A Comprehensive Income Dataset

Bipartisan Policy Center
September 6, 2018
Bruce D. Meyer, Univ. of Chicago, NBER, AEI, U.S. Census Bureau

Joint with
Carla Medalia, U.S. Census Bureau
Amy O’Hara, Stanford University
Derek Wu, University of Chicago

Disclaimer: Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.
A Comprehensive Income Dataset (CID)

- Combine the advantages of three sources of data
- **Surveys**: rich demographics, random sample of population, flexible platform
- **Tax Data**: accuracy, broad coverage (with information returns)
- **Program Data**: accuracy, needed to evaluate programs, eligibility and other information not available on surveys
Goals of the CID

- Improve Census Bureau Surveys
- Enhance tax administration
- Provide the information to better examine public policies and support research
Survey Data Alone Inadequate

- Severe under-reporting of many types of income
  - Government transfers
  - Private pensions
- Most of bias in government program dollars in surveys due to measurement error: Meyer, Mok, Sullivan (JEP Fall 2015)
- One can compare weighted survey reports to administrative totals
- Gives net reporting, net bias (false positives can cancel out false negatives)
Table 1
Proportional Bias in Survey Estimates of Mean Program Dollars and Months Received, by Program and Survey, 2000–2012

<table>
<thead>
<tr>
<th></th>
<th>AFDC/ TANF</th>
<th>FSP/ SNAP</th>
<th>Social Security</th>
<th></th>
<th></th>
<th></th>
<th>UI</th>
<th>WC</th>
<th>NSLP</th>
<th>WIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dollars</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACS</td>
<td>−0.519</td>
<td>−0.458</td>
<td></td>
<td>−0.165</td>
<td>−0.299</td>
<td>−0.046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE</td>
<td>−0.767</td>
<td>−0.587</td>
<td></td>
<td>−0.149</td>
<td>−0.214</td>
<td>−0.283</td>
<td>−0.583</td>
<td>−0.618</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPS</td>
<td>−0.500</td>
<td>−0.417</td>
<td></td>
<td>−0.086</td>
<td>−0.187</td>
<td>−0.162</td>
<td>−0.325</td>
<td>−0.541</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSID</td>
<td>−0.619</td>
<td>−0.308</td>
<td></td>
<td>−0.086</td>
<td>−0.176</td>
<td>−0.322</td>
<td>−0.360</td>
<td>−0.646</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIPP</td>
<td>−0.357</td>
<td>−0.170</td>
<td></td>
<td>−0.070</td>
<td>−0.146</td>
<td>0.164</td>
<td>−0.388</td>
<td>−0.651</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Months</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−0.154</td>
<td>−0.261</td>
<td>−0.372</td>
<td></td>
</tr>
<tr>
<td>CPS</td>
<td>−0.453</td>
<td>−0.422</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−0.147</td>
<td>−0.154</td>
<td>−0.397</td>
<td>−0.503</td>
</tr>
<tr>
<td>PSID</td>
<td>−0.574</td>
<td>−0.297</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−0.114</td>
<td>−0.121</td>
<td>−0.502</td>
<td>−0.470</td>
</tr>
<tr>
<td>SIPP</td>
<td>−0.232</td>
<td>−0.165</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.008</td>
<td>0.041</td>
<td>0.023</td>
<td>0.141</td>
</tr>
</tbody>
</table>

Source: Meyer, Mok and Sullivan JEP (2015)
Tax or Program Data Alone Inadequate

- **Tax Data**
  - Lack demographics
  - Lack program information for in-kind benefits
  - Lack information on non-taxable income
  - Not for economic unit
  - Miss nonfilers

- **Program Data**
  - Limited history and demographics
  - Miss non-recipients
## Data

<table>
<thead>
<tr>
<th>Source type</th>
<th>Phase I</th>
<th>Phase II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Survey of Income and Program Participation (SIPP)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>American Community Survey (ACS)</td>
<td></td>
</tr>
<tr>
<td>Tax data</td>
<td>Forms 1040, W-2, 1099-R</td>
<td>Better 1040 extracts, more extensive info returns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tax credits (e.g. EITC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unemployment Insurance (UI)</td>
</tr>
<tr>
<td>Federal programs</td>
<td>SSA: Social Security and Supplemental Security Income</td>
<td>VA: Veterans Benefits</td>
</tr>
<tr>
<td></td>
<td>HUD: Federal housing assistance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HHS: Medicare and Medicaid enrollment, TANF</td>
<td></td>
</tr>
<tr>
<td>State programs</td>
<td>Public Assistance (e.g. TANF, GA)</td>
<td>More Public Assistance, SNAP, WIC, LIHEAP</td>
</tr>
<tr>
<td></td>
<td>SNAP, WIC</td>
<td>Workers Compensation</td>
</tr>
<tr>
<td></td>
<td>LIHEAP</td>
<td></td>
</tr>
</tbody>
</table>
Linking

- Records are linked using an anonymized version of the SSN called a Protected Identification Key (PIK)
  - Usually can attach a PIK to 99 percent of admin data
    - Records usually have SSN that has been validated
  - Usually can attach a PIK to 90-97 percent of households in survey data
Uses of CID: Survey Improvement

- Assess extent and nature of errors
- Improve imputation of missing values
- Substitute administrative values for survey ones
- Augment surveys, avoid asking some questions
Uses of CID: Tax Administration, Policymaking

- Tax administration (IRS)
  - EITC takeup
  - Identifying areas of noncompliance
- Forecasting and simulation of changes in programs and taxes
Uses of CID: Research

- Measure income distribution, poverty, inequality
- Estimate effects of social insurance and means-tested transfers
- Simulate changes in programs and taxes
The Poverty Reduction Due to Social Security and Means-Tested Transfers

Meyer and Wu (forthcoming)

- Worst case scenario for value of the CID because uses better survey data than any new data
Poverty Rate Reduction from Combined vs. Survey Data: OASDI, SSI, SNAP, PA
The Use and Misuse of Income Data and the Rarity of Extreme Poverty in the United States

Meyer, Mooers and Wu (2018)
Share of Households in Extreme Poverty After Corrections, by Household Type

- Reported Cash
- After In-Kind Transfers
- After Wage/Salary Hours
- After Wage/Salary and Self-Employment Hours
- After Substantial Assets
- After Admin Tax Income
- After Admin Tax + Transfer Income

Extreme Poverty Rate

All Household Types
- Elderly
- Multiple Adults
- Single Individuals
- Multiple Parents
- Single Parents
Future Plans for the CID

- Provide products to federal and state agencies
  - Work with constituencies to design them
- Better documentation of data and methods
- Advisory board
- Possible Census adoption of CID as regular program