Using Administrative Data to Improve the SIPP

Graton Gathright
US Census Bureau

WSS Conference on Administrative Records
September 18, 2014

Disclaimer

My remarks today do not necessarily reflect the official position of the US Census Bureau.

Using Administrative Data to Improve the SIPP

1. SIPP background

2. Challenges to using admin data for SIPP improvement

3. Using linked SIPP
   a) Evaluation of 2014 SIPP
   b) Public-use linked data: SIPP Synthetic Beta
   c) Model-based imputation of missing data in 2014 SIPP

Survey of Income and Program Participation (SIPP)

- Authorized by Title 13
- Sponsored and conducted by Census
- Focuses on intra-year dynamics of earnings, income, and receipt of program benefits
- First panel in 1984, 14 multi-year panels since (lasting 2-6 years each)
- 10,000 to 50,000 household per panel (50,000 in 2014 Panel)
- Nationally representative
- Over-samples low-income
- 2014 Panel changed to 1 year reference period (from every 4 months)

United States
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

4

Challenges to using admin data to improve SIPP

1. Data availability
2. Data acquisition
3. Data quality
4. Data linking
5. Data access
6. Data protection

Admin data linked to SIPP

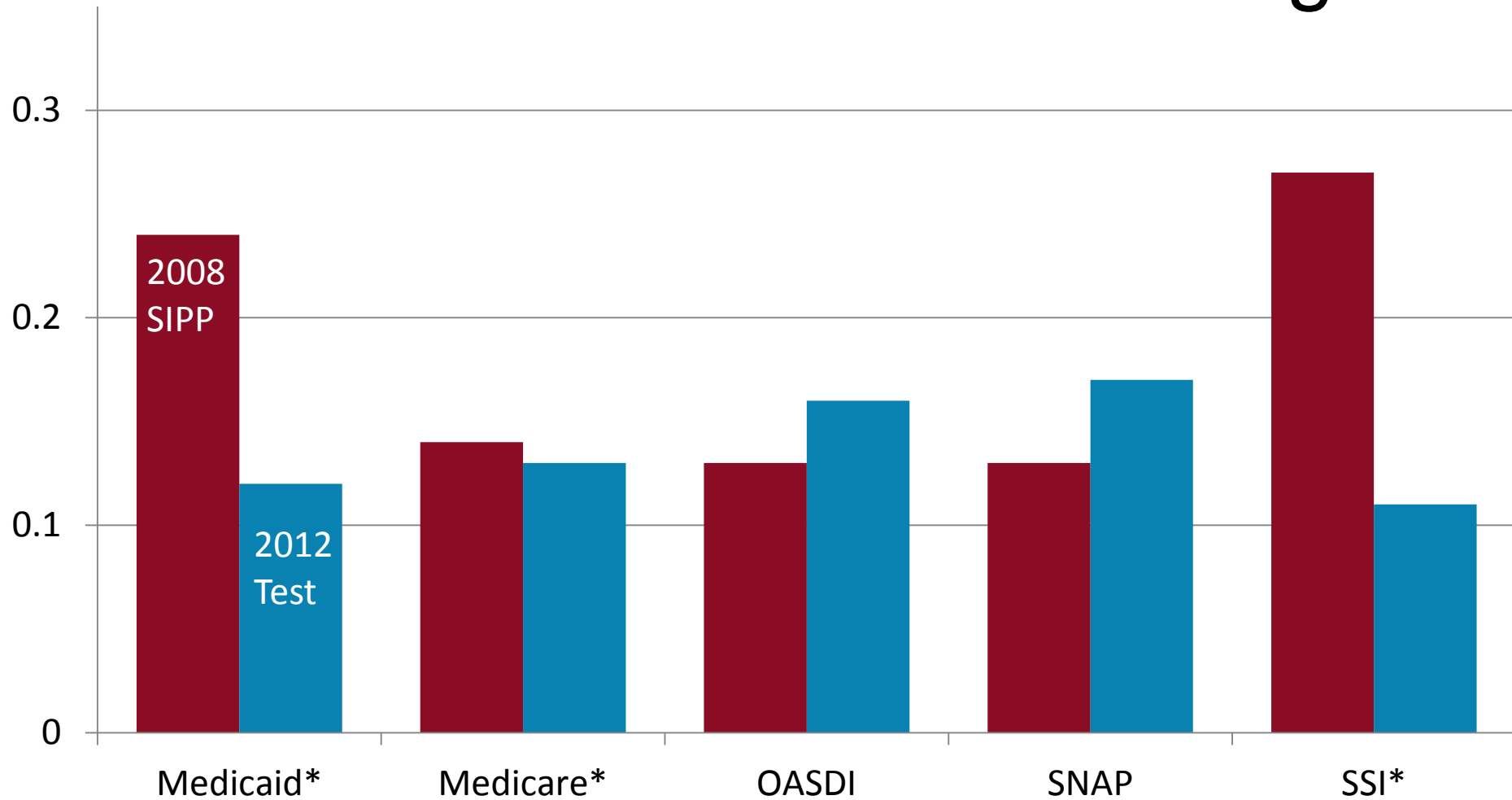| Data	Topic(s)	Source	Years available (as of September 2014)
DER/SER	Jobs and earnings	SSA/IRS	1978-2012
MBR/PHUS	Social Security applications, eligibility, & payments	SSA	1962-2013
SSR	Supplemental Security Income	SSA	1962-2013
NY SNAP/TANF	SNAP and TANF participation and payment for NY	NYS OTDA	2009-2012
MSIS	Medicaid	CMS	to 2012
EDT	Medicare	CMS	to 2013
PICS & TRACS	Housing subsidies	HUD	to 2013
IRS Form 1040	Income	IRS	2004-2012
IRS Form 1099R	Retirement income distributions	IRS	2004-2012
Business Register	Employer characteristics	IRS and others	to 2012

