

# Is the Consumer Expenditure Survey Representative by Income?

**John Sabelhaus**

Federal Reserve Board

**David Johnson**

**Stephen Ash**

U.S. Census Bureau

**David Swanson**

**Thesia Garner**

**John Greenlees**

**Steve Henderson**

U.S. Bureau of Labor Statistics

Presentation prepared for the CRIW Conference on *Improving the Measurement of Consumer Expenditures*, December 3, 2011. Views here do not represent those of the Federal Reserve Board, the U.S. Census Bureau, or the U.S. Bureau of Labor Statistics.

# Motivation

- BLS tables show CE is missing about one-third of comparable NIPA PCE spending totals
- Ratio got worse in 90s, but stable since 2003
- Assuming PCE correct, two sources of error:
  1. Higher income consumer units (who presumably spend more) under-represented in CE sample
  2. Under-reporting of total spending by at least some CE respondents

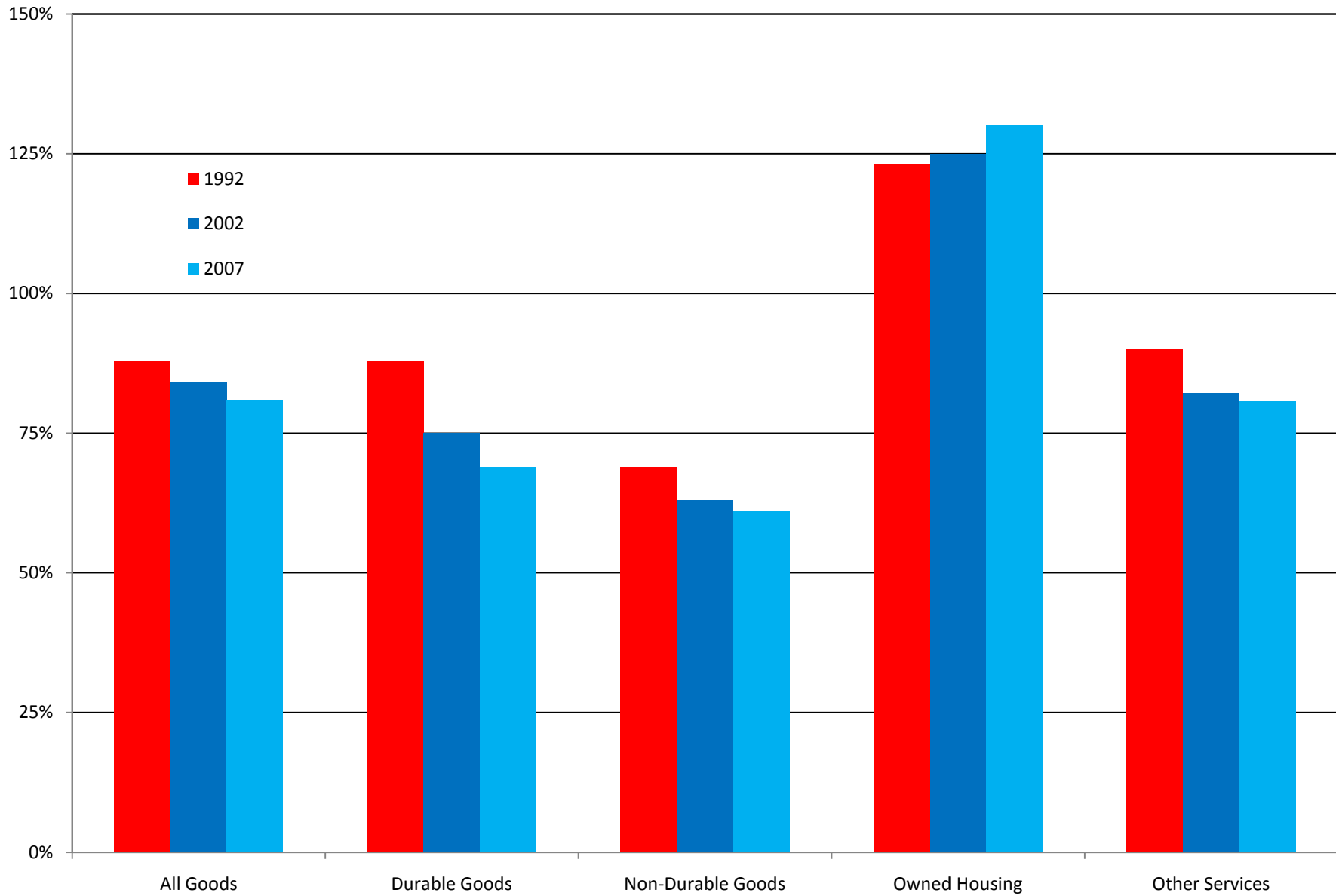
# Contributions of this Paper

- New test of survey response rates by income
  - Link sampled CE units to their zip-code level adjusted gross incomes (AGI) from public-use IRS data
  - Response rates are lower at highest AGI percentiles
  - Ratio of mean CE income to AGI falls with AGI
- But, differential non-response by income may possibly explain half of the aggregate shortfall
  - Some respondents under-reporting total spending
  - Implications for inequality, saving, consumption taxes

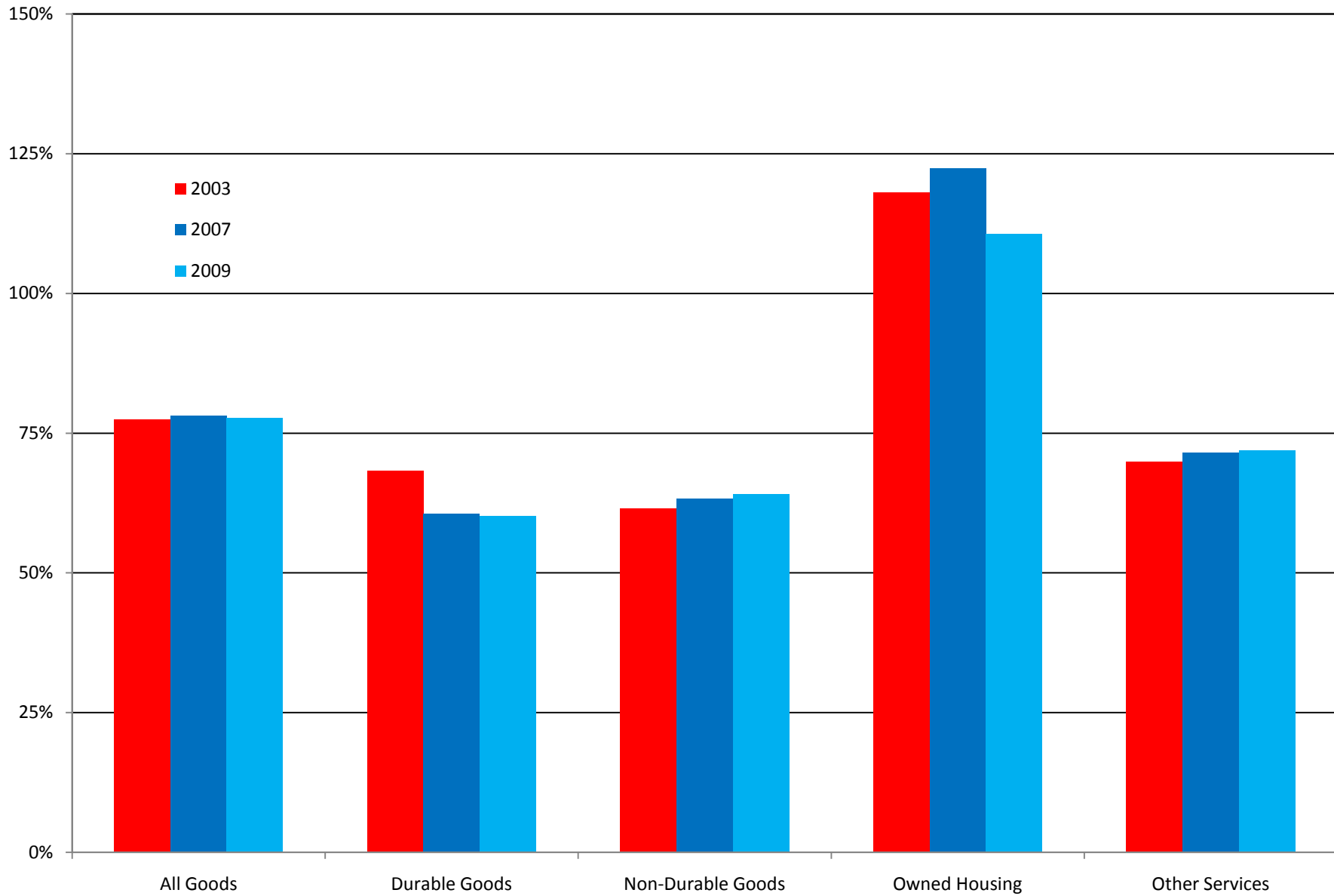
# Background: CE to PCE Ratios

- Need to adjust data sets for comparability of populations and spending concepts
- Observe persistent differentials by types of spending (especially owned housing services)
- Some decline in CE/PCE during 1990s, but
- Ratios relatively stable since 2003

