Small-Area Population Forecasting: A Holistic Theory-Based Spatio-Temporal Approach

Guangqing Chi and Paul R. Voss
Applied Population Laboratory
Department of Rural Sociology
University of Wisconsin-Madison

Extended abstract

Existing demographic forecasting techniques for small areas[1] generally are not guided by theoretical models or demographically-driven conceptual schemes. Most forecasting techniques were developed and refined for relatively large geographic areas (counties and larger areas) where the components of population change are dealt with separately in an age-specific context and a period-cohort perspective. Moreover, non-demographic factors generally are ignored in traditional forecasting methodologies for all areas, large and small. Furthermore, nearly all population forecasting methods ignore the "neighborhood" context[2] in which local demographic trends actually are played out. Thus it is not surprising that existing multiple regression models do not outperform simple extrapolation techniques when forecasting populations for small geographic areas.

We assert, and herein propose a rigorous test of the assertion, that multiple regression approaches to small-area population forecasting can be improved in three integrated ways: (1) by examining, holistically rather than partially, the relationships between population change and relevant non-demographic factors from disciplines normally not involved in formal population forecasting efforts; (2) by explicitly incorporating spatial process effects into the model; and (3) by conducting analyses and forecasts at a fine geographic level where environmental (natural and built) and geophysical effects on population change can be better captured and modeled.

In particular, we propose a theory-based, spatio-temporal approach to small-area population forecasting. The central research question of the proposed study is to test our belief that this approach is superior to existing small-area population forecasting techniques. Several

[1] In this study, small areas refer to any geographic units below the county level. Some examples are minor civil divisions, census tracts, block groups, partial block groups, and blocks.
[2] The neighborhood context is not wholly ignored in traditional models given that controls to higher geography are usually applied. However, our use of the term "neighborhood" anticipates the incorporation into our models of explicit spatial effects.

sub-questions are addressed along the way. First, when reasonably good data are at hand, what particular variables appear to be stronger predictors of population change? Second, in a single forecasting model, how should one best handle the obviously large number of variables derived from several population and non-population related disciplines? Third, does spatial autocorrelation in the variables or in the modeled residuals affect the statistical analysis? Fourth, if so, will the incorporation of the spatial effects into the multiple regression specification improve the forecasting model?

Relevant data from several disciplines are brought together using GIS and related software tools. A regression model incorporating spatio-temporal population effects and other neighbor characteristics is applied to examine population change since 1970 and to prepare a 1990-based forecast of population for 2000 at the minor civil division (MCD) level in Wisconsin. 39 independent variables are used to develop five indices: (1) *demographics* (local demographic characteristics), (2) *accessibility* (transportation infrastructure), (3) *developability* (the potential for land conversion and development), *desirability* (a measure of the area's attractive power from the presence or absence of natural amenities), and (5) *livability* (a measure of socioeconomic conditions suitable for living). The indices are generated by using principal component analysis and the ModelBuilder function in ArcGIS. The proposed approach first estimates the relevance of various covariates for predicting population change for the period 1980-1990 at the MCD level. For each MCD, the population growth rate for 1980-1990 is regressed on its growth rate for 1970-1980, its various characteristics in 1980, growth rates for 1970-1980 among neighboring areas, and neighborhood characteristics in 1980. The estimated regression coefficients and spatial parameters are then used for projecting MCD populations in 2000. The statistical quality of the projections is measured against the 2000 census counts, with the principal comparison being against the state's official MCD extrapolation-based projections for 2000.

The findings strongly support our hypotheses. The five categories of variables are significant in explaining population change. Each of the five indices has strong spatial autocorrelations, and the incorporation of spatial effects improves the overall performance of the estimation model. However, the proposed population forecasting approach does not achieve a remarkable advantage over the extrapolation projection.

Despite that, the proposed approach is necessary for future small-area population forecasting as local change can occur rapidly and existing forecasting models for local areas are almost nonexistent. For one thing, environmental factors almost certainly play an important role in affecting small-area population change, and to the possible extent should be incorporated into forecasting models. Natural disasters such as Hurricane Katrina, the South Asian Tsunami, and viral pandemics such as Avian Flu and SARS likely can never be predicted despite the immense need to make inroads in that direction. However, some environmental influences on populations (e.g., wetlands, land slope and elevation, lakes, and forests) can be measured, and their relationships to population and population change can be parameterized and brought explicitly into forecasting models.

In addition, political contexts, locally, such as "Smart Growth" legislation, zoning, comprehensive land use policies, minimum lot size and setbacks, and such matters as growth policy and migration restrictions can, and should, be included in forecasting models. Exclusively demographic perspectives and methodologies are incapable of capturing most non-demographic forces. Finally, spatial effects should be considered in population forecasting. These have been well developed in the fields of econometrics and regional science, but have not yet found their way into the demographic literature. Growth and decline in a place is not independent of change in neighboring areas. This reality must be brought into our forecasting models – recognizing, of course, that potential endogeneity will require attention in model estimation.

This study provides a nascent spatio-temporal approach to population forecasting by taking a theory-based approach, rather than a data-driven approach. The research (1) offers an interdisciplinary theory-based approach to population forecasting, (2) develops indices to handle the large number of explanatory variables and different disciplines' approaches to modeling population change, (3) integrates environmental modeling and population regression forecasting into a single effort, (4) revises conventional spatial econometric models to a spatial-temporal approach for forecasting purposes, (5) builds a Wisconsin place-based longitudinal (1970, 1980, 1990, and 2000) database which contains demographic, social, economic, institutional, environmental, geophysical, infrastructure, and legislative information, and (6) finally and hopefully, better meets policy demands for improved population forecasts.