Examining the Relationship between Household Income and Response Rates in the Consumer Expenditure Survey

Lauren Vermeer
U.S. Bureau of Labor Statistics, 2 Massachusetts Ave NE, Washington, DC 20212

Abstract
Income is an important analysis variable in household surveys. In 2013, BLS found a variable on a publicly available dataset from the Internal Revenue Service (IRS), which gave the average adjusted gross income (AGI) in most US zip codes. This zip-code level AGI variable was found to be correlated with the Consumer Expenditure Survey's (CE) response rates. The variable was introduced into CE's nonresponse adjustment procedure, the first to be added from outside the survey's sampling frame. It is important to periodically check and confirm variables' functionality in survey procedures, so the variable was reexamined in 2023. This paper describes both the reexamination process and the results, which show that the relationship between income and households' response rates is less stable than previously determined.

Key Words: income, response rates, nonresponse adjustment

1. Introduction

The Consumer Expenditure Survey (CE) is a nationwide household survey which collects and analyzes expenditure data to see how U.S. households spend their money. The CE consists of two separate surveys, a quarterly Interview survey (CEQ) and a two-week Diary survey (CED). The U.S. Census Bureau collects the data for both surveys under contract with the U.S. Bureau of Labor Statistics (BLS).

Sabelhaus et al. (2013) used zip code level income data from the IRS and showed that income was correlated with CE response rates. The study concluded that households in higher-income zip codes had low response rates and households in lower-income zip codes had high response rates. However, in recent years, this relationship may have changed. Periodic research and reexamination into trends are needed to keep sampling methods current and accurate. Results from our reexamination into this matter are documented and explained below.

Using data from the CE, the IRS, and the American Community Survey (ACS), we conducted a statistical investigation to observe trends in the relationship between response rates and income from surveyed households. Data from the CE provided values on response rates, and data from the IRS and ACS provided values on household income. Although Sabelhaus et al. (2013) infers a stable relationship between CE response rates and income, our research indicates this relationship has become unstable, especially in recent years.

2. Background on CE, and Income's Role in the CE

To capture US expenditure data for the CE, a random sample of residential addresses are selected from the Census Bureau's Master Address File, and then residents from the selected addresses are interviewed, and their expenditure information is collected. Residents who complete the interview are considered respondents and are referred to as consumer units (CUs), where there is usually one CU per household. When expenditure information cannot be collected, it is referred to as a

noninterview. A noninterview can happen for a variety of reasons, from resident nonparticipation (nonrespondent) to the selected address being unoccupied, nonresidential, or nonexistent. For the scope of this research, we are focusing only on respondents who complete their interview and nonrespondents who choose not to complete their interview.

Every survey has respondents and nonrespondents, and most surveys utilize adjustment procedures to remove bias generated by the nonrespondents. The CE has a nonresponse adjustment procedure which uses the traditional cell method with a set of four cell-defining variables (region, average income by zip code according to the IRS, household size, and number of contact attempts) which increases the weights of the respondents to account for the nonrespondents.

Income serves an important role as one of the four cell-defining variables in the CE nonresponse adjustment procedure, as it is the only variable of the four to come from an external source – the IRS. Examining whether past trends between response rates and income still hold is key to the variable's effectiveness. In the CE, the four variables that are used to define cells for the nonresponse adjustment procedure are divided into three or four categories. Specifically for income, there are three categories: Lower 10%, Middle 80%, and Upper 10%. Each CU is categorized into one of these three income categories – based on the average adjusted income (AGI) in their zip code, according to the IRS.

Income and expenditures generally have a direct relationship – as income increases expenditures increase, and as income decreases expenditures decrease. Expenditures are the crux of CE data collection, so any factors that influence a CU's expenditures are of strong interest. The current research into the relationship between income and expenditures will highlight the performance of the income variable within the CE.

3. Procedure

To obtain the most realistic and practical results, research was conducted in alignment with the current production setup. This means we divided the sample households into the same three income categories used by CE (Lower 10%, Middle 80%, Upper 10%), and when using the IRS data, we used a two-year lag in relation to the CE data[1]. Analyzed data spanned the following years: 2015 – 2022 (CE), 2013 – 2019 (IRS), 2015 – 2019 (ACS)[2]. (CE's timeline served as the marker years for any calculated values.)

