

## Research Spotlight

# BEA's County-Level Personal Income and Employment Estimates

## An Enhanced Geocoding Methodology

By *Mauricio Ortiz and Lisa Ninomiya*

STARTING IN THE EARLY 1970s, the Bureau of Economic Analysis (BEA) began regularly releasing county-level estimates of personal income and employment. These estimates have since become a critical component of BEA's broad suite of regional economic statistics, which are widely used by government officials, economists, business analysts, and others.<sup>1</sup>

As part of its efforts to continuously improve its regional economic measures, BEA has often confronted a common challenge: the limited availability of geographically detailed source data. Such challenges tend to become more acute as the size of the geographic region becomes smaller and more focused.

This issue has been particularly problematic in estimating nonfarm proprietors' income and employment, which are key components of personal income and total employment at the county level.

This Research Spotlight briefly discusses the following:

- The evolution of the data sources and methods used by BEA to geographically locate the economic activity of nonfarm unincorporated businesses in specific counties
- The shortcomings of these data and methods
- The experimental results of using nine-digit ZIP codes instead of five-digit ZIP codes to estimate county-level nonfarm proprietors' income and employment
- The next steps in BEA's efforts to more broadly incorporate nine-digit ZIP codes

The research suggests that adopting a methodology that relies on nine-digit ZIP codes instead of five-digit ZIP codes to more accurately capture nonfarm proprietors' income and employment in specific counties would only modestly affect BEA's estimates of personal income and total employment. However, the potential revisions to estimates of county-level nonfarm propri-

etors' income and employment are significant, with small counties more significantly affected than large counties.<sup>2</sup>

### Historical background

Historically, BEA's nonfarm proprietors' income and employment estimates were based on Internal Revenue Service (IRS) tabulations of individual income tax returns, which reported noncorporate business income. Address information provided on the returns was used to assign—or to “geocode”—income to specific geographic areas. The tabulations, however, provided limited income detail and were not available to BEA on a timely or regular basis.

Beginning in 1987, BEA began using data from IRS Form 1040 Schedule C (Sole Proprietorships) and Form 1065 (Partnerships), along with the five-digit ZIP codes in taxpayer addresses, to estimate nonfarm proprietors' income and employment at the state and county levels. BEA receives all Schedule C and Form 1065 returns for a given tax year. BEA geocodes each return to a specific geographical unit, such as a county, based on the five-digit ZIP code in the address on the tax forms.

This methodology marked an improvement over the initial methodology; it provided more income detail and more consistent geographic detail over time among other things. However, the use of five-digit ZIP codes for geocoding purposes is not without some key drawbacks.

### Shortcomings of five-digit ZIP codes

The United States Postal Service (USPS) introduced five-digit ZIP codes in 1963 to facilitate nationwide mail delivery. While these ZIP codes have been widely

1. For example, in fiscal year 2014, \$366.5 billion in federal funds were distributed in accordance with [BEA regional estimates](#).

2. Small counties are those with populations of less than 50,000 people. Medium counties are those with populations that are equal to, or more than, 50,000 and less than 500,000. Large counties are those with populations that are equal to, or more than, 500,000.

used by marketing analysts and others to delineate standard geographic areas, these codes were never meant to define spatial boundaries. They are more accurately defined as categories for grouping mailing addresses.

In most cases, ZIP codes resemble spatial areas in that they comprise spatially clustered street ranges. However, in some areas—rural areas, for example—ZIP codes are essentially collections of lines or rural delivery routes. In areas where there is no mail delivery, such as desert or mountain areas, ZIP codes may not be defined at all.

There are two major complications when working with ZIP codes to geocode data for estimation purposes. The first is that ZIP codes can change over time. The USPS can change the boundaries of an existing ZIP code, or there may be a need for mail delivery that did not previously exist (because of economic growth, for example). As a result, a location can be assigned a new ZIP code or a ZIP code that was previously used elsewhere. When using ZIP codes in time series analysis, these changes over time must be accounted for to ensure consistent estimates.

A second complication is that five-digit ZIP codes may cross county lines. A county may encompass one or more whole five-digit ZIP codes or one or more partial five-digit ZIP codes. BEA's current methodology, however, generally assigns nonfarm proprietors' income and employment from a given five-digit ZIP code to a single county.

In 1983, the USPS introduced nine-digit ZIP codes, which do not cross county lines, presenting obvious advantages for estimating county-level nonfarm proprietors' income and employment. However, in order to use the nine-digit ZIP code to geocode and to tabulate the IRS data, BEA needed to wait until the IRS adopted the use of nine-digit ZIP codes on tax forms. BEA also needed to obtain, and annually update, a correspondence file with county federal information processing standards (FIPS) codes assigned to every nine-digit ZIP code.