Outline

1. SIPP background

2. Challenges to working with linked SIPP

3. Using linked SIPP

   a) **Evaluation of 2014 SIPP**

   b) Public-use linked data: SIPP Synthetic Beta

   c) Model-based imputation of missing data in 2014 SIPP

Evaluation of the 2014 SIPP Redesign

	Survey	Topics
Stinson, Gathright, and Skog (2012)	2010 Field Test	Earnings, number of jobs, industry
Gathright, Stinson, and Reeder (2011)	2010 Field Test	Social Security, SSI, SNAP, TANF, Medicare
Gathright, et al. (2013)	2011 and 2012 Field Tests	Earnings, jobs, Social Security, SSI, SNAP, TANF, Medicare, Medicaid
Census Bureau evaluation	2014 SIPP Wave 1	Social Security, SSI, SNAP, TANF, Medicare, Medicaid
CNSTAT evaluation	2014 SIPP Wave 1	Social Security, SSI, SNAP, TANF, Medicare, Medicaid

False negative rates for 2008 SIPP and 2012 Test of 2014 Redesign

Outline

1. SIPP background

2. Challenges to working with linked SIPP

3. Using linked SIPP

    a) Evaluation of 2014 SIPP

    b) Public-use linked data: SIPP Synthetic Beta (SSB)

    c) Model-based imputation of missing data in 2014 SIPP

The SIPP Gold Standard and the SIPP Synthetic Beta (SSB)

SIPP Gold Standard File (GSF)
- Nine harmonized SIPP panels (selected variables)
- Admin earnings histories for 1951-2012
- Admin SSA benefits histories for 1962-2013

SIPP Synthetic Beta (SSB)
- Reflects relationships between GSF variables
- Public-use
- Protects confidential data (cannot be linked to SIPP)
- Serves as platform for remote analysis of GSF
- 16 multiple imputation implicates

Design Decisions:
SSB Contents – SIPP variables

- Demographic Variables
  - five category education
  - Black, Hispanic
  - birth month/year, death month/year
  - disability limits work or prevents work
  - total number of children in family, number of biological children, birth date of first and last child
  - history of marital events up to 4 marriages, age at which marital events occur
  - foreign born, decade arrive in the United States

- Economic – monthly indicators
  - Weeks with job, weeks with pay
  - Usual hours worked
  - Survey-reported earnings
  - Total personal income
  - Health insurance coverage, any and employer-provided

