A Quantitative Analysis of the Relative Effects of the Components of Change Used in Postcensal County Population Estimates

Sam Davis
Population Estimates Branch
Population Division
U.S. Census Bureau

Extended Abstract

Introduction

The Population Division produces annual postcensal estimates of county population using a component method which employs various sources of administrative records data. We know that each of the six components used to produce the estimates have a different impact on the population change of different counties. We are also able to use current displays of the components to compare the relative effects of the various components between to counties or among a small group of counties. What we have not obtained is a quantitative summary of the relative impact of each component. Having such a summary will allow us to do two things: 1) know how much impact a component has on groups of counties and all counties, and 2) for each county we can know how much effect each component has on that county's population change and how it compares with other counties.

There are two major uses of the knowledge which is generated by this quantitative analysis. The first is to give a more focused understanding of the population change for counties. While the components for individual counties are readily available, a summary enabling one to compare a county with the nation, regions, and other counties will provide a better understanding of the basis of the population change. The other use relates to the collection and review of the various inputs used to measure the components. Knowing how much influence a particular piece of input data has on the overall estimates will facilitate determining the resources that should be reserved for the collection and quality assurance of the data. This knowledge can help inform us of the impact of possible changes in the nature of input data and changes in the availability of data sources.

Estimates Procedure and Components

The basic procedure for applying the component method for population estimates involves taking the population at the beginning of a period and adding the net components to obtain the population at the end of the period. In our current estimates process the components we use are net natural change (births minus deaths), net internal migration, and net international migration. Some of these broad components are divided into subcomponents to use available data and adapt to different situations. Internal migration is obtained from IRS data for the population under 65 years of age and from Medicare enrollment data for the population 65 and over. The net internal migration also includes an estimate of the net change in the Group Quarters population. Net International Migration includes both net change in the foreign born population and net movement of the military population from abroad. Another component represented by change due to challenges and special censuses is not examined here due to the very small number of cases so far this decade.

Relative Influence of Components, the Postcensal Estimates

To measure the relative influence of components, we will look at the impact of each component on the beginning population for the period. For this situation the beginning population is the April 1, 2000 Census figure from the base population used for the vintage 2003 estimates. The components are for the period from April 1, 2000 to July 1, 2003. For each component we will obtain its percent of the total population. Since it is possible that changes in sign from year to year could dampen the total net effect of a component, we will use the sum of the absolute annual components over the period. We will then identify which component has the largest absolute percent of the beginning population. To determine the dominance of the largest component over the others, we will take the difference between the largest component percent and the lesser components. This will leas to grouping counties by the predominant component of change, by the difference between the largest component and the lesser components, and by positive and negative values of the components. From this various summaries and

description of the influence of the components on the total population can be produced. The basic tabulations will show the average percent of total population for each component by sign and magnitude group (First, second, third, etc.).

While we already know that natural increase leads in overall dominance among components, this analysis will allow us to identify those counties where it is more dominant, those where it is barely dominant, and those where other components are dominant, either positively or negatively. In the case of components which have little effect on a large number of counties (e.g. Group Quarters change), we will be able to identify in a quantitative manner where the counties are that are influenced substantially by the "minor" components.

In order to identify the influence of data used for internal migration, we have to look at the estimates by age, under 65 and 65 and over. For the under 65 population, we will be assessing the influence of the IRS data, and for the 65 and over population it is the Medicare data we will be examining.

Components During the 1990s, the Evaluation Context

To examine the influence of components on the estimates during the 1990s, we take a different approach. The absolute percent error for 1990-based county estimates for April 1, 2000 has already been computed using estimates produced by the routine component method. For each component will compute a set of estimates ignoring that component by setting that component to zero. We will then carry out the same analysis of the average absolute error from the Census 2000 results. The change in error for each component from the existing estimates will be used as a measure of that component's influence on the quality of the estimates. The increases (or decreases) in error can then be compared among the components as they are omitted from the computation. We will examine which counties are affected the most by which component's omission, and the degree to which their estimates are affected.