## Ratio of CE to Comparable PCE, Garner et al (2009)



## Ratio of CE to Comparable PCE, Latest BLS Published



# CE Income Distribution

- Compare CE income distribution (units and dollars) to CPS, SCF, and SOI data in 2006
  - CPS is most like the CE in terms of coverage and questions, but focused on income, not expenditures
  - SCF over-samples wealthiest households, but otherwise income concepts similar to CE and CPS
  - SOI is administrative data, but concepts diverge
- Compare counts and dollars; can't distinguish between missing very high income and under-reporting of income for those households

# Missing Income in the CE

- Numbers of units (\$50k<, \$50-100k, \$100k+) similar across the four data sets
  - CPS/SCF slightly more units \$100k+, SOI fewer
  - SOI is tax units, includes dependent filers
- Dollars of income much more divergent
  - Overall \$0.5 (CPS) to \$1.7 (SCF) trillion missing
  - Most income missing \$100k+; offset slightly by higher CE aggregate incomes below \$100k
  - SOI shows \$1.0 trillion missing \$100K+, even though non-taxable incomes missing



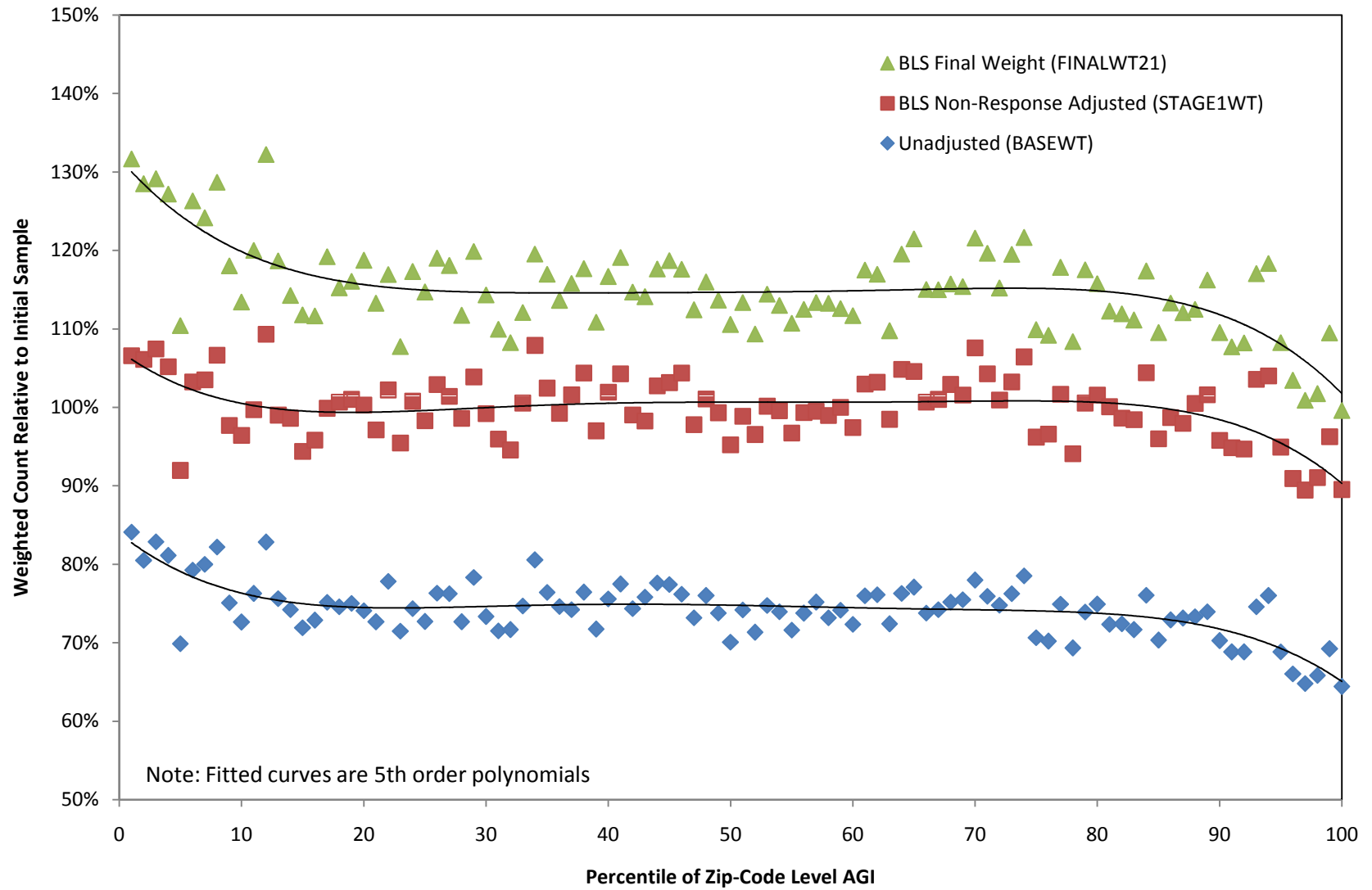
# Response Rates by Income

- Data set begins with 104,830 sampled CE units for 2007 and 2008
  - Exclude benchmark interviews, sample is 83,366
  - 61,456 respondents; 74 percent response rate
- Link all sampled units to 5 digit zip-code level average AGI from IRS public-use data
  - Sort all sampled units by zip-code average AGI
  - We are NOT testing sampling rates by AGI
  - Analyze response rates within AGI percentiles

# Using Zip-Code Level AGI

- AGI is less than or equal to household income
  - AGI excludes non-taxable transfers, interest (bias)
  - Dependent filers are separate units (bias)
  - Excludes non-filers (possible bias)
- Is 5 digit zip-code homogeneous enough?
  - We find significant differences in mean AGI across percentiles of AGI, especially top zip-codes
  - We also compare means within percentiles
- In any case, zip-code AGI good for ranking

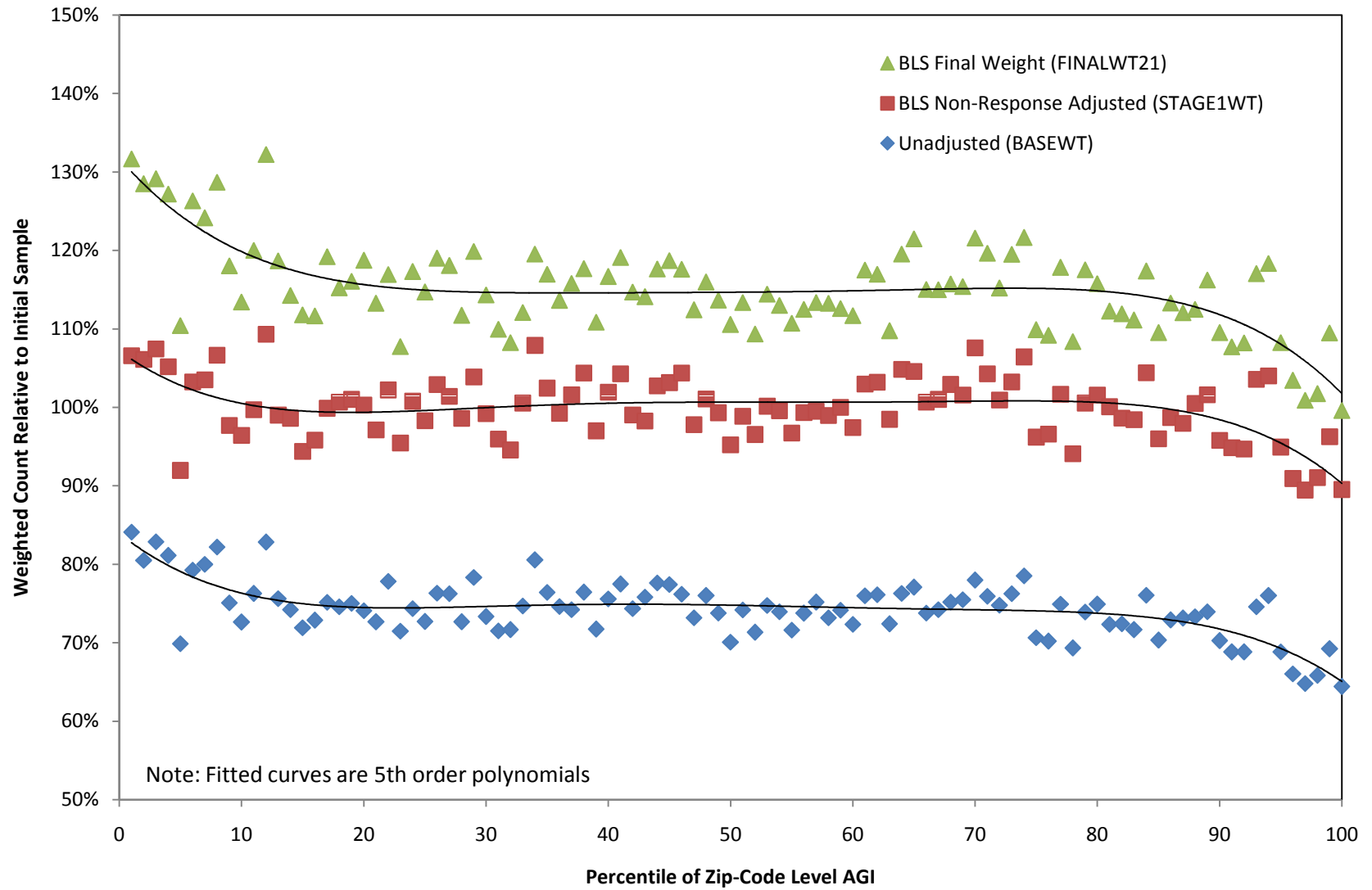
**Figure 1. Consumer Expenditure Survey (CE) Response Rates by Zip-Code Level Adjusted Gross Income (AGI) Percentile**