- Wealth and Pension
  - Total net worth, home ownership indicator, home equity, non-housing wealth
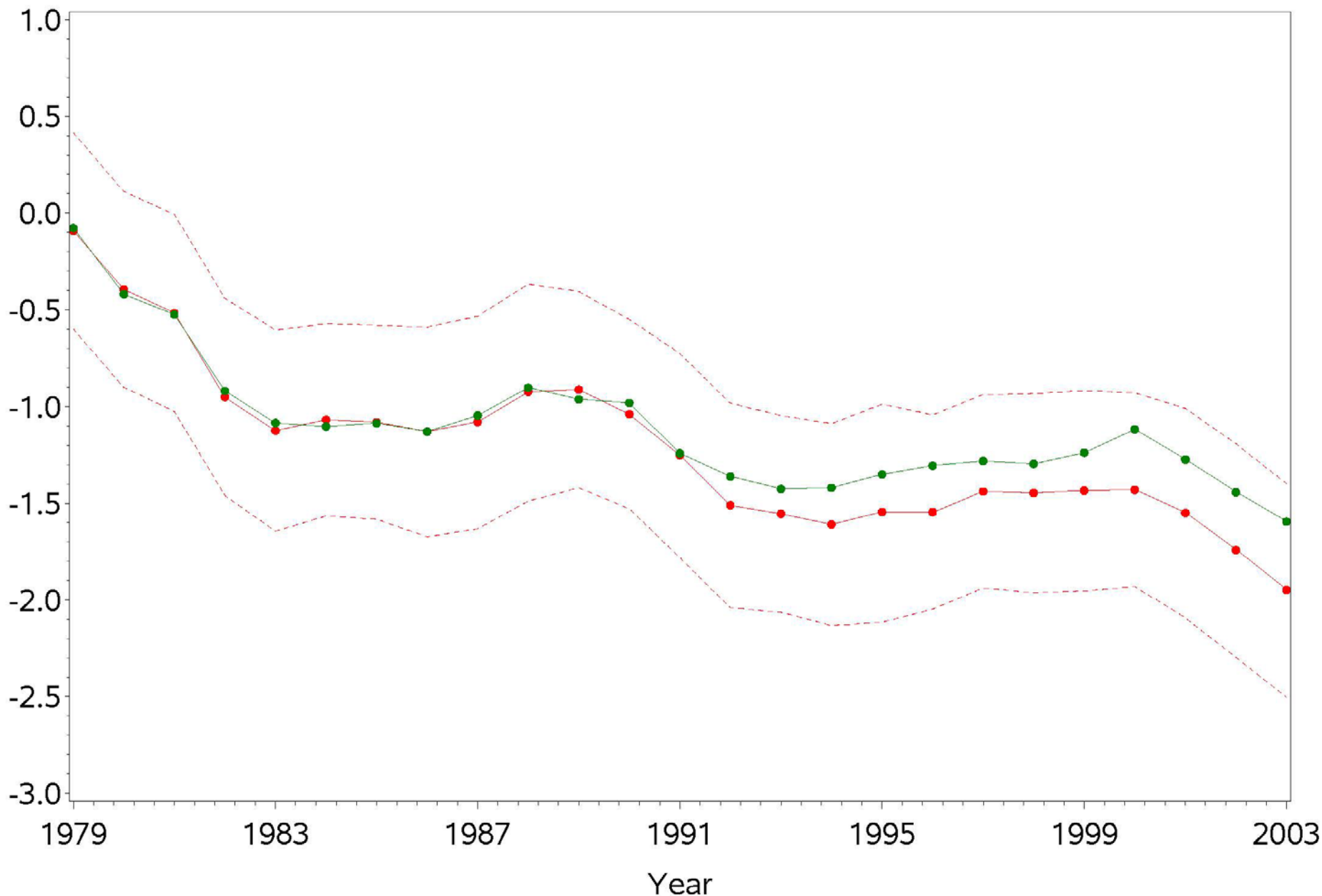 - Defined benefit plan, defined contribution plan

United States™
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

13

Design Decisions:
SSB Contents – administrative variables

- Summary Earnings Record Extract 1951-2011
  - Annual total FICA covered earnings (capped)
  - Annual pattern of quarters worked 1951 – 1977
  - Total earnings from 1937 to 1950

- Detailed Earnings Record Extract 1978-2011
  - Wages, Tips, and other compensation (Box 1 Form W-2) (uncapped)
  - Deferred Wages (Box 13 Form W-2) – e.g. 401(k) contributions
  - Split into earnings from FICA covered employment and earnings from non-FICA covered employment
  - Summed across all employers

- Master Beneficiary Record – OASDI program
  - Indicators for type of benefit
  - Date began receiving each type of benefit
  - First amount received