Each survey's yearly response rate was calculated over an eight-year time frame, spanning from 2015-2022. For each year, a R^2 value was calculated using the following formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{100}(r_i - \hat{r}_i)^2}{\sum_{i=1}^{100}(r_i - \bar{r})^2}$$

where is r_i is the response rate for all the sample CUs within the ith income percentile; $\hat{r}_i$ is the modeled response rate based on the spline (piecewise linear) regression model fit to the data; and $\bar{r}$ is the overall (average) response rate for all CUs in the sample.

This R^2 value serves as a comparison tool, to assess whether a relationship exists between response rates and the respondent's income. Due to privacy reasons, we are considering the respondent's

[1] The two-year lag is in refence to the IRS data. For example, we matched 2015 CE data with 2013 IRS data, we matched 2016 CE data with 2014 IRS data, and so forth.

[2] The ACS file has a single five-year estimate, which covers years 2015 – 2019.

income as the average income of the sample households within a zip code (IRS) or tract (ACS). R^2 values are between 0 and 1, with values near 1 indicating a stronger correlation between response rates and income, and with values near 0 indicating a weak correlation between response rates and income. An R^2 value was calculated per year for both surveys, resulting in eight R^2 values for the Interview survey, and eight R^2 values for the Diary survey. This was done for one IRS and two ACS datasets, for a total of 48 (= 8 x 2 x 3) R^2 statistics. Additionally, a spline regression graph was also created, to provide a further analysis of the income variable. For each external dataset, income data was split into 100 percentiles, and ranked, before calculating the response rate.

When examining CE data against IRS data, income was available as a single variable in the IRS data – the average annual adjusted gross income. This allowed one representative income to be assigned to each zip code within the CE data and allowed for a straightforward comparison. When examining CE data against ACS data, income was available as two variables in the ACS data – the median household income and the average household income. This case provided the opportunity for two representative incomes to be assigned to each tract within the CE data. Two variables presents two possibilities for sorting – with median income listed first and average income listed second, and with average income listed first and median income listed second. In research, it is important to test multiple methods if they are available, so both options were examined. To help keeps these cases straight, the instance where median income is first is referred to as ACS 1, and the instance where average income is first is referred to as ACS 2.

In the Sabelhaus study, the Lower 10% income group had a high response rate, and the Upper 10% income group had a low response rate. To confirm that the relationship between income and CE response has become unstable, we should find that data used in our study behaves differently than data used in Sabelhaus et al. (2013) – meaning that we should see low response rates in the Lower 10% income group, and high response rates in the Upper 10% income group to confirm an unstable relationship.

4. Results

Table 1 displays the 48 R^2 values found in our investigation over the observed eight-year span (2015 – 2022). Additionally, the values are separated by their data source, either IRS, ACS 1, or ACS 2. The R^2 values indicate the strength of the relationship between CE data, and the respective data source. Higher values indicate a stronger relationship between response rates and income, while lower response rates indicate a weaker relationship between response rates and income.

When working with data from social sciences, R^2 values that are close to 1 are rare, which explains why no values in the below table approach 1. The highest R^2 value for the Interview survey is 0.2296, which is seen in the IRS data in 2015. The highest R^2 value for the Diary survey is 0.3630, which is seen in the IRS data in 2020. Looking at the average R^2 values for all three datasets, the IRS dataset average R^2 values are the highest, of about 0.15 for both the Interview and Diary surveys. The average R^2 values for ACS 1 and ACS 2 are similar, between about 0.05-0.07 which makes sense seeing as the root data is the same, with a difference in sort order of the income variables. (Note: Due to the COVID-19 pandemic, a break in any pattern for 2020 is both expected and observed).

Additionally, some aspects of the calculated data were unique, for example only occurring in one of the three datasets. First, R^2 values from Interview x ACS 1 data were the lowest amongst all the tested data, as shown in **Table 1**.

