Journal of Official Statistics, Vol. 38, No. 4, 2022, pp. 1125–1144, http://dx.doi.org/10.2478/JOS-2022-0048

Analyzing the Association of Objective Burden Measures to Perceived Burden with Regression Trees

December 2022

Daniel K. Yang[1] *and Daniell S. Toth*[1]

Higher levels of perceived burden by respondents can lead to ambiguous responses to a questionnaire, item nonresponse, or refusals to continue participation in the survey which can introduce bias and downgrade the quality of the data. Therefore, it is important to understand what might influence the perception of burden in respondents. In this article, we demonstrate, using U.S. Consumer Expenditure Survey data, how regression tree models can be used to analyze the associations between perceived burden and objective burden measures conditioning on household demographics and other explanatory variables. The structure of the tree models allows these associations to easily be explored.

Our analysis shows a relationship between perceived burden and some of the objective measures after conditioning on different demographic and household variables and that these relationships are quite affected by different respondent characteristics and the mode of the survey. Since the tree models were constructed using an algorithm that accounts for the sample design, inferences from the analysis can be made about the population. Therefore, any insights could be used to help guide future decisions about survey design and data collection to help reduce respondent burden.

Key Words: nonparametric; nonresponse; respondent burden; sample design; survey data.

1. Introduction

The Consumer Expenditure Survey (CE) is a national survey conducted by the U.S. Bureau of Labor Statistics (BLS) to collect data on how American households spend their money. The collected data are used to estimate consumer expenditures, which are published twice a year, as well as to annually produce public-use microdata files to allow researchers to do their own analyses. This is the only federal survey that provides information on U.S. consumer expenditures as well as household income and demographic characteristics, making the data it collects critically important to government and private agencies examining the association of consumer expenditures and income to household characteristics. This type of analysis is used by economic policy makers to understand the effects of policy changes on households among diverse socioeconomic groups.

Importantly, CE data are inputs for producing the Consumer Price Index (CPI), a Principal Federal Economic Indicator, used by The Federal Reserve to help set U.S. monetary policy. The data are used to construct new "market baskets" of goods and services, determine the relative importance of components, and to derive cost weights for

[1] U.S. Bureau of Labor Statistics Office of Survey Methods Research, 2 Massachusetts Avenue Suite 5930, NE Washington D.C. 20212, U.S.A. Emails: yang.daniel@bls.gov and toth.daniell@bls.gov

the market baskets used in the calculation of the CPI. CE data are also used by the Department of Commerce for calculating the Supplemental Poverty Measure, by the Department of Agriculture for estimating the cost of raising a child, by the Internal Revenue Service for calculating alternate sales tax standard deductions and by the Department of Defense for determining cost-of-living allowances for military personnel.

Because of the essential role that CE data play in setting policy and in the managing of the U.S. economy, it is imperative that the quality of the data be maintained at the highest level possible. Lower response rate is one way that the quality of survey data can be degraded. A low response rate can potentially introduce response bias as well as increase the variability of statistics obtained from the data (Groves 2006). In addition, low survey response rates erode user confidence in the data.

For these reasons, the BLS has put a great deal of effort into maintaining a high response rate for its surveys. These efforts include the introduction of computer-assisted personal interviewing and an instrument to track interviewer contacts, as well as a redesign of the survey (Edgar et al. 2013a,b). Despite these changes the CE has observed a decay of the response rate to its surveys over the last two decades.

Figure 1 shows the response rates of the CE interview survey falling from a rate of around 80% to a rate of less than 50% over a 21 year period. Falling response rates is hardly unique to CE. Czajka and Beyler (2016) found that response rates were declining at a similar rate for all the U.S. federal surveys that they studied including the National Health and Nutrition Examination Survey and the Current Population Survey. The large dip in the CE response rate that occurred late in 2013 was due to the disruption of data collection and nonresponse follow-up efforts caused by the government shutdown.

One factor that may have an effect on response rates of a survey is the amount of burden a survey puts on respondents. Burden is something that is difficult to bear, worrisome, stressful, or oppressive. In survey research, burden is often thought of as the collection of all costs that the survey respondent incurs for responding to the survey, including loss of

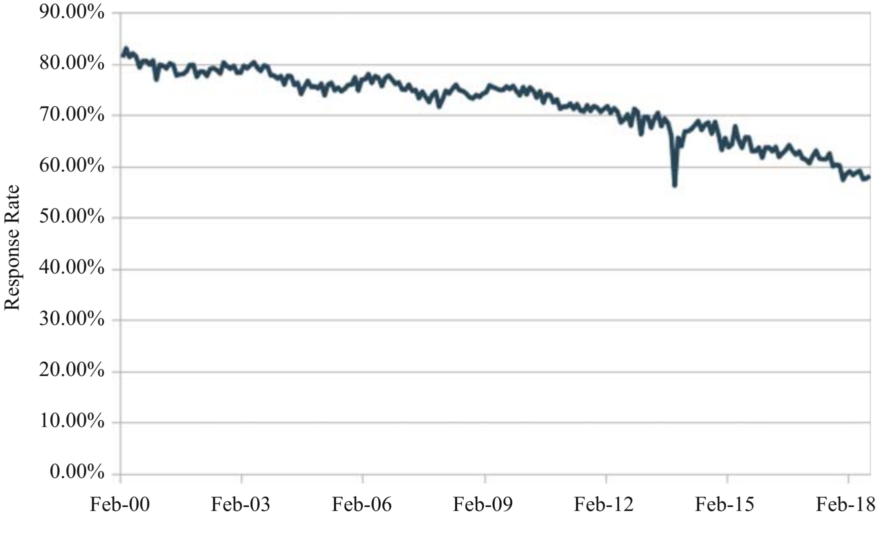

Fig. 1. Response rate of CE interview survey from january 2000 to june 2020.

time, exertion of effort, and stress associated with providing the requested information (Ashmead et al. 2017).

While how much stress or worry is felt by a respondent will depend on the individual, there are objective measures of burden that can be collected for each survey interview. These include length of the interview, number of questions asked, and whether information from records were required to answer the questions. Because of its breadth, the CE Quarterly Interview Survey (CEQ) is likely burdensome based on these measures, taking almost an hour to complete with many questions requiring the respondent to look through receipts or other records to answer.

Though these objective measures do not measure the actual burden felt by a respondent (perceived burden), one would expect that they are related and therefore have an effect on the response rate. However, evidence of the effect objective measures of burden have on the response rate of a survey have been mixed in the literature. Indeed, Bogen (1996) reviewed several observational and experimental studies of the relationship between questionnaire length and the response rate and found evidence both supporting and refuting a relationship. For instance, Lynn (2014) found no evidence that the initial interview length affected the participation rate for subsequent interviews while Galesic and Bosnjak (2009) did find evidence that the length of the questionnaire affected response rates and data quality. Nevertheless, time required to take a survey is definitely a cost that potential respondents weigh when deciding whether to participate in a survey and must be accounted for when measuring burden. Interview length was a significant input variable in the respondent burden model of Fricker et al. (2014).