# BLS Post-Stratification

- Unit non-response at higher incomes affects final sample if other controls uncorrelated
  - BLS non-interview adjustment by region, tenure, CU size, race (64 cells) creates STAGE1WT
  - Final calibration for age, race, tenure, region, and rural/urban (24 cells) creates FINALWT21
- The two sets of adjustments shift response by AGI line up, but tails are same
  - *Relative* response rate 90 percent for top vingtile

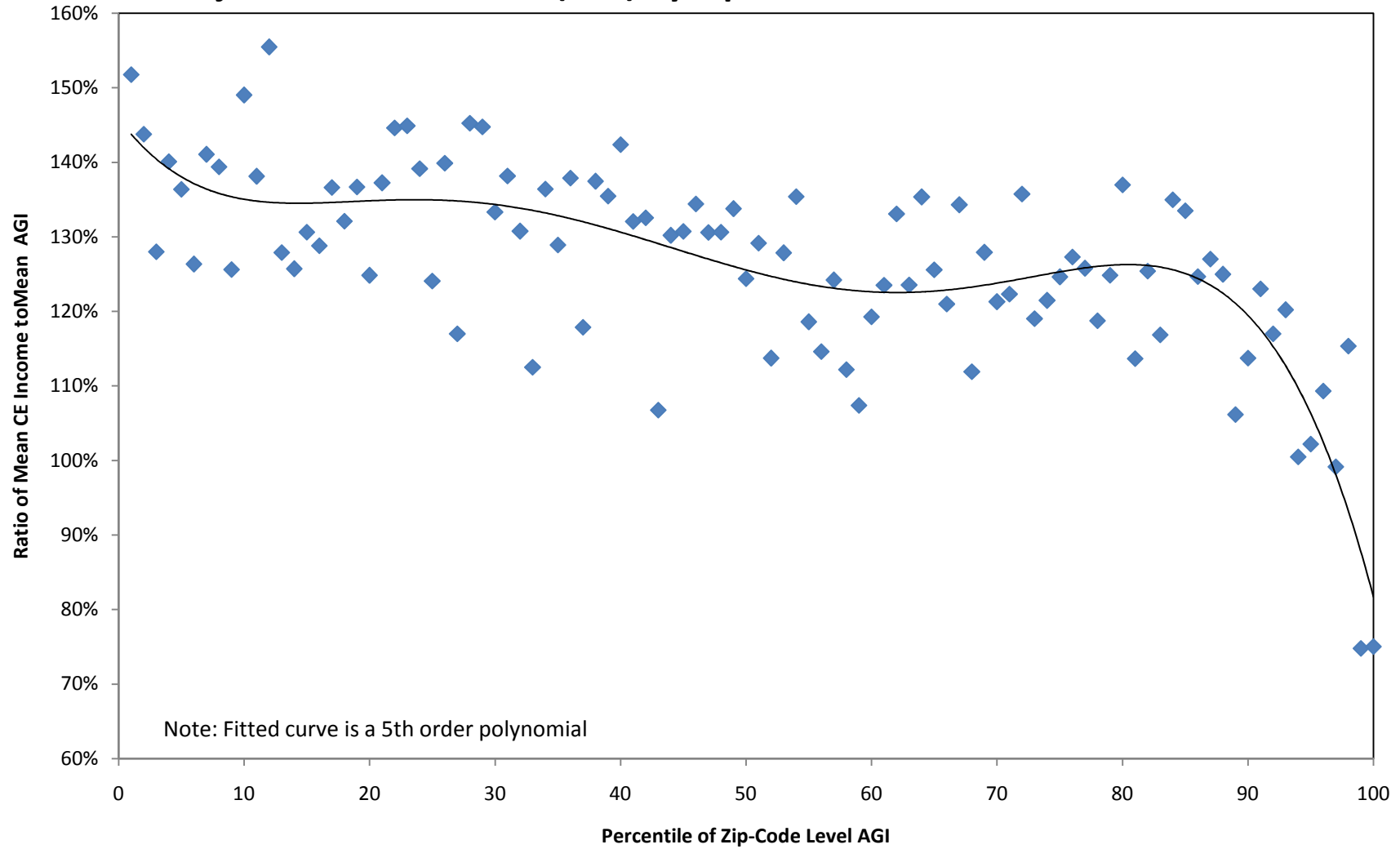
**Figure 1. Consumer Expenditure Survey (CE) Response Rates by Zip-Code Level Adjusted Gross Income (AGI) Percentile**



# Comparing Means

- Second step is comparing AGI for all sampled units to CE household income for respondents
- Conceptual issues about AGI more of a factor
  - Overall, CE household income 20 percent higher
  - CE income 40 percent higher at bottom, falls steadily, plummets to 75 percent lower at top
- Outstanding question is whether “wedge” between AGI and household incomes varies with percentiles of AGI

**Figure 2. Ratio of Mean Consumer Expenditure Survey (CE) Income to Adjusted Gross Income (AGI) by Zip-Code Level AGI Percentile**



# Unit Non-Response and Income

- Even if zip-code level AGI is just a good sorting variable, clear evidence that response is lower for very highest income families
- Visual evidence confirmed by probit estimates with existing post-stratification variables; AGI is a significant predictor of response
- Using this for alternative survey non-response adjustment will require reconciling AGI and household income



