the predictive distribution by simulating the posterior distribution first and then the conditional distribution; that is, successively draw $\widetilde{\mathbf{Y}}_{t+s}$ from

$$\widetilde{\mathbf{Y}}_{t+s} \sim N\left(\widetilde{\mathbf{X}}_{t+s}\beta + (\mathbf{I}_N - \rho\mathbf{W}_N)^{-1}\mu, \sigma_\varepsilon^2[(\mathbf{I}_N - \rho\mathbf{W}_N)'(\mathbf{I}_N - \rho\mathbf{W}_N)]^{-1}\right) \tag{2.21}$$

using the MCMC samples of model parameters drawn above.

It seems not to be straightforward to compare classical and Bayesian estimator and predictor since Bayesian relies on sampling methods as discussed above. However, the comparison is possible by comparing estimation and prediction results. Chapter 3 includes the discussion.

Evidently, Bayesian inference made possible by the development of the MCMC sampling methods, which is usually evaluated with the expression of “MCMC has revolutionized Bayesian practice”. On computational perspective, however, Bayesian seems to have to another issue to develop efficient computer algorithms, especially to handle complex models with large datasets.

This spatial econometric model is an example. In programming of this particular econometric model, we need to consider at least the following two things: (1) For prediction, Equation (2.21) contains inverse matrix calculations of matrices including $(\mathbf{I}_N - \rho \mathbf{W}_N)$, whose dimension depends on the number of spatial cross-sectional units (i.e. the number of cities in the world), and (2) For estimation, Equation (2.20) contains the determinant calculations of the matrix, $|\mathbf{I}_n - \rho \mathbf{W}|$, whose dimension is the number of observations by the number of observations.

Recall that the matrix $(I_n - \rho \mathbf{W})$, in which $W = diag(W_1, \cdots, W_T)$, is a block-diagonal $n \times n$ matrix with n denoting the total number of observation. one computationally efficient way to calculate the determinant of the matrix is to use the following two mathematical propositions: (1) $| I_n - \rho \mathbf{W} | = \prod_{i=1}^{n} (1 - \rho \lambda_i)$ with λ_i being the i-th eigenvalue of the spatial weight matrix $\mathbf{W}$, which was proved by Ord (1975) and (2) the eigenvalues of the block-diagonal matrix $\mathbf{W} = diag(\mathbf{W}_1, \cdots, \mathbf{W}_T)$ are those of the diagonal blocks $\mathbf{W}_1, \cdots, \mathbf{W}_T$. The determinant calculation of the matrix receives spatial econometrician's attention. See Smirnov and Anselin (2001), Anderson et al. (1999), Barker et al. (2001), Barry and Pace (1999) Pace and LeSage (2000), among others.

In the case of parallel programming with multi-processors, we need to additionally consider the following element: parallel computing of MCMC can have the problem of "data dependence" (or "data dependency") due to its Markov chain properties. That is to say, to generate the next value, it requires one realized in the previous step. We also need to note that parallel computing can be less efficient for small size problems since parallel computing requires message passing among processors which don't have "shared memory". This thesis no longer discusses about it. Instead, see Kontoghiorghes (2006) for more details on parallel computing in statistics.

2.A Overview of Markov chain Monte Carlo estimation

Let us consider Markov chain Monte Carlo (MCMC) estimation methods in statistics and econometrics. Let $\pi(x)$ be a probability density function of a continuous random variable X. Statistical inference involves evaluating integrals

$$I = \int_{\Omega} h(x)\pi(x)dx \tag{2.22}$$

of some function $h(x)$ with respect to the probability density function $\pi(x)$ with sample space Ω. A simple Monte Carlo method uses the estimate $\hat{I}_1 = N^{-1}\sum_{i=1}^{N} h(x_i)$ using random samples $x_{(1)}, x_{(2)}, \cdots, x_{(N)}$ from $\pi(x)$.

This method is useful since there are well-established methods to draw random samples of standard probability distributions. See Tanizaki (2004) for random number generators. Most mathematical and statistical computer software packages provide functions to draw random samples of well-known distributions. Provided that $\sigma_g^2 = \mathrm{Var}(h(x))$,

$$\sqrt{N}(\hat{I} - \mathrm{E}h(x)) \overset{d}{\to} N(0, \sigma_g^2)$$

The problem, however, arises in cases that the probability distribution involved is not of standard form and is of high dimension, and is known up to the normalizing constant; As will be seen, Bayesian inference also involves such unfavorable situations. Markov chain Monte Carlo (MCMC) methods simulate a Markov chain whose distributions converge to stationary distribution $\pi(x)$.

Markov chain

A Markov chain is a sequence of random variables $\{X^{(t)} : t \geq 0\}$ on state space Ω with the transition probability $P(x^{(t)}, A)$ defined by

$$P(x^{(t)}, A) \equiv Prob(X^{(t+1)} \in A | X^{(t)} = x^{(t)}),\ x^{(t)} \in \Omega, A \subset \Omega.$$

It is the conditional probability of moving from a state $x^{(t)}$ in t-th iterate to the set A in $(t+1)$th-iterate. We assume that the chain is homogeneous; the transition probabilities do not depend on index t. The transition probability has properties that $P(x^{(t)}, \Omega) = 1$, and $P(x, \{x\}) \neq 0$ (the chain can stay in the same state).

When state space Ω is finite, the transition probability is defined as a transition matrix with elements $Prob(X_{t+1} = x^{(t+1)} | X_t = x^{(t)}), x^{(t)},\ x^{(t+1)} \in \Omega$. When Ω is continuous, the transition probability is defined such that

$$P(x^{(t)}, A) = \int_A k(x^{(t)}, x^{(t+1)}) dx^{(t+1)}$$

with a transition probability density function $k(x^{(t)}, x^{(t+1)})$. Given an initial state distribution of the chain $X^{(0)}$, the conditional distribution of $X^{(s)}$ given $X^{(0)}$ is written as $P^s(X^{(0)}, A) = Prob(X^{(s)} \in A | X^{(0)} = x^{(0)})$ where P^s denotes the s-th iterate of the kernel P.

Consider a state density function $\pi(x)$ on Ω satisfying the condition $\pi = \pi P$ in finite case and $\pi(x') = \int k(x,x')\pi(x)dx$ in continuous case. The density $\pi(x)$ is called *stationary* (or *invariant*) distribution of the chain. Here, we use the terms density and distribution interchangeably. The invariant distribution π is an equilibrium distribution of the chain if

$$\lim_{s \to \infty} P^s(x, A) = \pi(A)$$

in which $\pi(A) = \int_A \pi(x)dx$. If a chain has a proper invariant distribution π and it is *irreducible* and *aperiodic* [4], then π is the unique invariant distribution and is also the equilibrium distribution of the chain (Tierney, 1994: p.1712). By the theorem, Markov chain Monte Carlo estimation methods use samples from the transition density k instead of direct sampling from π.

Note that the theorem does not depend on the initial value and distribution of a chain. Markov chain theory is concerned with the speed of the convergence, *mixing time* of a chain. The convergence is also an important implementation issue in MCMC methods. Also, note that the MCMC samples are correlated, and thus estimate of the standard deviation of an estimate and assessment of the error of an estimate may require more care than with independent samples (Hastings, 1970).

Once we get MCMC samples from a transition kernel P with unique invariant distribution π, we are interested in the behavior of the sample average

$$\hat{I}_2 = T^{-1} \sum_{t=1}^{T} x^{(t)} \tag{2.23}$$

The ergodic theorem for Markov chains is required for law of large numbers and central limit theorem. Consider an irreducible and aperiodic Markov chain $X^{(t)}$ with unique stationary distribution π. the sample average above converges almost surely to the expectation of the function with respect to the stationary distribution; The law of large numbers holds for any ergodic chain (Tierney, 1994: p.1717). The idea of the ergodic theorem for Markov chains is that "chain averages equal state averages".

[4] A Markov chain with invariant distribution π is irreducible if, for any initial state, it has positive probability of entering any set to which π assigns positive probability. A chain is periodic if there are portions of the state space it can only visit at certain regularly space intervals;otherwise, the chain is aperiodic.

Markov chain Monte Carlo

Let us now consider the reverse problem we are interested in for MCMC methods: Given a *target* distribution π on Ω, can we construct an ergodic Markov chain with the stationary distribution π? Consider a Markov chain with a transition probability density expressed as, for some function $p(x^{(t)},x^{(t+1)})$,

$$k(x^{(t)},x^{(t+1)}) = p(x^{(t)},x^{(t+1)}) + r(x^{(t)})\delta_{x^{(t)}}(x^{(t+1)}) \tag{2.24}$$

in which $p(x^{(t)},x^{(t+1)}) = 0$ if $x^{(t+1)} = x^{(t)}$, $r(x^{(t)}) = 1 - \int_{\Omega} p(x^{(t)},x^{(t+1)})dx^{(t+1)}$ is the probability that the chain remains at the previous state $x^{(t)}$ in the $(t+1)$-th iterate, and, by construction, $\delta_{x^{(t)}}(x^{(t+1)}) = 1$ if $x^{(t+1)} = x^{(t)}$ and 0 otherwise. The corresponding transition probability is expressed as

$$P(x^{(t)},A) = \int_A p(x^{(t)},x^{(t+1)})dx^{(t+1)} + r(x^{(t)})I_A(x)$$

If the function $p(x^{(t)},x^{(t+1)})$ satisfies the *reversibility condition* or *detailed balance*

$$\pi(x^{(t)})p(x^{(t)},x^{(t+1)}) = \pi(x^{(t+1)})p(x^{(t+1)},x^{(t)}),$$

then π is the stationary density of the chain (Tierney, 1994). Note that the left-hand side is the unconditional probability to move from $x^{(t)}$ to $x^{(t+1)}$ where $x^{(t)}$ is generated from $\pi(\cdot)$ and right-hand side is the unconditional probability to move from $x^{(t+1)}$ to $x^{(t)}$ where $x^{(t+1)}$ is generated from the same density $\pi(\cdot)$. This condition gives us a sufficient condition for $p(x^{(t)},x^{(t+1)})$ to be satisfied. The transition density implies that, given a state $X^t = x^{(t)}$, movement from the state $x^{(t)}$ to other state $x^{(t+1)}$ ($x^{(t+1)} \neq x^{(t)}$) is determined by the function $p(x^{(t)},x^{(t+1)})$ and transition to the same state (($x^{(t+1)} \neq x^{(t)}$)) occurs with probability $r(\cdot)$.

Metropolis-Hastings algorithm

The Metropolis-Hastings kernel is

$$p_{MH}(x^{(t)},x^{(t+1)}) \equiv q(x^{(t)},x^*)\alpha(x^{(t)},x^*) \tag{2.25}$$

with

$$\alpha(x^{(t)},x^*) = min\left\{\frac{\pi(x^{(*)})q(x^*,x^{(t)})}{\pi(x^{(t)})q(x^{(t)},x^*)},1\right\}$$

and $q(x^{(t)}, x^*)$ is any transition probability density. Metropolis-Hastings algorithm is defined by the function called *candidate-generating* or *proposal* density. In the random-walk Metropolis-Hastings algorithm, for instance, candidates are generated from the process $x^* = x^t + z$ with a normal distribution z. If $q(x^{(t)}, x^*) = q(x^*, x^{(t)})$ like the above random-walk process, the acceptance probability simplifies to

$$\alpha(x^{(t)}, x^*) = min\{\pi(x^*)/\pi(x^{(t)}), 1\}.$$

Metropolis-Hastings algorithm does not require the normalizing constant part of π since it appears in both numerator and denominator of $\alpha(x^{(t)}, x^*)$, making the algorithm attractive. The Metropolis-Hastings algorithm works as follows: Given a realized state $X^{(t)} = x^{(t)}$ in t-th iterate, draw a sample point x^* from a pre-specified proposal density $q(x^{(t)}, x^*)$, and move to the point $X^{(t+1)} = x^*$ with probability $\alpha(x^{(t)}, x^*)$ and remain in the previous state $X^{(t+1)} = x^{(t)}$ with probability $1 - \alpha(x^{(t)}, x^*)$.

Algorithm 1 (Metropolis-Hastings algorithm):

- Choose a starting value $x^{(0)}$.
- Generate x^* from the candidate-generating density $q(x^{(t)}, x^*)$
- Draw u from uniform distribution $U[0, 1]$.
- Calculate $\alpha(x^{(t)}, x^*)$.
- If $u < \alpha(x^{(t)}, x^*)$, set $X^{(t+1)} = x^*$, else set $X^{(t+1)} = x^{(t)}$.

Gibbs Sampler

Let the random vector X be comprised of $(X_1, X_2, \cdots, X_d)$ with some arbitrary d blocks of random variables. As Geman and Geman (1984) show, the Gibbs sampler kernel with invariant distribution $\pi(x), x = (x_1, \cdots, x_d)$ is the production of the full conditional distributions defined by

$$\pi(x_j | x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_d), j = 1, \cdots, d.$$

The Gibbs sampler kernel is

$$k_G(x^{(t)}, x_1^{(t+1)}) \equiv \prod_{j=1}^{d} \pi(x_j^{(t+1)} | x_1^{(t+1)}, \cdots, x_{j-1}^{(t+1)}, x_{j+1}^{(t)}, \cdots, x_d^{(t)})$$

with $r(x^{(t)}) = 0$ (the probability to remain at same state is zero). When $x = (x_1, x_2)$, the Gibbs sampler kernel is

$$P_G(x_1^{(t)}, x_2^{(t)}, x_1^{(t+1)}, x_2^{(t+1)}) \equiv \pi_{X_1|X_2}(x_1^{(t+1)}|x_2^{(t)})\pi_{X_2|X_1}(x_2^{(t+1)}|x_1^{(t+1)})$$

Algorithm 2. The Gibbs sampler algorithm takes the from:

- Choose starting values $x_2^{(0)}, \cdots, x_d^{(0)}$.
- Generate
  $x_1^{(t+1)} \sim \pi(x_1|x_2^{(t)}, x_3^{(t)}, \cdots, x_d^{(t)})$.
  $x_2^{(t+1)} \sim \pi(x_2|x_1^{(t+1)}, x_3^{(t)}, \cdots, x_d^{(t)})$.
  $x_3^{(t+1)} \sim \pi(x_3|x_1^{(t+1)}, x_2^{(t+1)}, x_4^{(t)} \cdots, x_d^{(t)})$.
  $\vdots$
  $x_d^{(t+1)} \sim \pi(x_d|x_1^{(t+1)}, x_2^{(t+1)}, \cdots, x_{d-1}^{(t+1)})$.

The Gibbs sampler is widely used in estimation of statistical and econometric models.

Chapter 3

Baseline Models

The objective of this chapter is to establish a basic framework for econometric modeling of developing-country city population growth. Based on this, Chapter 4 develops more advanced models.

This chapter investigates international-level city population data and develops several basic methods for city population forecasting. The city growth models to be examined include both classical and Bayesian econometric models for panel data. The methods themselves are of considerable interest but the most significant feature is that it draws upon a newly assembled and comprehensive cities database for thousands of individual cities in nearly all of the world's developing countries, combining time-series information from the UN Population Division and spatial information from the Global Urban-Rural Mapping Project (GRUMP) housed at CIESIN (Center for International Earth Science Information Network) at Columbia University's Earth Institute, to develop probabilistic forecasts of city growth.

Limiting the scope of our analysis to low- and middle-income countries (to be termed "developing countries" in the remainder of this thesis), I transform the city population data into time-series of city growth rates, and link to these growth rates more aggregated levels of total fertility rates and child mortality rates. In estimating models of city growth and forming probabilistic forecasts, we will first employ classical and Bayesian models for longitudinal data in which each city's growth trajectory is assumed to be independent (given covariates) of the trajectories for other cities.

The next section introduces the newly combined UN and GRUMP cities database. Section 3.2 develops and tests our simple city growth models, followed by forecasts of city growth rates based on the models in Section 3.3.

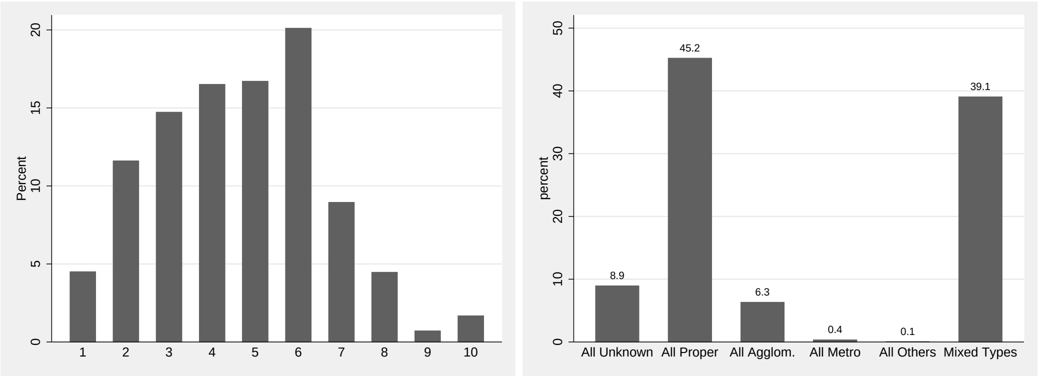
3.1 The UN and GRUMP cities data

3.1.1 The UN tabular data

The United Nations cities data (United Nations, 2008b) take the form of a panel dataset, containing city population counts for individual cities over time, generally recorded at irregular intervals. It is worth noting the UN's data collection process: the main source of city population data is population censuses conducted at different times by national statistical units that use different definitions of the underlying data on urban populations, which implies that the city population data available for different countries vary in terms of both their underlying definitions and their time references (United Nations, 2004).

Because countries take population censuses at different times, the actual dates of observation vary from city to city. Figure 3.1(a) summarizes the number of observations available on a per-city basis for the cities of developing countries. As can be seen, the number of observations varies by city and around 20 percent of the cities are observed six times. Time intervals (not shown in the figure) between records within a city vary and most common time interval is 10-year interval which reflects the fact that population census often take place every 10 years in most countries.

City population is one of the statistical variables that are difficult to compare at the international level since multiple social, economic, administrative, and political judgements come into play in the formation of such city definitions. As described in United Nations



(a) Number of records per city

(b) Statistical concepts on a per-city basis

Figure 3.1: UN cities database 2007 version (United Nations, 2008b): population records on some 3,000 cities in developing countries, 1950 - 2007

(2002), each country reports its city population data with various definitions and criteria. Confronting city population data with the variety of criteria, the United Nations endeavors to record each population count with three "statistical concepts" that serve to define city boundaries; city proper, urban agglomeration and metropolitan region. The United Nations favors the agglomeration concept which reflects population of urbanized areas and where possible, data are adjusted by UN staff to conform to the agglomeration concept—but of course this is not always possible.

Indeed, as Figure 3.1(b) shows for the cities with two or more entries in the database, in only a small percentage of cases—6.4 percent—are all of the city's records expressed in terms of urban agglomerations. The city proper is by far the more common concept in these data, with the populations of 45.2 percent of cities being consistently recorded in this way. For another 8.9 percent of cities, no information is available on the concept in which population is reported for any of the recorded dates, while in the remaining 39.1 percent of cities, the city's population time-series mixes two or more boundary concepts.

The difficulties stemming from such mixed time-series are illustrated in Figure 3.2. The series for Cuiabá, the capital city of Mato Grosso state in Brazil, begins with three

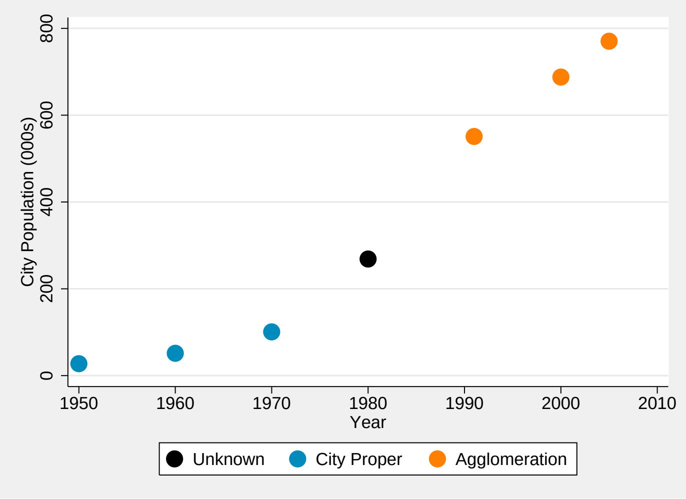


Figure 3.2: City population time-series for Cuiabá, Brazil

entries expressed in terms of city proper, followed by one of unknown type and a final three records couched in terms of urban agglomeration. In such mixed cases, it is certainly not obvious how to define a rate of population growth for spells of time that begin with one boundary concept but end with another. Neither is it obvious whether growth rates for city propers or urban agglomerations are strictly comparable with each other or with the rates for metropolitan areas. Despite on-going United Nations efforts to maintain consistency in each new revision of *World Urbanization Prospects*, there is an irreducible minimum of boundary-related variation in these data and far more heterogeneity remains in the city time-series than is commonly realized.

