Small-area Population Forecasting: Borrowing Strength across Space and Time

Guangqing Chi1* and Paul R. Voss2
1Department of Sociology and Social Science Research Center, Mississippi State University, USA
2Odum Institute for Research in Social Science, The University of North Carolina at Chapel Hill, USA

ABSTRACT

Demographic forecasting techniques do not perform particularly well for small areas. In this study we propose a spatio-temporal regression approach for small-area population forecasting that borrows strength from what has happened nearby and what has happened in the past. In particular, a regression model incorporating temporally lagged neighbour growth and neighbour characteristics is applied to examine population change at the minor civil division (MCD) governmental level in Wisconsin, USA since 1960. For each MCD, the population growth rate for 1980–1990 is regressed on its growth rate for 1970–1980, its various characteristics in 1980, neighbour growth rates for 1970–1980, and neighbour characteristics in 1980. The estimated regression coefficients are then used for projecting population in 2000. Accuracy of the forecasts is measured against the 2000 Census counts. The state’s official MCD projections for 2000, which were based on an extrapolation projection – the most often used traditional population forecasting approach for small geographic areas – are taken as the ‘gold standard’ against which improvements through the regression formulations are sought. The projection evaluations reveal mixed results and do not suggest unambiguous preference for the spatio-temporal regression approach or the extrapolation projection. We discuss several reasons for the inability of our spatio-temporal forecasting model to outshine the simple extrapolation forecast. Although this is disappointing, the proposed approach is more solidly theoretically grounded and provides useful information to policy and decision makers at the community level regarding the consequences of various developmental strategies being adopted by themselves as well as by their neighbours. Copyright © 2010 John Wiley & Sons, Ltd.

Received 23 November 2009; revised 1 March 2010; accepted 8 March 2010

Keywords: population forecasting; small area; spatio-temporal regression; neighbour growth; neighbour characteristics

INTRODUCTION

Small-area population forecasts are deemed essential for sound local planning and decision making. However, demographic forecasting techniques generally do not work particularly well for small areas.1 We posit three reasons for this. First, most demographic forecasting models (e.g. the traditional cohort component model) have been developed and refined for relatively large geographic areas (counties and larger areas) where the components of population change can be dealt with separately in an age-specific context. For small areas, the requisite

Earlier versions of this paper were presented at the annual meeting of British Society for Population Studies, 18–20 September 2006, Southampton, UK and the annual North American meeting of the Regional Science Association International, 10–12 November, 2005, Las Vegas, NV.

*Correspondence to: Guangqing Chi, Department of Sociology and Social Science Research Center, Mississippi State University, PO Box C, Mississippi State, MS 39762, USA. E-mail: gchi@ssrc.msstate.edu

Copyright © 2010 John Wiley & Sons, Ltd.
data are simply too thin to support forecasts of this kind. Second, for small sub-county areas (especially in sparsely settled rural landscapes), non-demographic factors, generally ignored in traditional forecasting methodologies, assume a level of importance greater than whatever demographic forces appear to be at work. Environmental constraints and land use restrictions are two examples. Third, nearly all population forecasting methods ignore the ‘neighbourhood’ context in which local demographic trends are cast in ways totally oblivious to the surrounding demographic metabolism.

Keeping these reasons in mind, we revisit the regression approach for population forecasting. The use of familiar standard regression models in the production of population forecasts for sub-county geographic entities has been addressed in many books (e.g. Pittenger, 1976; Smith et al., 2001; Alho and Spencer, 2005) and journal articles (see Wilson and Rees 2005 for a summary of the literature). Yet it is our view that most applied demographers making population forecasts for small areas have largely ignored regression forecasting approaches. One justifiable reason for this appears to be that thus far, standard regression models do not outperform much simpler extrapolation techniques.

We argue that standard regression approaches for small-area population forecasts can be improved by incorporating spatial effects (including explicit population spillovers and legacy effects arising from neighbouring growth rates and characteristics) into the model. In this study, relevant data from several disciplines are brought together using Geographic Information Systems (GIS) and related software tools. Spatio-temporal regression models are specified to estimate the relevance of various covariates for predicting population change. These covariates are then used in parallel specifications to generate the population forecasts. The central research question is whether this approach is superior to existing small-area population forecasting models. Two additional questions are addressed along the way: Does spatial autocorrelation in the variables or in the modelled residuals affect the statistical analysis? If so, will the incorporation of the spatial effects into the standard regression improve both the specification and the population forecasts that result from the approach?

To answer these questions, we use data related to population change at the minor civil division (MCD) level in Wisconsin, USA since 1960. We propose a comprehensive approach for examining the relationships between population change and factors that include not only socio-economic variables but also infrastructure and geophysical factors. For each MCD, the population growth rate for 1980–1990 (the dependent variable) is regressed on several variables covering the 1970–1980 period and stock measures from the 1980 census. Additional variables include those from environmental, local infrastructure, and policy databases. Variables are selected based on both theoretical and empirical significance. Moreover, we introduce and formally test a revised regression specification that brings in explicit ‘neighbour’ influences through spatial regression techniques. Ultimately, this model is expanded to include neighbour growth rates for 1970–1980 (temporally lagged spatial population effects) and neighbour characteristics in 1980 (temporally lagged neighbour characteristics). The regression coefficients are estimated and used to forecast the population growth rate in 1990–2000. Finally, the population forecasts for 2000, thus derived, are compared with actual population counts from the 2000 census to calculate several measures of error. Similar error measures are also calculated for the ‘official’ 2000 population projections made by demographers in the Wisconsin state demographic agency, the Demographic Services Center of the Wisconsin Department of Administration (WIDOA).

