

Comparison Between the Demographic Composition of Matched IRS Exemptions and 2010 Census Household Population Counts

Authors: Joseph Bowman, Esther R Miller

I. Introduction

The U.S. Census Bureau's Population Division develops population estimates using a component of population change method. The domestic migration component is indirectly measured by matching two consecutive years of IRS tax return exemptions. Implicit is the assumption that people who do not file taxes migrate at the same rate as tax filers and their associated exemptions. However, the tax filing population may or may not be representative of the population as a whole.

This paper provides a comparison of matched IRS exemptions to the enumeration of the United States population in Census 2010. To do this we calculate coverage rates. Coverage rates are the percent at which the matched IRS universe covers the Census 2010 household population counts (i.e., the number of matched IRS exemptions divided by census household population counts). We examine differences in coverage rates by subnational geography and demographic characteristics.

II. Data Sources

Data for the coverage research are drawn from three types of sources: decennial census files, Internal Revenue Service (IRS) 1040 tax returns, and the Census Bureau's Person Demographic Characteristics File (PDCF).

A. Census 2010 Counts

The Constitution requires the Census Bureau to enumerate all residents in the United States every 10 years. The most recent enumeration is April 1, 2010. The total population includes people enumerated in both housing units and Group Quarters (GQ) facilities. Because the universe of the domestic migration component is the household population, we only use Census 2010 household population counts.

B. Internal Revenue Service (IRS) 1040 tax returns

An IRS 1040 tax return contains data for a filer, spouse/dependents (if applicable), and one address. The Census Bureau converts returns into individual person records and geocodes each person's address to a county Federal Information Processing Services (FIPS) code. It then matches two consecutive years of tax exemptions by a Protected Identification Key (PIK).¹ This allows us to compare county FIPS codes across years and assign migration status. For this research, we use tax year 2009 (TY09) and tax year 2010 (TY10).

C. Person Demographic Characteristics File (PDCF)

The IRS 1040 data do not contain demographic characteristics for individuals. Instead, demographic characteristics are assigned to exemptions by matching IRS records to the PDCF by PIK. The PDCF is created by combining data from the Social Security Administration's Numident file, the 2000 and 2010 Decennial Census files, and other administrative record data. The Numident file contains a PIK, birth date, and sex for every SSN ever assigned by the Social Security Administration. Race and Hispanic origin are assigned to PIKs using data from the 2000 and 2010 Censuses and other administrative records data.

III. Domestic Migration Universe

In order to determine a record's migration status, a PIK must appear in two consecutive years. We refer to these records as matched IRS exemptions. We then match these records to the PDCF. Those that do not appear in two consecutive years, or do not have a PDCF record, are excluded from the migration universe.

¹ A Protected Identification Key (PIK) is assigned to a Social Security Number (SSN) to protect confidentiality. The same PIKs are assigned to all administrative records acquired by the Census Bureau. PIKs can be matched across administrative records.

IV. Coverage Rate Computations

Coverage rates are the percent at which the matched IRS universe covers the Census 2010 household population counts. Prior to estimating domestic migrants, we calculate the household population by subtracting the group quarters population from the resident population. Migrants are estimated by applying the migration rates (and proportions), developed from tabulations of the matched TY09-TY10-PDCF universe, to the household population. Coverage rates are calculated as:

$$\textit{Coverage Rate} = \frac{\textit{Matched IRS Exemptions}}{\textit{Household Population}}$$

V. Results

There are differences in coverage rates by geography and demographic characteristics. We plan to present overall state and county coverage rates spatially using GIS software. Demographic characteristics (e.g., age, sex, race, and Hispanic origin) will be presented using a combination of tables and charts (e.g., graphs, boxplots, etc.)