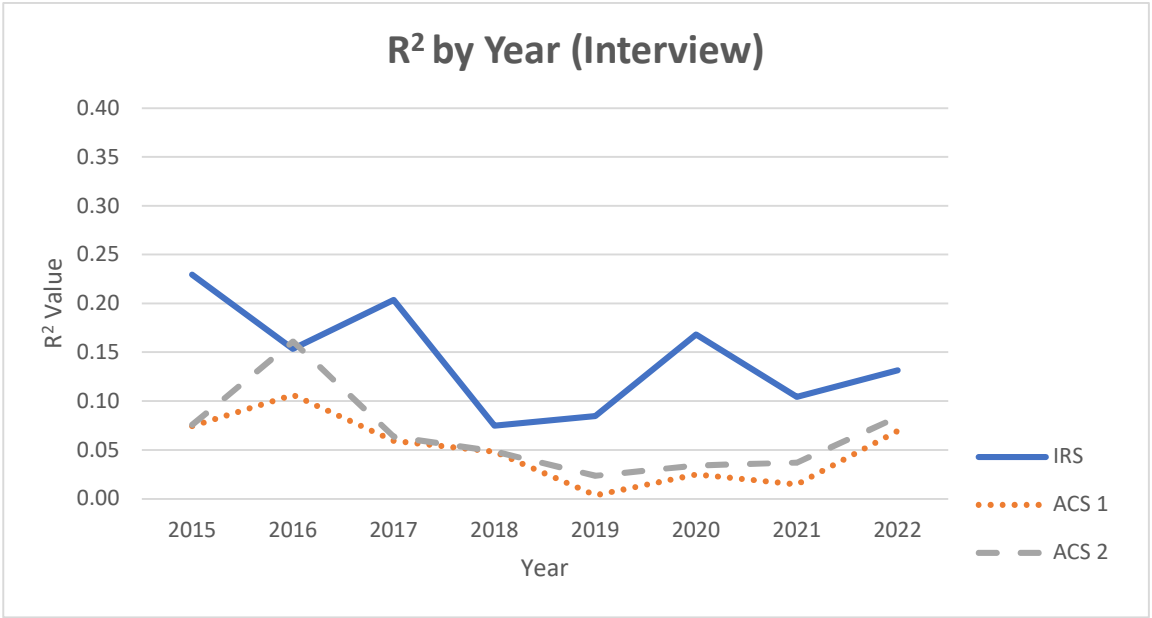
Table 1: R^2 Results – by External Data Source and CE Survey

	IRS		ACS 1		ACS 2	
Year	**Interview**	**Diary**	**Interview**	**Diary**	**Interview**	**Diary**
2015	0.2296	0.0274	0.0744	0.0339	0.0757	0.0095
2016	0.1535	*0.0189*	0.1064	0.1162	0.1610	0.0962
2017	0.2035	0.2350	0.0595	0.0411	0.0638	0.0974
2018	*0.0750*	0.2133	0.0481	*0.0197*	0.0487	0.0359
2019	0.0846	0.1267	*0.0036*	0.0390	*0.0236*	*0.0082*
2020	0.1680	0.3630	0.0251	0.1398	0.0342	0.1593
2021	0.1042	0.0833	0.0147	0.1027	0.0371	0.0287
2022	0.1314	0.0966	0.0695	0.1121	0.0837	0.0549
μ	0.1437	0.1455	0.0502	0.0756	0.0660	0.0613
σ	0.0519	0.1098	0.0344	0.0466	0.0437	0.0526

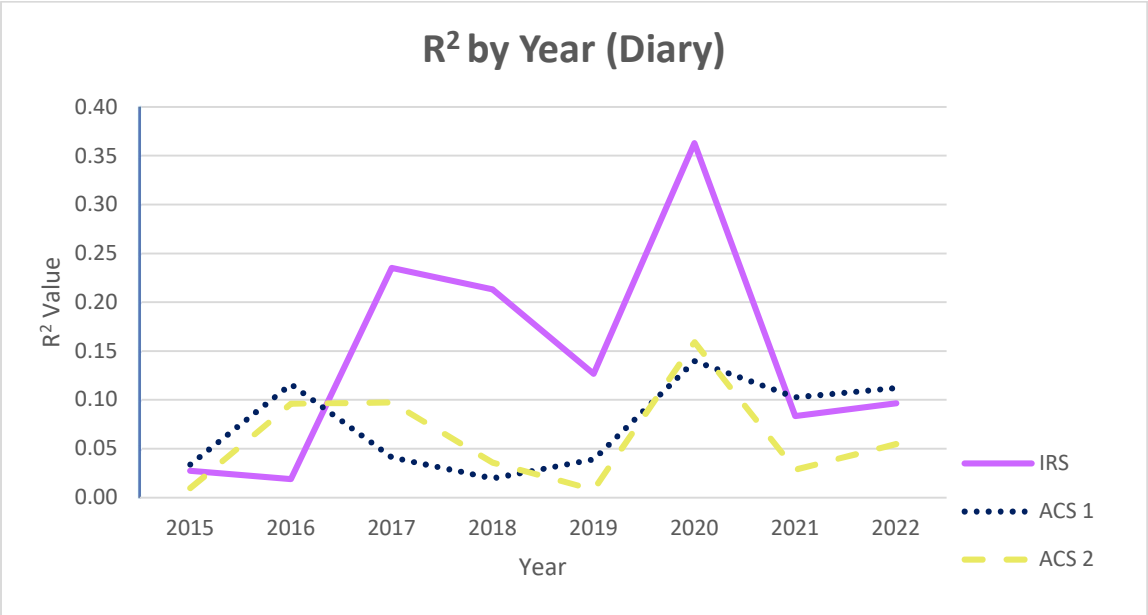
Key: Underline indicates the highest R^2 value in each column, *italics* indicates the lowest R^2 value in each column