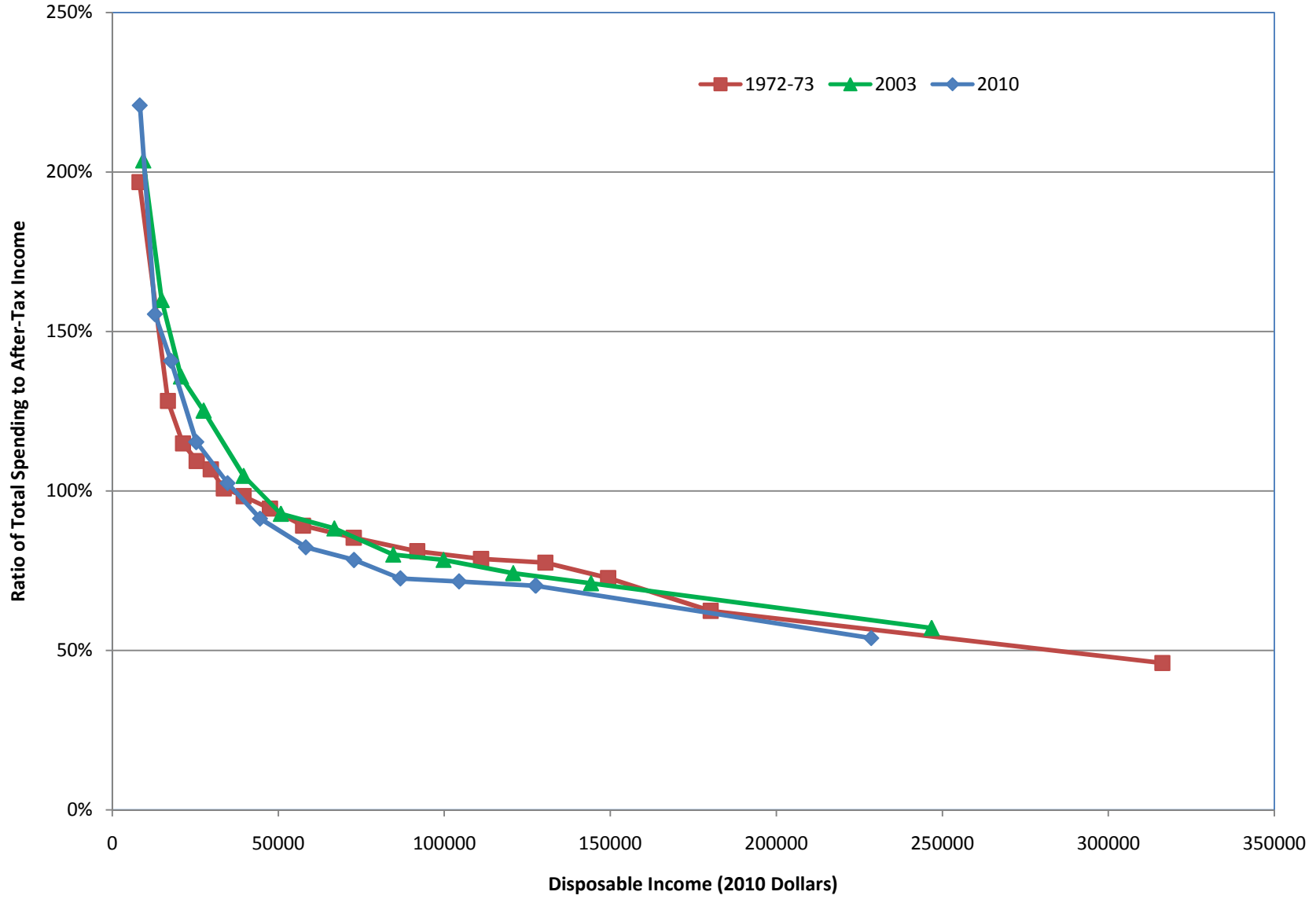
# Implications of Differential Non-Response

- Back of the envelope calculation
  - SCF: CE might be missing \$1.7 trillion \$100k+
  - CE ratio of spending to income 61 percent \$100k+
  - Implies \$1.0 trillion “missing” consumption
- Total CE spending is \$5.8 trillion, so “missing” consumption is at most 17% of CE total
- Thus, if wealthiest families included in CE, could resolve at most half the CE/PCE shortfall

## Respondent Under-Reporting

- Published BLS tables show aggregate spending to after-tax income was 83 percent in 2006
- Problems with BLS tables
  - Not a saving concept, taxes poorly measured
  - Adjustments might explain several percentage points of gap, real ratio may be 90 percent
  - But, ratio should be 100 percent or more
- Ratio of spending to after-tax income was 84 percent in 2003, 89 percent in 1972-73

**Figure 3. Expenditure to After-Tax Income Ratios in Published CE data**



# Spending to Income Ratios

- In all years, seems like spending to income ratios too high at bottom, too low at top
- Implies income under-reporting for some, sorted to bottom of distribution tables
- Also implies under-reporting of spending for at least some households, probably at top
- Simple adjustments to spending (proportional scaling) probably inappropriate

# Research Implications

- If CE just under-represents the very top
  - CPI biased only if their spending basket different
  - Measures of resource distribution will not be complete, but unbiased for relevant incomes
- But more complex under-reporting of both income and spending raises questions about
  - Saving rates across groups
  - Distributional burden of consumption taxes
  - Consumption versus income inequality

# BLS/Census To-Do List

- Create more comparable expenditure and after-tax income concepts for published tables; use model-based income taxes instead of reported
- Reconcile CE income and AGI concepts, re-do both parts of the new experiment
- Extend analysis back in time to see if very high incomes less represented now. How does that correlate with CE/PCE patterns over time?
- Extend to other data sets like CPS