These mixed results may be attributed, in part, to the fact that it is the perceived burden that affects a potential respondent's decision whether or not to participate and how a respondent reacts to the objective measures of burden vary for different respondents (Sharp and Frankel 1983). The amount of burden felt (or perceived) by a respondent is likely not perfectly correlated with the objective measures of burden like length and difficulty of the survey, but rather an interaction between these measures and characteristics of the respondent. For example, though Fekete et al. (2017) do not find a strong correlation between objective and subjective burden, they find that high subjective burden was linked to poorer general health. Though not tested in their report, this result could indicate that respondents with health issues are more sensitive to the objective measures of burden than other respondents.

Though a short survey is likely to have less perceived burden than a long survey in general, which surveys are judged to be short, and which are judged to be long will depend on the individual doing the judging, on their current circumstances, on their interests and on the topic of the survey, among other things. Likewise, which questions are difficult to answer can vary considerably among people depending on their household and personal situation at the time they are participating in the survey. Therefore, a measure of the amount of perceived burden a survey is likely to impose along with the objective measures of burden is needed to determine how the response rate of a survey is likely to be affected by the objective measures of burden.

Another factor found to influences the relationship between perceived burden and quantitative measures of burden is the participant's impression of the importance of the survey or it's salience. Salience is defined as "the quality of being particularly noticeable or important" (Cannell et al. 1981). Survey researchers have long known that how salient the topic of a survey or particular questions are to a respondent can affect the response rates. For instance, Bradburn (1978) found that a boring questionnaire may drive respondents away, while an interesting one may motivate them.

Bradburn (1978) explained that survey interviews are social interactions and researchers must first understand the respondent's motivation to participate in a survey interview. He recommended using a sense of civic duty and knowledge about the importance of the survey as motivators for participation, putting emphasis on how the questionnaire design can contribute or mask burden of a survey; noting that burden appeared to become more tolerable for respondents who are persuaded that the collected data are crucial.

Groves et al. (2000) found evidence that respondents maybe more willing to participate in a burdensome cognitive activity for a salient topic than for a non-salient one. Groves et al. (2001) used measures of opportunity cost and salience as components of social interaction. Likewise, Connelly et al. (2003) studied mail survey response rates and found the salience of the study topic was among the five most significant explanatory variables in their model for response rate. They found a 25% increase in response rates for the highly salient survey compared to the non-salient survey. Fricker et al. (2014) also suggests that the perception of burden was associated with non-salient (less motivated) topics.

For example, participating in a survey on early childhood education that takes a half hour to complete might feel like no burden to a respondent who is concerned about their young children's education, but may seem like a big inconvenience to someone without young children. Therefore, Cannell et al. (1981) suggest that helping a respondent recognize that the intent of the survey is important to them could mitigate survey burden.

For a given survey, the amount of perceived burden that responding to it will induce depends on the current circumstances of the individual, the quantitative measures of burden for the survey, as well as the survey's salience to that individual. This suggests that to estimate the likely burden of a survey, one must account for the characteristics of the individual respondent as well as the objective measures of burden and that this relationship between the factors could be complicated.

Getting an accurate understanding of how perceived burden for a given respondent is affected by changes to the survey design is very important to CE. This could help drive future changes to the survey that reduce burden and slow or stop the decline of response rates. Besides reducing response rates, there is evidence that higher levels of burden are associated with measurement error which also impact data quality (Abayomi et al. 2018; Ashmead et al. 2017). Therefore, understanding and measuring what impacts the burden felt by the respondent is necessary for mitigating the burden put on respondents which is critical for maintaining the quality of the data and the sustainability of the survey.

In order to properly understand the relationship between perceived burden and the objective measures of burden using data collected as part of the survey, one must account for both the differences in respondents as well as the survey's sample design. In this article we demonstrate the use of regression tree models to study relationships between perceived burden and objective burden measures conditioned on characteristics of the survey respondent using CE data. Regression trees are an easily interpreted nonparametric conditional model type that can make it easy to understand interaction effects. In this case the interactions between objective and subjective burden and survey, respondent, and household characteristics. To obtain these models, we use the $\mathscr{R}$ package **rpms** which builds regression trees that account for the sample design in their estimation (R Core Team 2020; Toth 2020).

When trying to assess the amount of perceived burden a survey puts on a participant, objective measures such as measures of the time and effort required to complete the survey are often used as a proxy (Rolstad et al. 2011). Our analysis will show that the effect of these objective measures of burden on the perceived burden are different for different households in the CE interview survey. Our findings also show the effect of respondent characteristics and the mode of the survey on these relationships.

If the proposed models could "predict" the perceived burden outcome variable accurately conditionally on the household data collected during the first wave of collection, the predicted values could be used to warn of respondents who are likely to experience high levels of burden which could lead to nonresponse or data quality issues in future waves. This could potentially allow survey administrators to intervene before the next data collection or make changes to collection procedures to head off potential issues.

The rest of the article is organized as follows. The CE survey data and how it is collected is described in Section 2. This section includes a description of the variables in the data set used for the analysis, including the created perceived burden measure. Section 3 contains the regression tree model analysis showing how the relationship between subjective burden and the objective burden measures depend on the household and demographic characteristics of the respondent. A discussion of the study results and their potential utility in future survey designs is contained in Section 4.

2. CE Interview Survey Data

The CE data on household spending, demographic and socioeconomic characteristics are collected through two separate household surveys, the CEQ and the Diary Survey. While the Diary Survey collects data on smaller purchases and irregular expenditures, the CEQ is designed to collect data on large and recurring expenditures that consumers can be expected to recall for at least three months, such as rent and utilities. Together, the data from the two surveys cover the complete range of consumer expenditures.

The interview survey is conducted through a structured questionnaire using one of two collection modes: personal visit or telephone. Households are selected to be in the panel using a two-stage cluster sample of addresses where the clusters are geographical regions defined by groups of counties. The 91 clusters are selected using a PPS (probability-proportional-to-size) sample with 23 certainty units and then addresses are randomly selected within each chosen cluster.

Data are collected from the sample of households over four waves. During the first wave, a field representative collects the demographic and social-economic characteristics of the household and the spending during the previous month to use as a baseline. Expenditure data is collected for each household in the sample using a multiple panel questionnaire during the second, third and the final waves. See Yang (2019) and U.S. Bureau of Labor Statistics (2018) for more information on the sample design and data collection procedures.

The response rate is defined as the proportion of eligible sampled housing addresses from which usable interviews were obtained. A sampled housing address is determined to

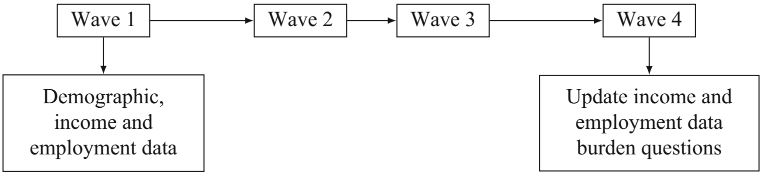

Fig. 2. Illustration of the CE interview survey process.

be ineligible if the house is vacant, under construction, destroyed, abandoned, converted to nonresidential use, or contains temporary residents.