Another issue in the UN cities database is lack of information on cities geographic location (usually expressed as latitude and longitude) and their spatial extents. The traditional data collection method is to create tabular data by compiling census data (human population) for administrative units but the development of Geographic Information System makes it possible to create geographically-referenced data for mapping and spatial analysis of the census data. It also serves as a way to realize the conceptualization of city and measurement of city population beyond conventional urban-rural dichotomy (Champion and Hugo, 2004; Montgomery and Balk, 2008).

3.1.2 The GRUMP geospatial data

A need for spatial demographic data is initiated by Global Demography Project (Tobler et al., 1995) in which demographic data to be referenced to a uniform coordinate system (such as latitude and longitude quadrilaterals) rather than a tabular data organized at administrative units. The Gridded Population of the World (GPW) project (Deichmann et al., 2001; Balk and Yetman, 2004) combines spatial administrative boundary data with administrative-level population census data and allocates the population of each administrative unit uniformly over grid cells that fall into the unit. [1]

The Global Rural-Urban Mapping Project (Balk et al., 2005; Balk, 2009) aims to improve the GPW's population distribution raster data by taking urban and rural areas of each administrative unit into account. To do this, GRUMP creates input datasets which are invaluable themselves; this involves gathering information on place names, geographic locations, and population counts of human settlements of 5,000 persons or more from extensive external data sources, detecting the physical extent of urbanized areas derived mainly from satellite images of stable night-time lights, and identifying the urban area in terms of its place name and population.

The GRUMP project combined the United Nations cities database with the GRUMP georeferenced datasets. The combined UN-GRUMP data. The availability of such data will shed light on patterns of urbanization. is illustrated in Figure 3.3 with an example of Cuiabá,

[1] Spatial data contain spatial/geometry location and attribute information of geographic locations. By types of data format, it is categorized as either raster data format or vector data format. Vector data represent geographic features as discrete points, lines, or polygons. An example is the GPW's administrative boundary data. Raster data represent the landscape as a rectangular matrix of square cells. An example is the GPW's population grid at spatial resolution of 30 arc-seconds. In a grid at 30 arc-second resolution, each grid cell size is 30 by 30 second in unit of latitude and longitude, which covers an area of approximately 1 square kilometer at the equator.

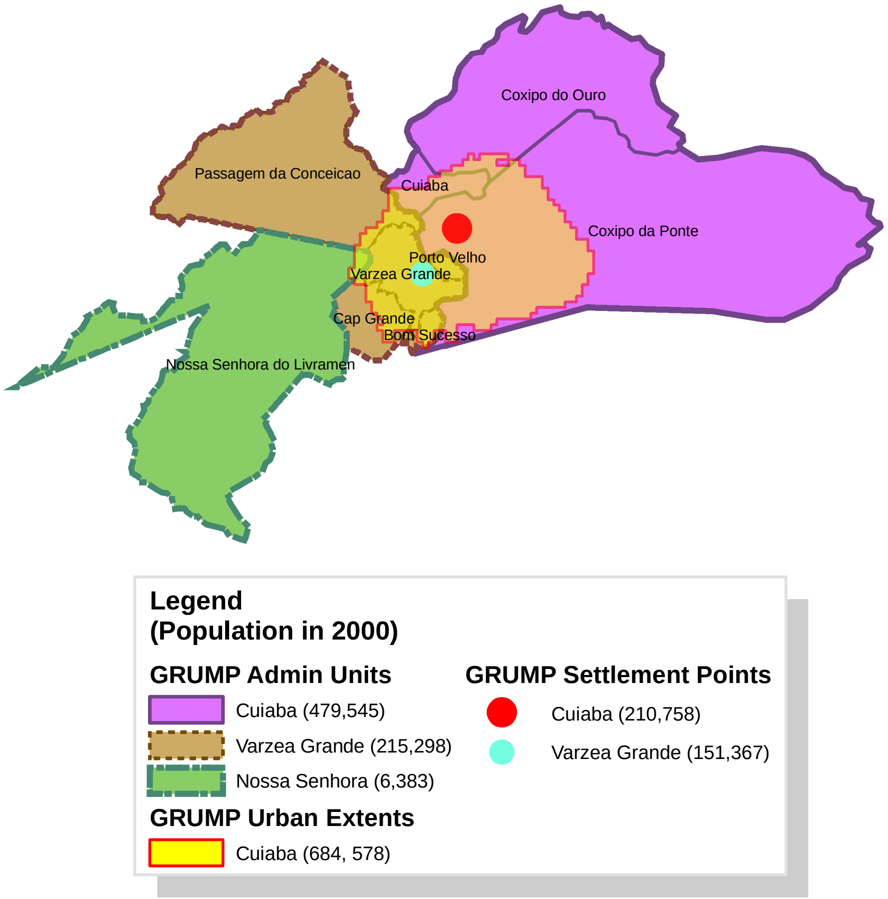
Brazil, in which the Cuiabá urbanized area detected by stable night-time light is overlaid with surrounding three administrative units (Cuiabá, Varzea Grand, and Nossa Senhora along with their sub-units) and two settlements across the administrative units (expressed as points with different colors and sizes depending on their population size in 2000). In 2000, the Cuiabá urbanized area (filled with yellow and surrounded in red) has 684,570 persons. In addition, the Cuiabá settlement (expressed as the red point) has a population of 210,758 and the total population of the three administrative units is 701,226.

In the United Nations records, a population count as reported by national authorities is available for the urban agglomeration of Cuiabá in 2000 (recall Figure 3.2), in which at the time there resided some 687,835 persons according to these authorities. Unfortunately, as has often been the case, the national authorities did not describe the boundaries of this agglomeration in sufficient detail for it to be mapped. The GRUMP program addresses this deficiency by providing an explicitly spatial view of the extent of the Cuiabá urban agglomeration.

It is worth noting the importance of the combination. Since the UN city population data are georeferenced through the link with the GRUMP data, it can systematically incorporate other city-level information if such data are available in the form of geospatial data. Geospatial analysis makes that possible. This is where we need geospatial data and analysis. In Chapter 4, we will see how the availability of such geospatial data and analysis will shed light on patterns of urbanization.

3.2 Basic modeling of city growth

To analyze and forecast growth of cities such as Cuiabá, we first translate each city's series of population counts into a series of growth rates—this can be done for cities with three or more population records—and then link to these growth rates information on more aggregated levels of urban total fertility and child mortality rates.

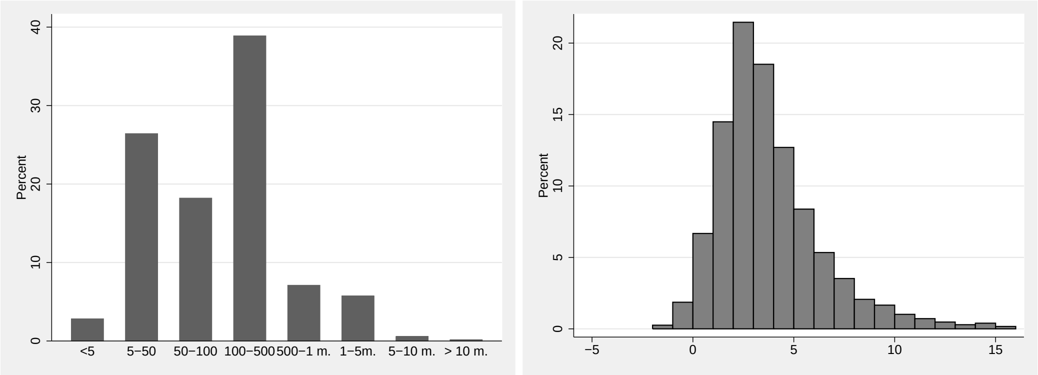


* Cuiada UN poplation in Uran Agglomeration in 2000 is 687,835
* Varzea Grande is not listed in UN cities database

Figure 3.3: Administrative units and urban settlements around Cuiabá urbanized area in Brazil with population for each entity. Combined UN-GRUMP cities data.

The urban demographic rates that we will employ are mainly derived from two of the three major international survey programs of the past thirty-five years, the World Fertility Surveys of the late 1970s and early 1980s and the Demographic and Health Surveys program, which began in the mid-1980s and continues to the present. The World Fertility Surveys (WFS) program contributes 38 surveys for which urban (and rural) rates can be estimated at the level of sub-national regions; the Demographic and Health Surveys (DHS) provide an additional 164 surveys covering some 71 countries. (We are in the process of preparing data from the third of these large demographic programs, the Multiple Indicator Cluster Surveys, which has been in operation from the late 1990s.) The surveys from which the urban demographic rates are calculated supply (in general) reliable estimates for the sub-national regions in which a city is situated, but as mentioned they cannot provide meaningful estimates at the individual city level. Nevertheless, as will be seen, urban demographic rates estimated on a more aggregated basis prove to be powerful influences on city-level growth rates.

Formation of a city and its growth are main topics in urban economics. The economic theory of city growth provides several mechanisms of both city formation and growth through the tradeoff between agglomeration economies (i.e. economics of scale) and diseconomies (i.e. congestion costs)(Henderson, 1974, 2005). The growth of a city is affected by geography and human capital both of which are key elements determining the size of agglomeration economies through productivity growth and standards of living (Glaeser et al., 1992; Glaeser and Shapiro, 2001; Beeson et al., 2001; Black and Henderson, 2003; Henderson, 2005; Shapiro, 2006; da Mata et al., 2007: Among others). A city with high productivity and/or standards of living will attract firms and consumers, which leads to flows of migration into the city.



(a) City population records

(b) City growth rates

Figure 3.4: Distributions of city population records and city growth rates, All cities in developing countries, 1950 - 2007

3.2.1 City population growth rates

For each city in the UN cities database, I have converted the available population data into measures of city growth rates g_{i,t_0}, with growth over the period t_0 to t_1 defined in continuous terms and estimated as $g_{i,t_0} = (\ln P_{i,t_1} - \ln P_{i,t_0})/(t_1 - t_0)$. Figure 3.4(a) depicts the distribution of city population counts used to calculate the growth rates. Of the population counts recorded in the UN cities database, about 40 percent are one between 100,000 and 500,000. There are also population counts below 5,000.[2] Figure 3.4(b) shows the city growth rates for all cities and time periods from 1950. The median growth rate recorded is 3.24 percent and the mean is 3.86 percent. By region, Africa has the highest median growth (3.88 percent), followed by Latin American (3.24 percent) and Asia (3.06 percent). As the figure shows, there are instances of city population decline evident in these data as well as cases of rapid growth at rates of 10 percent and above.

[2] The UN monitors all cities of 100,000 population and above; when a given city crosses this threshold, the Population Division endeavors to reconstruct its history.

3.2.2 Model specification

A simple fertility-based panel data regression model of city growth [3] is set out as equation (3.1),

$$g_{i,t} = \alpha + \beta \mathrm{TFR}_t + \delta q_t + \mathbf{D}'_{i,t}\gamma + v_{i,t}. \tag{3.1}$$

In this equation the i subscript denotes the i-th city and t is a point in time; $g_{i,t}$ is the estimated city population growth rate at that time; and the fertility and mortality components of growth are represented by the urban total fertility rate TFR_t and q_t, the urban child mortality rate. The vector $\mathbf{D}_{i,t}$ is a set of dummy variables indicating the start-of-period and end-of-period units in which the city's population is recorded. In the Cuiabá example shown in Figure 3.2, these dummy variables would take into account the fact that in the early 1970s, one era of growth began with the population recorded in terms of the city proper but ended with a count expressed in unknown units.

Of course, growth models including observed city-specific explanatory variables will generally be preferred to those without such variables, provided that city-specific observables are either fixed over time or can be forecast with reasonable confidence. To show how our approach generalizes to include observed city-specific explanatory variables, we will develop in the next chapter an expanded model of city growth in which city i's population size exerts an influence on its growth rate.

In what follows, we explore two specifications of $v_{i,t}$, the regression disturbance term. The first is a *random effects* specification in which the disturbance term is represented as a composite $v_{i,t} = u_i + \varepsilon_{i,t}$, containing one component, u_i, that is specific to city i and whose value can be estimated as $\hat{u}_i$. In this approach, u_i is assumed to be uncorrelated with the other right-hand side explanatory variables (e.g., TFR_t and q_t). Our second specification

[3] Three components of urban growth are natural increase, net migration, and reclassification (i.e. spatial expansion). The city growth modeling is based on the United Nations (1980) and Chen et al. (1998)'s finding that, in developing countries, about 60 percent of the urban growth rate is attributed to natural growth, the difference between urban birth and death rates. This model is extended in the next chapter.

is a *fixed effect* specification in which the disturbance term also takes the composite form $v_{i,t} = u_i + \varepsilon_{i,t}$, but in which u_i is allowed to be correlated with other right-hand side variables. As in the random-effects approach, the value of u_i can be estimated (using techniques similar though not necessarily identical to those applied in the random-effects method). This specification will prove useful when city-specific endogenous explanatory variables are introduced in the model.

To estimate the models, we consider both classical and Bayesian methods. For classical generalized-least-squares (GLS) estimation, there are well-established counterparts in Bayesian approach and, for both random-effects and fixed-effects models, its posterior distribution can be simulated with Gibbs sampling algorithm (Geman and Geman, 1984; Gelfand and Smith, 1990: among others.). The Gibbs sampler simulates the posterior distribution of parameters indirectly by breaking the parameters into blocks, deriving the distribution of each block conditional on the other parameters and data, and successively drawing samples from the conditional posterior distributions (see Appendix 3.B for the details). The reason for the blocking is that the posterior distribution itself is difficult to simulate.

3.3 City growth forecasts

Forecasts with national vital rates

As covariates, we first use national-level estimates and forecasts of total fertility rates and child mortality rates. The UN maintains a large program in which it forecasts fertility and mortality rates at the national level, to date these forecasts of demographic rates have not figured explicitly into the UN's companion projections of city size and growth (United Nations, 2008a).

Table 3.1 presents the basic regression models, with classical and Bayesian ordinary least squares estimates shown in the first two columns, followed by the fixed-effects and

Table 3.1: Classical and Bayesian city growth regression models with national total fertility rates and child mortality rates as covariates, All cities in developing countries, 2005-2007.

	OLS		Fixed-Effects		Random-Effects	
	Classical	Bayesian	Classical	Bayesian	Classical	Bayesian
Total Fertility Rate	0.700	0.700	0.856	0.857	0.740	0.740
(*Z* statistic)	(24.08)	(24.09)	(19.13)	(18.68)	(24.01)	(24.26)
Child Mortality Rate	-0.004	-0.003	-0.006	-0.005	-0.004	-0.004
	(-5.62)	(-5.61)	(-5.25)	(-5.48)	(-6.12)	(-6.13)
Constant	0.766	0.766	0.175		0.641	0.640
	(6.42)	(6.46)	(0.78)		(4.89)	(4.90)
σ_u					1.015	1.014
					(22.95)	(24.70)
σ_ε	3.200	3.200	3.035	3.037	3.028	3.030
		(148.83)		(132.50)	(133.14)	(150.73)

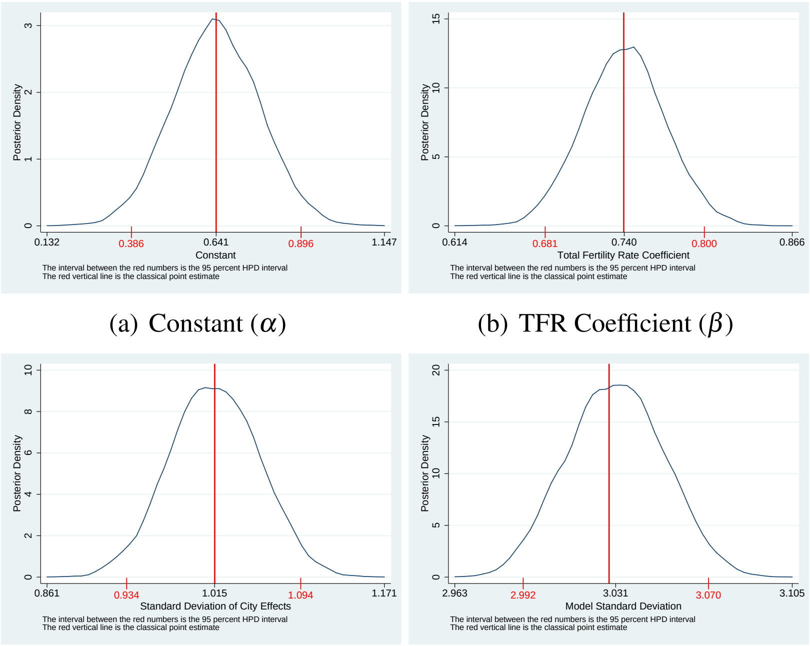
See Appendix3.B on how to implement Bayesian estimation. Bayesian estimates are posterior means obtained by averaging each of 10,000 Gibbs samples after discarding first 40,000 samples as burn-in. For comparison with classical estimates, the Z-statistics are obtained by dividing posterior means by posterior standard errors. The Monte Carlo standard errors (not shown in the table) are all less than 0.01. Convergence of Gibbs samples are tested with the Geweke (1992)'s and Gelman and Rubin (1992)'s test statistics.

random-effects models. In the Table, Bayesian point estimates are posterior means which are obtained by averaging their resulting samples from the Gibbs Sampler. Figure 4.2 shows posterior distribution of some coefficients for random-effects model along with its corresponding classical estimate.

As can be seen in Table 3.1, the coefficients on the total fertility rate is highly significant, with an increase of 1 child in the TFR implying increases in city growth rates ranging from 0.700 to 0.857 percentage points, depending on the model. The child mortality rate (the variable is coded in terms of deaths per 1000 children) has a smaller effect on city growth, but the coefficient attains statistical significance. The results of the fixed-effect specification are especially striking, given that such models include a great number (i.e. the number of cities, about 2,500) of city-specific dummy variables (whose effects are expressed in the $\hat{u}_i$) and yet exhibit large and statistically significant TFR coefficients. Indeed, the fixed-effects estimate of the total fertility rate coefficient is by far the largest in this set of estimates.

The Bayesian point estimates have almost same results as their classical counterparts for all the three models. Figure 4.2 confirms the fact. The figures show posterior distributions of some coefficients for random-effects models along with classical estimates (vertical red line). The classcial estimates lie in the highest probable region of the posterior distributions of these Bayesian counterparts. Also, the figures show the 95 per cent highest posterior density (hpd) intervals, intervals between numbers in red. For instance, the 95 per cent hpd interval for the TFR coefficient is (0.681, 0.800).

Table 3.2 shows coefficients of a set of dummy variables indicating the start-of-period and end-of-period units in which the city's population is recorded where the baseline category is urban agglomeration at start and end of spell. 5,575 cases of growth rates are calculated from population defined as city proper at start and end of spell, which makes up the highest proportion. Growth rates based on the proper-proper definition is higher than those based on the baseline category by roughly 0.3 percent, which is of statistical significance. The coefficient of the proper-agglomeration dummy variable is 1.573 and



(a) Constant (α)

(b) TFR Coefficient (β)

(c) Standard deviation (σ_u^2)

(d) Model standard deviation (σ_ε^2)

Figure 3.5: Posterior distributions of selected parameters, Bayesian random-effects models

Table 3.2: Growth rates and changes in definition, Relative to the baseline category with city defined in terms of urban agglomeration at start and end of spell, Classical random-effects Models.

City Definitions	Cases	Classical	(Z-stat)
Unknown-Unknown	2,565	0.819	(6.68)
Unknown-Proper	723	1.090	(7.03)
Unknown-Agglomeration	140	0.361	(1.28)
Unknown-Metro. Region	83	-0.360	(-1.00)
Proper-Unknown	242	1.432	(6.35)
Proper-Proper	5,575	0.293	(2.66)
Proper-Agglomeration	125	1.573	(5.34)
Agglomeration-Unknown	40	-0.974	(-1.93)
Agglomeration-Proper	43	-0.289	(-0.59)
Agglomeration-Metro. Region	16	1.223	(1.55)
Metro. Region-Metro. Region	115	0.130	(0.38)
Others-Others	22	3.482	(4.84)

significant.

In the analysis below, we forecast city population growth based on random-effects model which also come in classical and Bayesian varieties. We will take the United Nations point forecasts of national total fertility rates and child mortality rates as given.[4] Given data to period t, the Bayesian forecasts of city growth rates are obtained by simulating the following (conditional) distribution of future growth rate $g_{i,t+s}$ in period $t+s$, [5]

$$g_{i,t+s} \sim N(\alpha + \beta TFR_{t+s} + \delta q_{t+s} + \hat{u}_i, \sigma_\varepsilon^2\ )$$

with the Gibbs samples of a set of parameters α, β, δ and σ_ε^2 successively and then by averaging of the realized values of $g_{i,t+s}$. The estimate $\hat{u}_i$ is the Bayesian point estimate

[4] This assumption can be relaxed to allow for forecast errors in future fertility and mortality.