In the following sections, we first overview the standard regression approach for population forecasting, address the spatial effects, and propose our spatio-temporal regression approach for small-area population forecasting. Second, we introduce the context of the study and the data that relate to population change at the MCD level in Wisconsin since 1960 and address the troublesome issue of MCD boundary changes over time. Following this is the ‘Analytical Approaches’ section that introduces four population forecasting approaches – extrapolation projection, standard regression, regression with temporally lagged neighbour growth, and finally, a regression specification with both temporally lagged neighbour growth and neighbour characteristics. Here we also discuss some useful adjustment procedures for the population forecasts and
evaluate the four approaches. We conclude with our principal findings followed by a summary and discussion.

THE SPATIO-TEMPORAL REGRESSION APPROACH FOR POPULATION FORECASTING

Existing Regression Forecasting Models

To begin, we define what we mean by regression forecasting. The standard regression model can be expressed in matrix notation as

$$Y = X\beta + \varepsilon$$ (1)

where \(Y\) is a column vector of \(n\) observations on the dependent variable, \(X\) is an \(n\) by \(p\) matrix of independent variables, \(\beta\) is a \(p\) by 1 vector of coefficients, and \(\varepsilon\) is a column vector of \(n\) error terms. A standard regression forecasting model proceeds in two steps. In Step One, it establishes a relationship between the dependent variable and the independent variables by estimating the \(\beta\) vector. Under rather strict assumptions, unbiased and efficient estimates of the vector of parameters can be achieved using the least squares estimator. When the assumptions are not met, alternative estimation procedures (e.g. maximum likelihood, instrumental variables, or generalised method of moments) generally will suffice. The dependent variable is examined and possibly transformed to ensure univariate symmetry; independent variables are carefully chosen so as not to be highly intercorrelated; and for each independent variable, linearity against the dependent variable is examined and possibly corrected via transformation. The model must meet other rather strict assumptions so that the resulting parameter estimates are unbiased and efficient. Our independent variables are fixed observations at time \(t\), or observations over the interval \((t - 10, t)\), and other initial conditions at time \(t + 10\) to forecast population change over the decade \((t + 10, t + 20)\).

$$Y_{t, t+10} = Y_{t, t+10} \hat{\alpha} + X_{t, t+10} \hat{\beta}$$ (3)

The critical assumption is that the relationships between the independent variables and the dependent variable, established in the base period, remain constant over time and can thus be used to forecast change in a decade 10 years beyond the base period where the relationships (the estimated \(\beta\) vector) are established.

While the regression approach for population estimating has been implemented by a number of demographers for small areas, the interest in regression forecasting models appears to be low for three reasons. First, the assumption that the parameters estimated from the regression model based on historic data remain constant into the future is of dubious validity, and surely will hold better for some time periods than others. Second, cohort component methods have long been preferred by demographers for population forecasting, especially for larger units of geography. Third, the most important reason is that regression forecasting models have not practically performed well in the past. We argue, however, that another look at standard regression approaches for small-area population forecasts is warranted because such models can be improved by incorporating neighbour growth and neighbour characteristics.

It is noted that we deliberately exclude six types of models from the regression forecasting specification due to their limitations in small-area population forecasting, although these six approaches have certain strengths and have been used for various purposes of population forecasting as discussed below. The time-series models have been used for population forecasting at national, state, and county levels but not at sub-county levels due to the lack of an appropriate time series (Tayman et al., 2007). The post-censal population estimation models rely on contemporaneous systematic indicators, which are more subject to errors (Swanson and Beck, 1994). The conditional probabilistic models can quantify the uncertainty range and measure the sensitivity of specific alternatives’ sequences, but they are noted for being mechanistic and for ignoring non-demographic factors (Lutz and Goldstein, 2004). Integrated land use models consider a
variety of population change factors and use advanced spatial analysis and statistics techniques but require high data capability and well-grounded expertise in spatial analysis and statistics (Tayman, 1996). The spatial microsimulation methods provide spatially disaggregated microdata that can be aggregated to any geographical level and enable assessment of ‘what-if’ questions, but like the integrated land use models they require high data capability and extensive knowledge of spatial analysis and statistics (Ballas et al., 2005). Grid cell-based population forecasting uses more timely remote sensing data but is expensive to develop and implement (Riahi and Nakicenovic, 2007). The literature covering the six alternative approaches is large. See Chi (2009) and Wilson and Rees (2005) for detailed discussions and references.

Spatial Effects

Nearly all existing forecasting models for small areas, even though widely disparate in their individual methodologies, share a single common shortcoming: they treat each unit of geography (e.g. a census tract, an MCD, or a small city) as an independent, stand-alone entity rather than as an entity surrounded by other geographic areas with which they interact (e.g. through commuting patterns, shopping patterns, residential preferences, etc.). In fact, however, population growth in one unit of geography (the ‘focal’ unit) can be shown to be correlated (autocorrelated) with its neighbouring units. This observed pattern is supported by at least three theories. Tobler’s (1970) First Law of Geography states that everything is related to everything else, but nearer ones more so. The spatial diffusion theory of population geography argues that population growth forces will spread (spill over) to surrounding areas (Boyce, 1966; Morrill, 1968; Thrall et al., 2001), implying that population growth should demonstrate patterns of spatial autocorrelation. Regional economic theories such as growth pole theory apply spread and backwash notions to explain the mutual geographic interdependence of economic growth and development, which, in turn, influences population change (Hartshorn and Walcott, 2000; Richardson, 1976).

