Metropolitan and Nonmetropolitan Counties: Evaluation of Postcensal Population Estimates and Census 2010 Results

By

Ben Bolender
Population Division
U.S. Census Bureau

Irene Dokko
Population Division
U.S. Census Bureau

Jamie Lewis
Population Division
U.S. Census Bureau

For submission as a paper presentation to the annual meeting of the Population Association of America, San Francisco, CA May 3-5, 2012

BACKGROUND

The purpose of this paper is to compare the results of the U.S. Census Bureau's postcensal population estimates series with the population counts from Census 2010 across metropolitan, micropolitan, and nonmetropolitan counties. Office of Management and Budget (OMB) county classifications of metropolitan status go beyond simple categorization by population size. For one thing, they depend primarily on the location of urbanized areas (densely populated geographies of a certain population size) instead of overall population within the county borders. Further, these classifications include factors that indicate linked economic systems, such as shared commuting zones. As such, this analysis expands upon a simple size grouping scheme to look for impacts arising from a broader definition of the differences between metropolitan and nonmetropolitan counties.

The postcensal population estimates program calculates population figures for the nation, states, and counties by demographic characteristics on an annual basis. The basic principle is that we start with the previous decennial Census counts and estimate forward using administrative record inputs roughly based on the demographic balancing equation. In its simplest form, this means that the population at a given time equals the population from the base (the previous Census) plus births, minus deaths, plus net migration. National estimates are produced using birth and death records, military population data for those within the national borders and those overseas, estimates of international migration, and group quarters information. State and county estimates account for all of these (both individually and through higher geographic level controls) and additionally include the effects of domestic migration.

This analysis uses the OMB classifications (based on Census 2000 data) as a filter to separate the county level data into groupings based on metropolitan category (metropolitan, nonmetropolitan, and micropolitan). These classifications represent a "starting point" for counties in the decade, based on information similar to what would be available now as the new, 2010-based postcensal estimates series begins. The postcensal estimates series is predicated upon a population base of the previous decennial Census count. We classify counties based on their metropolitan category at the start of the decade since our overall ability to estimate the population greatly depends on the count as tabulated at that time.

METHODS

The estimates are compared to Census 2010 values using a variety of "measures of accuracy." First, numeric differences and percent differences are presented at the national and state level. We then present several measures which were selected prior to the beginning of the analysis.

Mean Absolute Percent Error (MAPE) = $((\sum ((|\text{Estimate} - \text{Census}|)/\text{Census}))/N)*100$

This measure takes the absolute value of the difference between the estimate and the Census value for each evaluation geography, divides that by each respective Census value, sums these up, divides by the number of evaluation geographies, and multiplies the end result by 100. The goal is to provide a relative measure of error. It ranges from zero to positive infinity and represents the average error across cases,

regardless of sign. Also, this is one of the most commonly used techniques available, and is therefore readily understood and easily accessible.

Mean Algebraic Percent Error (MALPE) = ((∑((Estimate – Census)/Census))/N)*100

Similar to the MAPE, this measure takes the difference between the estimate and the Census value for each evaluation geography, divides that by each respective Census value, sums these up, divides by the number of evaluation geographies, and multiplies the end result by 100. The goal is to identify systematic bias and provide an alternative for a relative measure of error. Its main value is that it preserves the sign of the error, allowing us to assess whether the estimates were generally higher or lower than the Census count.

Root Mean Squared Error = SQRT(∑ ((Estimate – Census)^2)/N)

This measure squares the difference between the estimate and the Census number for each evaluation geography, sums these values across evaluation geographies, divides by the number of evaluation geographies, and finds the square root of this value. It presents an alternative measure that places greater emphasis on large numeric errors versus mean absolute errors. We can interpret it in a similar fashion to a standard deviation being in roughly the same units as the values being evaluated.

Extreme Percent Error = Number of percent errors above a certain threshold

Unlike the other measures, this is a numeric value that relies upon an arbitrarily set threshold (5% and 10%). In short, the percent error is computed by dividing the difference between the estimate and Census values for a given area by the Census value for that area and multiplying by 100. The end measure simply represents a count of how many evaluation geographies in the summary area exceeded a particular threshold in their absolute percent error of the estimate. It provides an intuitive measure of the distribution of errors. It may also be presented as the percent of smaller geographies that fall outside this tolerance range within a larger geography (e.g. counties within a state).

Total Absolute Error of Shares = ∑ | ((Estimate/∑Estimate) – (Census/∑Census) |

This measure finds the proportion of each estimate to the total estimate for the summary geography and subtracts the proportion of the Census value and the total Census value for the summary geography. The absolute value of these proportional differences across evaluation geographies is then summed to the level of the larger geography. The goal is to provide a measure of the total amount of distributional error. Its interpretation is similar in some respects to an index of dissimilarity. It represents roughly the proportion of people who would need to be re-distributed in order to match the comparison distribution.

RESULTS

We first take a broad look at the overall results of the estimates evaluation project. Numeric and percent differences are displayed for the nation and states. Measures of accuracy are also presented for

all counties (evaluated at the national level). This provides a general backdrop for the more focused discussion on the distribution of measures by county metropolitan type.

We then move to show how the estimates compare to Census values in metropolitan, micropolitan, and nonmetropolitan counties. Measures of accuracy for these groups are compared to measures in county classifications based on size and growth as well. Overall, we look to see whether there are substantive differences in our ability to estimate the population based on the additional factors associated with metropolitan/nonmetropolitan classification beyond those seen if the focus was limited only to general population size measures.