[5] We forecast city growth rates with the urban agglomeration unit. Note that the agglomeration unit is the baseline category in our model. If city i's last population count is not measured in the agglomeration unit, we adjust it with coefficient of its corresponding dummy variable to forecast with the agglomeration unit at its first forecast period.

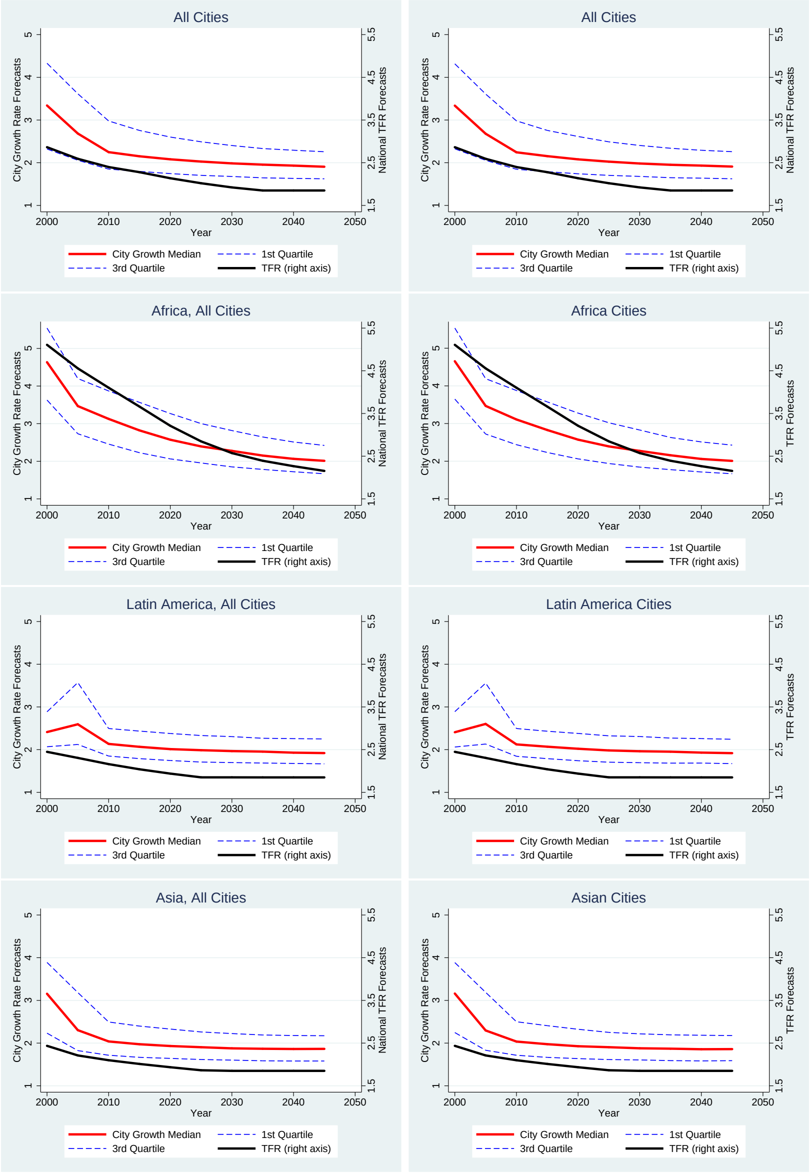
obtained by averaging its samples. With the resulting samples of future city growth rate, the Bayesian forecasts emerge naturally in probabilistic terms. Classical point forecast of city growth rate for city i in period $t+s$ is

$$\tilde{g}_{i,t+s} = \hat{\alpha} + \hat{\beta} TFR_{t+s} + \hat{\delta} q_{t+s} + \tilde{v}_{i,t+s}$$

in which the symbol '˜' denotes a forecast value and the symbol 'ˆ' denotes an estimated quantity based on data up to period t. Forecast error variance also can be derived. The Goldberger (1962)'s best linear unbiased predictor is used here.

Figure 3.6 shows classical and Bayesian forecasts of city growth rates with the random-effects models of table 3.1 along with the UN's forecasts of national-level total fertility rates, which extend to 2045-50. The median city growth forecasts are shown in the figure along with the 25th and 75th city growth rates for all cities and by region. Also, the median national total fertility rates are shown in the figure. Our classical and Bayesian city growth forecasts suggest the gradual decline in city growth rates, as national fertility rates to decline over time, especially in Africa. The classical forecasts show almost same patterns as the Bayesian forecasts.

In our models, the forecasted decline in city growth is wholly attributable to declines in future fertility and mortality (the mortality effect by itself would imply rising rates of city growth as child death rates fall, but in our models these mortality effects are overwhelmed by the effects of falling fertility). Linking the United Nation's the two large programs of population projections, we have uncovered strong evidence supporting the use of total fertility rates in econometric models of city growth and in the forecasts based on these models. Our city growth forecasts are consistent with, and indeed largely based upon, the UN forecasts of fertility and mortality rate declines at the national level. In the next session, we reconfirm our results with the preferable sub-national urban total fertility rates and child mortality rates as covariates instead of the national rates.



(a) Classical

(b) Bayesian

Figure 3.6: Classical and Bayesian forecasts of city growth rates (in terms of urban agglomeration) with random-effects models in Table 3.1, along with the UN's forecasts of national total fertility rates. 2000–2045. Median growth rates are shown along with 25th and 75th percentiles (left axis) and median national total fertility rates are shown (right axis).

Forecasts with urban vital rates

The demographic materials in this analysis are drawn from over 200 surveys fielded in developing countries from the mid-1970s to the present. The World Fertility Surveys (WFS) program contributes 38 surveys for which urban (and rural) rates can be estimated at the level of sub-national regions; the Demographic and Health Surveys (DHS) provide an additional 164 surveys covering some 71 countries; and the second and third rounds of the Multiple Indicator Cluster Surveys (MICS) are expected to add as many as fifty surveys to the total by the end of 2010. Still, not every developing country is represented in this body of surveys, and we will be filling in the record with estimates from other sources. Also, reliable forecasts of urban fertility rates and child mortality rates are not available.

Confronted with this situation, I test our city growth rate models with both the *observed* urban rates and our *estimated and forecasted* urban rates. To do this, we first linked the observed urban total fertility rates and child mortality rates with its national counterparts, separately, and test association between urban rates and national rates with simple regression analysis. In this paper, the WFS data on Urban TFRs are taken from Ashurst et al. (1984) and the DHS data on Urban TFRs and Q5 are taken from DHS website [6]. The resulting

[6] Source: Macro International Inc, 2008. MEASURE DHS STATcompiler. http://www.measuredhs.com,July28th,2008

three models are as follows[7]:

$$\text{Model 1}: \text{ Urban TFR} = 0.311 + 0.702 * \text{National TFR} \tag{3.2}$$
$$\text{Urban Q5} = 12.04 + 0.705 * \text{National Q5}$$
$$\text{Model 2}: \text{ Urban TFR} = 1.228 + 0.681 * \text{National TFR} - 0.018 * \text{Time} \tag{3.3}$$
$$\text{Urban Q5} = 32.111 + 0.706 * \text{National Q5} - 0.425 * \text{Time}$$

Model 3 :

$$\text{Urban TFR} = -1.389 + 1.198 * \text{National TFR} + 0.039 * \text{Time} - 0.011 * (\text{National TFR*Time}) \tag{3.4}$$
$$\text{Urban Q5} = -18.586 + 1.162 * \text{National Q5} + 0.646 * \text{Time} - 0.009 * (\text{National Q5*Time})$$

Using the three models above, I estimated and forecasted urban rate trajectories from 1950 to 2045 for all cities in UN database. As will be shown, the projected urban total fertility rates and child mortality rates are different depending on the model and I will see how city growth rates forecasted with my city growth models respond to the different trajectories of future urban TFRs.

Table 3.3 presents classical estimates of my random-effects and fixed-effects city growth models with the *observed* urban total fertility rates and child mortality rates in the first two columns, followed by those with *estimated* urban total fertility rates and child mortality rates derived with the above models. As can be seen, I lost many observations with the *observed* urban rates but the significance of TFR coefficients in never lost. The TFR coefficients are highly significant regardless of the different estimates of urban TFR. The results in the table show that when urban total fertility rates increase by one child, this is associated with an increase in city growth rates ranging from 0.875 to 1.611 percentage points. It reconfirms strong association between city growth rates and total fertility rates

[7] All of the regressions gain statistical significance. The number of observations is 204 for Urban TFR models and 106 for Urban Q5 models. 1950 is set as 1 in the Time variable

and the association is robust regardless of whether I use the observed urban rates or the three different estimated urban rates.

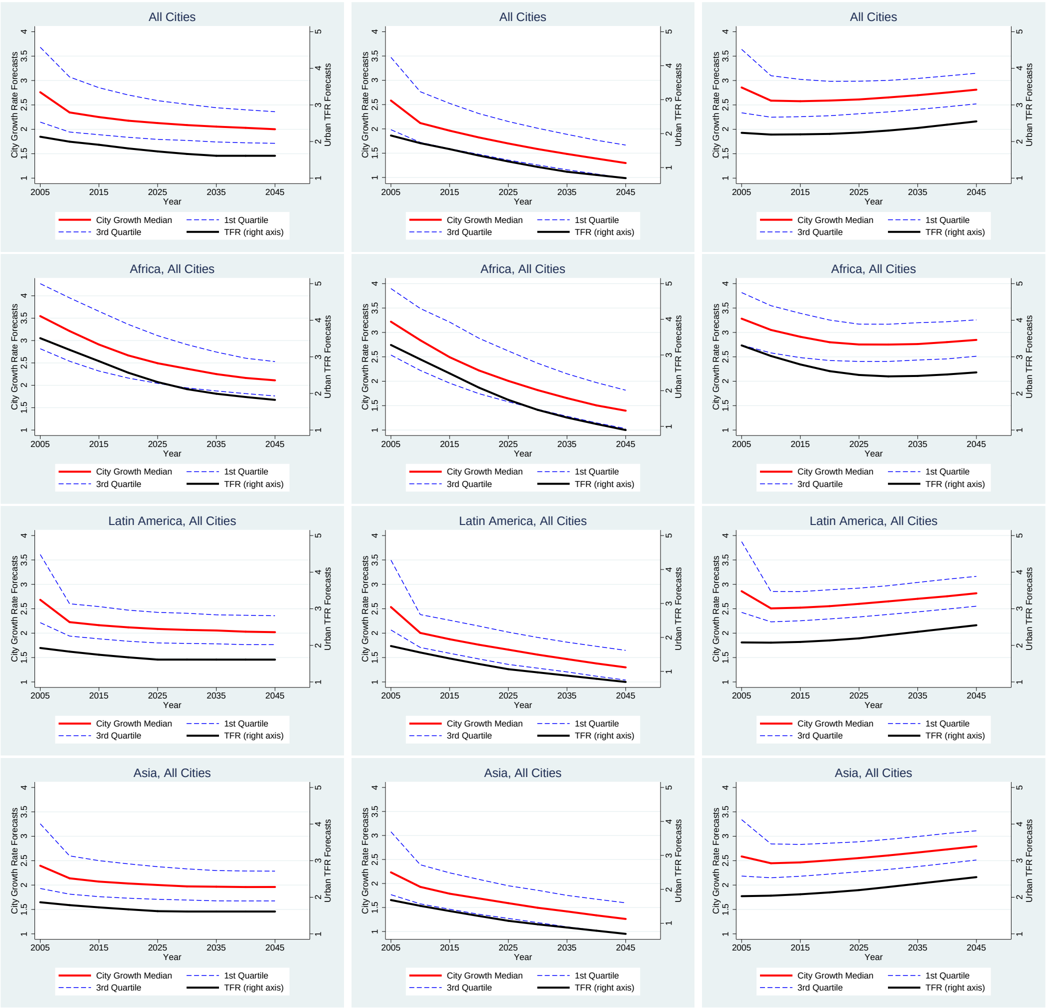
The estimates are then extrapolated to forecast city growth using the classical forecast method described above. In Figure 3.7, I compare forecasted city growth rates with random-effects models in Table 3.3 for three different projected urban rates. It clearly shows that how city growth forecasts response to the different trajectories of future urban fertility rates. As can be seen, model 1 for the urban TFR, equation (3.2), exhibits a moderate decline of the future urban TFR and model 2, equation (3.3), exhibits a more steep decline in the urban TFR. The decline in city growth rates is steeper with model 1 than with model 2. Unlike the other models, model 3, equation (3.4), exhibits an increase in the urban TFR and the forecasted city growth rates increase with it. These results reconfirm the fact that, in our simple growth model, the forecasted city growth is wholly attributable to future fertility.

The strong association between city growth and the total fertility rate uncovered in this chapter has good policy implications for city-growth which is almost universally ignored. Many developing-country policy-makers have expressed greater concern about rates of city growth in their countries than about national population growth, and they have not infrequently acted on such concerns with aggressive tactics aimed to expel slum residents and repel rural-to-urban migrants.

It is therefore surprising how little attention has been paid to a growth-rate policy of a very different character: urban voluntary family-planning programs. Over the past half-century, such programs have compiled an impressive record across the developing world in facilitating fertility declines and reducing unwanted fertility. Such family-planning programs offer an effective and humane alternative to ineffective and brutalizing measures that have been applied all too often.

Table 3.3: Classical city growth regression models with observed and projected urban total fertility and child mortality rates as covariates. .

	Observed		Projected (Model 1)		Projected (Model 2)		Projected (Model 3)	
	RE	FE	RE	FE	RE	FE	RE	FE
Urban TFR	1.058	1.611	1.055	1.212	1.015	1.202	0.875	0.966
	(10.44)	(6.65)	(24.01)	(19.10)	(23.79)	(19.66)	(24.72)	(20.20)
Urban Q5	-0.001	0.004	-0.006	-0.008	-0.009	-0.013	-0.007	-0.009
	(-0.40)	(0.76)	(-6.15)	(-5.14)	(-8.02)	(-7.94)	(-8.69)	(-8.29)
Constant	-0.023	-3.457	0.400	-0.070	0.674	0.366	0.981	0.757
	(-0.06)	(-3.66)	(2.95)	(-0.31)	(5.21)	(1.71)	(7.84)	(3.57)
σ_u	0.868	2.190	1.017	1.861	1.050	1.870	1.033	1.877
	(6.78)		(23.03)		(24.01)		(23.53)	
σ_ε	2.581	2.533	3.027	3.034	3.014	3.019	3.028	3.034
	(45.16)		(133.14)		(133.00)		(133.22)	
log-likelihood	-4,025	-3,500	-28,488	-26,597	-28,463	-26,550	-28,463	-26,550
Number of observations	1,666	1,666	11,059	11,059	11,059	11,059	11,059	11,059



(a) Projected (Model 1)

(b) Projected (Model 2)

(c) Projected (Model 3)

Figure 3.7: Classical forecasts of city growth (in terms of urban agglomeration) based on random-effects models in Table 3.3 with projected urban total fertility rates and child mortality rates. 2005–2045.

3.A Linking the UN data with the GRUMP data

This appendix explains how the United Nations cities database is linked to GRUMP georeferenced data. The UN cities database contains the populations of nearly 5,400 cities around the world and the GRUMP settlements data contain nearly 67,000 settlements. The matching of UN and GRUMP records in the two enormous datasets must be carried out mainly on the basis of the city names that are available in both databases. Whereas the GRUMP database was constructed on an explicitly geographic basis, with close attention given to finely disaggregated administrative unit populations and boundaries, the UN cities database has not been organized geographically at a comparable level of detail. In the UN cities database, for example, about half of Africa cities (468 out of 933 cities) have no information on administrative unit names. In addition, despite on-going international efforts for standardization of geographic names, there are substantial differences in spelling of city and administrative unit names between the two databases.

Our approach is based on a combination of exact string matching and (where this fails, as it often does) alternative approximate string matching approaches using what is termed "fuzzy logic". The exact string matching is based on city name, administrative name and country. The inclusion of administrative unit and country is necessary to prevent mismatches of different places with same city name (Homonym). In addition, city and administrative names are expressed without special accents, codes, and spaces to put them in a common format to the extent possible before matching is carried out. Both UN and GRUMP datasets have some alternative city and administrative names and thus matching can be carried out successively using all combinations of the UN and GRUMP city and administrative names.

For the cases which are not matched in the exact-match algorithms, manual matching is basically the next step to perform second attempt at matching. Although there is no way such that this second attempt at matching can be fully programmable, there are many algorithms and programs of approximate string matching which can be used to make the manual matching as easy as possible, among others, Navarro (2001); Schnell et al. (2004). I use the Lavenshtein edit distance which is defined as the minimum number of insertions, deletions, or substitutions of a single character which are needed to transform one string into the other. The steps for the matching as follows: for each of the UN unmatched cities, (1) successively calculate the Lavenshtein distance of GRUMP cities of the country in which the UN city is located, (2) link the UN city to the GRUMP cities and order the GRUMP cities by its Lavenshtein distance rank and (3) manually check whether each link is correct with available information including population on the UN and GRUMP cities. The matching was programmed with Fortran 95 and Stata.

3.B Implementing the Bayesian models

In this chapter, data analysis and classical econometric analysis are done with a commercial statistical software Stata. Stata has good capacities for data handling and classical panel data analysis. However, it does not provide Bayesian analysis. Though Stata supports programming, I have explored other programming language for Bayesian implementations. This was not an easy task.

The basic Bayesian models had been programmed and implemented with computer programming languages R, Matlab, WinBUGS but I finally settled with Fortran 95. Bayesian MCMC methods are computationally intensive especially in implementing complex models with large-scale data. In the sense, R and Matlab were not appropriate to use even for these simple Bayesian models due to the computational burden unless it uses parallel programming capacity with multi-processors.

WinBUGS, designed solely for Bayesian analysis, has its unique syntax of programming which is easy to learn and use. In addition, the computation speed is as fast as Fortran 95. However, as explicitly specified in its manual, the downside of WinBUGS is that it has some restrictions in modeling. Unfortunately, due to the restriction, WinBUGS programming of these models for unbalanced panel data is not possible (See the WinBUGS manual `http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf`). For readers interested in R, see `http://mcmcpack.wustl.edu/index.html` which is about a recently-developed R package for MCMC methods.

In what follows, I derive the Gibbs samplers of the Bayesian models using parameters familiar with classical statistical methods (i.e. Variance instead of precision. Some Bayesian books use precision instead of variance). Based on the derivations, Fortran 95 programs were written and used for the analysis.

Consider a vector $\mathbf{g}_i$ that has n_i entries, which constitute city i's full record of population growth rates. There are N such cities and therefore $\sum_{i=1}^{N} n_i$ records in total. For the i-th city, the growth model of equation (3.1) is expressed here in the simplified form

$$\mathbf{g}_i = \mathbf{X}_i\theta + \iota_{n_i}u_i + \varepsilon_i,$$

in which the matrix $\mathbf{X}_i$, whose dimensions are $n_i \times k$, contains all of the covariates and θ denotes the corresponding parameters. The specification includes ι_{n_i}, a vector of n_i ones, which inserts the city-specific effect u_i into the growth rate specification for each time period covered in city i's time-series.

Random-effects models

In the Bayesian approach, statistical inference is based on the posterior distribution of the parameters, which we denote by $p(\text{parameters}|\text{data})$, this being proportional to the prior distribution $\pi(\cdot)$ assumed for the parameters multiplied by $l(\cdot)$, the likelihood function. The unobservable effects u_i in the model are treated as if they were N additional parameters to be estimated.

The Bayesian approach to random-effects models employs a prior distribution for u_i that is expressed in a hierarchical form, whereas in the fixed-effects case the prior is not hierarchically structured (Chib, 1996). For the random-effects model, the prior is usually specified in terms of the normal distribution, with $u_i \sim \mathscr{N}(0, \sigma_u^2)$ for each i, and the σ_u^2 parameter of this distribution is itself assumed to be taken from an inverted gamma distribution $iG(h0/2, p0/2)$ in which $h0$ and $p0$ are the *hyperparameters* whose values are established by the researcher.[8] For θ and σ_ε^2 (the variance of $\varepsilon_{i,t}$), we use independent priors, that is, $\theta \sim \mathscr{N}(\theta_0, \mathbf{M}_0^{-1})$ and $\sigma_\varepsilon^2 \sim iG(v_0/2, s_0/2)$, with θ_0, $\mathbf{M}_0$, v_0, and s_0 being the hyperparameters. The posterior distribution of the Bayesian random-effects model is represented in the general form

$$p(\theta, \sigma_u^2, \mathbf{u}, \sigma_\varepsilon^2 | \mathbf{g}, \mathbf{X}) \propto l(\mathbf{g}|\mathbf{X}, \theta, \sigma_u^2, \mathbf{u}, \sigma_\varepsilon^2) \cdot \pi(\theta, \sigma_u^2, \mathbf{u}, \sigma_\varepsilon^2),$$

in which the vector $\mathbf{g}$ (of dimension $\sum_i n_i$) and the $\mathbf{X}$ matrix ($\sum_i n_i \times k$) contain the data, and $\mathbf{u}$ is a vector of the unobservable city-specific effects. The posterior can be simulated by using the Gibbs sampling algorithm (Geman and Geman, 1984; Gelfand and Smith, 1990). The Gibbs sampler simulates the posterior distribution of parameters indirectly by separating the parameters into blocks, deriving the distribution of each block conditional on the other parameters and data, and successively drawing samples from the conditional posterior distributions.