The past decade has seen substantial increase in the application of spatial statistics, GIS, and remote sensing techniques to sociological and demographic studies thanks to the upsurge in the availability of geographically referenced data, the development of user-friendly spatial data analysis software packages, and increased computing power combined with affordable computers (Chi and Zhu, 2008). Spatial dynamics of population change has been formally incorporated into demographic models and empirical studies (for a summary of the literature, see Fossett, 2005; Entwisle, 2007; Reibel, 2007; Voss, 2007). However, these techniques have not yet found their way into population forecasting models where forecasts for a focal unit explicitly recognise the forecasts (and possibly other covariates) for neighbouring units. Because most existing spatial regression models incorporate only same-period (i.e. cross-sectional) spatial effects, they cannot be used effectively to project future population. In this study, we propose a temporally lagged spatial regression approach to overcome this challenge.

The Spatio-Temporal Regression Forecasting Approach

In order to test for spatial dependence effects, spatial lag and spatial error models can be specified and estimated (Voss and Chi, 2006). However, these so-called simultaneous autoregressive regression (SAR) models account for same-period spatial effects rather than temporally lagged spatial effects. For example, in the traditional SAR model, the weighted neighbour population growth rates in 1980–1990 might be used for each focal unit as an independent variable to explain population growth rate in 1980–1990. This imposes difficulty for population forecasting because the goal is to project future population rather than simply examine the same-period population relationships. Based on the SAR models, Elhorst (2001) developed a general first-order serial and spatial autoregressive distributed lag model, which considers both the same-period and temporally lagged spatial effects. By dropping the same-period variables, the Elhorst model becomes our spatio-temporal econometric forecasting model (Eq. 4).

\[ Y_{t+10} = \alpha X_t + \beta Y_{t-10} + \lambda W + \rho + \epsilon \]  

The right side of the model includes a term for the temporally lagged population growth rate \((Y_{t-10})\), explanatory variables \((X_t)\), spatially and
temporally lagged neighbour population growth rates \( (WY_{t-10}) \), and spatially lagged neighbour characteristics \( (WX) \).

Here we need to pay attention to the \( n \) by \( n \) matrix of spatial weights \( (W) \). A spatial weights matrix is necessary for spatial data analysis of areal units, but there is scant theory by which to choose an optimal set of weights for the neighbourhood influence structure in a given application. The weights matrix is defined exogenously, and it behooves the analyst to compare several weights matrices in order to select a defensible one (Anselin, 2002). Practically, we can create and compare several weights matrices and select one based on the levels of the coefficient of spatial autocorrelation achieved and on the levels of statistical significance attained. In an early study (Chi and Zhu, 2008), the 7 nearest neighbour weights matrix was found to provide the highest spatial correlation of population growth out of 40 different types of weights matrices tested. Thus, we elected to use this weights matrix for this study which declares each focal area to be influenced by a weighted average of conditions among the seven nearest neighbouring units (based on an inter-centroid distance test).\(^6\) The \( k \)-nearest neighbour structure has often been shown to be superior to mere distance or continuity weights in areal-based spatial data analysis (Anselin, 2002).

DATA

In this study, we focus on the state of Wisconsin in the USA to test our spatio-temporal forecasting approach. Wisconsin is a so-called strong MCD state because its MCDs are functioning governmental units with elected officials, service provision obligations, and taxing authority. MCDs in Wisconsin contain a mix of low-density rural land, many small towns and villages, and a few large cities and surrounding suburbs – much as found elsewhere in the country. The MCD geography consists of non-nested, mutually exclusive, and exhaustive political territory with an average size of 29.56 square miles and 2920 persons in 2000. The advantages of using MCDs are their relevance to planning and public policy making and availability of data from the national census. In most parts of the state, census tracts have average sizes rather similar to MCDs and might serve as alternative units of analysis. However, census tracts are geographic units delineated largely for statistical purposes and only rarely have political roles analogous to that found in functioning governmental units such as MCDs.

Various data are used for each of the four population forecasting approaches discussed in the following section. Overall, the data include population data from decennial censuses from 1960 to 2000, highway expansion data from 1970 to 1990 at five-year intervals provided by the Wisconsin Department of Transportation, natural amenity and geophysical characteristics provided by the Wisconsin Department of Natural Resources and the US Geological Survey, and many socio-demographic and economic factors derived from the census data. The boundaries, and even the names, of MCDs in Wisconsin have not been stable over the study period. Boundaries change, new MCDs emerge, old MCDs disappear, names change, and status in the geographic hierarchy may shift – e.g. towns become villages, villages become cities. In order to adjust the data for these changes, new MCDs must be merged into the original MCDs from which they emerge. Disappearing MCD problems can be solved by dissolving the original MCDs into their current ‘home’ MCDs, and several distinct MCDs must be dissolved into one super-MCD in order to establish a consistent data set over time. We carried out these necessary adjustments using standard GIS functionality. The final number of geographic units used in our analysis is 1837.

ANALYTICAL APPROACHES

Four population forecast approaches of increasing complexity are examined and compared in this analysis. The first approach is the extrapolation projection prepared by the state demographic agency (WIDOA, 1993) and will be used as a baseline projection against which to evaluate the performance of the spatio-temporal regression approach. The second is a standard linear regression approach. The third is a regression approach incorporating temporally lagged spatial population growth rates (i.e. growth rates from the previous decade for neighbouring MCDs). The motivation is that growth rates in communities will induce spillover effects in surrounding communities, although such influences are not instantaneous. Lagging these effects in
time has intuitive, if not fully theoretical, justification. The fourth is a regression approach incorporating all the above in addition to temporally lagged neighbour independent variables (characteristics from the previous decade for neighbouring MCDs). This approach is motivated out of recognition that the drivers of population change are not limited solely to factors within community boundaries but, rather, that there are forces on the ground that need not, and do not, respect often arbitrary political boundaries.

These four population forecast approaches are used first to project the 2000 population for all MCDs in Wisconsin based on 1960–1990 population data. Second, the MCD projected populations are proportionally adjusted so as to sum to their corresponding independently derived county population projections. This adjustment step ensures uniformity among the four methods in terms of total county population (a commonplace and reasonable approach) and permits a clear examination of inter-MCD variability among the projections. Finally, the adjusted population projections are evaluated by comparing them to the actual 2000 population.