For many years, these conditions remained unmet, while BEA began to explore improvements to the geocoding process. Because nonfarm proprietors' income is a large component of personal income and because the amount reported on tax returns has a high variance, any change to the geocoding process can cause significant revisions. This consideration counseled patience in solving the matter of geocoding.

Subsequently, the USPS made available its ZIP+4 product, which made possible the creation of a correspondence file with county FIPS codes assigned to every nine-digit ZIP code. The ZIP+4 product provides a monthly file that matches every nine-digit ZIP code in

the United States to a county. Thus, using this ZIP+4 file for the matching tax year to tabulate IRS Schedule C and Form 1065 eliminates both the problem of five-digit ZIP codes that cross county lines and the problem of new or changing ZIP codes.

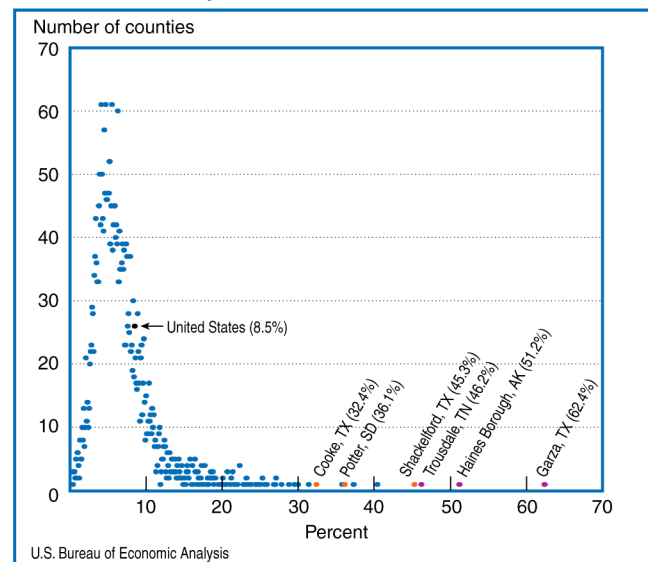
BEA has obtained from the USPS a set of annual ZIP+4 files with nine-digit ZIP codes matched to county FIPS codes for 2001 to the present. BEA has also reached an agreement with the IRS to retabulate the Schedule C and Form 1065 returns for 2001 forward with the nine-digit ZIP codes from taxpayer and partnership addresses. These data for 2013 form the basis of the experimental estimates discussed in this article.

## Results

Assigning nonfarm proprietors' income and employment according to nine-digit ZIP codes instead of five-digit ZIP codes results in only modest revisions to BEA's county-level estimates of personal income and total employment for 2013. However, the revisions to estimates of county-level nonfarm proprietors' income and employment are widespread and significant, with small counties more affected than large counties.

In the currently published 2013 estimates, nonfarm proprietors' income accounted for 8.5 percent of personal income nationally. At the county level, nonfarm proprietors' income's share of personal income exhibited large variation. The share ranged from as little as 0.3 percent in Forsyth, Georgia, in Wheeler, Georgia, and in Sussex, Virginia to as much as 62.4 percent of personal income in Garza, Texas, 51.2 percent in Haines Borough, Alaska, and 46.2 percent in Trousdale, Tennessee (chart 1).

**Chart 1. Published Nonfarm Proprietors' Income as a Percent of County Total Personal Income, 2013**



Nonfarm proprietors employment accounted for 21.1 percent of total employment nationally in 2013. Again, at the county level, nonfarm proprietors employment's share of total employment varied widely. The share ranged from as little as 2.9 percent in Ohio, West Virginia to as much as 75.9 percent of total employment in Trousdale, Tennessee, 74.9 percent in Chase, Kansas, and 73.9 percent in Haines Borough, Alaska (chart 2).

Recreating the estimates using 2013 nine-digit ZIP codes to geocode and to retabulate the IRS source data results in significant revisions to county-level estimates of nonfarm proprietors' income in certain cases. For example, for Trousdale, Tennessee, nonfarm proprietors' income estimate was revised down 89.3 percent, and the personal income estimate was revised down 41.2 percent (table 1). Its nonfarm proprietors' income share of personal income fell to 8.4 percent from 46.2 percent (chart 3).