Using the block θ, σ_u^2, $\mathbf{u}$, and σ_ε^2, Gibbs samples of the Bayesian random-effects model are drawn in the following way. Define $\mathbf{M}_1 = \mathbf{M}_0 + \sigma_\varepsilon^{-2}\mathbf{X}'\mathbf{X}$ and let the $\sum_i n_i \times 1$ vector $\mathbf{g}^* = (\mathbf{g}_1^{*'}, \ldots, \mathbf{g}_N^{*'})'$ with $\mathbf{g}_i^* = \mathbf{g}_i - \iota_{n_i} u_i$. We draw θ from

$$\theta \sim \mathscr{N}\left(\mathbf{M}_1^{-1}\left(\mathbf{M}_0\theta_0 + \sigma_\varepsilon^{-2}\mathbf{X}'\mathbf{g}^*\right), \mathbf{M}_1^{-1}\right),$$

and draw σ_u^2 from

$$\sigma_u^2 \sim iG\left(\frac{N + h_0}{2}, \frac{\mathbf{u}'\mathbf{u} + p_0}{2}\right).$$

[8] The inverted gamma distribution of z, denoted by $z \sim iG(\alpha, \beta)$, is

$$p(z; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{-\alpha}} z^{-\alpha-1} e^{-\frac{\beta}{z}},$$

with $0 < z < \infty,$and $\alpha > 0, \beta > 0$.

Finally, we draw u_i from

$$u_i \sim \mathscr{N}\left(\frac{\sigma_u^2 \sum_i (\mathbf{g}_i - \mathbf{X}_i \theta)}{\sigma_\varepsilon^2 + n_i \sigma_u^2}, \frac{\sigma_u^2 \sigma_\varepsilon^2}{\sigma_\varepsilon^2 + n_i \sigma_u^2}\right)$$

and draw σ_ε^2 from

$$\sigma_\varepsilon^2 \sim iG\left(\frac{\sum_i n_i + v_0}{2}, \frac{(\mathbf{g}^* - \mathbf{X}\theta)^{'} (\mathbf{g}^* - \mathbf{X}\theta) + s_0}{2}\right).$$

Fixed-effects models

Consider the fixed-effects model of city growth, which for the i-th city is

$$\mathbf{g}_i = \mathbf{Z}_i \delta + \iota_{n_i} u_i + \varepsilon_i,$$

in which $\mathbf{Z}_i$ contains only the covariates that vary with time, with δ denoting their parameters. The individual effects u_i can be viewed as city-specific intercepts, whose values summarize all of the time-constant attributes of the city that affect its growth rates. For the prior distribution of the u_i, we assume that $u_i \sim \mathscr{N}(u_{0,i}, \sigma_{u_{0,i}}^2)$, in which both $u_{0,i}$ and $\sigma_{u_{0,i}}^2$ are hyperparameters. Note that in contrast to the random-effects model, here the variances of the u_i are not hierarchically structured, implying that in theory, all the individual effects are realizations taken from separate and distinct distributions. The posterior distribution of the Bayesian fixed-effect model is written as

$$p(\delta, \mathbf{u}, \sigma_\varepsilon^2 | \mathbf{g}, \mathbf{Z}) \propto l(\mathbf{g} | \mathbf{Z}, \delta, \mathbf{u}, \sigma_\varepsilon^2) \cdot \pi(\delta, \mathbf{u}, \sigma_\varepsilon^2)$$

Assume as before that $\delta \sim \mathscr{N}(\delta_0, \mathbf{M}_0^{-1})$ and $\sigma_\varepsilon^2 \sim iG(v_0/2, s_0/2)$. The Gibbs sampler can be used to draw samples from the distribution above using a blocking scheme. Using the three blocks δ, $\mathbf{u}$, and σ_ε^2, the Gibbs samples of the Bayesian fixed-effects model are drawn as follows. Again let $\mathbf{M}_1 = \mathbf{M}_0 + \sigma_\varepsilon^{-2} \mathbf{Z}^{'} \mathbf{Z}$ and $\mathbf{g}^* = (\mathbf{g}_1^{*'}, \ldots, \mathbf{g}_N^{*'})^{'}$ with $\mathbf{g}_i^* = \mathbf{g}_i - \iota_{n_i} u_i$. We draw δ from

$$\delta \sim \mathscr{N}\left(\mathbf{M}_1^{-1}\left(\mathbf{M}_0 \delta_0 + \sigma_\varepsilon^{-2} \mathbf{Z}^{'} \mathbf{g}^*\right), \mathbf{M}_1^{-1}\right).$$

We draw u_i from

$$u_i \sim \mathscr{N}\left(\frac{\sigma_\varepsilon^2 u_{0,i} + \sigma_{u_{0,i}}^2 \sum_i (\mathbf{g}_i - \mathbf{Z}_i \theta)}{\sigma_\varepsilon^2 + n_i \sigma_{u_{0,i}}^2}, \frac{\sigma_{u_{0,i}}^2 \sigma_\varepsilon^2}{\sigma_\varepsilon^2 + n_i \sigma_{u_{0,i}}^2}\right).$$

Finally, draw σ_ε^2 from

$$\sigma_\varepsilon^2 \sim iG\left(\frac{\sum_i n_i + v_0}{2}, \frac{(\mathbf{g}^* - \mathbf{Z}\delta)'(\mathbf{g}^* - \mathbf{Z}\delta) + s_0}{2}\right).$$

I have used FORTRAN 95 programs to estimate these Bayesian models. In our analysis, we specify vague priors for the hyperparameters; that is, $\theta_0 = \mathbf{0}$, $\mathbf{M}_0 = 1^{-5}\mathbf{I}$, $h_0 = 0$, $p_0 = 0$, $v_0 = 0$, and $s_0 = 0$ for the random-effect model and $\delta_0 = \mathbf{0}$, $\mathbf{M}_0 = 1^{-5}\mathbf{I}$, $v_0 = 0$, and $s_0 = 0$, $u_{0,i} = 0$, $\sigma^2_{u_{0,i}} = 10^5$ for all i in the fixed-effects model.

Chapter 4

Spatial Econometric Models

This chapter analyzes how current urban populations are distributed by ecological environments and how the patterns evolve over time. To assess the risks that global climate change presents for the city and town dwellers of poor countries, it is vitally important to know who lives where—that is, to know enough about the locations of the people who will be facing climate change, and the types of settlements in which they live, for the most vulnerable among them to be identified and given priority.

To do that, for the first time, city population data are situated in three ecological zones: the low-elevation coastal zone; drylands ecosystems; and inland water systems. This is made possible with (1) recent developments of geospatial data handling and analysis and (2) newly-assembled, comprehensive cities database by integrating demographic and ecological data which come from various sources in various forms. The data sources used in this study are summarized in Appendix 4.B.

To estimate city growth rates, I use spatial econometric models to test the hypothesis that population growth of a city is affected by not only the city's characteristics but by also those of its neighboring cities.

4.1 Urban populations by ecological zones

This section analyzes the distribution of current urban populations by ecological zones. Before we present and discuss the results, a brief discussion is needed of the ecozone variables used in this analysis. We situate each city in relation to three such zones: the low-elevation coastal zone; drylands ecosystems; and inland water systems. Note that these zones are not mutually exclusive; for instance, a given city can be located in both the low-elevation coastal zone and a drylands zone.

4.1.1 Ecological zones

Drylands According to the Middleton et al. (1997), the drylands ecosystems consist of dry subhumid zone, semi-arid zone, arid zone, and hyper-arid zone, which is measured by the degree of aridity and is based on annual precipitation and temperature (See Appendix 4.B for the details). Figure 4.1 shows the drylands ecosystems (by its sub-aridity zones) along with the GRUMP's national boundaries. As will be shown soon, we can see how many urban dwellers in developing countries are exposed in the drylands environment with water shortages and how city growth depends on the degree of aridity.

Water stress in drylands ecosystems has important implications that reach beyond access to drinking water as such. Especially in sub-Saharan Africa, a number of cities have become dependent on hydro-power for much of their electricity (Showers, 2002; Muller, 2007). As Showers (2002: 639) describes it, hydroelectric power is "a major source of electricity for 26 countries from the Sahel to southern Africa, and a secondary source for a further 13. . . . Hydroelectric dams are, however, vulnerable to drought when river flows are reduced. Cities and towns in countries from a wide range of climates were affected by drought induced power shortages in the 1980s and 1990s." Furthermore, "In several nations urban areas receive electricity from hydropower dams beyond their national boundaries . . . National drought emergencies, therefore, can have regional urban repercussions.

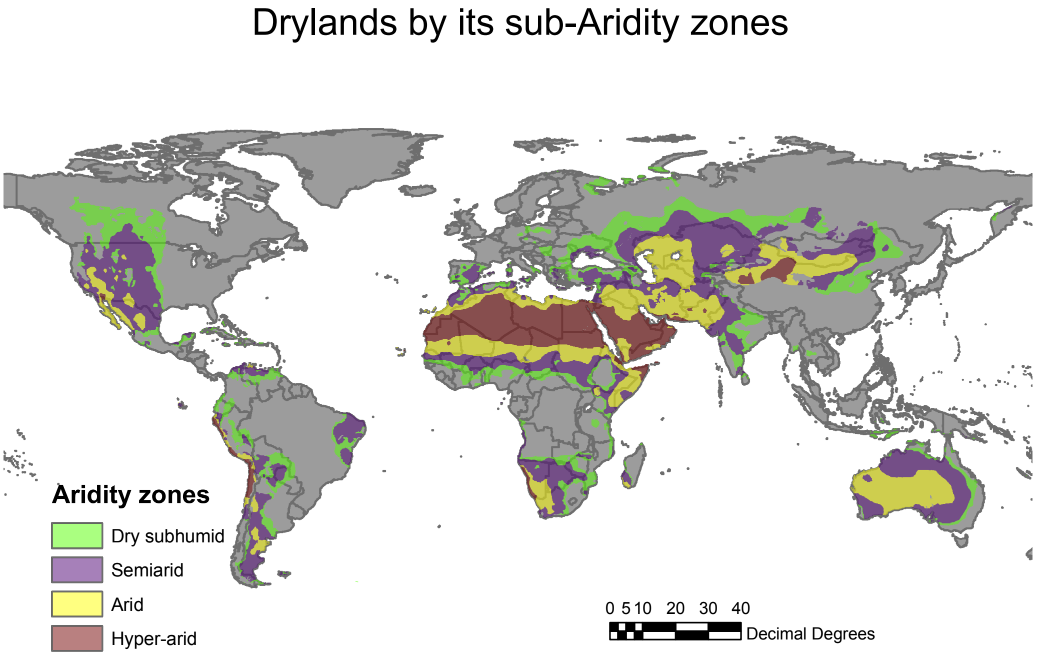


Figure 4.1: Visualization of drylands geospatial data along with national boundaries

Lomé and Cotonou suffered when interior Ghana's drought reduced power generation at the Akosombo Dam." (Showers, 2002: 643).

Safriel et al. (2005: 650) discuss other likely impacts of climate change in drylands ecosystems, including reductions in water quality and a higher frequency of dry spells that may drive farmers to make greater use of irrigation, with implications especially for coastal drylands: "Since sea level rise induced by global warming will affect coastal drylands through salt-water intrusion into coastal groundwater, the reduced water quality in already overpumped aquifers will further impair primary production of irrigated croplands." The productivity consequences may have the effect of increasing the costs of production in agriculture, which may in turn cause agricultural prices to rise, reduce employment and earnings, and possibly encourage both circular and longer-term migration to urban areas (Muller, 2007; Adamo and de Sherbinin, 2008).

Low-elevation coastal zone McGranahan, Balk, and Anderson (2007) defines the LECZ (low elevation coastal zone) as land area contiguous with the coastline up to a 10-metre rise elevation, based on the measure from the Shuttle Radar Topography Mission (SRTM) elevation data set. In some places, mostly the mouths of major rivers such as the Amazon in Brazil and the Yenisey river in Russia, the LECZ extends well beyond 100 kilometres inland, although for most of its extent, the zone is much less than 100 kilometres in width.

According to current forecasts, sea levels will gradually but inexorably rise over the coming decades, and this will place large coastal urban populations under threat around the globe. Richard B. Alley et al. (2007) foresee increases of 0.2 to 0.6 meters in sea level by 2100, a development that will be accompanied by more intense typhoons and hurricanes, storm surges, and periods of exceptionally high precipitation. Many of Asia's largest cities are located in coastal areas that have long been cyclone-prone. Mumbai saw massive floods in 2005, as did Karachi in 2007 (Kovats and Akhtar, 2008; Bank, June 2008). Storm surges and flooding also present a threat in coastal African cities (e.g., Port Harcourt, Nigeria, and

Mombasa, Kenya; see Douglas et al. (2008) and Awuor et al. (2008)) and in Latin America (e.g., Caracas, Venezuela, and Florianópolis, Brazil; see Hardoy and Pandiella (2009)).

Inland water The inland water classification used in our study comes from the level 3 data of the Global Lakes and Wetlands Database, which was assembled from various spatial data sources and geo-processing by Lehner and Döll (2004). A succinct summary is given in `http://www.worldwildlife.org/science/data/WWFBinaryitem8606.pdf`. The inland water zone includes lakes (including both natural lakes and manmade reservoirs), rivers, and several types of wetlands. As will be shown soon, more than half of the cities in our analysis are situated in any type of the inland water zone. Furthermore, the cities in the inland water zone are more likely to grow faster than other cities.

The ecozone data in the form of geospatial raster format are integrated to the combined UN-GRUMP cities data to generate the ecozone variables below used in the analysis below through geoprocessing. For the geoprocessing, Python scripts are used with ArcGIS's python geogrocessing module to automate the work. The *zonal statistics* method, one of main methods used in the analysis, is used to identify whether a city is located in one of the econzones. The scripts are available up on request.

4.1.2 Analysis

Table 4.1 shows the distribution of urban population by city-size ranges in Asia, and Table 4.2 re-expresses these data by showing the percentage of all Asian urban dwellers in a given city-size range who live in these zones. Tables 4.3 and 4.4 present the figures for Africa and South America. These tables show that drylands are home to about half of Africa's urban residents irrespective of city size, and even greater percentages-ranging from 54 to 67 percent—in the important case of India. In South America and China, however, much lower percentages of all urban dwellers live in drylands. For all of the regions considered here, significant numbers and percentages of urban residents live in the LECZ, although

Table 4.1: Distribution of the Asian urban population and land area in the LECZ and drylands, by population size ranges. Population in thousands (000s) and land area in square kilometers. (Size and area in 2000, estimated using GRUMP methods.)

	Number of	All Ecozones		Drylands		LECZ	
City Population	Cities	Population	Area	Population	Area	Population	Area
All Asia							
Under 100,000	10,582	341,000	446,295	142,000	219,204	27,200	28,753
100,000–500,000	1,470	301,000	279,866	122,000	141,552	37,000	26,061
500,000–1 million	180	124,000	94,797	48,500	46,348	15,700	8,689
1 million+	200	722,000	327,318	229,000	128,032	174,000	59,873
India							
Under 100,000	2,845	77,100	113,396	51,700	76,986	2,839	3,733
100,000–500,000	300	59,300	53,033	38,300	33,703	4,473	2,898
500,000–1 million	33	22,200	13,785	13,100	7,005	896	699
1 million+	37	126,000	41,800	68,500	24,355	29,400	4,321
China							
Under 100,000	5,711	198,000	167,796	58,000	54,829	15,700	11,040
100,000–500,000	690	141,000	81,895	40,300	30,713	15,300	6,803
500,000–1 million	81	56,400	29,438	13,100	9,502	8,406	3,164
1 million+	76	221,000	80,575	60,000	26,700	58,700	19,198
Asia Other Than India and China							
Under 100,000	2,026	65,900	165,102	32,300	87,389	8,661	13,980
100,000–500,000	480	100,700	144,938	43,400	77,137	17,227	16,361
500,000–1 million	66	45,400	51,574	22,300	29,841	6,398	4,827
1 million+	87	375,000	204,943	100,500	76,977	85,900	36,354

Table 4.2: Percentages of the Asian urban population and land area in the LECZ and drylands, by population size ranges. Population in thousands (000s) and land area in square kilometers. (Size and area in 2000, estimated using GRUMP methods.)

	Drylands		LECZ	
City Population	Population	Area	Population	Area
All Asia				
Under 100,000	41.6	49.1	8.0	6.4
100,000–500,000	40.6	50.6	12.3	9.3
500,000–1 million	39.2	48.9	12.7	9.2
1 million+	31.7	39.1	24.1	18.3
India				
Under 100,000	67.1	67.9	3.7	3.3
100,000–500,000	64.5	63.6	7.5	5.5
500,000–1 million	59.1	50.8	4.0	5.1
1 million+	54.2	58.3	23.2	10.3
China				
Under 100,000	29.3	32.7	8.0	6.6
100,000–500,000	28.5	37.5	10.8	8.3
500,000–1 million	23.2	32.3	14.9	10.7
1 million+	27.2	33.1	26.6	23.8
Asia Other Than India and China				
Under 100,000	49.0	52.9	13.1	8.5
100,000–500,000	43.1	53.2	17.1	11.3
500,000–1 million	49.1	57.9	14.1	9.4
1 million+	26.8	37.6	22.9	17.7

Table 4.3: Distribution and percentages of the African urban population and land area in the LECZ and drylands, by population size ranges. Population in thousands (000s) and land area in square kilometers.(Size and area in 2000, estimated using GRUMP methods.)

	Number of	All Ecozones		Drylands		LECZ	
City Population	Cities	Population	Area	Population	Area	Population	Area
Under 100,000	3,247	61,800	123,359	29,800	67,017	3,820	5,042
100,000–500,000	301	61,400	58,417	27,800	28,854	6,870	4,695
500,000–1 million	32	22,100	13,050	10,700	7,107	3,531	1,788
1 million+	42	130,000	56,985	61,700	28,686	17,300	4,787

	Drylands		LECZ	
City Population	Population	Area	Population	Area
Under 100,000	48.3	54.3	6.2	4.1
100,000–500,000	45.3	49.4	11.2	8.0
500,000–1 million	48.4	54.5	16.0	13.7
1 million+	47.5	50.3	13.3	8.4

Table 4.4: Distribution and percentages of the South American urban population and land area in the LECZ and drylands, by population size ranges. Population in thousands (000s) and land area in square kilometers. (Size and area in 2000, estimated using GRUMP methods.)

	Number of	All Ecozones		Drylands		LECZ	
City Population	Cities	Population	Area	Population	Area	Population	Area
Under 100,000	2,739	45,000	170,998	12,300	49,244	2,055	7,179
100,000–500,000	198	40,200	68,926	14,300	28,964	2,890	4,974
500,000–1 million	28	19,900	23,257	6,220	6,627	1,946	1,956
1 million+	34	111,000	71,677	25,500	20,234	10,800	5,844

	Drylands		LECZ	
City Population	Population	Area	Population	Area
Under 100,000	27.4	28.8	4.6	4.2
100,000–500,000	35.6	42.0	7.2	7.2
500,000–1 million	31.2	28.5	9.8	8.4
1 million+	22.9	28.2	9.7	8.2

the figures are lower than the drylands figures. Among all urbanites residing in cities of 1 million or more, the percentages in the LECZ range from 9.7 percent in South America to 26.6 percent in China. It is worth noting that China and India are not exceptions: the pattern found in Asian cities apart from China and India is similar to these large countries. Dryland cities have smaller fractions of their population, as compare to their land area, whereas LECZ cities contain more persons relative to land.

We have shown elsewhere that the average urban population density (total population divided by the total land area for each city) in LECZ cities is greater than in dryland or all other cities; and this is especially apparent for cities above 1 million in size. However, even a city with much of its land area and population in the LECZ, need not have all of its area and population in the zone. Thanks to the unprecedented spatial detail of our dataset, we can examine city population density in more depth than has previously been possible for a large subset of the data (where the spatial resolution is high).

Considering cities with any land area within the LECZ, we refine our estimates of population density by calculating the density in the LECZ portion of the city and compare that with the density in the non-LECZ portion. We further compare these estimates with the population density of cities that have no land area whatsoever in the LECZ. This analysis can only be done on a subset of the data in which the spatial resolution (in terms of the number and geographic size of the city administrative units) is high enough to provide within-city variation: we include urban areas comprised of more than one administrative unit (or parts thereof).

The top panel of Table 4.5 displays the results for Africa, (all of) Asia, and South America. In Africa and Asia, we find that LECZ cities, and the portions of such cities actually in the LECZ, exhibit substantially higher population densities. The densities found in the LECZ portion of cities is 43 percent and 20 percent, respectively, above the densities found in the non-LECZ portions of the cities. In South America, cities located (wholly or in part) in the LECZ are denser but there is not much difference in density evident between

Table 4.5: City population density in persons per square kilometer, by ecozone and city population size ranges, all regions. Figures are for cities that intersect more than one administrative area; cities contained within a single administrative area are omitted.

	Cities Outside LECZ	Cities Fully or Partly in LECZ	
Region		LECZ Density	Other Density
Africa	620	2,406	1,680
Asia	1,473	1,827	1,525
South America	661	1,079	1,003

	Cities Under 1 Million			Cities Over 1 Million		
	Cities Outside LECZ	Cities Fully or Partly in LECZ		Cities Outside LECZ	Cities Fully or Partly in LECZ	
Region		LECZ Density	Other Density		LECZ Density	Other Density
Africa	542	1,274	872	2,705	4,294	2,960
Asia	1,313	1,463	1,136	2,413	3,518	3,125
South America	560	805	678	1,251	1,665	1,676

the LECZ and non-LECZ portions of these cities. It might be thought that the increased population density in the LECZ is simply a reflection of the presence of large cities in this zone. The bottom panel of Table 4.5 provides evidence to the contrary. For cities above and below 1 million persons, we find equally strong evidence, particularly in Africa where the ratio in mean densities within and beyond the LECZ in LECZ-cities is comparable in large and small cities alike. Only in South American large cities do we find densities that are alike irrespective of whether the land area is within the LECZ or is beyond it. In Asia, we also find that for smaller and medium-sized cities that are entirely outside of the LECZ have densities that are lower than those found in the LECZ land portion of LECZ cities, but higher than the non LECZ-land portion of LECZ cities.

In summary, we have found evidence suggesting that LECZ cities tend to be more densely populated, even within the LECZ portion of cities that are only partly contained within the zone. Dense population, especially in coastal areas, has some important implications for climate change adaptation and mitigation but greater density is not always desirable Dodman (2008). Denser cities may (depending on many factors, including the quality of urban governance and management) economize on the use of scarce resources, including those of the ecozones within and nearby the city, and may produce less by way of climate-damaging emissions; but denser cities also present governments with health and management challenges, especially in large cities that lack adequate infrastructure.

4.2 Extended modeling of city growth

We have seen how urban settlements are currently distributed according to ecological zone—but will these patterns be substantially reshaped as cities and towns continue to grow? To generate forecasts of city population growth, we now turn to the city time-series supplied by the United Nations. Table 4.6 shows the number of UN-recorded cities in each of the ecozones we consider. (The inland water zone is included here along with the

Table 4.6: Number of cities in inland water, LECZ, and dryland ecozones. Dryland consists of dry subhumid, semiarid, and arid; the last category includes hyper-arid.

Region	Inland Water	LECZ	Dry Subhumid	Semi-arid	Arid	N
Africa	325	165	143	95	68	720
Latin America	257	163	88	56	27	466
Asia	808	406	265	279	120	1,233
Total	1,390	734	496	430	215	2,421

low-elevation coastal zone and drylands.) Table 4.7 displays the combinations of LECZ and drylands ecozones that are found in our data.

Table 4.7: Number and percentage of cities by LECZ and aridity.

	All Regions		Africa		Latin America		Asia	
	N	%	N	%	N	%	N	%
LECZ								
Humid	463	19.12	81	11.25	83	17.74	299	24.25
Dry sub-humid	162	6.69	45	6.25	55	11.75	62	5.03
Semi-arid	49	2.02	16	2.22	14	2.99	19	1.54
Arid	60	2.48	23	3.19	11	2.35	26	2.11
Non-LECZ								
Humid	817	33.75	333	46.25	214	45.73	270	21.90
Dry sub-humid	334	13.80	98	13.61	33	7.05	203	16.46
Semi-arid	381	15.74	79	10.97	42	8.97	260	21.09
Arid	155	6.40	45	6.25	16	3.42	94	7.62
Total	2,421	100	720	100	468	100	1,233	100

4.2.1 Econometric specifications

A panel data regression model of city growth is set out as,

$$g_{i,t} = \beta_0 + \beta_1 \mathrm{TFR}_{i,t} + \beta_2 \mathrm{Q}_{i,t} + \mathbf{D}'_{i,t}\gamma + \mathbf{X}'_{i,t}\delta + v_{i,t}. \tag{4.1}$$

In this equation $g_{i,t}$ is the estimated population growth rate for city i at time t, and the fertility and mortality components of growth are represented by the urban total fertility rate $\mathrm{TFR}_{i,t}$ and $Q_{i,t}$, the urban child mortality rate. We include in $\mathbf{X}_{i,t}$ a set of dummy variables recording city i's population size, which as we will see, turns out to be an important influence on the rate of population growth. We also include here the ecosystem indicators, which we discuss in more detail in the next section. The vector $\mathbf{D}_{i,t}$ contains a set of dummy variables indicating the start-of-period and end-of-period units in which the city's population is recorded. In the Cuiabá example of Figure 3.2, these dummy variables would take into account the fact that in the early 1970s, one era of growth began with the population recorded in terms of the city proper but ended with a count expressed in unknown units. In principle, of course, a number of additional city-specific explanatory variables could be introduced to explain city growth. Variables that are fixed over time present no particular difficulties. Those that change with time, however, would themselves need to be forecast in the process of generating city growth forecasts.

A word is in order on two further aspects of the regression specification. First, although the UN Population Division has a long-standing research program in which it estimates and forecasts a number of demographic rates at the national level, including total fertility and child mortality rates, it does not produce separate urban and rural estimates. Although we have derived estimates of these rates from countries with a World Fertility Survey—see Ashurst et al. (1984) for urban total fertility rates—or a Demographic and Health Survey [1], a number of countries have participated in neither of these programs. To estimate urban fertility and mortality rates for these cases, therefore, we have used descriptive regressions in which the available urban rates are regressed upon the UN's national-level estimates of the rates—published for all countries, with forecasts to 2050—together with time trends and interactions of time with the UN's national estimates. The descriptive models are given

[1] Source: Macro International Inc, 2009. MEASURE DHS STATcompiler. `http://www.measuredhs.com`, September 5th, 2009.

Table 4.8: Descriptive regressions of observed urban rates on the UN's estimated national rates. `Time` variable set to 1 for 1950.

Urban TFR	Constant	National TFR	Time	TFR · Time	R^2
(N=308)	0.035	0.766			0.788
	(0.32)	(33.68)			
	-0.027	0.769	0.001		0.788
	(-0.15)	(32.13)	(0.42)		
	-2.422	1.235	0.055	-0.011	0.805
	(-4.90)	(13.32)	(5.13)	(-5.19)	

Urban Q5	Constant	National Q5	Time	Q5 · Time	R^2
(N=186)	3.828	0.768			0.870
	(1.42)	(35.15)			
	30.808	0.764	-0.550		0.875
	(3.05)	(35.50)	(-2.76)		
	5.650	0.992	-0.031	-0.004	0.876
	(0.27)	(5.97)	(-0.07)	(-1.38)	

in Table 4.8. Using the three models shown in the table, we generated predicted values for urban rates from 1950 to 2045 for all cities in the combined cities database. This procedure is admittedly something of a stop-gap measure, on which we will rely only while we comb the literature for credible series of urban demographic rates for countries lacking WFS and DHS surveys. These imputed urban fertility and child mortality figures generally appear reasonable, but obviously more research to refine the estimates is in order. We have tested our city growth rate models with both an *observed* urban rates sample restricted to cities in countries with a WFS or DHS survey and compared the results to those obtained by using a larger sample with *estimated* urban rates, finding few differences of note.

We also need to address the properties of $v_{i,t}$, the regression disturbance term. An error-components specification provides a sensible entry-point for our analysis. In such specifications, the disturbance term is represented as a composite, $v_{i,t} = u_i + \varepsilon_{i,t}$, containing one component, u_i, that is specific to city i and whose value can be estimated as $\hat{u}_i$. In a random-effects error components model, u_i is assumed to be independent of (or at a minimum, uncorrelated with) the other right-hand side explanatory variables. A second specification is the so-called *fixed effect* specification in which the disturbance term takes the same algebraic form $v_{i,t} = u_i + \varepsilon_{i,t}$, but in which u_i is allowed to be correlated with other right-hand side variables. As in the random-effects approach, the value of u_i can be estimated (using techniques similar to those of the random-effects method). This specification proves useful when city-specific endogenous explanatory variables (such as city size) are introduced in the model.

4.2.2 Spatial linkages among cities

Cities do not stand isolated from each other; they are linked through many sorts of networks involving migration, trade, information exchange, and the like. These interactions may induce a correlation among the growth rates disturbance of the cities that are linked within a spatial network. In a spatial econometric model, the regression disturbance term takes the composite form

$$v_{i,t} = \rho \sum_{j \neq i} w_{i,j} v_{j,t} + \varepsilon_{i,t}$$

with $\varepsilon_{i,t} = u_i + \eta_{i,t}$. In this specification, the disturbance $v_{i,t}$ for city i is directly linked, via $\rho w_{i,j}$, to $v_{j,t}$, its counterpart for city j. The spatial autocorrelation coefficient ρ and a pre-specified spatial weight $w_{i,j}$ determines the size and direction of the relationship. The spatial error specification implies that spatial correlation present in both the city-specific effects component u_i and the remainder error component $\eta_{i,t}$. In this way, the disturbance term exhibits both spatial and temporal correlation (Kapoor et al., 2007; Baltagi et al., 2006). Stacking all disturbances first over city i for each time and then (in blocks) over time t yields

$$\mathbf{v} = \rho \mathbf{W}_n \mathbf{v} + \varepsilon \qquad (4.2)$$

In this equation, $\mathbf{W}_n$ is a block diagonal matrix each of whose blocks is the spatial weight matrix whose diagonal elements are zeroes and whose off-diagonal elements, when multiplied by ρ, establish the connections between the spatial units i and j insofar as their disturbance terms are concerned. [2].

Broadly speaking, there are three general types of weight matrices that can be considered: $\mathbf{W}$ can be a symmetric matrix, or the asymmetric result of row-standardizing a symmetric matrix, or it can be fundamentally asymmetric. The row-standardized specification is used so often in spatial econometrics that it has almost become the default. In this approach, each weight $w_{i,j}$ is scaled by $\sum_j w_{i,j}$, so that for every i the scaled weights sum to one. Standardizing the weights in this way often makes good sense. The popularity of row-standardization has been further enhanced by the numerical properties that it imparts to the scaled matrix, as will be seen. However, the approach does not apply to all cases of interest. We obviously cannot row-standardize spatially isolated observations for which $\sum_j w_{i,j} = 0$, and the conversion of weights to averages is not always the right thing to do. Unfortunately, the literature rarely develops results for fundamentally asymmetric weight matrices, mainly because numerical issues arise with such weights that need to be studied on a case-by-case basis. In addition to these features of the $\mathbf{W}$ weight matrix, the nature of its elements also matters, in that computational short-cuts become available when these elements take a zero–one, “binary” form.

Our classical estimator of the city growth model with spatial correlation is that devised by Baltagi et al. (2007), and we have also developed a Bayesian version of that model. In the Bayesian implementation, the posterior distribution of the parameters is simulated by the Metropolis-within-Gibbs algorithm (Tierney, 1994) with the following five blocks: $(\alpha, \beta, \delta, \gamma)$, u, σ_u^2, σ_η^2, and lastly ρ itself. The posterior distribution of ρ conditional on the

[2] The block diagonal matrix W_n depends on whether panel data is balanced or not: For balanced panel data, each block is same $W_n = I_T \otimes W$ in which $\otimes$ denote the kronecker product; For unbalanced panel data, however, each block can be different, depending on availability of data. That is, $W_n = diag(W_1, W_2, \cdots, W_T)$. See Chapter 2 for the details

other parameter blocks and data is not available in closed form; it must be simulated using the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). Although classical econometric models can be estimated by generalized least squares or the generalized method of moments without the log-determinant term that enters the likelihood function. Bayesian models require this term, however, and thus face the numerical difficulties involved in calculating it for models of the size with which we are dealing; see Cliff and Ord (1981); Pace and LeSage (2004).

We have tested these spatial econometric models with row-standardized spatial weights based on distance $d_{i,j}$ between city centroids. The spatial weights are specified as row-standardized version of inverse distance, $w_{i,j} = d_{i,j}^{-1} / \sum_{j=1}^{N_t} (d_{i,j}^{-1})$ where N_t is the number of city observations at time t. This specification implies that the linkage between the growth rate disturbance terms of cities i and j grows weaker the more distant the two cities are. Distances are expressed in kilometers.

4.3 Results

4.3.1 Random and fixed-effect city growth results

The results are shown in Table 4.9 for all UN cities, and region-specific results are provided in Appendix 4.A. The results for ecozone indicate that cities in the inland water zone grow relatively faster than other cities, the difference amounting to about 0.38 to 0.40 percentage points in the pooled results. The effect is also significant and of roughly the same size across regions, as shown in Appendix 4.A. The effects of the low-elevation coastal zone and drylands are more difficult to interpret owing to the need to consider interaction terms. In the models with all cities pooled in the analysis, cities in the LECZ but not in the drylands tend to grow more slowly, with Asia presenting a partial exception. However, as confirmed by Wald tests (not shown), LECZ cities that are also in the drylands tend to grow faster, a finding that is especially clear for coastal Asian cities that are situated in semi-arid or even drier environments.

Urban fertility rates display very strong positive effects on city growth rates in the pooled results of Table 4.9, which indicate that a decline of 1 child in the urban TFR is associated with a drop of 0.87 to 1.00 percentage points in the city growth rate, a quantitatively important impact. The effects of urban fertility are also highly significant in the fixed-effect models, where fertility rates have an even larger influence than is evident in the random-effect models. Urban fertility also emerges as quantitatively important in the region-specific results (also in Appendix 4.A), where the models for Africa exhibit random-effect coefficients of about 0.64 for the urban total fertility rate, Latin America's coefficient is 0.79, and the coefficient for Asian cities is 0.93, the largest among the regions. Child mortality rates show the expected negative sign in the pooled results (Table 4.9) and in Asia (the Appendix 4.A) but are insignificant in Latin America and take a positive sign in the African results. In the pooled results and also across regions, larger cities tend to grow more slowly than do cities under 100,000 population (which is the omitted category in the regression specification), and the effect is important in quantitative terms as well as being highly significant statistically. Controls for changes in the statistical concept for which city population is recorded—city proper, agglomeration, etc (including whether the concept was unknown)—make a statistically significant difference as a group (results not shown) but the details are complicated.

Table 4.9: Regressions with estimated urban vital rates, all UN cities (*Z*-statistics in parentheses).

	OLS	Random-Effects	Fixed-Effects
Start-of-period Urban TFR	0.816	0.873	0.999
	(18.92)	(19.28)	(16.24)
Start-of-period Urban Q5	-0.005	-0.007	-0.016

Continued on next page

... table 4.9 continued

	OLS	Random-Effects	Fixed-Effects
	(-5.59)	(-6.99)	(-9.92)
Inland water	0.381	0.403	
	(5.80)	(5.20)	
LECZ	-0.209	-0.264	
	(-2.33)	(-2.47)	
Dry subhumid	-0.654	-0.651	
	(-6.40)	(-5.44)	
Semiarid	-0.452	-0.449	
	(-4.94)	(-4.13)	
Arid and above	-0.382	-0.403	
	(-2.97)	(-2.68)	
LECZ * Dry subhumid	0.685	0.683	
	(4.12)	(3.43)	
LECZ * Semiarid and above	0.630	0.613	
	(3.48)	(2.84)	
100 <= City Size < 500	-0.805	-0.905	-1.573
	(-10.73)	(-11.51)	(-14.37)
500 <= City Size < 1,000	-1.000	-1.311	-3.026
	(-7.18)	(-8.95)	(-14.42)
City Size >= 1,000	-1.270	-1.594	-3.993
	(-8.35)	(-9.33)	(-13.72)
Unknown-Unknown	0.500	0.530	0.672
	(4.46)	(4.19)	(2.61)
Unknown-Proper	0.899	0.777	0.668

Continued on next page

... table 4.9 continued

	OLS	Random-Effects	Fixed-Effects
	(5.88)	(4.86)	(2.52)
Unknown-Agglomeration	0.277	0.362	0.721
	(0.97)	(1.29)	(2.21)
Unknown-Metro. Area	-0.292	-0.293	0.084
	(-0.79)	(-0.80)	(0.19)
Proper-Unknown	1.202	1.027	0.772
	(5.19)	(4.39)	(2.43)
Proper-Proper	-0.007	-0.089	-0.086
	(-0.07)	(-0.76)	(-0.36)
Proper-Agglomeration	1.520	1.413	1.182
	(5.03)	(4.76)	(3.55)
Agglomeration-Unknown	-0.857	-0.784	-0.492
	(-1.67)	(-1.56)	(-0.90)
Agglomeration-Proper	-0.494	-0.465	-0.402
	(-1.00)	(-0.96)	(-0.76)
Agglomeration-Metro. Area	1.463	1.212	0.423
	(1.83)	(1.54)	(0.49)
Metro. Area-Metro. Area	0.314	0.294	0.280
	(0.99)	(0.85)	(0.53)
Others-Others	2.714	2.608	1.340
	(3.93)	(3.62)	(1.12)
Constant	1.870	1.967	2.773
	(11.32)	(10.77)	(10.14)
σ_u		1.032	

Continued on next page

... table 4.9 continued

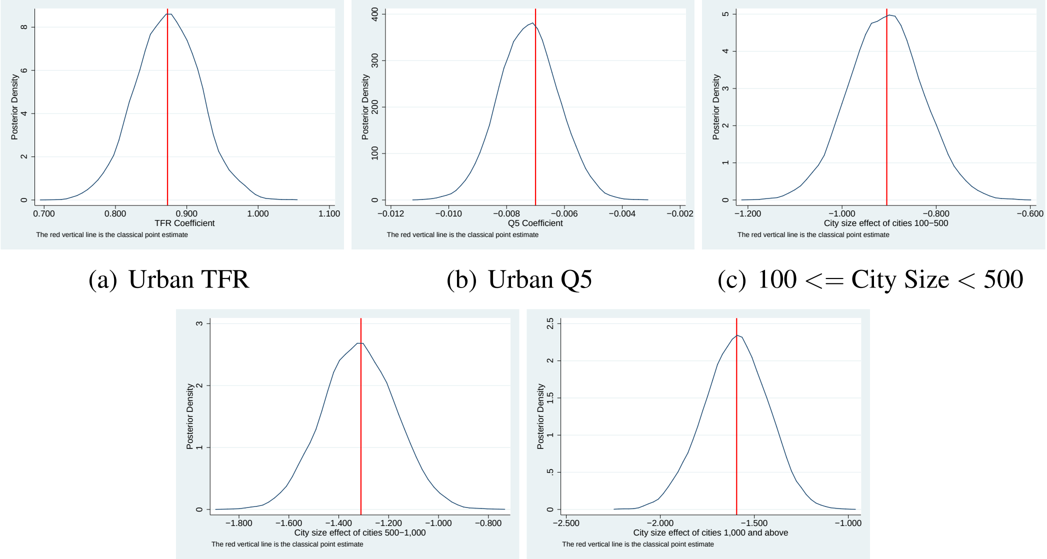
	OLS	Random-Effects	Fixed-Effects
		(23.11)	
σ_e		3.001	
		(130.90)	

Our Bayesian implementations of these models produce posterior distributions of the model parameters instead of the point estimates generated by the classical approach. Figure 4.2 shows that the posterior distributions emerging from the Bayesian approach are approximately centered on the classical estimates of the corresponding parameters (See Table 4.9). The posterior distributions are drawn from their Markov chain Monte Carlo samples after checking convergence of the sample. See also Appendix 4.A for the Bayesian point estimates (i.e. the posterior mean calculated as the mean of the posterior distribution) and Bayesian Z-statistic (calculated by dividing the posterior mean by the posterior standard error) for both the random- and fixed-effects models.