**Extrapolation Projection: Baseline Population Projection**

We use an extrapolation projection as the baseline projection against which to evaluate the accuracy of our proposed spatio-temporal regression forecasting approach. This is justified by the fact that these extrapolations presently serve as the 'official' projections used in Wisconsin for planning purposes. Population projections based on some form of extrapolation of the past into the future is an established and fundamental population forecast technique used for small geographic areas in many states of the USA for many years. For example, in 1993 the state demographic agency projected the Wisconsin 2000 population at the MCD level using census population data from 1960, 1970, 1980, and 1990.

The first step of the extrapolation projection (originally suggested by Voss and Kale, 1985) is to calculate a weighted average annual population change rate (Eq. 5):

\[
G = \left[ \frac{P_{90} - P_{80}}{10} + \frac{P_{80} - P_{70}}{20} + \frac{P_{70} - P_{60}}{30} \right] / 3
\]

where \( G \) is the weighted average annual numerical population change and \( P_{60}, P_{70}, P_{80}, \) and \( P_{90} \) are population counts from the corresponding census years. The formula calculates the average annual population change rate over three time periods, all ending in 1990, with the more recent data receiving greater influence in the projected growth.

The projected 2000 population equals the 1990 population plus 10 times \( G \) (Eq. 6). These projections, based on the 1960–1990 data, were available to us. However, the way in which the boundaries of MCDs are adjusted by the demographic agency is not quite the same as our method. Therefore, in order to make the MCDs consistent for the four approaches, we had to modify the 2000 population projections made by the state agency so that they conformed to our adjusted MCD boundaries.

\[
P_{90} = P_{90} + 10G
\]

**Regression Forecasting Approaches**

Regression forecasting approaches assume that the factors affecting population change have constant effects on population change over time. Therefore, we can first use historical data to estimate these effects and then apply these effects (via the estimated parameters) to project future population. In this study we test three regression approaches: a standard regression that considers only population change factors, a regression that considers temporally lagged neighbour growth, and a regression that also considers temporally lagged neighbour characteristics (spatial-temporal spillover among independent variables). The first two regression approaches are for comparison purposes; the last approach is our preferred spatio-temporal regression forecasting approach.

**Model 1: Standard Regression**

The first step is to use a standard linear regression model to build relationships between population change and relevant covariates. Current regression approaches for population forecasting generally consider core demographic factors (age distributions, for example) in explaining population change, but these factors tend to be chosen from an unnecessarily narrow demographic perspective rather than a perspective informed by other theories and potential data sets (Chi, 2009).
Population growth and decline are influenced not only by demographic characteristics and socio-economic characteristics but also by physical infrastructure, environmental characteristics, cultural resources, and potential legal constraints (for a detailed discussion of these population change factors, see Chi, 2009). These factors can and should be incorporated into regression models for population forecasting.

In our regression model, we attempt to incorporate as many relevant variables derived from population growth theories and empirical studies as data availability allows. The dependent variable is the MCD population growth rate, which is expressed as the natural log of the later census population over the earlier census population to achieve the desired bell-shaped distribution and better linearity with the independent variables. The independent variables include population growth rate from 1970–1980 and 31 additional variables from the 1980 Census, Wisconsin Department of Natural Resources, Wisconsin Department of Transportation, and other sources (see Table 1). Some of these independent variables are transformed to achieve maximum conformance to the linear model assumptions.

The estimation model (Model 1 Estimation) is specified as

$$Y_{80,90} = Y_{70,80}^\alpha + X_{80}^\beta + \varepsilon$$

where $Y_{70,80}$ is a column vector of temporally lagged growth rate from 1970 to 1980, $\alpha$ is the coefficient expressing the conditional relationship of $Y_{70,80}$ with $Y_{80,90}$, $X_{80}$ is a matrix of independent variables in 1980, and $\beta$ is a column vector of regression coefficients expressing conditional relationships of each variable in $X_{80}$ with the dependent variable. The intercept (constant term), which represents the overall growth rate and is identical for all MCDs, is excluded from the regression model to eliminate the background change in population redistribution processes. The exclusion of the intercept can force the overall growth rate into the coefficients of independent variables (Chi, 2009).

In the second step, insignificant independent variables, or those showing lower levels of significance, are discarded from the regression model; this process continues until only four or five variables are left in the model (population growth rate from the early decade is always kept due to its theoretical importance). There are two reasons for this step. One is that one or more of the independent variables, net of the others, may not be statistically significant in explaining population change. The other is that the reduced variable set reduces the extent of multicollinearity, often a problem in the regression forecasting context (Armstrong, 2001).

In the third step, we employ the parameters derived from the second step and the variables for the period 10 years later to build a projection model (Model 1 Projection, Eq. 8):

$$\hat{Y}_{90,00} = Y_{80,90}^\hat{\alpha} + X_{80}^\hat{\beta}$$

where $Y_{80,90}$ is the temporally lagged growth rate from 1980 to 1990, $X_{80}$ are the independent variables for 1990, and $\hat{\alpha}$ and $\hat{\beta}$ are estimated parameters from step two. From the left-hand term in Eq. 8, we can calculate the projected 2000 population.

**Model 2: Regression with Temporally Lagged Neighbour Growth Rates**

Model 2 adopts the same approach as Model 1, except that Model 2 includes an explicit spatial spillover representing a weighted average of neighbour population growth rates one decade earlier. The estimation model (Model 2 Estimation, Eq. 9) then is

$$Y_{80,90} = Y_{70,80}^\alpha + X_{80}^\beta + WY_{70,80}^\lambda + \varepsilon$$

The spatio-temporal growth influence is represented by the $WY_{70,80}$ term. Following Model 1, we assign the parameters estimated from the Model 2 Estimation into the corresponding projection model (Model 2 Projection, Eq. 10):

$$\hat{Y}_{90,00} = Y_{80,90}^\hat{\alpha} + X_{80}^\hat{\beta} + WY_{80,90}^\hat{\lambda}$$

**Model 3: Regression with Temporally Lagged Neighbour Growth and Neighbour Characteristics**

Model 3 shares the approach adopted for Model 2, the only difference being that Model 3 takes into account not only the spatio-temporal lag of growth rates but also additional neighbour characteristics for the base estimation period. The variables representing neighbour characteristics are created by averaging the neighbours' values of independent variables corresponding to those finalised in Model 1. In stepwise fashion, the four
independent variables representing neighbour characteristics are added to Model 3, and the four are iteratively tested and eliminated until one strong neighbour characteristic remains. The Model 3 Estimation (Eq. 11) is specified as

\[ Y_{80,90} = Y_{70,80} + X_{80} \beta + WY_{70,80} \lambda + WX_{70,80} \rho + \varepsilon \]  

(11)

The neighbour characteristic set is represented by the \(WX_{70,80}\) term, and the strength of these neighbouring effects is expressed in the column vector

Table 1. Variable measures and data sources.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population density</td>
<td>The natural log of the number of persons per squared kilometres</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Young</td>
<td>Square root of the proportion of young population (ages 12–18)</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Old</td>
<td>Square root of the proportion of old population (ages 65+)</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Blacks</td>
<td>Square root of the proportion of black population</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Hispanics</td>
<td>Square root of the proportion of Hispanic population</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>College population</td>
<td>Square root of the proportion of college population</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>High school degree</td>
<td>Proportion population (age 25+) with high school degrees</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>Proportion population (age 25+) with bachelor’s degrees</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Stayers</td>
<td>Proportion non-movers (age 5+)</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Female-headed families</td>
<td>Proportion female-headed families with children under 18 years old</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Income</td>
<td>Square root of median household income</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Unemployment</td>
<td>Square root of unemployment rate</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Retail</td>
<td>Square root of proportion of workers in retail industry</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Agriculture</td>
<td>Square root of proportion of workers in agricultural industry</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Public transportation</td>
<td>Square root of proportion of workers using public transportation to work</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Commute time to work</td>
<td>Square root of proportion of workers travelling 30 minutes or more to work</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>House value</td>
<td>Square root of median house value</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>New housing</td>
<td>Proportion housing units (40 years or fewer)</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Seasonal housing</td>
<td>Square root of proportion of seasonal housing units</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Public water</td>
<td>Proportion housing units using public water</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>Public sewer</td>
<td>Proportion housing units using public sewer</td>
<td>Census 1980 STF3A</td>
</tr>
<tr>
<td>County seat</td>
<td>County seat status (dummy variable)</td>
<td>State of Wisconsin Blue Book 2001–02</td>
</tr>
<tr>
<td>Buses</td>
<td>Having urban buses or not (dummy variable)</td>
<td>State of Wisconsin Blue Book 1989–90</td>
</tr>
<tr>
<td>Major roads</td>
<td>Square root of total lengths of major roads</td>
<td>National Atlas of the U.S.</td>
</tr>
<tr>
<td>Natural amenities 1</td>
<td>Natural amenities more on river lengths and proportion of forest areas, generated by principal factor analysis</td>
<td>ArcIMS servers: <a href="http://maps.dnr.state.wi.us">http://maps.dnr.state.wi.us</a> and <a href="http://maps.botany.wisc.edu">http://maps.botany.wisc.edu</a>, and the U.S. Geological Survey (USGS) 1:100,000 Hydro Digital Line Graphs</td>
</tr>
<tr>
<td>Natural amenities 2</td>
<td>Natural amenities more on proportion of water areas, generated by principal factor analysis</td>
<td>ArcIMS servers: <a href="http://maps.dnr.state.wi.us">http://maps.dnr.state.wi.us</a> and <a href="http://maps.botany.wisc.edu">http://maps.botany.wisc.edu</a>, and USGS 1:100,000 Hydro Digital Line Graphs</td>
</tr>
<tr>
<td>North rural areas</td>
<td>North rural areas or not (dummy variable)</td>
<td>USGS 1:100,000 Hydro Digital Line Graphs and the U.S. Office of Management and Budget</td>
</tr>
<tr>
<td>Highway expansion 1</td>
<td>MCDs within 10 miles of highway expansion finished five years before population change period</td>
<td>Wisconsin Department of Transportation</td>
</tr>
<tr>
<td>Highway expansion 2</td>
<td>MCDs at a range of 10–20 miles from highway expansion finished five years before population change period</td>
<td>Wisconsin Department of Transportation</td>
</tr>
<tr>
<td>Highway expansion 3</td>
<td>MCDs within 10 miles of highway expansion finished just before population change period</td>
<td>Wisconsin Department of Transportation</td>
</tr>
<tr>
<td>Highway expansion 4</td>
<td>MCDs at a range of 10–20 miles from highway expansion, finished just before population change period</td>
<td>Wisconsin Department of Transportation</td>
</tr>
</tbody>
</table>

MCD, minor civil division.
ρ (in our Model 3 with only one spatially lagged independent variable, this vector degenerates to a single scalar parameter). Correspondingly, the Model 3 Projection (Eq. 12) is specified as
\[
\hat{Y}_{90,0} = Y_{90,90} + \alpha + X_{90} \beta + WY_{90,90} \lambda + WX_{90} \rho
\] (12)