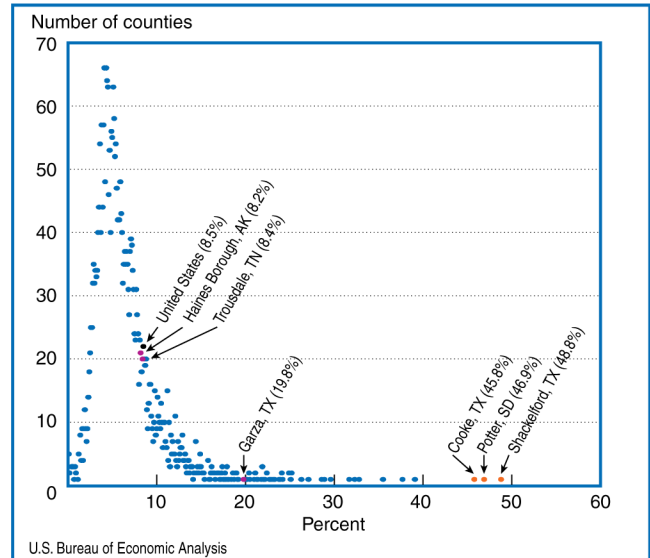
Similarly, for Haines Borough, Alaska, nonfarm proprietors' income estimate was revised down 91.5 percent, and the personal income estimate was revised down 46.8 percent. Its nonfarm proprietors' income share of personal income fell to 8.2 percent from 51.2 percent. For Garza, Texas, nonfarm proprietors' income estimate was revised down 85.1 percent, and the personal income estimate was revised down 53.2 percent. Its nonfarm proprietors' income share of personal income fell to 19.8 percent from 62.4 percent.

For Trousdale, Tennessee, Haines Borough, Alaska, and Garza, Texas—which are small counties—the current five-digit ZIP code based geocoding procedure has not properly responded to changing ZIP code assign-

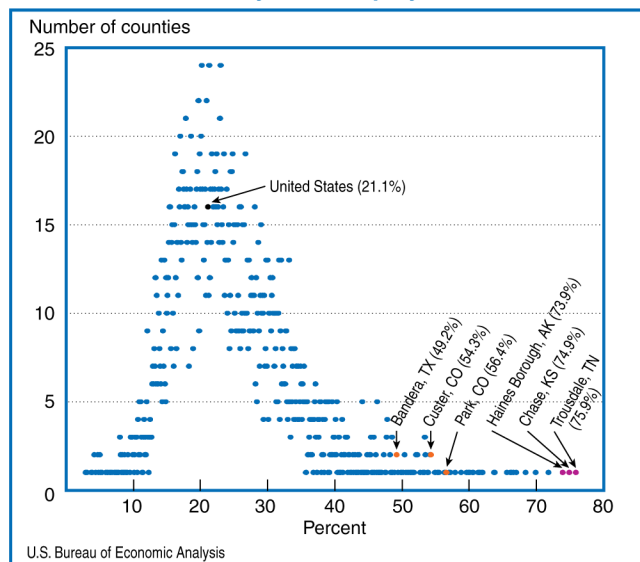
ments, resulting in misallocated income from zip codes no longer included within county borders. The nine-digit ZIP code geocoding process corrects the problem but results in large revisions to estimates of nonfarm proprietors' income. More typical results are the ones experienced by three other small counties—Shackelford, Texas, Potter, South Dakota, and Cooke, Texas (charts 1 and 3).

For Haines Borough, Alaska, nonfarm proprietors employment estimate was revised down 78.9 percent, and the total employment estimate was revised down 58.4 percent (table 2). Its nonfarm proprietors employment share of total employment fell to 37.4 percent from 73.9 percent (chart 4).

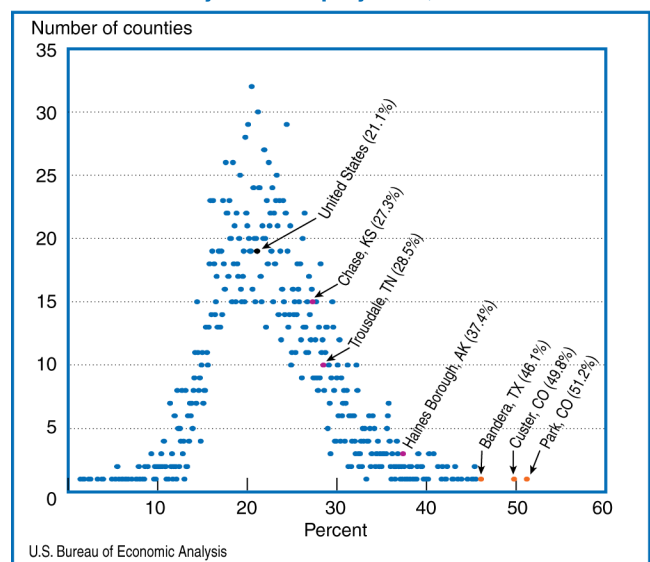
**Chart 3. Revised Nonfarm Proprietors' Income as a Percent of County Total Personal Income, 2013**