4.3.2 Estimation with spatial linkages

Table 4.10 shows classical and Bayesian random-effects city growth models with spatial correlated errors. The generalized method of moments and Markov chain Monte Carlo estimation methods are used for classical and Bayesian approaches, respectively.[3]

[3] I used FORTRAN 95 programs. See chapter 2 for algebraic formulae. For Bayesian, I specify vague priors for the hyperparameters; that is, $\theta_0 = \mathbf{0}$, $\mathbf{M}_0 = 1^{-5}\mathbf{I}$, $r_0 = 0$, $p_0 = 0$, $v_0 = 0$, and $s_0 = 0$ and a uniform prior over $(-1, 1)$ for ρ. In practice, to draw ρ, I used the natural logarithm of $p(\cdot)$ which includes the log-determinant, $\ln|B|$. Ord (1975) showed that $| \mathbf{I} - \rho \mathbf{W}_n | = \prod_{i=1}^{n}(1 - \rho\lambda_i)$ with λ_i being the i-th eigenvalue of the spatial weight matrix $\mathbf{W}_n$ of dimension n. It is computationally efficient to use the fact that the eigenvalues of the block-diagonal matrix $\mathbf{W}_n = diag(\mathbf{W}_1, \cdots, \mathbf{W}_T)$ are those of the diagonal blocks $\mathbf{W}_1, \cdots, \mathbf{W}_T$. The samples drawn from this Metropolis-within-Gibbs procedure are used to estimate the model. In a machine equipped with Intel Xeon 3.40GHz processor, it took 5.8 minutes to draw every 1,000 samples for the model



(a) Urban TFR (b) Urban Q5 (c) 100 <= City Size < 500

(d) 500 <= City Size < 1,000) (e) City Size >= 1,000)

Figure 4.2: Bayesian posterior distributions of the random-effects model

The classical and point estimates of spatial correlation coefficient are 0.619 and 0.671, respectively, implying the existence of spatial correlation in our error term of our city growth model. Table 4.10 show that, among other things, the national TFR coefficient are still significant, allowing for spatial correlation in the regression error terms.

Our results show that consideration of spatial correlation should be taken into account to estimate and forecast econometric models of city population growth. It is surprised to see that few studies consider how to model city-network effects on city growth. Importance of the interactions among cities are increasing with migration and trade among cities in both intercountry and international levels. As Voss et al. (2006) argue, when there is reason to suspect that spatial error correlation exist, models that do not take it into account will likely be biased in terms of coefficient standard errors, thus contaminating inference and causing forecast error variances to be calculated incorrectly.

with the total number of observations being 10,766, the number of cross-section units being 2,408, and the number of covariates being 25.

Table 4.10: Classical and Bayesian panel data city growth regression models with spatially correlated errors. Spatial weights are given by distance-based ones, $1/d_{ij}$, in which d_{ij} denotes the Haversine great-circle distance (in kilometers) between cities i and j. Spatial weights are row-standardized.

	Classical GMM		Bayesian MCMC	
	Coeff.	(Z-stat.)	Coeff.	(Z-stat.)
Start-of-period Urban TFR	0.841	(12.03)	0.843	(11.23)
Start-of-period Urban Q5	-0.005	(-3.53)	-0.005	(-3.31)
Inland water	0.384	(5.20)	0.385	(5.19)
LECZ	-0.069	(-0.64)	-0.053	(-0.48)
Dry subhumid	-0.407	(-3.28)	-0.387	(-3.07)
Semiarid	-0.266	(-2.17)	-0.249	(-1.99)
Arid	-0.179	(-1.09)	-0.162	(-0.96)
LECZ * Dry subhumid	0.515	(2.71)	0.502	(2.56)
LECZ * (Semiarid or arid)	0.546	(2.62)	0.537	(2.50)
100<=City Size<500	-0.788	(-10.00)	-0.785	(-10.17)
500<=CIty SIze<1,000	-1.247	(-8.79)	-1.262	(-8.82)
City Size >=1,000	-1.516	(-9.29)	-1.533	(-9.32)
Unknown-Unknown	0.305	(2.15)	0.300	(2.11)
Unknown-Proper	0.525	(2.97)	0.504	(2.83)
Unknown-Agglomeration	0.339	(1.20)	0.346	(1.23)
Unknown-Metro.Area	-0.019	(-0.04)	-0.008	(-0.01)
Proper-Unknown	0.658	(2.51)	0.628	(2.37)
Proper-Proper	-0.220	(-1.88)	-0.231	(-1.95)
Proper-Agglomeration	1.592	(5.22)	1.598	(5.25)
Agglomeration-Unknown	-0.967	(-1.95)	-0.959	(-1.93)
Agglomeration-Proper	-0.629	(-1.35)	-0.631	(-1.35)
Agglomeration-Metro.Area	0.925	(1.22)	0.895	(1.18)
Metro.Area-Metro.Area	0.434	(1.32)	0.437	(1.31)
Others-Others	2.950	(4.28)	2.971	(4.28)
Constant	1.891	(7.91)	1.887	(7.38)
ρ	0.619		0.671	(33.41)
σ_u	0.843		0.880	(19.81)
σ_η	2.904		2.892	(132.18)

4.4 City growth forecasts

With lagged city size in the model, forecasts of city growth must be made recursively. The growth rate forecast $\tilde{g}_{i,t}$ for period t and $t+1$ implies a forecast for city i's population size as of time $t+1$, or $\mathrm{P}_{i,t+1}$, which then goes on to influence the growth rate $\tilde{g}_{i,t}$ forecast for the period $t+1$ to $t+2$.

The forecasts of city growth based on these regressions are summarized in Figure 4.3 for all regions, and separately in Figure 4.6 in Appendix 4.A for each of the three main regions. These figures show the (implied) projection of urban fertility rates (values are displayed on the right axis of each figure) as well as the median forecast of city growth rates and the 25th and 75th percentiles. The projected decline in urban fertility is the dominating factor—it brings about reductions in the median growth rate forecast from nearly 4 percent in 2000 to a level just above 2 percent as of 2045. A similar pattern is seen in the forecasts based on region-specific models (Figure 4.6 in Appendix 4.A) and in the forecasts according to LECZ and drylands ecozones (Figure 4.7), with urban fertility again being the main force projected to drive down city growth rates in the future. It is, however, worth asking whether even by 2045, African urban TFRs are likely to reach the level of 1.5 children that has been projected, which may well be over-optimistic.

Figure 4.4 shows our another forecasting exercise along with the UN forecasts for cities over 750,000 and above in which UN forecasts are available. These forecasts, based on the fixed-effects model, are much closer to the UN's forecasts of growth rates, and owing to the inclusion of a negative city size feedback effect, growth rates of cities over 750,000 and above are lower than the case where all cities are considered.

We have demonstrated that it is a simple matter to reconcile the main features of our city growth forecasts with those of the United Nations, by introducing lagged city size into the specifications. To be sure, it is not at all obvious that reconciliation of these forecasts should be our aim. Too much doubt has been cast on the validity of the UN forecasts to

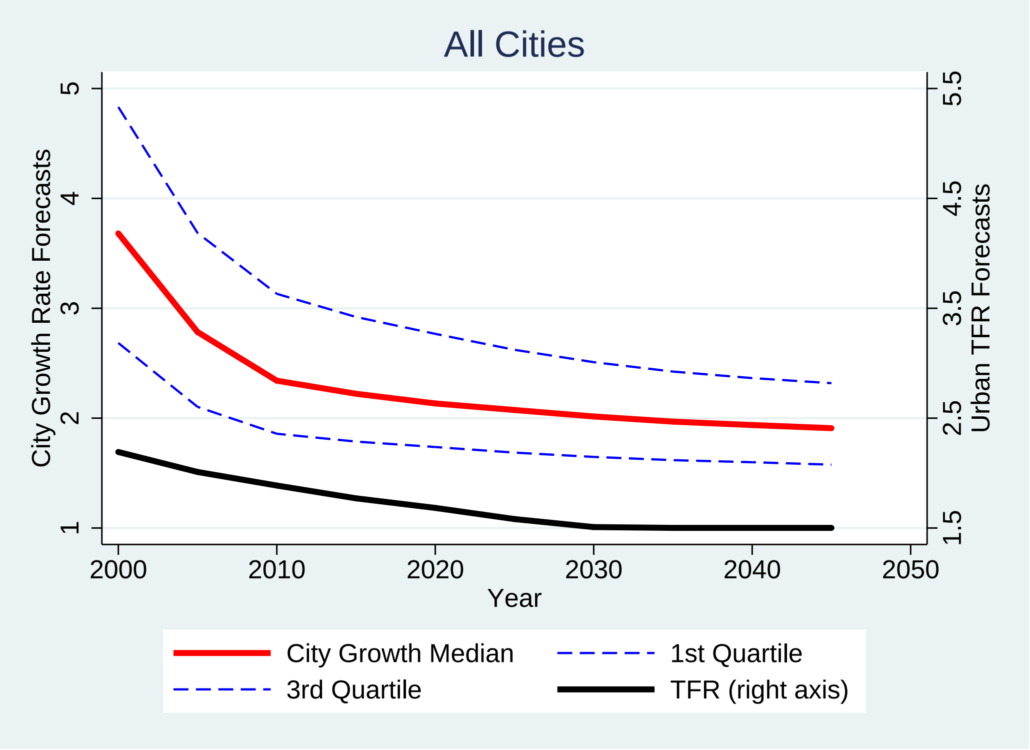


Figure 4.3: Forecasts of city growth rates conditional on UN projections of fertility and mortality

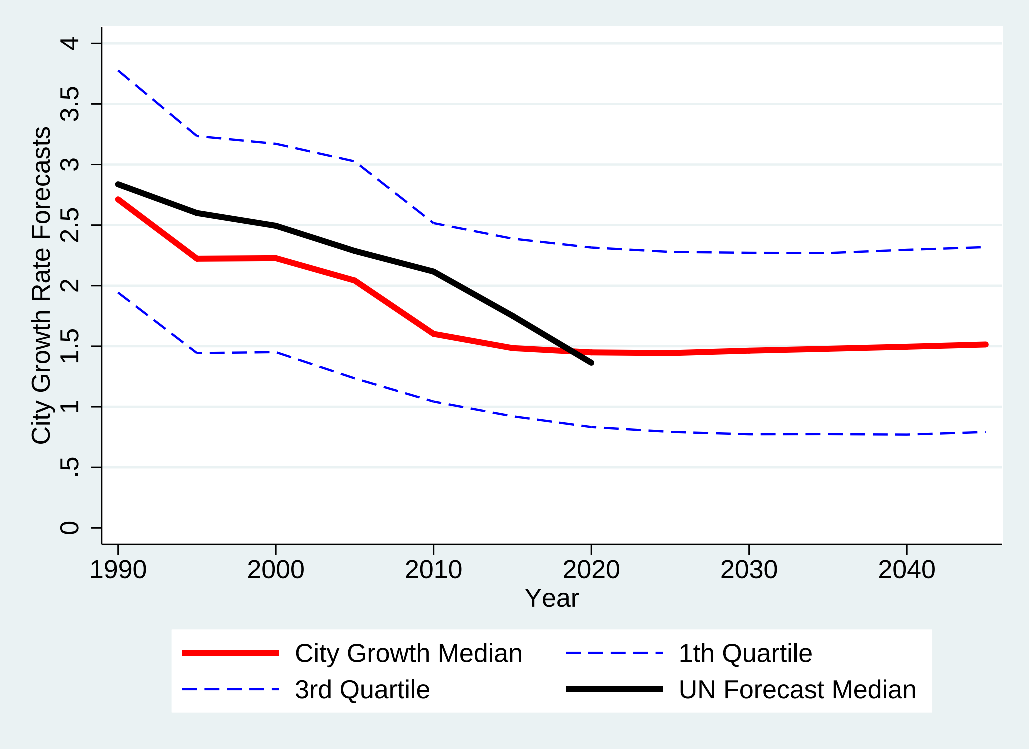
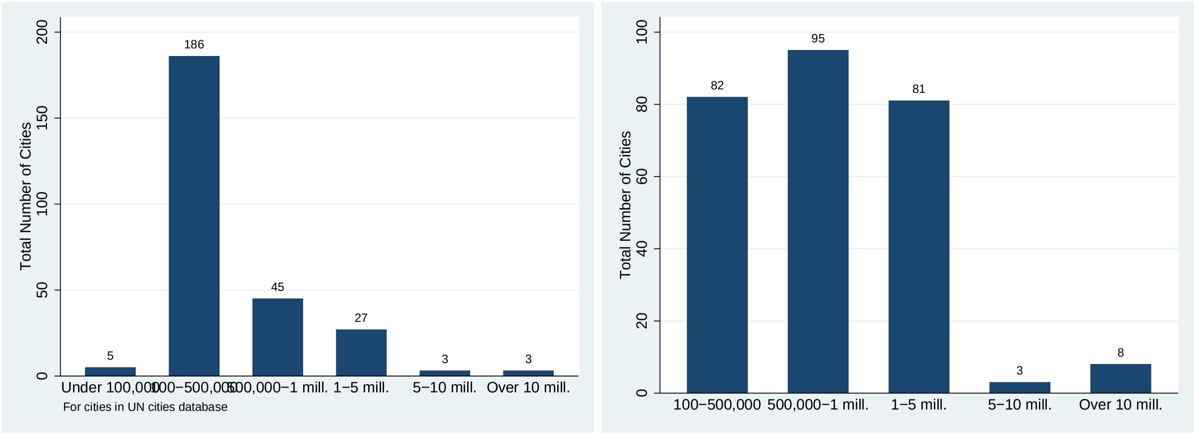


Figure 4.4: Forecasts of city growth rates conditional on UN projections of fertility and mortality, Cities over 750,000 and above where UN forecasts are available.

adopt them, uncritically, as the standard for comparison.

We further our forecasting exercise by calculating city sizes given forecasted growth rates based on the random-effect model. Figure 4.5 shows a graph with city size forecasts in 2050 for India, along with 2000 estimates. We can expect hundreds of middle-sized cities along with several cities with 10,000 million persons and above in India if the estimated vital rates and ecozone effects on city growth are assumed to hold in the future.



(a) 2000 estimates

(b) 2050 forecasts

Figure 4.5: Year 2000 estimates and year 2050 forecasts of city size in India, conditional on UN projections of fertility and mortality. Limited to cities in UN cities database that crosses UN's threshold of 100,000 persons and above.

4.A Supplementary regression results

Table 4.11: Ordinary least-squares and random-effect estimates by region, using estimated urban fertility and mortality rates. Models without controls for city size.

	Africa		Latin America		Asia	
	OLS	RE	OLS	RE	OLS	RE
Urban TFR	0.697	0.702	0.840	0.970	0.986	1.101
	(5.98)	(6.05)	(8.92)	(9.81)	(16.78)	(17.47)
Urban Q5	0.007	0.007	0.005	0.001	-0.011	-0.013
	(3.06)	(3.03)	(1.44)	(0.30)	(-8.81)	(-9.58)
Inland Water	0.345	0.342	0.396	0.411	0.242	0.243
	(2.08)	(2.06)	(4.07)	(2.88)	(2.71)	(2.38)
LECZ	-0.453	-0.465	-0.494	-0.507	0.135	0.098
	(-1.82)	(-1.86)	(-3.59)	(-2.53)	(1.08)	(0.69)
Dry subhumid	-0.697	-0.699	-0.129	-0.130	-0.251	-0.242
	(-2.60)	(-2.61)	(-0.69)	(-0.47)	(-1.82)	(-1.55)
Semiarid	-0.599	-0.598	-0.317	-0.318	0.080	0.060
	(-2.30)	(-2.30)	(-1.95)	(-1.33)	(0.62)	(0.41)
Arid and above	-0.545	-0.544	-0.109	-0.087	0.171	0.098
	(-1.68)	(-1.68)	(-0.48)	(-0.26)	(0.94)	(0.48)
LECZ*Dry subhumid	0.878	0.890	0.575	0.584	0.355	0.363
	(1.94)	(1.96)	(2.23)	(1.55)	(1.56)	(1.37)
LECZ* (> Semiarid)	0.048	0.058	0.735	0.735	0.767	0.694

Continued on next page...

... table 4.11 continued

	Africa		Latin America		Asia	
	OLS	RE	OLS	RE	OLS	RE
	(0.11)	(0.13)	(2.70)	(1.83)	(2.91)	(2.31)
Unknown-Unknown	-0.067	-0.064	0.212	0.196	0.832	0.931
	(-0.15)	(-0.14)	(1.37)	(0.99)	(5.85)	(5.96)
Unknown-Proper	1.047	1.048	0.628	0.710	1.075	1.003
	(2.31)	(2.32)	(2.74)	(2.92)	(5.49)	(4.95)
Unknown-Agglomeration	0.095	0.102	-0.332	0.019	0.382	0.430
	(0.10)	(0.11)	(-0.96)	(0.06)	(0.97)	(1.09)
Unknown-Metro. Area	-1.676	-1.670	-0.607	-0.812	-0.270	0.066
	(-0.58)	(-0.58)	(-1.97)	(-2.61)	(-0.23)	(0.06)
Proper-Unknown	1.202	1.205	1.499	1.452	0.628	0.510
	(2.01)	(2.03)	(6.08)	(5.65)	(1.21)	(0.99)
Proper-Proper	0.231	0.228	0.385	0.431	-0.005	-0.011
	(0.69)	(0.69)	(2.50)	(2.22)	(-0.04)	(-0.08)
Proper-Agglomeration	2.002	2.004	1.364	1.079	0.942	0.950
	(3.13)	(3.15)	(1.99)	(1.69)	(2.25)	(2.30)
Agglomeration-Unknown	-1.723	-1.719	-0.109	0.596	-0.919	-0.935
	(-1.86)	(-1.87)	(-0.14)	(0.82)	(-0.91)	(-0.94)
Agglomeration-Proper	-2.557	-2.560	0.222	1.136	0.353	0.395
	(-2.10)	(-2.11)	(0.16)	(0.90)	(0.61)	(0.69)
Agglomeration-Metro. Area	-0.193	-0.191	0.594	-0.002	3.009	2.692
	(-0.07)	(-0.07)	(0.75)	(-0.00)	(2.22)	(2.01)
Metro. Area-Metro. Area	-0.334	-0.328	0.037	-0.308	-0.163	-0.094

Continued on next page...

... table 4.11 continued

	Africa		Latin America		Asia	
	OLS	RE	OLS	RE	OLS	RE
	(-0.37)	(-0.36)	(0.12)	(-0.86)	(-0.25)	(-0.14)
Others-Others	0.000		-4.136	-1.434	3.534	3.516
	(.)		(-1.77)	(-0.66)	(5.23)	(5.02)
Constant	0.705	0.697	0.727	0.548	0.918	0.749
	(1.35)	(1.34)	(3.72)	(2.40)	(4.69)	(3.53)
σ_u		0.205		1.168		0.840
		(.)		(18.34)		(12.82)
σ_ε		4.043		2.025		2.893
		(72.22)		(64.19)		(93.97)

Table 4.13: Regressions with national vital rates, all UN cities

	Model 1		Model 2	
	OLS	Random-Effects	OLS	Random-Effects
National TFR	0.697	0.751	0.593	0.634
	(22.92)	(23.52)	(18.92)	(19.28)
National Q5	-0.004	-0.005	-0.004	-0.005
	(-5.81)	(-6.73)	(-5.59)	(-6.99)
Inland Water	0.266	0.257	0.381	0.403

Continued on next page...

... table 4.13 continued

	Model 1		Model 2	
	OLS	Random-Effects	OLS	Random-Effects
	(4.08)	(3.38)	(5.80)	(5.20)
LECZ	-0.235	-0.285	-0.209	-0.264
	(-2.60)	(-2.67)	(-2.33)	(-2.47)
Dry subhumid	-0.656	-0.649	-0.654	-0.651
	(-6.38)	(-5.45)	(-6.40)	(-5.44)
Semiarid	-0.499	-0.487	-0.452	-0.449
	(-5.41)	(-4.51)	(-4.94)	(-4.13)
Arid and above	-0.424	-0.432	-0.382	-0.403
	(-3.27)	(-2.88)	(-2.97)	(-2.68)
LECZ * Dry subhumid	0.726	0.737	0.685	0.683
	(4.33)	(3.73)	(4.12)	(3.43)
LECZ * Semiarid and above	0.654	0.626	0.630	0.613
	(3.59)	(2.92)	(3.48)	(2.84)
100,000–500,000			-0.805	-0.905
			(-10.73)	(-11.51)
500,000–1 million			-1.000	-1.311
			(-7.18)	(-8.95)
Over 1 million			-1.270	-1.594
			(-8.35)	(-9.33)
Unknown-Unknown	0.684	0.777	0.500	0.530
	(6.13)	(6.21)	(4.46)	(4.19)
Unknown-Proper	1.180	1.119	0.899	0.777

Continued on next page...

... table 4.13 continued

	Model 1		Model 2	
	OLS	Random-Effects	OLS	Random-Effects
	(7.77)	(7.07)	(5.88)	(4.86)
Unknown-Agglomeration	0.230	0.346	0.277	0.362
	(0.80)	(1.22)	(0.97)	(1.29)
Unknown-Metro. Area	-0.416	-0.384	-0.292	-0.293
	(-1.12)	(-1.04)	(-0.79)	(-0.80)
Proper-Unknown	1.590	1.508	1.202	1.027
	(6.88)	(6.48)	(5.19)	(4.39)
Proper-Proper	0.362	0.347	-0.007	-0.089
	(3.64)	(3.08)	(-0.07)	(-0.76)
Proper-Agglomeration	1.602	1.556	1.520	1.413
	(5.27)	(5.20)	(5.03)	(4.76)
Agglomeration-Unknown	-1.016	-0.932	-0.857	-0.784
	(-1.97)	(-1.84)	(-1.67)	(-1.56)
Agglomeration-Proper	-0.348	-0.293	-0.494	-0.465
	(-0.70)	(-0.60)	(-1.00)	(-0.96)
Agglomeration-Metro. Area	1.235	0.988	1.463	1.212
	(1.53)	(1.25)	(1.83)	(1.54)
Metro. Area-Metro. Area	0.056	-0.003	0.314	0.294
	(0.18)	(-0.01)	(0.99)	(0.85)
Others-Others	3.462	3.445	2.714	2.608
	(5.00)	(4.79)	(3.93)	(3.62)
Constant	0.934	0.817	1.958	2.050

Continued on next page...