**Population Projection Adjustments**

In practice, population projections are often adjusted for the purpose of improving forecasting accuracy. Adjusting population projections can involve many steps. Two major considerations are (i) modifications to rein in severely abnormal change rates and (ii) adjustment (or control) to county projections (Voss and Kale, 1986; WIDOA, 2004). Modification of abnormal change rates is used to soften the occasional high population change rates that emerge when making small-area projections. On the basis of prior research, and recognizing that extreme growth rate in small areas often result from one-time epiphenomenal occurrences, it is anticipated that these adjustments will improve the overall performance of population forecasts (Voss and Kale, 1985). If the population change rate in an MCD is unusually high relative to neighbouring MCDs, the projected rate is softened under the assumption that rapid population change cannot be sustained for long periods. In order to identify unusually high population change rates (in absolute terms), an average annual population change rate (\textit{Mean}) for each MCD is determined by considering all MCDs in the county, and a mean and standard deviation (\textit{SD}) of local change is obtained. If an MCD falls outside the range (\textit{Mean} – 1.5 × \textit{SD}, \textit{Mean} + 1.5 × \textit{SD}), it will be assigned the nearer value of the two ends.

In addition, it is possible (although unlikely) for a projected MCD population for small areas to become negative in the long term. In such instances, we adjust the MCDs’ projected populations to be zero at exactly 80 years from the projection base (in this instance 2070). This readily constrains the population from becoming zero within the time frame of our projection horizon. The MCD projections are then finally adjusted by comparing the sum of the MCD population projections with their parent county population projections. Empirically, county-level projections are more accurate than sub-county-level projections, and common practice is to control the sum of the MCD projections to a projection for the parent county.8

**Projection Evaluations**

There are a number of measures for evaluating the accuracy of population forecasts (Long, 1995; Keilman, 1999; Smith et al., 2001). Two of the most commonly used measures are the mean algebraic percent error (MALPE) and the mean absolute percent error (MAPE). The MALPE is a measure where the positive and negative values can offset each other, so it is used mainly as a measure of bias (Smith, 1987; Tayman and Swanson, 1996). A positive MALPE indicates an overall upward bias in the model, and a negative MALPE indicates the reverse (Eq. 13). In contrast, the MAPE is a measure where positive and negative values do not offset each other (Eq. 14).

\[
MALPE = \frac{1}{n} \sum \left( \frac{\text{Forecasted} - \text{Observed}}{\text{Observed}} \right) \times 100 \quad (13)
\]

\[
MAPE = \frac{1}{n} \sum \left| \frac{\text{Forecasted} - \text{Observed}}{\text{Observed}} \right| \times 100 \quad (14)
\]

\[
RMSPE = \sqrt{\frac{1}{n} \sum \left( \frac{\text{Forecasted} - \text{Observed}}{\text{Observed}} \right)^2} \times 100
\]

The four projections are also compared on the basis of population size and growth, both by the MALPE and the MAPE. As commonly found in a large literature, the statistical quality of a projection set is a function of population size and growth rate – forecast accuracy increases as population growth rate (in absolute terms) decreases, and as population size increases until reaching a threshold level that varies with numerous factors (Smith, 1987). The evaluations based on population size and growth provide further details into

---

Copyright © 2010 John Wiley & Sons, Ltd. Popul. Space Place 17, 505–520 (2011)
DOI: 10.1002/ppsp
FINDINGS

Estimations of Regression Models

Following extensive exploratory analysis and specification testing, five independent variables are finally chosen for Model 1 estimation (Table 2): population growth rate in previous decade, population density, young (population age structure), house value, and extent of new housing (see Table 1 for full variable descriptions and source). The five variables are then used in Models 2 and 3. Model 2 has another independent variable, the average population growth rate of neighbours (a measure of spatial spillover). Model 3 has yet one additional independent variable, the average proportion of young (age structure) in neighbours.

Table 2 summarises the estimates of the standard regression model (Model 1), the regression with neighbour growth (Model 2), and the regression with neighbour growth and neighbour characteristics (Model 3). All independent variables are significant in explaining population growth in all three models. The goodness-of-fit measures indicate that Models 2 and 3 are superior to Model 1 and that Model 3 outperforms Model 2. These goodness-of-fit measures show that the incorporation of temporally lagged neighbour growth and temporally lagged neighbour characteristics improves the estimation of the regression model. The robust Lagrange Multiplier (error and lag) tests on the residuals suggest that there is spatial lag dependence in Model 1 and spatial error dependence in Model 2; neither of them remains in Model 3. This again indicates that Model 3 outperforms Models 1 and 2 in terms of estimation.

Evaluations of Population Forecasts

After estimation of the models, the three regression approaches are then used to make three sets of population projections for 2000. The three regression approaches are evaluated by...
comparing them to the extrapolation projection (comparisons based on basic descriptive statistics, population size, and population growth rate). First, the basic descriptive statistics do not suggest a strong preference for the spatio-temporal regression approach or extrapolation projection (see Table 3). Neither the regression models nor the extrapolation projection uniformly achieves higher accuracy than the others in terms of bias or precision. The three regression approaches generate very similar MALPEs and MAPEs, although the proposed spatio-temporal regression approach, augmented with a spatially lagged covariate (Model 3), slightly outperforms the other two approaches. The three regression approaches also have similar distributions of projection errors, although again, Model 3 appears to have a slight edge. The regression approaches produce MALPEs closer to zero but higher MAPEs and RMSPEs than the extrapolation approach. This indicates that the regression approaches are less biased but also less precise than the extrapolation approach. The differences are marginal, however. Based on projection error distribution, the regression approaches produce slightly fewer projections with small errors (−10% to 10%) and slightly more projections with high errors (−10% or less and 10% or more).