**Chart 2. Published Nonfarm Proprietors Employment as a Percent of County Total Employment, 2013**



**Chart 4. Revised Nonfarm Proprietors Employment as a Percent of County Total Employment, 2013**



Similarly, for Chase, Kansas, nonfarm proprietors employment estimate was revised down 87.4 percent, and the total employment estimate was revised down 65.4 percent. Its nonfarm proprietors employment share of total employment fell to 27.3 percent from 74.9 percent. For Trousdale, Tennessee, nonfarm proprietors employment estimate was revised down 87.3 percent, and the total employment estimate was revised down 66.3 percent. Its nonfarm proprietors employment share of total employment fell to 28.5 percent from 75.9 percent.

The weaknesses in current geocoding procedures have also resulted in inflated estimates of nonfarm proprietors employment for Chase, Kansas. The nine-digit ZIP code geocoding process corrects the problem and results in better employment estimates. The revisions to nonfarm proprietors employment for Haines Borough, Alaska, Chase, Kansas and Trousdale, Tennessee, result in large revisions to estimates of total employment. More typical results are the ones experienced by Park, Colorado; Custer, Colorado; and Bandera, Texas, small counties, also highlighted in charts 2 and 4.

The effect of using nine-digit ZIP codes on personal income and total employment are much more modest. Only 80 counties or approximately 3 percent had revisions to personal income greater than plus/minus 10 percent (chart 5); 75 of these counties are small, and 5 are medium sized. For total employment, 538 counties, or about 17 percent, had revisions greater than plus/minus 10 percent (chart 6); 463 of these counties are

small, 72 are medium sized, and 3 are large.

**Next steps**

The statistics presented here are being released for evaluation and comment by data users. Subject to data users' evaluations and comments BEA intends to revise state and county estimates of nonfarm proprietors' income and employment back to 2001 based on nine-digit ZIP codes. Releasing a long span of retabulated data allows the data to be used for time series analysis, a valued aspect of BEA's regional statistics.

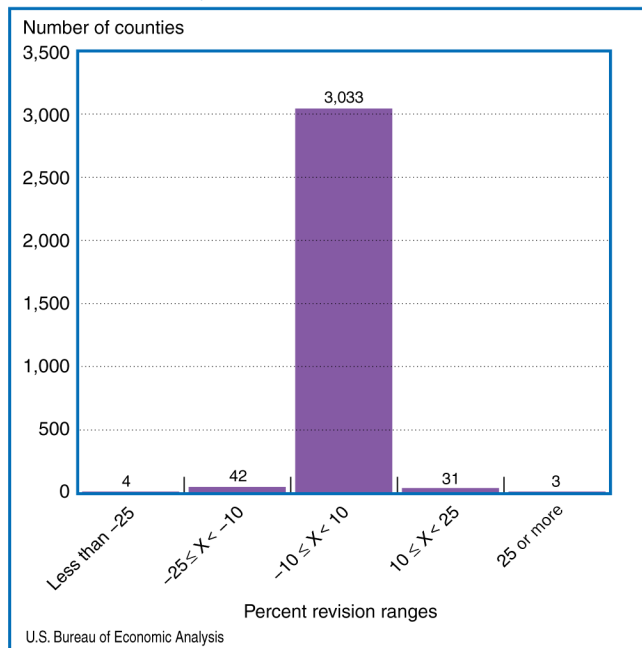
BEA is currently in the process of reacquiring 2001–2012 tax return information from the IRS, a necessary step toward incorporating nine-digit ZIP codes.

BEA will also take the opportunity to revisit its entire methodology for creating the nonfarm proprietors' income and employment estimates. Please note that the data in this article are experimental. When BEA makes the change to nine-digit ZIP codes and updates the nonfarm proprietors' methodology, the revised estimates may not match the estimates presented here.

The timing for incorporating the nine-digit source data into published statistics depends on the availability of resources at both the IRS, which must supply the Schedule C and Form 1065 returns for historic periods, and at BEA, which has to process, review, and publish the revised statistics. The process also calls for close collaboration between the two federal agencies to safeguard confidential taxpayer information.

A detailed timetable is currently under development.

**Chart 5. Revisions to County Estimates of Total Personal Income, 2013**



**Chart 6. Revisions to County Estimates of Total Employment, 2013**

