International Journal of Population Data Science

Journal Website: www.ijpds.org

Linking Survey and Administrative Data to Measure Income, Inequality, and Mobility

Medalia, C^{1*}, Meyer, B², O'Hara, A³, and Wu, D²

¹U.S. Census Bureau

²University of Chicago

³Stanford University

Income is one of the most important measures of well-being, but it is notoriously difficult to measure accurately. Income data are available from surveys, tax records, and government programs, but each of these sources has important strengths and major limitations when used alone. We are linking multiple data sources to develop the Comprehensive Income Dataset (CID), a restricted micro-level dataset that combines the demographic detail of survey data with the accuracy of administrative measures. By incorporating information on nearly all taxable income, tax credits, and cash and in-kind government transfers, the CID surpasses previous efforts to provide an accurate and comprehensive measure of income for the population of U.S. individuals, families, and households. We use models to evaluate differences across the data sources and explore imputation methods and trends over time. The CID can enhance Census Bureau surveys and statistics through investigating measurement error, improving imputation methods, and augmenting surveys with the best possible estimates of income. It can also be used to improve the administration of taxes by the Internal Revenue Service and forecast and simulate changes in programs and taxes. Finally, the CID has substantial advantages over other sources to analyze numerous research topics, including poverty, inequality, mobility, and the distributional consequences of government transfers and taxes.

*Corresponding Author:

Email Address: carla.medalia@census.gov (C Medalia)