Second, the regression models are compared with the extrapolation projection on the basis of population size in 2000 by MALPE and MAPE (Table 4). The results do not indicate a strong preference for the spatio-temporal regression or extrapolation projection. For MCDs with 250 or fewer persons (118 MCDs), all three regression methods outperform the extrapolation method, with the exception of Model 3. For MCDs with 251–2000 persons (1310 MCDs), the regression models provide slightly less bias but slightly less precise projections than the extrapolation projection. For MCDs with 2001 persons or more (409 MCDs), the regression models produce less accurate projections than the extrapolation projection, with the exception that Model 1 produces the least biased projection for MCDs with more than 20,000 persons.

Third, the regression models are compared with the extrapolation projection based on population growth rate from 1990 to 2000 by MALPE and MAPE (Table 5). Again, the results do not suggest a strong preference for the spatio-temporal regression or extrapolation projection. For MCDs gaining 5%–10% population (299 MCDs), the regression models produce more precise and less biased projections than the extrapolation method, and the spatio-temporal regression model generally outperforms the other two regression models. For MCDs gaining less than 5% or more than 10% population, the regression models generally underperform the extrapolation projection except that Model 1 and Model 2 produce more precise projections than
the extrapolation method for MCDs gaining 0%–5% population.

Overall, the projection evaluations reveal mixed results and do not suggest unambiguous preference for the spatio-temporal regression approach or the extrapolation projection. Although the results do indicate that our spatio-temporal regression approach significantly improves model estimation, the spatio-temporal regression approach does not outperform the extrapolation projection.

SUMMARY AND DISCUSSION

This study attempts to provide a theoretically driven spatio-temporal econometric approach for small-area population forecasting. We revise conventional spatial econometric models into a spatio-temporal specification for forecasting purposes by supplementing same-period spatial effects with temporally lagged spatial effects. Our approach seems eminently reasonable and is solidly undergirded by population-related theories and prior empirical studies that suggest strong correlation between population change and socio-economic and environmental factors, and other spatial and temporal spillover effects. However, the projection evaluations reveal mixed results and do not suggest unambiguous preference for the spatio-temporal regression approach over the extrapolation projection. Why then does such a model not win handily when compared with a simple atheoretical extrapolation of past trends into the future? Why do the projection evaluations not unambiguously reveal a clear preference for a theoretically grounded spatio-temporal regression approach – especially when the goodness of fit measures at the estimation stage of the projection process hint at the superiority of the spatio-temporal regression model?

One reply is less an answer than an observation. The forecasting literature is replete in its cautions that complex forecasting models are not necessarily superior to more simple models. Parsimony in this regard is a strong virtue (Armstrong, 2001; Smith et al., 2001). ‘However interdisciplinary we become, there are some clear limits to knowledge of the interrelations of the variables whose combined operation will bring about the future population’ (Keyfitz, 1981: 579). Less complex models are easier to understand, and they are more easily communicated to users of the forecasts. Yet our regression approach is not inordinately complicated and has, moreover, a much stronger appreciation of the presumed causes of population change and spatio-temporal spillovers. As a consequence, the lack of triumph over simple extrapolation is doubly discouraging.

Another, more easily understood reply is that our study covers a single state in the USA and a single episode in time, 1960–2000 – in which the state of Wisconsin witnessed strong cycles of growth that might, out of sheer unfortunate coincidence, have resulted in the failure of the space-time spillover forecasts to outshine the simple extrapolations. The decade of the 1980s was a decade of particularly slow population growth in Wisconsin, a decade sandwiched by stronger growth in the 1970s and 1990s. Replication of our approach (presently underway) might well reveal different results, even if only shifted one decade forward. That remains a test for another time, and perhaps another set of geographic areas.

Table 4. Evaluating population projections by population size in 2000.

<table>
<thead>
<tr>
<th>Population size (number of MCDs)</th>
<th>250 and less (118)</th>
<th>251–2000 (1310)</th>
<th>2001–20,000 (372)</th>
<th>20,001 and more (37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALPE (%)</td>
<td>MAPE (%)</td>
<td>MALPE (%)</td>
<td>MAPE (%)</td>
<td>MALPE (%)</td>
</tr>
<tr>
<td>Baseline projection</td>
<td>3.47</td>
<td>16.16</td>
<td>−3.99</td>
<td>9.41</td>
</tr>
<tr>
<td>Model 1</td>
<td>3.42</td>
<td>15.46</td>
<td>−3.51</td>
<td>10.30</td>
</tr>
<tr>
<td>Model 2</td>
<td>3.46</td>
<td>15.39</td>
<td>−3.50</td>
<td>10.30</td>
</tr>
<tr>
<td>Model 3</td>
<td>4.16</td>
<td>15.56</td>
<td>−3.50</td>
<td>10.28</td>
</tr>
</tbody>
</table>

MAPE, mean absolute percent error; MALPE, mean algebraic percent error; MCD, minor civil division.
In addition, the assumption of constant regression coefficients may be a serious flaw (Chi, 2009; Tayman and Schafer, 1985). Temporal instability in relationships may be undermining the promise of regression-based forecasting, the theoretical defence of the choice of variables notwithstanding. When population redistribution patterns differ in the estimate and the projection periods, the strength of the approach is obviously degraded. Wisconsin experienced ‘renewed metropolitan growth’ in the 1980s and reversed to ‘rural rebound’ in the 1990s (Johnson, 1999). The two processes are driven by different factors, and thus the effects and significance of some relevant variables on population change differ considerably. The elimination of the effect of coefficient drift clearly must be explored in future research of small-area population forecasting using regression approaches.