... table 4.13 continued

	Model 1		Model 2	
	OLS	Random-Effects	OLS	Random-Effects
	(6.68)	(5.30)	(12.06)	(11.40)
σ_u		0.993		1.032
		(21.96)		(23.11)
σ_ε		3.037		3.001
		(131.42)		(130.90)

Table 4.14: Regressions with observed urban vital rates, cities with such information available

	Model 1		Model 2	
	OLS	Random-Effects	OLS	Random-Effects
Urban TFR	0.525	0.564	0.428	0.465
	(5.91)	(6.16)	(4.87)	(5.13)
Urban Q5	0.003	0.004	0.002	0.002
	(1.36)	(1.52)	(0.86)	(0.96)
Inland Water	0.201	0.172	0.303	0.283
	(1.86)	(1.50)	(2.82)	(2.49)
LECZ	-0.465	-0.404	-0.419	-0.380
	(-3.14)	(-2.53)	(-2.87)	(-2.43)
Dry subhumid	-0.355	-0.315	-0.310	-0.285
	(-2.19)	(-1.87)	(-1.94)	(-1.72)

Continued on next page...

... table 4.14 continued

	Model 1		Model 2	
	OLS	Random-Effects	OLS	Random-Effects
Semiarid	-0.406	-0.341	-0.248	-0.202
	(-2.60)	(-2.08)	(-1.61)	(-1.26)
Arid and above	-0.418	-0.487	-0.278	-0.338
	(-1.64)	(-1.83)	(-1.11)	(-1.30)
LECZ * Dry subhumid	0.716	0.635	0.644	0.586
	(2.42)	(2.04)	(2.22)	(1.92)
LECZ * Semiarid and above	0.434	0.393	0.335	0.309
	(1.28)	(1.07)	(1.00)	(0.87)
100,000–500,00			-1.124	-1.061
			(-9.04)	(-8.38)
500,000–1 million			-1.059	-1.076
			(-4.83)	(-4.89)
Over 1 million			-1.298	-1.261
			(-5.50)	(-5.19)
Unknown-Unknown	1.012	1.306	0.846	1.092
	(4.91)	(6.10)	(4.03)	(5.03)
Unknown-Proper	0.439	0.655	0.063	0.243
	(1.87)	(2.81)	(0.27)	(1.03)
Unknown-Agglomeration	-0.056	0.408	-0.138	0.278
	(-0.12)	(0.94)	(-0.30)	(0.64)
Unknown-Metro. Area	-0.101	0.183	0.082	0.305
	(-0.12)	(0.23)	(0.10)	(0.39)

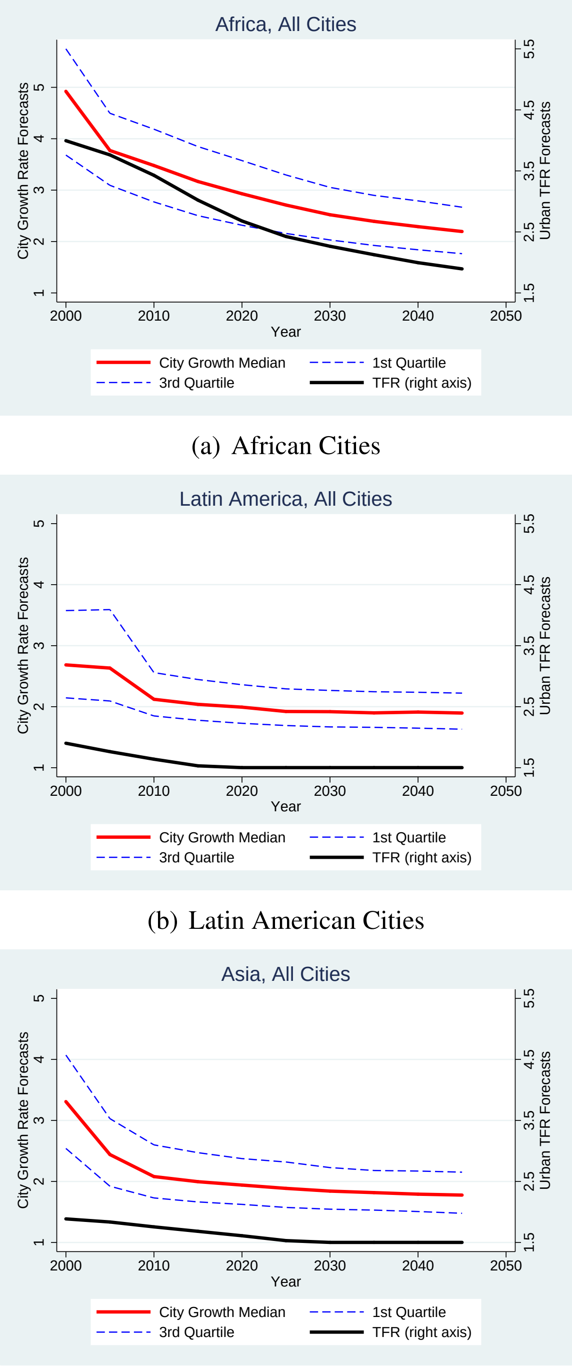
Continued on next page...

... table 4.14 continued

	Model 1		Model 2	
	OLS	Random-Effects	OLS	Random-Effects
Proper-Unknown	0.154	0.236	-0.052	-0.001
	(0.14)	(0.22)	(-0.05)	(-0.00)
Proper-Proper	0.189	0.187	-0.369	-0.350
	(1.09)	(1.05)	(-1.94)	(-1.80)
Proper-Agglomeration	0.392	0.374	0.455	0.395
	(0.64)	(0.63)	(0.75)	(0.67)
Agglomeration-Unknown	0.285	1.112	0.644	1.278
	(0.12)	(0.51)	(0.27)	(0.59)
Agglomeration-Proper	0.019	0.229	-0.719	-0.493
	(0.02)	(0.23)	(-0.73)	(-0.51)
Agglomeration-Metro. Area	0.338	0.341	0.403	0.426
	(0.36)	(0.39)	(0.44)	(0.49)
Metro. Area-Metro. Area	0.207	0.320	0.302	0.405
	(0.47)	(0.69)	(0.68)	(0.88)
Constant	1.212	0.991	2.547	2.319
	(4.97)	(3.90)	(8.87)	(7.76)
σ_u		1.424		1.347
		(17.02)		(15.51)
σ_ε		1.945		1.946
		(34.28)		(34.17)

Table 4.12: Fixed-effects city growth regression models, by region. Models without controls for city size.

	All	Africa	Latin America	Asia
Start-of-period Urban TFR	1.231	0.703	1.092	1.673
	(20.24)	(4.37)	(9.28)	(19.00)
Start-of-period Urban Q5	-0.010	0.013	-0.003	-0.024
	(-6.25)	(3.61)	(-0.63)	(-12.02)
Unknown-Unknown	1.227	0.463	0.105	1.535
	(4.73)	(0.61)	(0.29)	(4.37)
Unknown-Proper	1.079	1.235	0.637	0.954
	(4.02)	(1.73)	(1.73)	(2.57)
Unknown-Agglomeration	0.940	0.432	0.232	1.112
	(2.83)	(0.44)	(0.57)	(2.39)
Unknown-Metro. Area	0.077	-0.066	-1.055	1.561
	(0.17)	(-0.02)	(-2.47)	(1.21)
Proper-Unknown	1.343	1.191	1.277	0.311
	(4.18)	(1.39)	(3.36)	(0.50)
Proper-Proper	0.283	0.126	0.333	-0.068
	(1.16)	(0.21)	(0.96)	(-0.20)
Proper-Agglomeration	1.567	2.333	0.907	1.194
	(4.66)	(3.24)	(1.34)	(2.60)
Agglomeration-Unknown	-0.584	-1.378	1.147	-0.633
	(-1.06)	(-1.33)	(1.49)	(-0.59)
Agglomeration-Proper	-0.145	-3.001	1.859	0.427
	(-0.27)	(-2.20)	(1.41)	(0.66)
Agglomeration-Metro. Area	0.336	-0.051	-0.724	1.921
	(0.38)	(-0.02)	(-0.93)	(1.28)
Metro. Area-Metro. Area	-0.172	1.668	-0.995	0.364
	(-0.32)	(0.68)	(-2.09)	(0.40)
Others-Others	1.259	0.000	0.173	1.334
	(1.03)	(.)	(0.08)	(1.08)
Constant	0.119	-0.338	0.611	0.002
	(0.53)	(-0.45)	(2.07)	(0.01)



(a) African Cities

(b) Latin American Cities

(c) Asian Cities

Figure 4.6: Forecasts of city growth rates by region, conditional on UN projections of fertility and mortality

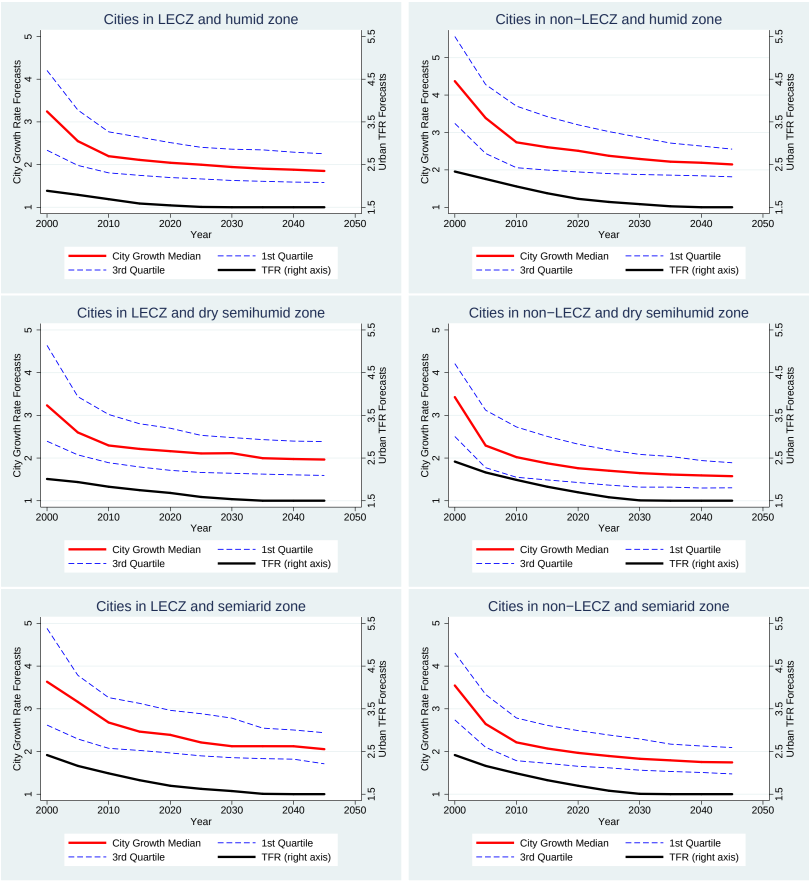


Figure 4.7: Forecasts of city growth rates by LECZ and aridity.

4.B Data sources

Table 4.15 summarizes sources of datasets used in this study which come from various sources with various data type. First of all, in the table, data type falls into data set in the form of tabular data (which is the “traditional” data storage method in a computer) and in the form of geographical data with geographic information data (GIS). Except for city populations, national and urban TFR and Q5 data, all the datasets used are geographically coded data. What follows is some explanations of each dataset.

City populations come from the United Nation’s cities data (United Nations, 2008b) which the United Nation Population Division gathers from each nation and uses for World Urbanization Prospects: The 2007 Revision.

City footprint is the urban extents of the Global Rural-Urban Mapping Project (Balk, 2009) by CIESIN at Columbia university. It is an attempt to delineate physical boundaries of urbanized area globally and the main input is the 1994-1995 stable nighttime light data obtained from the satellite image by the United States Air Force Defense Meteorological Satellite Program (DMSP) Operational Linescan System (See Small et al., 2005; Small, 2005: for the details).

Population grid [4] contains, in each cell of the grid, the number of people residing in the grid cell in 2000. The cell size is 1 kilometer by 1 kilometer (30-arc minute resolution). It is created by a mass-conserving algorithm called *GRUMPe* (Global Rural Urban Mapping Programme) with three input spatial data; human settlements data, urban extents data (mentioned above), administrative units data (See Balk, Pozzi, Yetman, Deichmann, and Nelson, 2005: for more details) Land area grid contains information on land area in each grid cell. (See Deichmann et al., 2001: for more details)

Inland Water grid delineates inland water system (i.e. lakes, reservoirs, rivers and so on) at 30-second resolution. It is the Global Lakes and Wetlands Database level 3 (GLWD-3) data created by (Lehner and Döll, 2004)

Low elevation coastal zone (LECZ) is defined as land area contiguous with the coastline up to a 10-metre rise elevation. It is based on the measure from the Shuttle Radar Topography Mission (SRTM) elevation data set. In some places, mostly the mouths of major rivers such as the Amazon in Brazil and the Yenisey river in Russia, the LECZ extends well beyond 100 kilometres inland, although for most of its extent, the zone is much less than 100 kilometres in width (McGranahan, Balk, and Anderson, 2007).

Finally, Aridity zones grid defines dryness of land. The main input datasets are rainfall

[4] Grid, as one of spatial data model, means Arc/INFO GRID format raster. Each grid cell has a certain value. All the grids used in this study has resolution a 30-arc second (i.e. each grid cell size represents 1 km by 1 km). `http://support.esri.com/index.cfm?fa=knowledgebase.techarticles.articleShow&d=30616`

Table 4.15: Sources of data

Data	Data type	Data source
City population	Tabular	World Urbanization Prospects (WUP) UN Population Division, United Nations (2008b)
City footprint	Geocoded	Global Rural-Urban Mapping (GRUMP) CIESIN at Columbia university
Population grid	Geocoded	Global Rural-Urban Mapping (GRUMP) Balk (2009)
Land area grid	Geocoded	Global Rural-Urban Mapping (GRUMP) Balk et al. (2005)
National TFR and Q5	Tabular	World Population Prospects (WPP) UN Population Division (http://esa.un.org/unpp/)
Urban TFR and Q5	Tabular	DHS*, WFS**, and Other demographic surveys For DHS, http://www.measuredhs.com/ For India, http://populationcommission.nic.in/birth.htm
Inland Water	Geocoded	Global Lakes and Wetlands Database (GLWD) level 3 Lehner and Döll (2004)
LECZ***	Geocoded	Global Rural-Urban Mapping (GRUMP) (McGranahan, Balk, and Anderson, 2007)
Aridity zones	Geocoded	World Atlas of Desertification UN Environment Programme (UNEP)

* Demographic and Health Survey

** World Fertility Survey

*** Low Elevation Coastal Zone

datasets (which are used for measuring mean annual precipitation, P) and, among other things, temperature datasets (used for measuring mean annual potential evapotranspiration, PET) over the period 1951-1980. Using the Aridity Index (AI), calculated as the ratio P/PET (mean annual potential moisture availability), (Middleton, Thomas, and UNEP, 1997) defines dryland as the following three aridity zones: (1) hyperarid ($AI <= 0.05$), (2) arid ($0.05 < AI <= 0.20$), (3) Semiarid ($0.20 < AI <= 0.50$), and (4) Dry subhumid ($0.50 < AI <= 0.65$).

4.C Geo-processing algorithms

GIS and geospatial programming is used to calculate city-level population and land area in total and in each ecozone for all countries worldwide. We will first look at an algorithm to calculate city-level population and land area in total and then one which takes into account each ecozone for the calculation. We used Python scripting languages with a Python module which ESRI's ArcGIS program provides for geoprocessing.

For the calculation of city-level population and land area, the input spatial data used are GRUMP's urban extents features, population, and land area grids. The key geo-processing for the calculation is to summarize values of population and land area, respectively, within cells which fall into each urban extent. The geo-processing is called the *zonal statistics*.

To use zonal statistics which is possible with raster-type (e.g. grid) spatial data, it is necessary to convert the urban extents in the form of vector data into a grid-type data in which each cell value represents urban ID for each urban extent. With the resulting urban extent grid, the existing population and land area grids, we can calculate zonal statistics of population and land area respectively.

What follows is the details on the geo-processing algorithm for city-level population and land area calculation: Assume there are *N* countries,

Step 1. Convert the existing urban extents data in the form of vector into a raster-type (i.e. grid) data in which cell values are urban IDs representing urban extents.

Step 2. For population, use the urban extents grid resulting from step 1 and the existing population grid to summarize values of population in cells falling into each urban extent (Zonal statistics for population).

Step 3. For land area, use the urban extents grid resulting from step 1 and the existing land area grid to summarize values of land area in cells falling into each urban extent (Zonal statistics for land area).

Step 4. Join the two resulting tables from step 2 and 3 by urban ID.

Step 5. Repeat from step 1 to step 3 for all the *N* countries.

Step 6. Finally, merge all the *N* country tables into a single table.

Along with the city-level population and land area in total, we also need city-level population and land area falling in each ecozone. This analysis is useful both to identify population and land area for each ecozone and to identify if a city is located in each ecozone or not. In this study, we use inland water, drylands (by its sub-category), and LECZ (low elevation costal zone). The key geo-processing for this analysis is, for example, to keep population values only in the cells falling in the ecozone. In other words, if a cell of population grid is outside of the ecozone, the cell has no value. This is called the *conditioning*. With the resulting new population grid conditioned, we can calculate zonal statistics with the urban extents grid above.

What follows is the details on the geo-processing algorithm to calculate city-level population and land area in each ecozone: Assume there are *N* countries and *k* global-level ecozone grids,

Step 1. Convert the existing urban extents data in the form of vector into a raster-type (i.e. grid) data in which cell values are urban IDs representing urban extents.

Step 2. Use the population grid and an ecozone grid to create a new *population_within_ecozone* grid which has values only in the cells falling in the ecozone (*Conditioning*).

Step 3. Use the land area grid and an ecozone grid to create a new *landarea_within_ecozone* grid which has values only in the cells falling in the ecozone (*Conditioning*).

Step 4. Use the urban extents grid resulting from step 1 and the *population_within_ecozone* grid from Step 2 to summarize population values in cells falling into each urban extent (Zonal statistics).

Step 5. Use the urban extents grid resulting from step 1 and the *landarea_within_ecozone* grid from Step 2 to summarize values of land area in cells falling into each urban extent (Zonal statistics).

Step 6. Repeat Step 2 and 5 for all the *k* econzones.

Step 7. Join the all the resulting tables by urban ID to create a country table.

Step 8. Repeat from step 1 to step 7 for all the *N* countries.

Step 9. Finally, merge all the *N* country tables into a single table.

Chapter 5

Conclusion

This last chapter concludes an international-level urbanization study in developing countries. Evidently, this study identified various unresolved issues on developing-country urbanization research ranging in data, modeling, and methodology. After concluding remarks in Section 5.1, Section 5.2 briefly discusses limitations of the study along with suggestions for future studies.

5.1 Concluding Remarks

Spatial econometrics serves as useful statistical tools for analysis of spatial dependence including spillovers, spatial interactions, social network effects, and peer effects. Thus, spatial econometrics has been increasing used in economic, demographic, political and social studies. However, it remains methodologically and computationally challenging especially with large-scale geospatial data since it requires specifying multidirectional relations among spatial units. In this thesis, I developed Bayesian MCMC estimation and forecasting methods of a panel data spatial econometric model when panel data are irregularly-spaced in the time dimension and, using a newly assembled cities database which made possible with geospatial data and analysis, applied to forecasting city population growth in developing countries to answer some questions.

United Nations forecasts of urban population growth suggest that over the quarter-century from 2000 to 2025, low- and middle-income countries will see a net increase of some 1.6 billion people in their cities and towns, a quantity that vastly outnumbers the expected rural population increase in these countries and which dwarfs all anticipated growth in the high-income countries (United Nations, 2008a). In the quarter-century after 2025, the UN foresees the addition of another 1.7 billion urban-dwellers to the populations of low- and middle-income countries, with the rural populations of these countries forecast to be on the decline. Where, precisely, will this massive urban growth take place? it is likely to be located in the regions of poor countries that would appear to be environmentally secure, or in regions likely to feel the brunt of climate-related change in the coming decades?

