

Notice: Important Changes to EITC Interactive Data

The Tax Year 2014 version of the EITC Interactive contains data that are substantially different from previous years’ data, due to changes in IRS SPEC’s suppression protocols. This document outlines the nature of the changes, their effect on the data, and provides some guidance about using the data to make comparisons.

**Changes to the Data**

The IRS made three major changes to the data this year:

1. **Only tax returns where the taxpayer’s adjusted gross income (AGI) was under $60,000 are included.** In previous years, the data included tax returns for all filers in a ZIP code, regardless of AGI.
2. **Data for ZIP codes that contain 100 or fewer returns are excluded.** Due to the $60,000 threshold described above, this change means that a ZIP code must have at least 100 filers with AGI under $60,000 in order to be included. Data for all other ZIP codes is not present. In previous years, ZIP codes that had at least 10 filers of any income level were included.
3. **Only variables where more than 20 tax returns were present were included in the data.** Cells that contained 20 or fewer returns are suppressed to zero. In previous years, the threshold for suppression was 10 tax returns.

**Effects of the Changes**

* **This year’s version of the EITC interactive only contains data covering EITC returns:**

This dataset no longer provides the size of the tax filing universe, due to the loss of data about filers with AGI over $60,000. As a result, the limited data on all tax filers is only available upon request (interested users should e-mail Natalie Holmes). Since filers with AGI over $60,000 generally are not eligible to claim the EITC, this particular change does not impact the EITC market segment as heavily.

* **Without data on all tax filers, it is not possible to compare the characteristics of the EITC market segment with those of the universe of tax filers.** For example, these data no longer can be used to determine what share of all filers claimed the EITC or what share of tax returns prepared by volunteers were EITC returns.
* **Even in the EITC market segment, data suppression has increased substantially:**

Unfortunately, the IRS does not provide a flag indicating that data has been suppressed, so it is difficult to precisely identify the full scope of suppression. However, in order to better understand the impact of the changes, we applied these new suppression protocols to the data from Tax Year 2013 and were able to identify several probable effects:

* + Complete suppression of ZIP codes is geographically concentrated in states with significant rural components, where ZIP codes contain fewer tax filers.
  + Even in ZIP codes where the total number of EITC filers is available, many key variables are newly suppressed. In particular, the variables for returns prepared by volunteers, filers who did not file last year, and the AGI categories are likely among the most affected.

**Comparisons Over Time and Across Geographies**

As a result of the changes in the universe of tax returns, comparisons between the TY2014 data and data from previous years are not valid. Even comparisons in the EITC market segment are unlikely to be valid, due to the higher suppression threshold. In addition, because the suppression protocols disproportionately impact less-populated places and because this effect may vary for different variables, comparisons across geographies should be made with caution.