The CE interview survey collects household expenditure data over four waves occurring every three months. In the first wave, demographic questions about the household are asked. Also in the first wave, income and employment information is collected. This information is then updated in the final wave of the survey. After completing the interview in the fourth wave, the respondent was asked four questions aimed at measuring the amount of burden they felt as a result of taking the survey. The burden questions are not part of the usual CE interview survey process and were only asked between the April 2017 and March 2018 study period. In general, the response rate of the interview survey tends to drop from wave 1 to wave 4.

The length and difficulty of a survey is likely to contribute to the attrition of respondents between the first and fourth waves (Kashihara and Ezzati-Rice 2004; Young et al. 2006; Gustavson et al. 2012). For instance, CE lost 22.5% of respondents to the interview survey between wave 1 and wave 4 of the survey in 2017 (Yang 2018). Because the attrition of respondents could impact data quality, it is important to try to reduce attrition by mitigating the burden that the survey imposes on respondents as much as possible through changes to the design and/or collection methods of the survey (Kashihara and Ezzati-Rice 2004; Baird et al. 2008; Cohen et al. 2013).

In order for CE to make meaningful changes to their data collection efforts, it is important to monitor and understand what causes a respondent to feel burdened by the survey and whether or not variables usually thought to be associated with burden, the objective measures of burden, are related to the amount of perceived burden actually felt by respondents. It is important to account for other survey features or respondent characteristics when assessing the relationship between perceived burden and the objective measures of burden because these can affect a respondent's experience of burden. Ignoring these other factors can lead to a misleading interpretation of how certain objective measures of burden are related to perceived burden including making them appear unrelated (Fricker et al. 2014).

We wish to understand how the objective measures of burden, which are obtained as part of the usual CE data collection process, are related to respondents' perceived burden in the CE Interview Survey conditioned on the household characteristics of the respondent. This analysis can help determine which objective burden measures are associated with perceived burden and by how much for different respondents.

One challenge in doing this type of analysis is that in general, surveys do not usually collect measurements of perceived burden (Bradburn 1978). Indeed, in their meta-analysis of studies that examined response rates in relation to a questionnaire's length, Rolstad et al.

(2011) found only three studies that had data that directly asked respondents which questionnaire they preferred and why. Thus, only three of the studies used data that measured respondent burden directly while 25 studies were found that examined the relationship indirectly by means of response rates.

Like most surveys, the CE interview survey does not typically include questions asking respondents directly about the burden they felt filling out the survey. In order to obtain data measuring this more directly, CE engaged in a study with its interview survey between April 2017 and March 2018, in which respondents in the final wave of data collection of the CE interview survey were asked to answer four questions at the end of the questionnaire that were designed to measure the respondent's perception of burden. Respondents were asked to choose answers among "Not at all", "A little", "Somewhat", "Very" or "Extremely" to questions about the amount of burden they felt filling out the survey, how difficult it was to fill out, how sensitive the questions were and if the survey was too long. To get the exact wording of the questions and the possible answers, see U.S. Bureau of Labor Statistics (2017). The relative frequency of responses to the four questions are given in Table 1. For this analysis we use the data from the 6,067 CE interview survey respondents who answered the burden questions at the end of the last wave of their data collection. We can see from these responses that more than half of respondents did not feel the questionnaire was too long or that the questions were very burdensome, difficult or sensitive.

The objective of the analysis demonstrated in this article is to understand the varying effect objective measures of burden have on perceived burden for different household types defined by their demographic and socioeconomic characteristics. To do this analysis, we will need a measure of the perceived burden for each respondent as well as variables that are considered objective measures of burden and variables capturing demographic and socioeconomic information about each respondent and their household.

To quantify the perceived burden felt by a respondent into a single value, we create a composite subjective burden score from the respondents' answers to the four burden questions. First, we assign numeric values to each of the answers of the perceived burden questions, ranging from five for the strongest response, "extremely," to one for the weakest, "not at all," and perform a principal component analysis on these values (Bollen et al. 2001, 2002). However, as Kolenikov and Angeles (2004, 2009) point out, principal component analysis is not well suited for ordinal data where the values are unlikely to

Table 1. The unweighted response rates for each of the response choices to questions on how burdensome, difficult, or sensitive the questions were to answer and how long the survey was in total.

Relative frequency					
Questions	Not at all	A little	Somewhat	Very	Extremely
Burdensome	34.3%	30.2%	24.1%	7.4%	4.1%
Difficult	44.7%	29.9%	20.3%	3.7%	1.3%
Sensitive	35.3%	26.5%	22.3%	10.4%	5.6%
	Very short	Somewhat short	Neither short nor long	Somewhat long	Very long
Length	4.9%	15.6%	41.8%	28.0%	9.8%

Table 2. Description of explanatory variables. For each numeric variable the mean, median, (standard deviation) and (range) are given. For each categorical variable the percentage of each response is given. All the statistics and percentages given in this table are unweighted.

Variable	Description	Descriptive statistics (unweighted)
NEXP	Number of expenditures reported	31.0, 33.8, (14.2),[0.0, 120.0]
TIME	Interview time (minutes)	63.4, 57.3, (31.7), [6.9, 374.5]
INC	Household income before tax (USD)	62,036.8, 42,000.0, (82,140.8), [-18,572.0, 2865,000.0]
PERL6	Number of people in the household less than 6 years old	0.2, 0.0, (0.5), [0.0, 4.0]
PERL18	Number of people in the household < 18 years old	0.6, 0.0, (1.0), [0.0, 11.0]
PERO64	Number of people in the household ≥ 64 years old	0.4, 0.0, (0.7), [0.0, 3.0]
NCHD	Total number of children	0.7, 0.0, (1.1), [0.0, 11.0]
NUMDK	Number of "Don't Know" responses	0.0, 57.3, (1.6), [0.0, 24.0]
NUMRF	Number of questions not answered	0.1, 0.0, (0.6), [0.0, 16.0]
FAMSIZ	Number of members in the household	2.4, 2.0, (1.5), [1.0, 15.0]
BOOK	If the respondent used the information booklet	Yes 38.0%, No 62.0%
RECS	If the respondent used the financial records	Yes 54.3%, No 45.7%
TENURE	Whether the household owns or rents their home	Homeowner 64.8%, Renter 36.9%
MORT	Whether the household has a mortgage	Yes 36.2%, No 63.8%
MODE	Interview mode	Phone 40.7%, Visit 59.3%
CREF	Respondent initially refused but was persuaded	Yes 13.6%, No 86.4%
DOOR	Door step concerns	No concerns 81.2%, Busy and logistics 8.2%, Privacy and government 7.5%, Other concerns 3.1%
HEDU	Highest education in the household	< high school 6.7%, High school 18.6% Some college 32.3% Bachelors and above 42.4%
FAMT	Family type	Married couple 49.7%, Single father 0.9%, Single mother 4.1%, Other 45.3%
URBAN	Urban or rural area	Rural 18.4%, Urban 81.6% Cannot be Determined 18.4%

follow a normal distribution, leading to estimates that are biased toward zero. For ordinal (or categorical) data where values are assigned in the manner here, Kolenikov and Angeles (2004) found that applying principal component analysis using a polychoric correlation matrix (Pearson and Pearson 1922; Olsson, 1979) rather than the standard correlation matrix corrects for that bias. Using the first component from this principal component analysis, as in Yang (2019), we obtain a single composite measure for each respondent that range in value from 3.4 to 17.0, with a median value of 7.6 and a mean of 8.0.