Although the proposed spatio-temporal regression approach for small-area population forecasting does not outperform extrapolation population projection, the spatio-temporal regression approach improves the estimation of the traditional standard regression approach for population projection by incorporating neighbour growth and neighbour characteristics. In addition, this approach evaluates an exhaustive list of variables for population forecasting. The approach builds on theoretical foundations that hypothesize strong correlations between population and socio-economic and environmental factors, neighbour growth, and neighbour characteristics. Existing standard regression forecasting approaches often use only a very limited set of variables. We argue that such an approach is not enough for finding the best variables in population forecasting. Forecast inaccuracy is due to the fact that population change is affected by many factors, some of which lie outside the demographic variables commonly used in population forecasting practice (Cohen, 1995). Thus, the spatio-temporal regression model proposed in this study is one step towards a more holistic examination of the relationship between population change and its relevant factors as well as towards a more informative population forecasting approach.

The innovations introduced in our spatio-temporal population forecasting method are partly in response to the needs of local planners and decision makers. Small-area population forecasting,

<table>
<thead>
<tr>
<th>Population growth rate (number of MCDs)</th>
<th>10% and more</th>
<th>5%–9.99%</th>
<th>0.1%–4.99%</th>
<th>0%</th>
<th>−0.01% to −0.5%</th>
<th>−5% to −10%</th>
<th>−10% to −15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>23.81</td>
<td>8.95</td>
<td>26.38</td>
<td>26.38</td>
<td>26.31</td>
<td>26.53</td>
<td>26.53</td>
</tr>
<tr>
<td>Model 3</td>
<td>23.81</td>
<td>8.95</td>
<td>26.38</td>
<td>26.38</td>
<td>26.31</td>
<td>26.53</td>
<td>26.53</td>
</tr>
</tbody>
</table>

MAPE, mean absolute percent error; MALPE, mean algebraic percent error; MCD, minor civil division.
such as that at the municipal level, is useful to
town, city, and village planning, as population
forecasts often are legally mandated prerequisites
for planning and policy analysis. Small-area
population growth trends also are an important
element in determining the demand for public
facilities and services as well as in distributing
fiscal and political resources. Although the spa-
tio-temporal regression approach as introduced
here does not dramatically outperform extrapo-
lation projection, the former does bring a signifi-
cant advantage to local planning and decision
making that cannot be provided by the latter.
Decision makers are often interested in ‘what-if’
scenarios (Land and Schneider, 1987). Our pro-
posed approach allows analysts to examine a
variety of consequences conditioned on the
change of covariates including change in neigh-
bouring places. The estimated relationships
between population change and relevant factors
can inform planners and decision makers of the
possible consequences of adopting various stra-
egies as well as suggest strategies to affect popu-
lation growth trends. The estimated relationships
between population change and neighbour char-
acteristics and dynamics can further inform plan-
ners and decision makers of the possible
consequences of contemplated actions – even
those taken by neighbouring communities.

NOTES

(1) In this study, ‘small areas’ refers to any geographic
units below county level. In the state of Wisconsin
(the study case), these small geographic units
include MCDs (with an average population size of
2920 in 2000), census tracts (4051), block groups
(1225), partial block groups (626), and blocks (119).
Their average population sizes are similar to those
in other states of the USA.

(2) In truth, neighborhood context is not wholly
ignored in traditional models given that controls
to higher geography are usually applied. For
instance, MCD populations projected by extrapola-
tion are often adjusted by forcing them to indepen-
dent population projections made for their parent
counties. However, this concession to neighbour-
hood context is more a matter of tidiness than it is
an acknowledgement of spatial population effects
and neighbour characteristics that we examine
in our models. See the ‘Analytical Approaches’
section for detailed explanations.

(3) A projection is taken to be value free – simply the
embodiment of one or more assumptions that
rather mechanically determine a future outcome.
A forecast, on the other hand, is a projection that is
considered the single most likely to occur, gener-
ally based on a set of assumptions and subjective
judgements. Nevertheless, we use ‘forecast’ and
‘projection’ interchangeably in this paper.

(4) These are clearly spelled out in the standard regres-
sion and econometrics literature. See, for example,
Draper and Smith (1998), Fox (1997), and Greene
(2000).

(5) In small areas, however, migration is substantially
more volatile than fertility and mortality. Any
abrupt changes in migration may not be as well
captured by cohort component methods as they
are by regression models.

(6) The neighbours considered in this study are con-
strained to the MCDs within the boundaries of
Wisconsin and do not include those in neighbour
states due to data unavailability. The neighbours
assigned to the MCDs near the border are not nec-
essarily the ideal neighbours based on spatial
weights matrices, and thus the assignment of
neighbours is vulnerable to the boundary effect.

(7) An extensive review of the relevant literature
results in more than 31 variables that significantly
affect population change theoretically or empiri-
cally (Chi, 2009). The 31 variables are chosen for
this research on the basis of a combination of
judgement, established theoretical or empirical
relationships, and the availability and quality of
data.

(8) In addition to the two major adjustments, practi-
cially we may need to consider further adjustments.
For example, if a restriction to population growth
is known, the projected population should be
adjusted to conform to the restriction. Such restric-
tions include zoning restrictions, land use plans,
and geophysical conditions that absolutely restrict
or limit growth. Such adjustments were not con-
ducted in this analysis, as the principal purpose
was to show the effects of borrowing strength over
space and time in the projection process.

REFERENCES

Ahlburg D. 1982. How accurate are the US Bureau of
the Census projections of total live births. Journal of
Forecasting 1: 365–374.

Alho JM, Spencer BD. 2005. Statistical Demography and

Anselin L. 2002. Under the hood: issues in the specifi-
cation and interpretation of spatial regression


Voss PR, Kale BD. 1985. Refinements to Small Area Population Models: Results of a Test Based on 128 Wisconsin Communities. Paper read at the Annual Meeting of
Population Association of America, 28–30 March, Boston.