The following two graphs display the history of the R^2 results over the observed years. Both graphs use the same data as found in **Table 1** but depict the strength of the relationship between CE data and either IRS or ACS data, and shows how this relationship has lost stability over the observed time frame (2015 – 2022).

Graph 1: History of Interview R^2 Values by Year
(IRS x ACS 1 x ACS 2)

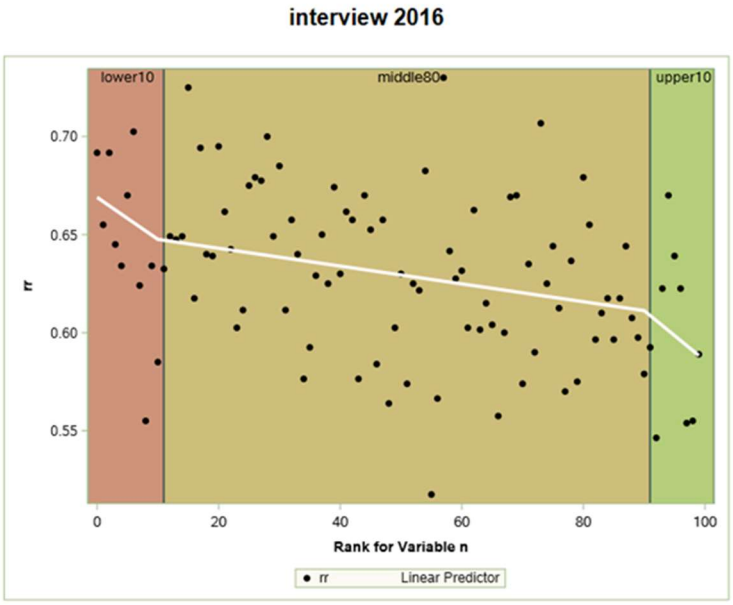


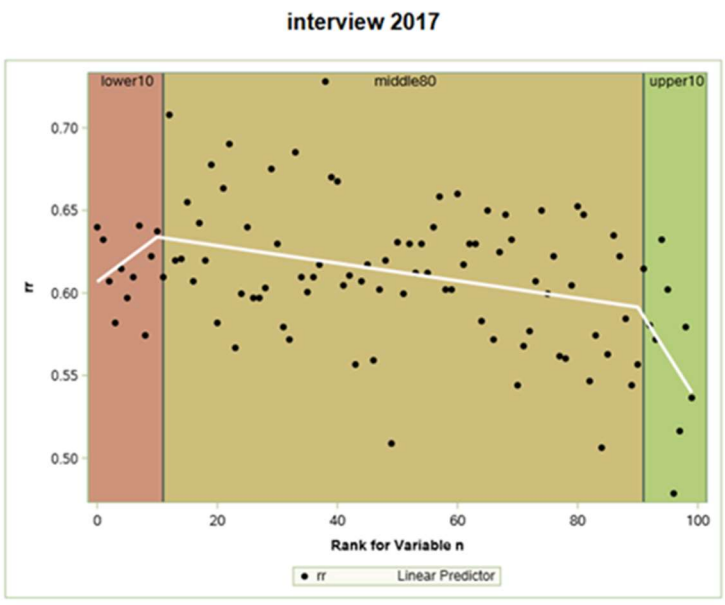
Graph 2: History of Diary R^2 Values by Year
(IRS x ACS 1 x ACS 2)



Using SAS®[3], spline plots were generated to show the relationship between CE's response rates and average income, distinctly by the three income groups. Every SAS plot is fitted with a piecewise line, clearly marking any instances where the slope changes between income groups. Distinguishing the three income groups within the SAS plots models the three income groups used to define cells within CE's nonresponse adjustment procedure. This similarity is important in keeping our testing as close to CE's production as possible. There are many plots to look at, however they all share slope changes between income groups, and instances where the relationship between income and response rates have changed over the years. For example, a change in the slope is observed in the Interview x IRS data for the Lower 10% income group, starting as a negative slope in 2016 and then changing to a positive slope the following year.