To help understand what can cause survey participants to feel burden, we model the relationship between this perceived burden measure and several variables collected as part of the survey. Some of these variable are usually associated with burden such as the number of expenditures the respondent had to report (*NEXP*) and interview time in minutes (*TIME*) or *BOOK* and *RECS* which record if the respondent used the information booklet and financial records respectively while answering the questions. These variables can take values of "Yes", or "No". The variable *NEXP* contain values that range from 0 to 120 with a median of 31 and mean of 33.76. The variable *BOOK* had a value of "No" 62% of the time while *RECS* had the value "No" only 46% of the time. These variables that measure the time and effort required to respond to the survey are often considered objective measures of burden.

Since the effect that objective measures of burden have on the amount of perceived burden can be different for different people in different households, we also include a number of household and demographic variables in the model to help us understand these differences. These variables include household income before tax (*INC*), whether the household owns or rents their home (*TENURE*), whether the household has a mortgage (*MORT*), number of people in the household less than six years old (*PERL6*), less than 18 years old (*PERL18*), and over 64 years old (*PERO64*) and the total number of children (*NCHD*).

Other variables that we include describe different aspects of the data collection process like interview mode (*MODE*), converted refusal indicator (*CREF*), and door step concerns (*DOOR*). The variable *MODE* is the mode of collection used to collect the survey data from the respondent and can be either a personal visit or telephone interview. Since the mode may affect the relationship between the objective measures of burden and perceived burden, we include this variable in the analysis.

The binary variable *CREF* indicates a respondent who initially refused to respond to the survey and has been persuaded by the interviewer. The categorical variable *DOOR* records any concerns that the respondent expressed to the interviewer before taking the survey. The interviewer can code the concerns of the respondent in variable *DOOR* as one of "no concerns", "busy and logistics", "privacy and government concerns", or "other concerns". The category "other concerns" is a catch all for several outcomes from the respondent saying they don't understand the survey to just shutting the door or hanging up the phone on the interviewer. Since the variables *CREF* (McDermott and Tan 2008) and *DOOR* are both associated with a respondents initial attitude toward the survey and thus affect their responses to the burden questions, these variables are also included. The variable *NUMDK* is the number of "Don't Know" responses and *NUMRF* is number of questions to which the respondent refused to answer.

There are several variables about the household collected in the CEQ that could be associated with how busy a respondent is or with their attitudes about government and privacy, so should be accounted for in the analysis. Among these is the variable *HEDU*, which is the highest education level among people in the household. This variable can take one of the values "less than high school", "high school", "some college", or "bachelors and above". The other variables of this type that we include are family type, *FAMT*, which can take one of the values "married couple", "single parent", or "other", *FAMSIZ*, which is the number of members in the household, and the variable *URBAN*, which records if the household is located in an urban or rural area.

3. Regression Tree Model Analysis

In order to understand how different aspects of a survey affect the perceived burden for different groups of respondents, we model the relationship between the perceived burden composite score and various characteristics of the survey, household, and respondent. A model that estimates the value of the perceived burden composite score conditioned on the values to the collected CEQ variables would allow us to see the effect of different measures conditionally on the data collected about the household. In addition, a model that allows us

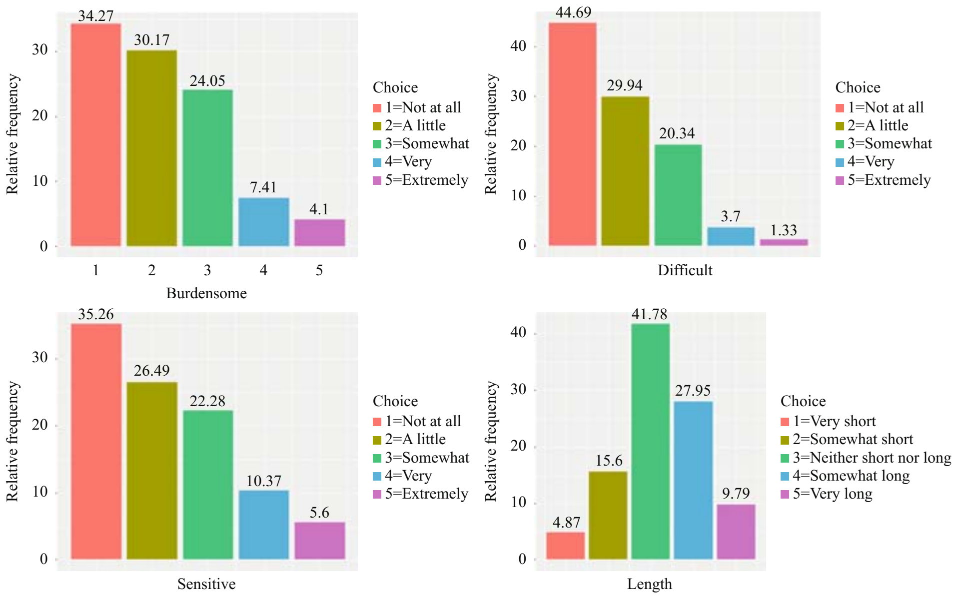

Fig. 3. Bar-charts showing the relative response rates of the choices for each of the burden questions.

to easily understand relationships between the conditional variables could help us better understand what drives the feelings of burden and potentially allow survey administrators to intervene before the next data collection or make changes to collection procedures to head off potential issues. For these reasons we use a recursive partitioning algorithm to create a tree model to do our analysis. These models partition members of the population by splitting them into sub-populations conditionally on their values of the independent variables which can lead to easy interpretation of the model (Toth and Phipps 2014). For examples, Phipps and Toth (2012) and Earp et al. (2018) used these types of models to understand establishment characteristics that affect responses to employment surveys.

The process is termed recursive because each sub-population may in turn be split an indefinite number of times until the splitting process terminates after a particular stopping criterion is reached (Hothorn et al. 2006). One can regard recursive partitioning as producing a model that "predicts" the value of a target variable ("leaf") based on input variables ("branch"). Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the end-nodes of the tree. Typically, within the tree structure, branches represent conjunctions of features that lead to the value of the given end-node.

Since CE data is collected from a sample drawn using a complex design and we would like to generalize our results to the population (Pfeffermann 1996; Pfeffermann and Sverchkov 1999), we use the package **rpms**, *Recursive Partitioning for Modeling Survey Data*, (Toth 2020), in $\mathscr{R}$ (R Core Team 2020) to estimate the models. This algorithm accounts for the survey design variables and sample weights during the recursive partitioning and parameter estimation to produce a design consistent model. We account for the sample design in our models by including the variables containing the design weights, *FINLWT21*, cluster identifiers, *PSU*, which are the primary sampling units and sample strata, *REGION*, (Northeast, Midwest, South, West), which are used to stratify the CEQ sample. Since this algorithm uses a design appropriate permutation test to test the statistical significance of each split, it allows us to specify a *p*-value for our analysis. For all models in this article, we specify a *p*-value threshold of 0.05 to test the significance of each split against.