In Asia, where a large share of the world's urban population growth is currently taking place, the cities in the low-elevation zone have grown faster to date than have those outside the zone. To explore the longer-term prospects, we have presented preliminary city population growth forecasts which suggest that rates of city growth are likely to decline as fertility rates decline, and which indicate that cities in the LECZ will eventually come to grow at about the same rates as elsewhere. Of course, the data and methods used to produce such forecasts need to be developed in much more depth. In particular, a way will need to be found to adjust the city growth estimates and forecasts to incorporate migration, which is largely induced by spatial differences in real standards of living. Historically, the lower transport costs provided by the LECZ have proven to be a powerful force attracting migrant labor and capital; in China and elsewhere, it remains to be seen whether climate change will introduce risks that offset the economic logic that has driven coastal development for millennia. Here as elsewhere, the adaptation policies and investments adopted by national and local governments will have a key role in shaping urban growth.

In the arid regions known as drylands, climate change will be manifested in complex ways, but it seems probable that in many places the net effect will be to increase water stress. The consequences are difficult to foresee, and as with coastal settlement, will depend in part on how people and their governments respond to scarcity. The drylands occupy substantially more land overall than the LECZ, and although population densities are generally lower, a larger share of urban dwellers live in drylands than in the low-elevation zone. There is also considerable variation in the dryland shares according to region. Our preliminary city growth results indicate that in Asia, Africa and Latin America, dryland city populations are growing significantly slower than is the case in other zones, although it seems that dryland cities which are also in the LECZ tend to grow somewhat faster. These findings will need to be revisited as data and methods improve.

5.2 Suggestions for Future Studies

This dissertation dealt with the population side of urbanization process, aiming at developing international-level estimation and forecasting methods of city growth, especially in developing countries. It identifies many unresolved issues and leaves them to future studies. These issues should be addressed to better understand urbanization patterns and trends, and their impacts on various aspects of our life and environments surround us.

As shown, it is challenging. In addition, Other measures of the population side such as population density should be considered about its implications on urbanization. Also, in a more broad sense, simultaneous approach considering multi-dimensions of urbanization process should be studied to have systematic analysis of complex urbanization process.

Measurement Issues

Measuring city populations Unless city populations are measured uniform definition of “city”, we cannot trace genuine urbanization patterns and trends. Since urbanization is closely related to various social, demographic, economic, environmental, and health issues, this study re-emphasize the issue. Urban researchers should pay particular attention to the issue of how to classify and measure human settlements as environments where we human live.

In that sense, The United Nations efforts for the issue are rather disappointing. The UN recognized the importance of urbanization trends and, over the 40 years, has been publishing the *World Urbanization Prospects* (which is now published bi-annually). By doing so, the UN has contributed global urbanization research. However, as shown in Section 3.1 in this study, more concerns and efforts among urban researchers, international organizations, and developing countries should be made on urbanization issues, especially on its data issue.

Active and innovative research on the issue is underway, utilizing new technologies like remote sensing and GIS. This innovative approach is one of research areas where international and multidisciplinary research including international and national multi-organizations is necessary. See, among others, Champion and Hugo (2004), de Sherbinin et al. (2002), and Hasse (2004) for more details.

One of recent interesting interesting project is e-Geopolis [1]. The project defines that "Urban agglomeration is a continuous built-up area where at least 10.000 inhabitants live. The continuity is defined by a maximum distance of 200 meters between two constructions." Using Google Earth and other satellite images, e-Geopolis identifies and draws physical extents for urban agglomerations. This project is still underway.

Modeling and Methodology issues

Modeling Limitations of information (i.e. data) give restrictions on modeling. I believe that the UN's current method based on simple deterministic mathematical interpolation and extrapolation was the best method at that time for international-level urban population projections. However, under rapid structural change including demographic transition, it is already shown that the UN's method exhibit systematic upward bias, especially in

[1] `http://www.e-geopolis.eu/`

developing countries which have been experiencing rapid change. New modeling was required.

As one of alternatives to the UN's method, this study proposed econometric modeling which is probabilistic, and takes into account of factors promoting and hindering city growth. Though this study put enormous efforts to data gathering and integration even using GIS data and geospatial programming, the estimation and forecasting results show that more information is necessary for more reliable forecasts of city growth and population. Research on factors affecting city growth should be continued. Case study is also suggested, from which we can learn lessons. Also, the results from Random-effects specification show that population time-series is too short to estimate latent city-specific effects which shrinks toward zero. New additions of population data from 2010 census are expected.

Model specification This study remains also unresolved issues on model specification. For instance, model coefficients from regional-specific regressions are different from those from the pooled regression. The effect of urban fertility rate vary across region on magnitude though it has positive, strongly significant effect on city growth. The LECZ effect are also different by region. This model specification might be tested along with model specification.

Model validation This study leaves model validation to one of future studies. As explained above, this study has some issues addressed before evaluating its performance. When it is ready, it would be one way to compare our results with the UN forecasts which is the only one comparable to our results. Though the UN forecasts are biased, there are no other city growth or population forecasts available for developing countries. In-sample predictions is another way to validating our various models though the in-sample prediction also is not straightforward due to unbalancedness of the city population data.

Bibliography

Karim M. Abadir and Jan R. Magnus. *Matrix Algebra*. Cambridge University Press, 2005.

Susana B. Adamo and Alexander de Sherbinin. The impact of climate change on the spatial distribution of populations and migration. Center for International Earth Science Information Network (CIESIN), Earth Institute at Columbia University, New York. Report prepared for the United Nations Population Division. January 2008, 2008.

E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

Luc Anselin. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988.

Luc Anselin. Spatial econometrics. A working paper, Center for Spatially Integrated Social Science, 1999.

Luc Anselin. Spatial econometrics in RSUE: Retrospect and prospect. *Regional Science and Urban Economics*, 37:450–456, 2007.

Hazel Ashurst, Sundat Balkaran, and John B. Casterline. Socio-economic differentials in recent fertility. World Fertility Survey *Comparative Studies* no. 42. Voorburg, Netherlands: International Statistical Institute, 1984.

Bettina Aten. Evidence of spatial autocorrelation in international prices. *Review of Income and Wealth*, pages 149–163, 1996.

Cynthia Brenda Awuor, Victor Ayo Orindi, and Andrew Ochieng Adwera. Climate change and coastal cities: The case of mombasa, kenya. *Environment and Urbanization*, 20(1): 231–242, 2008.

Pietro Balestra and Marc Nerlove. Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica*, 34:585–612, 1966.

Deborah Balk. More than a name: Why is urban population mapping a grumpy proposition? In P. Gamba and M. Herold, editors, *Global Mapping of Human Settlement: Experiences, Data Sets, and Prospects*, pages 145–161. Tayor and Francis, 2009.

Deborah Balk and Gregory Yetman. The global distribution of population: Evaluating the gains in resolution refinement. Working Paper, Center for International Earth Science Information Network (CIESIN), Columbia University, Available at: `http://sedac.ciesin.columbia.edu/gpw/docs/gpw3_documentation_final.pdf`, 2004.

Deborah Balk, Francesca Pozzi, Gregory Yetman, Uwe Deichmann, and Andy Nelson. The distribution of people and the dimension of place: Methodologies to improve the global estimation of urban extents. In *Proceedings of the Urban Remote Sensing Conference of the International Society for Photogrammetry and Remote Sensing*, 2005.

Badi H. Baltagi. On seemingly unrelated regressions with error components. *Econometrica*, 48(6):1547–1551, 1980.

Badi H. Baltagi and Young-Jae Chang. Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model. *Journal of Econometrics*, 62:67–89, 1994.

Badi H Baltagi and Dong Li. Prediction in the panel data model with spatial correlation. In Raymond J.G.M. Florax Luc Anselin and Sergio J. Rey, editors, *Advances in Spatial Econometrics: Methodology, Tools and Applications*, pages 283–295. Springer-Verlag, 2004.

Badi H. Baltagi and Qi Li. Prediction in the one-way error component model with serial correlation. *Journal of Forecasting*, 11:561–567, 1992.

Badi H. Baltagi and P.X. Wu. Unequally spaced panel data regressions with $ar(1)$ disturbances. *Econometric Theory*, 15:814–823, 1999.

Badi H. Baltagi, Peter Egger, and Michael Pfaffermayr. A generalized spatial panel data model with random effects. Presented in IZA seminar, 2006.

Badi H. Baltagi, Peter Egger, and Michael Pfaffermayr. Estimating models of complex FDI: Are there third-country effects? *Journal of Econometrics*, 140:260–281, 2007.

Badi H. Baltagi, Georges Bresson, and Alain Pirotte. Forecasting with spatial panel data. Technical report, IZA Discussion Paper No. 4242, 2009.

Sudipto Banerjee, Bardley P. Carlin, and Alan E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, 2004.

World Bank. *Climate-Resilient Cities: 2008 Primer*. Washington, DC, June 2008.

V. A. Barker, L. S. Blackford, J. Dongarra, J. Du Croz, S. Hammarling, M. Marinova, J. Waśniewski, and P. Yalamov. *LAPACK95 Users' Guide*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

Ronald Barry and R. Kelley Pace. A Monte Carlo estimator of the log determinant of large sparse matrices. *Linear Algebra and Its Applications*, 289(1–3):41–54, 1999.

Patricia E. Beeson, David N. DeJong, and Werner Troesken. Population Growth in U.S. Countries, 1840-1990. *Regional Science and Urban Economics*, 31:669–699, 2001.

Duncan Black and Vernon Henderson. Urban evolution in the usa. *Journal of Economic Geography*, 3:343–372, 2003.

Anne C. Case, Harvey S. Rosen, and Jr James R. Hines. Budget spillovers and fiscal policy interdependence. *Journal of Public Economics, 52, 1993, 285-307*, 52:285–307, 1993.

Tony Champion and Graeme Hugo, editors. *New Forms of Urbanization: Beyond the UrbanRural Dichotomy*. Ashgate, Aldershot, 2004.

Nancy Chen, Paolo Valente, and Hania Zlotnik. What do we know about recent trends in urbanization? In Richard E. Bilsborrow, editor, *Migration, Urbanization, and Development: New Directions and Issues*, pages 59–88. United Nations Population Fund (UNFPA), New York, 1998.

Siddartha Chib and Bradley P. Carlin. On MCMC Sampling in Hierarchical Longitudinal Models. *Statistics and Computing*, 9:17–26, 1999.

Siddhartha Chib. Inference in panel data models via Gibbs sampling. In László Mátyás and Patrick Sevestre, editors, *The Econometrics of Panel Data: A Handbook of the Theory with Applications Series*. Kluwer Academic Publishers, 1996.

Andrew David Cliff and J. Keith Ord. The problem of spatial autocorrelation. In A. J. Scott, editor, *London Papers in Regional Science, Studies in Regional Science*, page pp. 2555. Pion, London, 1969.

Andrew David Cliff and J. Keith Ord. *Spatial Processes: Models and Applications*. Pion, London, 1981.

Daniel da Mata, Uwe Deichmann, J. Vernon Henderson, Somik V. Lall, and Hyoung Gun Wang. Determinants of city growth in brazil. *Journal of Urban Economics*, 62:252–272, 2007.

Alex de Sherbinin, Deborah Balk, Karina Yager, Malanding Jaiteh, Francesca Pozzi, Chandra Giri, and Antroinette Wannebo. A CIESIN thematic guide to social science applications of remote sensing, 2002. URL `http://sedac.ciesin.columbia.edu/tg/guide_main.jsp`.

Uwe Deichmann, Deborah Balk, and Gregory Yetman. Transforming population data for interdisciplinary usages: From census to grid. Unpublished manuscript available on-line at: http://sedac.ciesin.columbia.edu/plue/gpw/GPWdocumentation.pdf, 2001.

David Dodman. Urban density and climate change. Technical report, Paper prepared for United Nations Population Fund (UNFPA), Analytical Review of the Interaction between Urban Growth Trends and Environmental Changes, December 2008.

Ian Douglas, Kurshid Alam, Maryanne Maghenda, Yasmin McDonnell, Louise McLean, and Jack Campbell. Unjust waters: Climate change, flooding, and the urban poor in africa. *Environment and Urbanization*, 20(1):187–205, 2008.

Peter Egger, Michael Pfaffermayr, and Hannes Winner. An unbalanced spatial panel data approach to us state tax competition. *Economics Letters*, 88:329–335, 2005.

J. Paul Elhorst. Specfication and estimation of spatial panel data models. *International Regional Sicence Review*, 26(3):244–268, 2003.

Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–511, 1992.

Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

John Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press, 1992.

Edward L. Glaeser and Jesse M. Shapiro. City growth and the 2000 census: Which places grew, and why. Brooking Center on Urban and Metropolitan Policy, May 2001.

Edward L. Glaeser, Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer. Growth in cities. *Journal of Political Economy*, 100(6):1126–1152, 1992.

Arthur S. Goldberger. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57(298):369–375, 1962.

Daniel A. Griffith. *Some guidelines for specifying the geographic weights matrix contained in spatial statistical models*, chapter 4, pages 65–83. CRC Press, 1996.

Jorgelina Hardoy and Gustavo Pandiella. Urban poverty and vulnerability to climate change in latin america. *Environment and Urbanization*, 21(1):203–224, 2009.

John Hasse. Shift in paradigm needed for urban spatial-temporal analysis and modeling. Panel Contribution to the PERN Cyberseminar on Urban Spatial Expansion, 2004. URL `http://www.populationenvironmentresearch.org/seminars112004.jsp`.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 51(1):97–109, 1970.

J. V. Henderson. The sizes and types of cities. *The American Economic Review*, 64(4): 640–656, 1974.

J. Vernon Henderson. Urbanization and growth. In Steven N. Durlauf By Philippe Aghion, editor, *Handbook of Economic Growth, Volumn 1*, chapter 24. Elsevier, 2005.

Cheng Hsiao. Panel data analysis – advantages and challenges. *Test*, 16:1–22, 2007.

George G. Judge, W. E. Griffiths, R. Carter Hill, and Helmut Lutkepohland Tsoung-Chao Lee. *Ithe Theory and Practice of Econometrics*. John Wiley, New York, 2nd edition edition, 1985.

Mudit Kapoor, Harry H. Kelejian, and Ingmar R. Prucha. Panel data models with spatially correlated error components. *Journal of Econometrics*, 140:97–130, 2007.

Harry H. Kelejian and Ingmar R. Prucha. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40(2): 509–533, 1999.

Erricos John Kontoghiorghes, editor. *Handbook of Parallel Computing and Statistics*. CRC Press, 2006.

Gary Koop, Dale J. Poirier, and Justin L. Tobias. *Bayesian Econometric Methods*. Cambridge University Press, 2007.

Sari Kovats and Rais Akhtar. Climate, climate change and human health in asian cities. *Environment and Urbanization*, 20(1):165175, 2008.

Lung-Fei Lee and William E. Griffiths. The prior likelihood and best linear unbiased prediction in stochastic coefficient linear models. *Working Papers in Econometrics and Applied Statistics No. 1, University of New England*, 1979.

Youngjo Lee. *Generalized Linear Models with Random Effects; Unified Analysis via h-likelihood*. Chapman & Hall/CRC, 2006.

Roger TH.A.J. Leenders. Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24:21–47, 2002.

Bernhard Lehner and Petra Döll. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, 296:1–22, 2004.

Cyrus C. MacDuffee. *The Theory of Matrices*. Chelsea, New York, corrected reprint of first edition in 1946 edition, 1933.

Jan R. Magnus. Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. *Journal of Econometrics*, 19:239–285, 1982.

Gordon McGranahan, Deborah Balk, and Bridget Anderson. The rising tide: Assessing the risks of climate change to human settlements in low-elevation coastal zones. *Environment and Urbanization*, 19(1):17–37, 2007.

David J. Mckenzie. Estimation of AR(1) models with unequally spaced pseudo-panels. *Econometrics Journal*, 4:89–108, 2001.

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

Nick Middleton, David Thomas, and UNEP. *World Atlas of Dessertification*. UNEP, 1997.

Mark R. Montgomery. The urban transformation of the developing world. *Science*, 319: 761–764, 2008.

Mark R. Montgomery and Deborah Balk. The urban transition in developing countries: Demography meets geography. In E. Birch and S. Wachter, editors, *Global Urbanization in the 21st Century*. University of Pennsylvania Press, 2008. forthcoming.

Mike Muller. Adapting to climate change: Water management for urban resilience. *Environment and Urbanization*, 19:99–113, 2007.

Gonzalo Navarro. A guided tour to approximate string matching. *ACM Computing Surveys*, 33(1):31–88, 2001.

Keith Ord. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70:120–126, 1975.

R. Kelley Pace and James P. LeSage. Closed-form maximum likelihood estimates for spatial problems. Department of Finance, Louisiana State University, Baton Rouge, LA, 2000.

R. Kelly Pace and James P. LeSage. Chebyshev approximation of log-determinants of spatial weight matrics. *Computational Statistics & Data Analysis*, 45:179–196, 2004.

Nathaniel L. Bindoff Zhenlin Chen Amnat Chidthaisong Pierre Friedlingstein Richard B. Alley, Terje Bertsen et al. Summary for policymakers: Contribution of working group i to the fourth assesment report. Technical report, Intergovernmental Panel on Climate Change, 2007. available at http://www.ipcc.chaccessed 7 November 2007.

Uriel Safriel, Zafar Adeel, David Niemeijer, Juan Puigdefabregas, Robin White, Rattan Lal, Mark Winslow, Juliane Ziedler, Stephen Prince, Emma Archer, Caroline King, Barry Shapiro, Konrad Wessels, Thomas Nielsen, Boris Portnov, Inbal Reshef, Jillian Thonell, Esther Lachman, and Douglas McNab. Dryland systems. In Robert Scholes Rashid Hassan and Neville Ash, editors, *Ecosystems and Human Well-being: Current State and Trends*, chapter 22. Island Press, Washington, DC, 2005.

Rainer Schnell, Tobias Bachteler, and Stefan Bender. A toolbox for record linkage. *Austrian Journal of Statistics*, 33(1&2):125–133, 2004.

Jesse M. Shapiro. Smart cities: Quality of life, productivity, and the growth effects of human capital. *Review of Economics and Statistics*, 88:324–335, 2006.

Thomas S. Shively. Testing for autoregressive disturbances in a time series regression with missing observations. *Journal of Econometrics*, 57:233–255, 1993.

Kate B. Showers. Water scarcity and urban Africa: An overview of urban-rural water linkages. *World Development*, 30(4):621–648, 2002.

Christopher Small. The global analysis of urban reflectance. *International Journal of Remote Sensing*, 26(4):661–681, 2005.

Christopher Small, Francesca Pozzi, and C. D. Elevidge. Spatial analysis of global urban extent from dmsp-ols night lights. *Remote Sensing of Environment*, 96:277–291, 2005.

Oleg Smirnov and Luc Anselin. Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach. *Computational Statistics & Data Analysis*, 35:301–319, 2001.

Hisashi Tanizaki. *Computational Methods in Statistics and Econometrics*. CRC, 2004.

Allan J. Taub. Prediction in the context of the variance-components model. *Journal of Econometrics*, 10:103–107, 1979.

Luke Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22: 1701–1762, 1994.

Waldo Tobler, Uwe Deichmann, Jon Gottsegen, and Kelly Maloy. The global demography project. Technical report, Technical Report 95-6, Natiional Center for Geographic Information and Analysis, Santa Barbara, 1995.

UNFPA. *State of World Population 2007: Unleashing the Potential of Urban Growth*. United Nations Population Fund, New York, 2007. George Martine, lead author.

United Nations. *Patterns of Urban and Rural Population Growth*. Number 68 in Population Studies. United Nations, Department of International Economic and Social Affairs, New York, 1980.

United Nations. *World Urbanization Prospects: The 2001 Revision. Special Tabulations*. United Nations, Department of Economic and Social Affairs, Population Division, New York, 2002.

United Nations. *World Urbanization Prospects: The 2003 Revision*. United Nations, Department for Economic and Social Information and Policy Analysis, New York, 2004.

United Nations. *World Urbanization Prospects: The 2007 Revision. Executive Summary*. United Nations, Department of International Economic and Social Affairs, 2008a.

United Nations. *World Urbanization Prospects: The 2007 Revision. CD-ROM Edition - Data in digital form (POP/DB/WUP/Rev.2007)*. United Nations, Population Division, Department of Economics and Social Affairs, New York, 2008b.

United Nations Statistical Division. *Standard Country or Area Codes for Statistical Use, Revision 4 (ST/ESA/STAT/SER.M/49/Rev.4, M.98.XVII.9)*. United Nations, 1999.

Paul R. Voss, David D. Long, Roger B. Hammer, and Samantha Friedman. County child poverty rates in the US: A spatial regression approach. *Population Research and Policy Review*, 25:369–391, 2006.

Tom Wansbeek and Arie Kapteyn. Estimation in a linear model with serially correlated errors when observations are missing. *International Economic Review*, 26:469–490, 1985.

Tom Wansbeek and Arie Kapteyn. Estimation of the error-components model with incomplete panels. *Journal of Econometrics*, 41:341–361, 1989.

Scott L. Zeger and M. Rezaul Karim. Gneralized linear models with random effects; a gibbs sampling approach. *Journal of the American Statistical Association*, 86(413):79–86, 1991.