[3] SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.





5. Analysis

Of the three tested data sources, IRS zip code data showed the strongest correlation with CE data. Overall, the average R^2 values were the largest for any tested dataset, with an average of 0.1455 for the Interview survey, and an average of 0.1525 for the Diary survey. This was more than double any of the averages seen in the ACS data. When using IRS data, CE Interview data is slightly more stable than CE Diary data, with less drastic changes. If a lower R^2 value occurs in one year, the following year typically has a higher R^2 by about 0.05. The highest R^2 value occurs in 2015, with values fluctuating in the subsequent years observed. IRS data captures one year of data and is published annually, while ACS data is a rolling five-year dataset. Although our research shows IRS data is more stable than ACS data, some volatility is still seen in IRS data. The frequency of the IRS data may be contributing to this volatility since trends cannot be established as well in a one-year time frame.

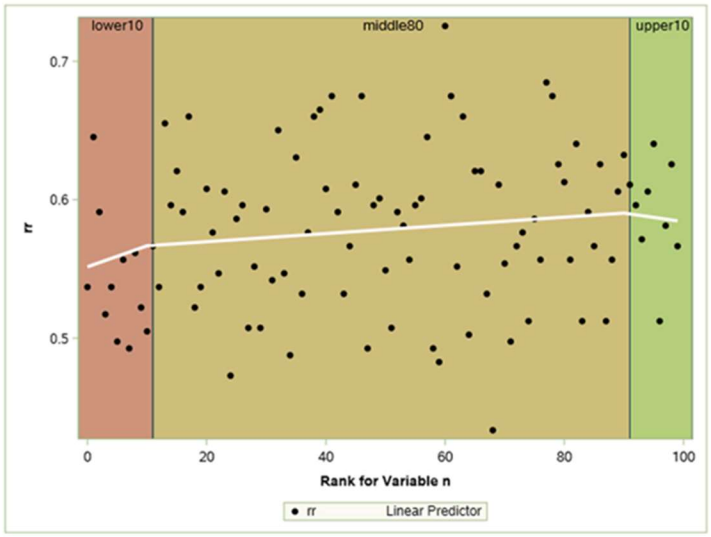
ACS 1 and ACS 2 contain the same data, with the differentiation being the sort order of the income variables. In certain years, the median income (ACS 1) had a higher R^2 value, while in other years, average income (ACS 2) had a higher R^2 value, and this can be noted in both CE surveys. The average R^2 value for ACS 1 and ACS 2 for both surveys ranges between ~0.05-0.07, which indicates that the variables perform relatively similarly, since their sort order did not drastically change the outcome. When individually ranking the four ACS dataset R^2 values, CE Interview x ACS 1 data contains the lowest R^2 values, indicating the weakest relationship. Well-defined maximums or minimums are not found in this dataset but can be found in other datasets. Since this set of R^2 values is the smallest, the ACS 1 line within **Graph 1** appears smooth although the values are slowly trending downward. Though the values were not large, this dataset is the only one to resemble a linear function.

In **Graph 2**, the ACS 1 line slightly resembles a quartic function, and has a local maximum which occurred in 2016. Of all the tested datasets, this is the only dataset which has a quartic function resemblance. The two maximum points in the data occur four years apart, with the minimum point in the dataset occurring directly in between the two maximum points. In this sense, it is unique to the other datasets. The correlation between CE Diary data against ACS 1 data (with median income sorted first) is slightly stronger than the Interview data, but has noticeable variance within it. While

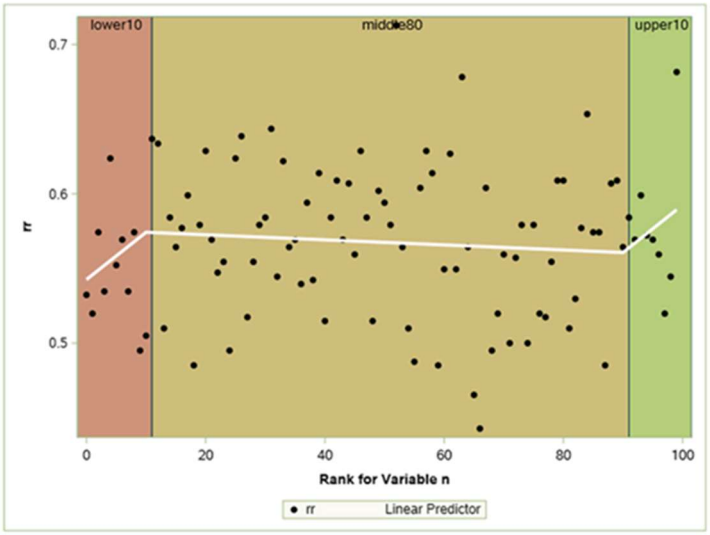
this pattern is not consistent relative to itself, it can be considered stable if the pattern is expected to repeat. Both Interview and Diary datasets using ACS 1 data would be potential candidates for future observation since they have relative patterns to them by resembling mathematical functions.

Overall, the relationship between response rates and income appears unstable. The SAS plots using IRS data show that in the Interview survey the response rate for the Lower 10% income group has changed, most notably between 2016 and 2017, and has remained that way. Prior to 2016, the slope was negative, and in 2017 it became positive. Conversely, in the Diary survey SAS plots show that the response rates for the Upper 10% income group have changed, most notably in 2016. The slope for all tested years for the Upper 10% income group was negative, except for 2016, when it became temporarily positive. Additionally, the Lower 10% income group sees a steeper negative slope throughout the years.

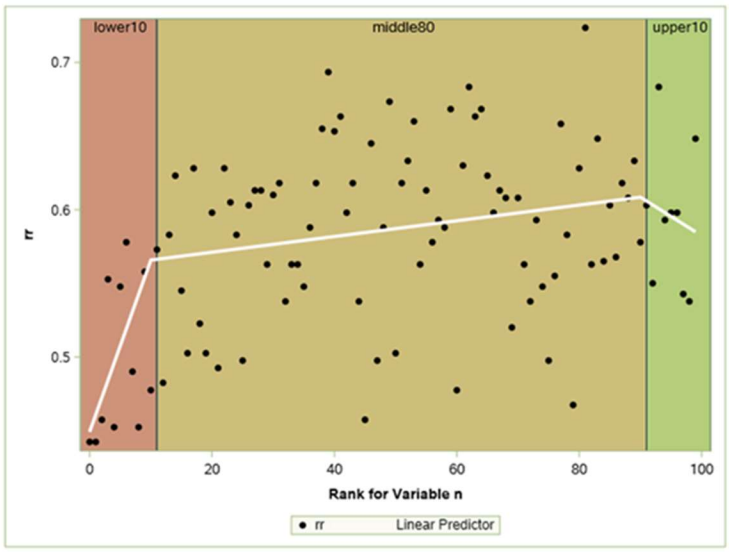
diary 2015



diary 2016



diary 2017



6. Conclusion

There are two main takeaways from this research. First, and most important, the results indicate that the relationship between response rates and income is unstable. In some cases, this shift occurs in a specific year and remains that way, and in other years, the shift occurs in one year, and changes back to its historic pattern. In instances where the change has remained, this is likely to be caused by a factor that is unchanging. In instances there the change has appeared and gone away, this is less stable and likely to be caused by a specific event, and when the event is over the slope pattern recovers. In either case, a change is observed.

Second, when comparing both the ACS and IRS datasets together, the IRS dataset has a stronger relationship with CE data, as indicated in the R^2 values. The IRS dataset produced average R^2 values of about 0.15 for both the Interview and Diary surveys, which is at least double any of the average R^2 values produced by the ACS datasets. This may be because the IRS data's time frame is ~1 year, vs. the ACS data data's time frame is five years. So, the IRS data contains only the most recent incomes, while the ACS data contains both recent and prior incomes.

The next step would be to monitor this data to see if any trends remain consistent, such as the cyclical pattern in the Diary survey x ACS data, or to see if there are additional slope changes. Pinpointing the cause of some of these slope changes also carries priority, to see if this is a factor that can be controlled. Overall, analysis into the data showed a change in the historic relationship between income groups and response rate, which provides the opportunity to react to the changing trends accordingly.

Disclaimer

References

Sabelhaus, John, et al. "Is the Consumer Expenditure Survey Representative by Income?" *National Bureau of Economic Research*, 31 Oct. 2013, www.nber.org/papers/w19589