3.1. Conditional Mean Tree Model

To understand the relationship between the value of the perceived burden score (*PB*) and variables that are usually thought to objectively measure burden, we first model *PB* conditionally on the values of those measures. Figure 4 shows the regression tree model of the mean of *PB*, conditioned on several objective burden measures and survey characteristics. The partitioning algorithm selected several different variables for splitting, *TIME*, *NUMEXP*, *CREF*, *BOOK*, *REGS*, *DOOR*, and *MODE*. These variables were identified by the recursive partitioning algorithm to significantly affect the amount of reported burden.

The model identifies whether the respondent expresses concern about their time, government, or privacy as the most influential variable on how much burden a respondent reports feeling. The recursive partitioning algorithm on the variable *DOOR*, where respondents with busy/logistics, privacy/government doorstep concerns reported the highest levels of perceived burden. This indicates that respondents in this group that have these initial opinions (about 15% of the sample), express a higher amount of burden on average than respondents without these concerns. Indeed, they have a *PB* value of 10.09, which is much higher than the 7.6 median and 8.0 mean for respondents overall. The model did not identify any other variable that had a significant effect on the perceived burden reported by this group, which could indicate that there is not much that could be done to change their experience of burden.

The left side of the tree model represents respondents who did not express these concerns. For these respondents, the amount of perceived burden reported was lower overall, but was influenced by the amount of time the survey took to complete; the longer the survey took the higher average reported burden. However, the effect that time has on the amount of perceived burden depends on the mode of the survey.

Respondents that completed the survey through an in-person interview reported a lower overall average amount of burden than respondents that completed the survey over the

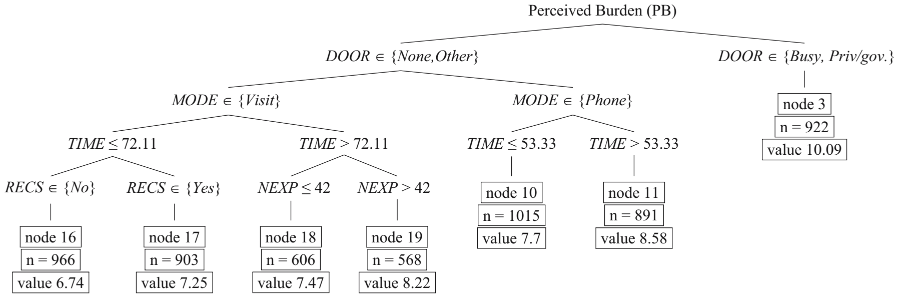

Fig. 4. Regression tree model of the conditional mean of PB: The partitioning algorithm selected the TIME; NUMEXP; CREF; BOOK; RECS; DOOR; and MODE variables for splitting. The mean of PB; conditioned on the splits using these variables, is given in each end node.

phone. For personal interview respondents, the average amount of burden did not significantly increase as long as the time to take the survey was less than 72 minutes, whereas the average reported burden had a significantly higher average for respondents taking the survey by phone after only a little more than 53 minutes. For respondents that had an in-person interview that took longer than 72 minutes, if the number of questions exceeded 42, they reported higher levels of burden than average. While all other respondents that had an in-person interview reported a lower-than-average level of burden, having to refer to records led to a higher reported burden than those who did not. We will see in the regression tree displayed in Figure 7 that the effect of using records on data is more nuanced.

3.2. Conditional Linear Tree Models

Though Figure 4 showed that several objective measures are indeed related to perceived burden, we would like to investigate the effect of each of these objective burden measures for different groups of respondents. Besides modeling the conditional mean of a variable in a tree model, the **rpms** package allows linear models to be fit conditionally on other variables in the tree model (Earp et al. 2018).

We use these models to investigate whether the effects of time, *TIME*, number of expenditures, *NEXP*, and use of records, *RECS*, is different for different groups of respondents. This is done by fitting a linear regression model between *PB* and one of these objective burden measures fit in each end node, while allowing the algorithm to split on the demographic variables when there is an estimated significant difference in the model parameters. By investigating these model parameters conditionally on the demographic information, we hope to understand how the effects of objective measures of burden varies for different types of respondents. Using this type of analysis, we consider how the survey length or the need to consult records or information booklets to answer questions affects perceived burden for different types of respondents.

To analyze the relative effect that time has on the perceived burden we fit the linear model $BURDEN = \beta_1 \times TIME$, while allowing the algorithm to split on any of the variables *MODE*, *INC*, *MORT*, *TENURE*. *PERL6*, *PERL18*, *PERO64*, *NCHD*, *HEDU*, *FAMT*, *FAMSIZ*, or *URBAN* at each step of the algorithm. This model is shown in Figure 5 We fit a linear model with no intercept because we hypothesize that a respondent to a hypothetical interview that took no time ($TIME = 0$ minutes) would report no burden ($PB = 0$). The lowest reported time in the data set is 6.9 minutes, while 75% of all interviews took over 40 minutes to complete.

The resulting regression tree model with this simple linear equation on interview length without intercept confirms what we saw in the previous model, that the effect of time depends on the survey mode (all coefficients are > 0). Though time leads to higher reported level of burden on average, the perceived burden score for a person responding to a personal visit increases by an average of less than one for each 11 minutes of survey length. Meanwhile, the effect of time for respondents answering the survey questions over the phone increases the average perceived burden score at a faster rate, depending on family income. For respondents with a reported family income below USD 25,000, the reported burden score increases by one for every 6.25 minutes they are on the phone, while for respondents with a reported family income greater than USD 25,000, the average

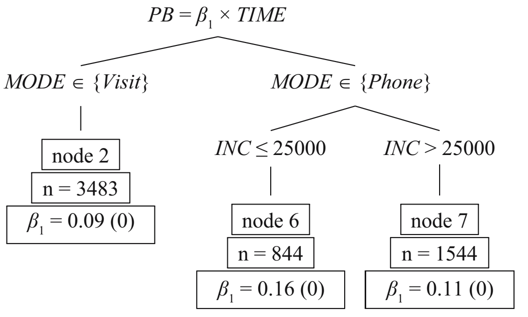

Fig. 5. $PB = \beta_1 \times TIME$: Regression tree model of the conditional relative effect that time has on perceived burden. In each end node the estimated coefficients are given with the (standard error) rounded to three digits.

reported burden score increases by one for each additional nine minutes spent on phone. This difference in rates of reported burden by income is likely due to a difference in the salience of the CE survey for these two groups. Respondents with reported family income below USD 25,000 are unlikely to have as much discretionary income and therefore not many expenditures that they might find interesting to report as compared to families with income greater than USD 25,000.

We also consider this type of tree model analysis for each of the variables *NEXP*, *RECS* and *BOOK* separately. Since each respondent answers demographic questions and questions about the household before the expenditure questions, there is burden associated with responding to the CEQ whether or not the respondent had any expenditures to report, used records, or used the information booklet, so for these variables, we use a linear effect with an intercept term.

Figure 4 shows that, like time, perceived burden increases when more expenditures are reported. Since the number of expenditures and time are correlated, it was natural to see if the effect of *NEXP* is different depending on the mode of survey collection or other demographic variables. To see this, we fit a linear model $PB = \beta_1 \times NEXP + \beta_0$ at each node while allowing the algorithm to split on all the splitting variables used in previous model. Note that, unlike the previous model of time in minutes, we fit a linear model on number of expenditures with a non-zero intercept term, because even if the respondent theoretically reported zero expenditures, they still have to answer questions and so would have burden. The resulting model is shown in Figure 6.