- Supplemental Security Record – SSI program
  - Date began receiving SSI benefit
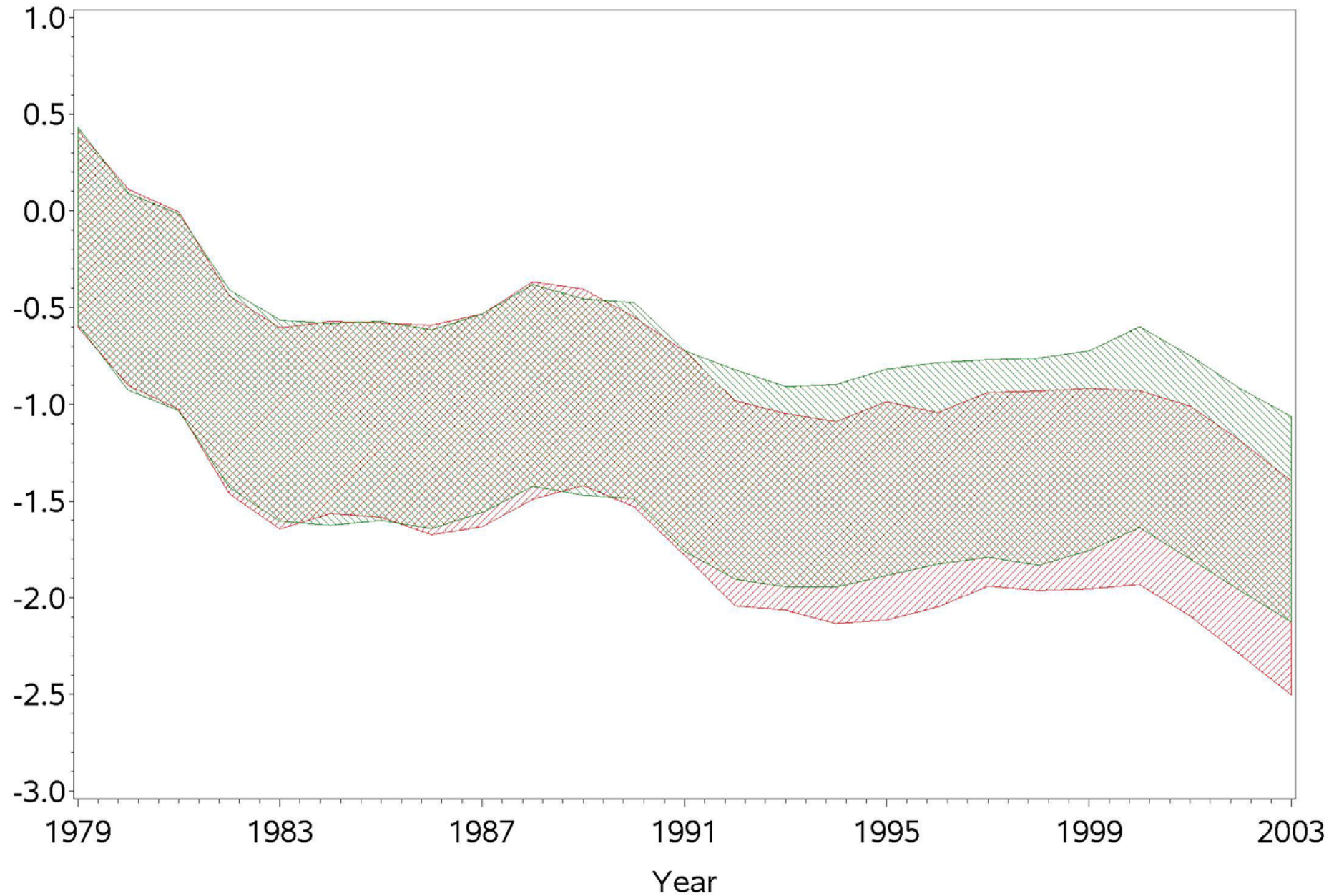 - First amount received

Log Earnings Relative to 1978 for Males Without H.S. Diploma
Comparison of Completed and Synthetic Data

Log Earnings Relative to 1978 for Males Without H.S. Diploma

Comparison of Completed and Synthetic Data

SSB Improvements

Version 6.0 (scheduled for release this winter)
  – Adds 1984 and 2008 panels
  – Adds CY2011 admin data
  – Adds disability benefits application data
  – Adds state of residence variable
  – Adds child sample persons

Future versions
  – Job-level earnings and employer characteristics
  – Parent-child links
  – Data from IRS Forms 1040 and 1099R
  – Additional SIPP variables, eg. Veterans benefits

Outline

1. SIPP background

2. Challenges to working with linked SIPP

3. Using linked SIPP

   a) Evaluation of 2014 SIPP

   b) Public-use linked data: SIPP Synthetic Beta

   c) Model-based imputation of missing data in 2014 SIPP

Imputation of missing data for 2014 SIPP using admin data

- Topic Flag Imputation
  - 44 topic flags in 2014 SIPP indicate whether a topic section is in universe for a sample person
  - Model-based imputation (but not multiple)
  - Models can include data from MBR, PHUS, SSR, and DER
- Challenges
  - Data acquisition
  - Approval process

Conclusion

Linked SIPP challenges
1. Data availability
2. Data acquisition
3. Data quality
4. Data linking
5. Data access
6. Data protection

Linked SIPP uses:
1. Evaluation of SIPP data quality, e.g. the 2014 redesign
2. Hybrid data products, e.g. GSF/SSB
3. Improvements, e.g. Improved imputation of missing data

Graton M. R. Gathright

Social, Economic, and Housing Statistics Division

US Census Bureau

301.763.7863

graton.m.gathright@census.gov

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU