Incorporating Spatial Effects into Hamilton-Perry Based Population Projections Improves Accuracy and Precision at the Census Tract Level.

Jack Baker[1], Xiaomin Ruan, Adelamar Alcantara

DRAFT: DO NOT CITE WITHOUT AUTHOR'S PERMISSION

For:
Population Association of America, Annual Meeting (2011)
Washington DC
March 31-April 4, 2011

Introduction. Limited data on components of demographic change (births, deaths, migration) constrain options for estimating and projecting sub-county populations. Hamilton and Perry (1962) were the first to suggest the simple idea of using a "short-cut" method based on intercensal *cohort change ratios* (the ratio of equivalent age classes at ten year intervals). Within the applied demography literature, a renewed interest in the method has been expressed repeatedly in recent years (Smith and Shahidullah, 1995; Smith et. al., 1999; George et. al., 2004; Ramaniuk, 2009; Swanson et. al., 2010). Previous evaluations have suggested that the method is capable of producing relatively accurate and unbiased population projections at the National, State, and Census tract levels (Smith and Shahidullah, 1995; Smith and Tayman, 2003); however, no study to date has attempted to address issues of spatial scale that are inherent in the method. The HP method is inherently *local*—the cohort change ratios that drive the model are estimated for each areal unit (here census tracts) in isolation. This avoids the liability of using a global model (such as a regression-based procedure) that may drastically over or under-estimate some tracts, but it also fails to leverage spatial clusterings of tracts and the inevitable spatial dependency of data aggregated to areal units such as census tracts (Voss, 2004; Frotheringham, 2003). This study attempts to strike a balance by extending the HP method to incorporate neighborhood groupings as spatial dependencies that are incorporated directly into the cohort-change ratios. This is achieved by estimating and applying simple spatial weights matrices within the algorithm. A series of age/sex specific population projections are made for 2000 using 1980 to 1990 intercensal trends to estimate cohort change ratios. These sets include both the standard, aspatial Hamilton-Perry projections, two spatially-weighted variants of the procedure, and a final spatially-weighted projection that incorporates controls to projections at higher geographic levels. **The results indicate strong support for the ideas that: (1) spatially-weighted extensions of the Hamilton-Perry method significantly reduce bias in population projections for Census tracts without sacrificing accuracy, and (2) that when proper controls are applied, may also significantly improve accuracy.**

The Hamilton-Perry Method. The foundation of the Hamilton-Perry method rests upon estimation of intercensal *cohort-change ratios*. Cohort-change ratios measure (intercensal) change in the population of corresponding age classes over a ten year period (Hamilton and Perry, 1962; Smith et. al., 1999; George et. al., 2004; Swanson et. al., 2010). For example, the ratio of individuals ages 15 to 19 in 2000 relative to those individuals aged 5 to 9 in 1990 constitutes a cohort change ratio. To project the 2000 population forward, in this scenario, the demographer simply multiplies the number of children ages 5 to 9 in 2000 against the estimated cohort-change ratio to estimate the population 15 to 19 years of age in 2010. The cohort-change ratio clearly measures a number of demographic processes. In this example, changes in the population within this 5 to 9 turns 15 to 19 year age class include a small number of deaths as well as individuals who entered or exited the age-class. Although it is not possible to decompose the components of change within the Hamilton-Perry method, the conceptual realization that it

[1] Address Correspondence to: Jack Baker, Ph.D., Senior Research Scientist, Geospatial and Population Studies, University of New Mexico. Jbaker4762@gmail.com.

measures aggregate processes of births, deaths, and migration has led previous demographers to point out that the method is conceptually equivalent to the well-known component method (Shyrock and Segal, 1973; Smith et. al., 1999; George et. al., 2004). The cohort-change ratios are used to project the population 10 and over forward in time; the population ages 0 to 4 and 5 to 9 are typically estimated using dependency ratios either from the previous census, or based on an extrapolation of previous intercensal trends (Smith et. al., 1999; George et. al., 2004).

A Simple Spatial Extension of the Hamilton-Perry Method. Statisticians dealing with spatial data have long contended that standard methods, which rely upon the computation of "global" statistics are insufficient for those interested in human geography (Frotheringham, 2003; Simpson, 1951). Spatial statisticians argue that spatial dependency in data—in which values of quantitites are more similar the closer they are in space (Tobler, 1970)—must be considered through the estimation not only of global statistics, but also of locally-important relationships. Spatial dependency in geographic data—to which population projections at the sub-county level inevitably are subject to—has been measured in a number of ways using *spatial autocorrelation* models (Cliff and Ord, 1972; Haining, 1979; Getis and Ord, 1992, 1995, 2001; Anselin, 1995, 1998; Rogerson, 1999). When spatial autocorrelation exists, standard statistical models are typically adjusted to account for this spatial dependency, often through adjusting parameter estimates through estimation of *spatial weights matrices*, which allow direct incorporation of spatial dependency into the estimation of statistical models (Frotheringham, 2003).

Within the field of applied demography, Voss (2004) has made the argument that the field is inherently spatial since the data used for analysis and modeling is typically aggregated within areal units. In spite of this, relatively few attempts have been made to incorporate spatial dependency into population estimation and projection methods. In the case of the Hamilton-Perry method, incorporation of spatial dependency may be achieved in a straightforward fashion by adjusting the estimated cohort-change ratios using a spatial weights matrix. At first blush, it may seem that this would be unnecessary since from one point of view, the Hamilton-Perry method incorporates spatial heterogeneity naturally because the cohort-change ratios are estimated for each census tract directly. We think that this may erase spatial grouping effects in which neighboring census tracts exhibit a spatial dependency which is, in fact, *erased within the aspatial utilization of the Hamilton-Perry method*. Incorporating a spatial weights matrix that allows clusters of census tracts to interact should improve accuracy and precision of projections, especially under scenarios where a group of census tracts are evolving together over time (consider a scenario of urban residence change in which retirees settle preferentially in neighboring census tracts over time, perhaps moving sequentially into the area over time so that not all tracts are settled simultaneously). Incorporation of a spatial weights matrix can potentially capture these effects and improve the accuracy and precision of population projections. A further improvement may be made by using the Hamilton-Perry method to estimate changes in the proportional population structure of census tracts, with spatial weighting, while controlling them to larger-scale projections that are likely to be more accurate small-area projections (Smith, 1985) and that may also be estimating trends in population distributions (a spatial statistic in and of itself).

In the case of the Hamilton-Perry method, however, the transition factors already represent *specific local statistics* that are, in fact, the parameters of interest since no global statistic is estimated in the procedure. Incorporating spatial effects in a Hamilton-Perry method, then, really means incorporating the values of surrounding neighbors into the estimation of cohort change ratios. In this case, then, a simple spatial weights matrix that captures the effects of contiguous neighbors on the cohort change ratios estimated for any given census tract should be adequate. Such a matrix would implicitly incorporate the conceptual idea that demographic patterns happen in geographic clusterings that may be captured not in census tracts in isolation. Spatial extensions would allow these neighborhood grouping effects to directly determine the cohort change ratios associated with each census tract.

As such, incorporation of these effects should improve the accuracy of projections based on spatially-weighted cohort-change ratios.

In this study, we compare aspatial to spatially-weighted variants of the Hamilton-Perry method for census tracts within Bernalillo County New Mexico. Using 1980 to 1990 trends to estimate cohort change ratios, we then move the 1990 age/sex structured population forward to 2000 and evaluate aspatial and spatial variants against the 2000 Census. We compare the model variants in terms of global measures of accuracy (mean absolute and absolute percentage errors—MAE and MAPE) and bias (mean algebraic and algebraic percentage error—MALGE and MALPE) both for the total population and within each age/sex category (five-year intervals, truncated at 80 years).

Validation Against Census 2000.

Table 1. Accuracy and Bias in Aspatial and Spatially-Weighted Variants of the Hamilton Perry Method

Census Tracts, 2000, Bernalillo County, New Mexico (Comparisons with Observed 2000 Census Counts)

Model	Global Mean Absolute Error	Global Mean Absolute Percentage Error (MAPE)	Global Mean Algebraic Error	Global Mean Algebraic Percentage Error (MALPE)
Aspatial Hamilton Perry	25	25.59%	-7	-10.08%
Queen Contiguity	28	31.55%	-4	-7.50%
Optimized Rook Contiguity	28	30.09%	-3	-4.00%
Shift-Share Controlled Model	Global Mean Absolute Error	Global Mean Absolute Percentage Error (MAPE)	Global Mean Algebraic Error	Global Mean Algebraic Percentage Error (MALPE)
Optimized Rook Contiguity	21	23.51%	-1	-3.12%

Table 1 reviews the global results (across all age/sex categories) obtained using aspatial and spatial variants of the Hamilton-Perry method. A clear result is that incorporation of spatial effects into the method reduces measures of bias (such as MALPE), while introducing only minor costs in accuracy. Spatial-weighting based on rook (geographies share an edge or corner) and queen (geographies share an entire edge only) contiguity schemes made little different in accuracy, but the rook variants displayed signifiantly lower bias in the validation projections. Controlling the rook-based model to a shift-share based on all census tracts dramatically improved *both accuracy and precision in the validation projections*. While the aspatial Hamilton-Perry model was actually more accurate than either Queen or Rook-based spatially-weighted projections, incorporation of this control reduced the global measures of absolute error significantly.

When viewed within age/sex classes, this progressive improvement is also observed. Table 2 (next page) reviews MAPE and MALPE measures between the aspatial Hamilton-Perry and the final, rook-based controlled estimate seen in the bottom row of Table 1. While substantial volatility remains across the age classes within each sex (this method does not solve the problem that larger errors will be observed in smaller geographies—see Smith and Shahidullah, 1995), it is clear that in most cases the spatial-weighted projections are less biased and more accurate than the aspatial alternative. Overall, among males 13 out of 17 age classes display a reduction in both absolute and algebraic percentage errors, indicating significant reductions in bias and increases in accuracy that range as high as 16 percent. Among females, a similar overall pattern exists with reductions in both errors in 12 of 17 classes ranging to as high as 15 percent. The overall averages of these averages (this is an acceptable measure—the sampling distribution of means has a long-standing history in statistics and is known to be normally distributed—see Samuels and Witmer, 1983; Pagano, 1998), also displays a remarkable reduction in both cases, around 7 percent in the absolute error and some seven to eight percent in the average percentage errors.

Table 2. Age-Specific Error and Bias, Aspatial and Spatial Variants of the Hamilton-Perry Method

Aspatial Hamilton-Perry Based Estimates			Spatially-Weighted Hamilton-Perry Based Estimates*		
Males	Global Mean Absolute Percentage Error (MAPE)	Global Mean Algebraic Percentage Error (MALPE)	**Males**	Global Mean Absolute Percentage Error (MAPE)	Global Mean Algebraic Percentage Error (MALPE)
0 to 4	20.10%	-10.73%	0 to 4	18.52%	-5.43%
5 to 9	18.37%	0.40%	5 to 9	22.35%	6.63%
10 to 14	34.58%	-22.70%	10 to 14	20.68%	-13.29%
15 to 19	22.36%	-4.04%	15 to 19	16.97%	1.68%
20 to 24	26.00%	16.45%	20 to 24	29.93%	14.55%
25 to 29	30.30%	-4.58%	25 to 29	36.56%	-1.67%
30 to 34	41.78%	-17.76%	30 to 34	32.84%	-10.53%
35 to 39	68.84%	-47.53%	35 to 39	26.19%	-31.50%
40 to 44	23.41%	-8.79%	40 to 44	18.25%	-1.37%
45 to 49	19.96%	-4.75%	45 to 49	14.79%	1.71%
50 to 54	18.89%	-1.65%	50 to 54	13.91%	6.37%
55 to 59	24.02%	-10.29%	55 to 59	12.68%	-0.96%
60 to 64	17.08%	1.86%	60 to 64	12.80%	7.00%
65 to 69	28.14%	-16.49%	65 to 69	16.66%	-7.77%
70 to 74	22.45%	-4.03%	70 to 74	49.93%	4.17%
75 to 79	50.31%	-46.27%	75 to 79	37.54%	-34.70%
80 Plus	33.40%	24.85%	80 Plus	21.69%	29.98%
Average	29.41%	-9.18%	**Average**	23.67%	-2.07%
Females	Global Mean Absolute Percentage Error (MAPE)	Global Mean Algebraic Percentage Error (MALPE)	**Females**	Global Mean Absolute Percentage Error (MAPE)	Global Mean Algebraic Percentage Error (MALPE)
0 to 4	16.93%	-9.01%	0 to 4	19.81%	-1.09%
5 to 9	21.63%	-7.19%	5 to 9	17.73%	2.19%
10 to 14	27.20%	-9.80%	10 to 14	28.09%	-6.57%
15 to 19	25.06%	-2.73%	15 to 19	18.46%	5.04%
20 to 24	24.94%	4.89%	20 to 24	27.95%	2.53%
25 to 29	30.96%	0.46%	25 to 29	30.27%	-2.33%
30 to 34	39.50%	-21.24%	30 to 34	29.62%	-12.45%
35 to 39	31.00%	-15.24%	35 to 39	44.04%	-5.14%
40 to 44	19.40%	-3.27%	40 to 44	20.99%	5.11%
45 to 49	21.82%	-8.37%	45 to 49	18.34%	1.61%
50 to 54	13.18%	1.90%	50 to 54	15.69%	10.82%
55 to 59	23.38%	-9.39%	55 to 59	16.44%	1.66%
60 to 64	21.11%	-2.11%	60 to 64	15.81%	6.75%
65 to 69	31.14%	-19.73%	65 to 69	17.19%	-8.59%
70 to 74	64.47%	-54.78%	70 to 74	10.48%	-45.38%
75 to 79	53.43%	-52.88%	75 to 79	35.22%	-37.43%
80 Plus	40.84%	-1.99%	80 Plus	31.06%	12.26%
Average	29.76%	-12.38%	Average	23.36%	-4.18%

* Rook-contiguity based variant controlled to shift-share tract total estimate for 2000.