Though the recursive partitioning algorithm found significant differences in average reported perceived burden for different groups of respondents, the effect of *NEXP* was small and about the same for every group. The differences are all in the intercept term. This indicates that the mode of the survey, owning or renting, and the number of children living in the household affects the amount of perceived burden. However, the results show that the perception of burden is not affected by the number of expenditures.

Next, we consider the effect that using records or the information booklet has on perceived burden. Since *RECS* and *BOOK* are indicator variables, fitting linear models $PB = \beta_1 \times RECS + \beta_0$ or $PB = \beta_1 \times BOOK + \beta_0$ at each node while allowing the algorithm to split on any of the variables used in the previous model, leads to an analysis of how the mean-shift effect of these variables changes for different groups of respondents. As in the previous model, Figure 6, we include an intercept term in the linear model which

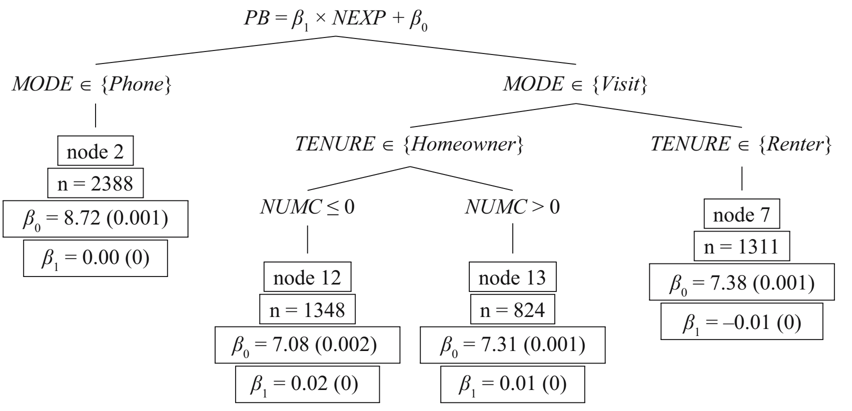

Fig. 6. $PB = \beta_1 \times NEXP + \beta_0$. Regression tree model of the conditional relative effect that the number of reported expenditures has on perceived burden. In each end node the estimated coefficients are given with the (standard error) rounded to three digits.

lets the algorithm differentiate between the effect of the variable being split on and the effect of the variables *RECS* or *BOOK*.

The algorithm did not find any differences in the effect that using the information booklet had on the amount of perceived burden reported. This is not surprising, since the indicator variable *BOOK* does not appear in the tree model relating reported values of perceived burden to objective burden measures (Figure 4). However, the model displayed in Figure 7, analyzing the effect that using records has on perceived burden shows some interesting differences between different groups of respondents.

In the model analyzing the effect of objective burden measures on perceived burden shown in Figure 4, the use of records was associated with a higher reported amount of burden, but only for respondents that responded to the survey through an in-person interview that lasted over an hour (72 minutes). When looking at the difference of this effect by itself among different groups of respondents (Figure 7), we see that though respondents answering the survey questions through a phone interview had higher reported perceived burden, the effect of using records was to lower the amount of perceived burden

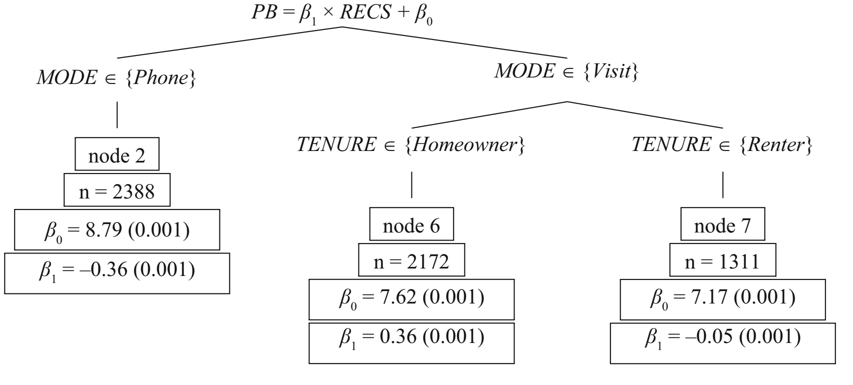

Fig. 7. $PB = \beta_1 \times RECS + \beta_0$. Regression tree model of the conditional relative effect that referring to records to answer survey questions has on the perceived burden. In each end node the estimated coefficients are given with the (standard error) rounded to three digits.

on average. For people responding to the survey through an in-person interview that owned their home, using records raised the reported amount of perceived burden on average. For renters responding to an in-person interview, the effect using records had on the reported perceived burden was negligible. Because consulting records could decrease measurement error in the data, these findings represent trade-offs in some cases between burden and potentially more accurate data. More study on the effect of using records on measurement error in the CEQ should be considered to undergo a complete cost-benefit analysis of asking respondents to consult their records.

The difference in the effect that using records has on the amount of perceived burden between the survey interview modes could potentially be explained by salience. Homeowners often have a larger number of expenditures to report that require records. Locating and using these records to answer questions requires more effort on the part of the respondent, so the positive coefficient is understandable. If the respondent is interviewed in person, it could potentially be more difficult for the respondent to refuse or make an excuse not to get their records. However, respondents interviewed by phone, who usually feel more burdened in general, could more easily say the records are not accessible if they are not interested. Therefore, a respondent's use of records could indicate that they are more interested in answering the survey and so feel less burdened than their counterparts who do not use records in this case.

4. Conclusion

In the above analysis, we demonstrate how regression tree models of the conditional mean can be used to assess the relationship between objective measures and perceived burden. Using respondent's answers to questions directly asking about the burden of CEQ, collected as part of a study by the CE program between April 2017 and March 2018, we were able to model the relationship between the objective measures usually collected as part of the survey and measures of perceived burden. This analysis involves converting the perceived data into a single composite measure obtained from the principal components, which allowed us to use regression tree analysis on the composite measure to see how different survey and respondent characteristics interact with the objective measures of burden to affect perceived burden.

Though there have been mixed findings in the literature on whether these objective measures of burden are related to perceived burden in general, our analysis shows a relationship between perceived burden and some of the objective measures of burden collected in the CEQ survey after conditioning on different demographic and household variables. The tree models also show that the relationship between objective and perceived burden measures are quite affected by the mode of the survey. In general, using a personal interview to collect data seems to ameliorate the effects that most objective measures of burden have on perceived burden.

Using tree models with conditional linear regression estimates at the end-nodes allows us to consider how individual objective measures of burden affect the perception of burden for different groups of respondents, conditioned on the mode of the survey. We modeled three objective measures of burden, interview length (*TIME*), number of expenditures (*NEXP*), and use of records (*RECS*), separately to see the relationship between the values of these objective measures and the amount of perceived burden. The regression tree model fitting the conditional linear model on *TIME* confirmed a positive association between the number of minutes it takes to complete the survey and the amount of perceived burden a respondent feels, but showed that relative change in perceived burden can be quite different for different groups of respondents. Unfortunately, the finding that door step concerns affect the amount of perceived burden, cautions us that there may be limits to how much changes to a survey or data collection can reduce the burden. This is because door step concerns are indicators of a negative initial attitude that could be very difficult to change.

Meanwhile, the other two tree models using conditional linear predictors, showed that the number of expenditures had almost no effect on perceived burden after conditioning on the survey mode and whether the respondent owned or rented their home, and that record usage can lead to an increase or decrease in the amount of reported burden depending on the mode and whether or not the respondent is a homeowner. The effects of these two variables are likely influenced by whether or not the survey is salient to the respondent. However, the study that collected this data on perceived burden did not directly ask the respondent about the salience of the survey, so we could not test this theory.

Another limitation of this analysis is that the data on burden was collected only from participants who completed the fourth interview. Since burden was not measured for anyone who dropped out before completing the final interview, the findings from this analysis could be misleading if the relationship between objective measures of burden and perceived burden are different for respondents and nonrespondents, even after conditioning on household and survey characteristics.

Despite some unavoidable limitations, the result of the study suggests that interview length, number of expenditures, door step concerns, survey mode, housing tenure and number of children affect perceived burden. Though the relationships between these variables and perceived burden can be complicated, using tree models helped us understand these relationships. By using the package **rpms** which allows us to account for the complex sample design of CEQ, we are able to generalize these results to the full population. This implies that these variables should be included in any model to predict a respondent's perceived burden outcome in future collections and possibly even different surveys.

When constructing a model for prediction, it is not necessary to restrict the model to only the most statistically significant effects as does our tree models nor must we restrict ourselves to models that are easily interpretable. Therefore, in future research, we would like to consider exploring the possibility of using design consistent random-forest models to predict respondent's perceived burden using objective burden measures and characteristics of the survey and respondent.

By constructing a model that can accurately estimate a respondent's anticipated perceived burden, the survey administrator could potentially make changes to the collection mode or survey design early in the data collection process to avoid levels of perceived burden that are likely to lead to nonresponse or possibly be used in an adaptive design. Some of these findings could also be used to guide future changes to the questionnaire or the administration of the survey. Using regression tree models for this analysis is a first step to understanding whether objective measures of burden actually affect perception of burden and to what degree.

5. References

Abayomi, E.J., S. Maliszewski, L. Kreiner, and T. Ballard. 2018. "They spoke, we listened: Reducing respondent burden using previously reported data. In proceedings of the Section on Government Statistics of the 2018 Joint Statistical Meetings, July 28–August 2, Vancouver, British Columbia, Canada. Available at: https://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/reports/conferences/JSM-2018/They_Spoke,_We_Listened-_Reducing_Respondent_Burden_Using_Previously_Reported_Data.pdf. (accessed May 2022).

Ashmead, R., E. Slud, and T. Hughes. 2017. "Adaptive Intervention Methodology for Reduction of Respondent Contact Burden in the American Community Survey." *Journal of Official Statistics* 33(4): 901–919. DOI: https://doi.org/10.1515/jos-2017-0043.

Baird, S., J. Hamory, and E.M. Miguels. 2008. *Tracking, attrition and data quality in the kenyan life panel survey round 1 (klps-1).* Institute of Business and Economic Research Center for International and Development Economics Research University of California, Berkeley Working Paper. Available at: http://emiguel.econ.berkeley.edu/assets/miguel_research/71/attrition_paper_FINAL-CIDER_aug08.pdf (accessed May 2022).

Bogen, K. 1996. "The effect of questionnaire length on response rates: a review of the literature." In Proceedings of the Survey Research Methods Section of the 1996 Joint Statistical Meetings, August 4–8, Chicago, Illinois, USA. U.S. Bureau of the Census. Available at: www.asasrms.org/Proceedings/papers/1996_177.pdf. (accessed May 2022).

Bollen, K.A., J.L. Glanville, and G. Stecklov. 2001. "Socioeconomic status and class in studies of fertility and health in developing countries." *Annual Review of Sociology* 27: 153–185. Available at: http://www.jstor.org/stable/2678618 (accessed May 2022).

Bollen, K.A., J.L. Glanville, and G. Stecklov. 2002. "Economic status proxies in studies of fertility in developing countries: Does the measure matter?" *Population Studies*, 56(1): 81–96. Available at: http://www.jstor.org/stable/3092943 (accessed May 2022).

Bradburn, N.M. 1978. "Respondent burden." In Proceedings of the Survey Research Methods Section of the American Statistical Association: 35: 35–40. American Statistical Association, Alexandria, Virginia, USA. Available at: http://www.asasrms.org/Proceedings/papers/1978_007.pdf. (accessed May 2022).

Cannell, C.F., P.V. Miller, and L. Oksenberg. 1981. "Research on interviewing techniques." *Sociological Methodology* 12: 389–437. DOI: https://doi.org/10.2307/270748.

Cohen, S.B., J.W. Cohen, and K. Davis. 2013 *Longitudinal design options for the medical expenditure panel survey insurance component.* Agency for Healthcare Research and Quality Working Paper. Available at: https://meps.ahrq.gov/data_files/publications/workingpapers/wp_13003.pdf (accessed May 2022).

Connelly, N.A., T.L. Brown, and D.J. Decker. 2003. "Factors affecting response rates to natural resource-focused mail surveys: Empirical evidence of declining rates over time". *Society & Natural Resources* 16(6): 541–549. DOI: https://doi.org/10.1080/08941920309152.

Czajka, J.L., and A. Beyler. 2016. "Background paper declining response rates in federal surveys: Trends and implications." *Mathematica policy research* 1: 1–86. Available at: https://aspe.hhs.gov/sites/default/files/private/pdf/255531/Decliningresponserates.pdf (accessed May 2022).

Earp, M., D. Toth, P. Phipps, and C. Oslund. “Assessing Nonresponse in a Longitudinal Establishment Survey Using Regression Trees.” *Journal of Official Statistics* 34(2): 463–481. DOI: https://doi.org/10.2478/jos-2018-0021.

Edgar, J., B. McBride, and A. Safir. 2013a. “Research highlights of the consumer expenditure survey redesign.” *Monthly Labor Review* 136:1. Available at: https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle = hein.journals/month136&section = 74 (accessed May 2022).

Edgar, J. D.V. Nelson, L. Paszkiewicz, and A. Safir. 2013b. *The gemini project to redesign the consumer expenditure survey: redesign proposal.* CE Gemini Project materials, U.S. Bureau of Labor Statistics. Available at: https://stats.bls.gov/cex/ce_gemini_redesign.pdf (accessed May 2022).

Fekete, C., H. Tough, J. Siegrist, and M.W.G. Brinkhof. 2017. “Health impact of objective burden, subjective burden and positive aspects of caregiving: An observational study among caregivers in switzerland.” *BMJ (British Medical Journal)* 7 (9). DOI: http://dx.doi.org/10.1136/bmjopen-2017-017369.

Fricker, S., T. Yan, and S. Tsai. 2014. Response burden: What predicts it and who is burdened out. In Proceedings of the American Association for Public Opinion Research (AAPOR) Annual Conference, May 15–18, 4568–4577, Anaheim, California, USA. Available at: http://www.asasrms.org/Proceedings/y2014/files/4002 98_500838.pdf (accessed May 2022).

Galesic, M., and M. Bosnjak. 2009. “Effects of questionnaire length on participation and indicators of response quality in a web survey.” *Public Opinion Quarterly* 73(2): 349–360. Available at: http://www.jstor.org/stable/25548084 (accessed May 2022).

Groves, R.M. 2006. “Nonresponse rates and nonresponse bias in household surveys.” *Public Opinion Quarterly* 70(5): 646–675. Available at: http://www.jstor.org/stable/4124220 (accessed May 2022).

Groves, R.M., D.A. Dillman, J.L. Eltinge and R.J.A. Little. 2001. *Survey Nonresponse*. New York: Wiley Inter-science.

Groves, R.M., E. Singer, and A. Corning. 2000. “Leverage-saliency theory of survey participation: Description and an illustration.” *The Public Opinion Quarterly* 64(3): 299–308. Available at: http://www.jstor.org/stable/3078721 (accessed May 2022).

Gustavson, K., T. von Soest, E. Karevold, and E Røysamb. 2012. “Attrition and generalizability in longitudinal studies: Findings from a 15-year population-based study and a monte carlo simulation study.” *BMC Public Health* 12(918): 1–11. DOI: https://doi.org/10.1186/1471-2458-12-918.

Hothorn, T., K. Hornik, and A. Zeileis. 2016. “Unbiased recursive partitioning: A conditional inference framework.” *Journal of Computational and Graphical Statistics* 15 (3): 651–674. Available at: http://www.jstor.org/stable/27594202 (accessed May 2022).

Kashihara, D., and T.M. Ezzati-Rice. 2004. “Characteristics of survey attrition in the household component of the medical expenditure panel survey (meps).” In Proceedings of the Survey Research Methods Section of the 2004 Joint Statistical Meetings, August 8–12: 3758–3765, Toronto, Ontario, Canada. Available at: http://www.asasrms.org/Proceedings/y2004/files/Jsm2004-000706.pdf. (accessed May 2022).

Kolenikov, S., and G. Angeles. 2004. *The use of discrete data in pca: Theory, simulations, and applications to socioeconomic indices*. Chapel Hill: Carolina Population Center, University of North Carolina: 20: 1–59. Available at: https://www.measureevaluation.org/resources/publications/wp-04-85 (accessed May 2022).

Kolenikov, S., and G. Angeles. 2009. "Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer?" *Review of Income and Wealth* 55(1): 128–165. DOI: https://doi.org/10.1111/j.1475-4991.2008.00309.x.

Lynn, P. 2014. "Longer interviews may not affect subsequent survey participation propensity." *Public Opinion Quarterly* 78(2): 500–509. Available at: http://www.jstor.org/stable/24545936 (accessed May 2022).

McDermott, N., and L. Tan. 2008. "The effect of refusal conversion on data quality in the consumer expenditure interview survey." *Consumer Expenditure Survey Anthology*: 23–32. Available at: https://stats.bls.gov/cex/anthology08/csxanth4.pdf (accessed May 2022).

U.S. Bureau of Labor Statistics. 2017. C*onsumer expenditure survey interview questionnaire (ceq).* Technical report, U.S. Bureau of Labor Statistics. Available at: https://www.bls.gov/cex/capi/2 017/2017-CEQ-CAPI-instrument-specifications.pdf (accessed May 2022).

U.S. Bureau of Labor Statistics. 2018. *Handbook of methods.* Technical report, U.S. Bureau of Labor Statistics. Available at: https://www.bls.gov/opub/hom/cex/home.htm (accessed May 2022).

Olsson, U. 1979. "Maximum likelihood estimation of the polychoric correlation coefficient." *Psychometrika* 44(4): 443–460. DOI: http://dx.doi.org/10.1007/BF02296207.

Pearson, K., and E.S. Pearson. 1922. "On polychoric coefficients of correlation." *Biometrika* 14(1/2): 127–156. DOI: https://doi.org/10.2307/2331858.

Pfeffermann, D. 1996. The use of sampling weights for survey data analysis. *Statistical Methods in Medical Research* 5(3): 239–261. DOI: https://doi.org/10.1177/096228029600500303.

Pfeffermann, D., and M. Sverchkov. 1999. "Parametric and semi-parametric estimation of regression models fitted to survey data." *Sankhyā* 61(1): 166–186. Available at: http://www.jstor.org/stable/25053074 (accessed May 2022).

Phipps, P., and D. Toth. 2012. "Analyzing establishment nonresponse using an interpretable regression tree model with linked administrative data." *The Annals of Applied Statistics* 6 (2): 772–794. Available at: http://www.jstor.org/stable/41713473 (accessed May 2022).

R Core Team. 2020. *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/. (accessed May 2022).

Rolstad, S., J. Adler, and A Rydén. 2011. "Response burden and questionnaire length: is shorter better? a review and meta-analysis." *Value in Health* 14(8): 1101–1108. DOI: https://doi.org/10.1016Zj.jval.2011.06.003.

Sharp, L.M., and J. Frankel. 1983. "Respondent burden: A test of some common assumptions." *The Public Opinion Quarterly* 47(1): 36–53. Available at: https://www.jstor.org/stable/2748704 (accessed May 2022).

Toth, D., and P. Phipps. 2014. "Regression tree models for analyzing survey response." In Proceedings of the Section on Government Statistics of the 2014 Joint Statistical Meetings: 339–351, Boston, Massachusetts, USA. Citeseer. Available at:

http://citeseerx.ist.psu.edu/viewdoc/download?doi = 10.1.1.696.8138&rep = rep1&type = pdf (accessed May 2022).

Toth, D. 2020. "*rpms: Recursive Partitioning for Modeling Survey Data*." Available at: https://cran.r-project.org/web/packages/rpms/rpms.pdf. R package version 0.4.0.

Yang, D.K. 2018. "Evaluating perceived burden of household survey respondents." In Presentation of the American Association for Public Opinion Research (AAPOR) Annual Conference, May 16–19, and DC AAPOR and Washington Statistical Society (WSS) Summer Conference July 16, Denver, Colorado, and Washington D.C. USA. Available at: http://app.core-apps.com/aapor_2018/abstract/e28a28d3-f1e7-40e9-b39b-a2cbc76e519, files.dc-aapor.org/slides/summer2018/Yang.pdf. (accessed May 2022).

Yang, D.K. 2019. "Assessing how a household survey is perceived by respondents." In Proceedings of the Section on Government Statistics of the 2019 Joint Statistical Meetings, July 27–August 1. 1–19. Denver, Colorado, USA. Available at: https://www.bls. gov/osmr/research-papers/2019/st190130.htm (accessed May 2022).

Young, A.F., J.R. Powers, and S.L. Bell. 2006. "Attrition in longitudinal studies: Who do you lose." *Australian and New Zealand Journal of Public Health* 30(4): 353–361. DOI: https://doi.org/10.1111/j.1467-842X.2006.tb00849.x.

Received February 2021
Revised November 2021
Accepted May 2022