Assessing How a Household Survey is Perceived by Respondents

November 2019

Daniel K. Yang*†

Abstract

Survey modifications could impact respondent burden. Increased burden could potentially lead to refusals in the following waves or inadequate answers to questions, which could in turn induce bias affecting overall data quality. Census has studied survey participant impressions of data security and privacy. Burden measurement would allow us to identify where interventions may be needed to offset the impact of respondents perception of burden and to mitigate burden-induced bias on data quality. During the 2012 and 2017 Consumer Expenditure Surveys (CE) Quarterly Interview, answers on perceived burden were collected at the end of the final interview wave. In this study, we will introducing a composite burden index score using a multivariate technique. We studied respondent burden proxy indicators by using the nonparametric recursive partitioning model under a complex survey design for the 2012 CE Quarterly Interview Survey. We will also present the results from the 2017 CE newly revised respondent burden questions.

Key Words: Data Collection, Respondent burden, Nonparametric, Recursive partitioning, Index, Indicator

1. Introduction

There have been sequential researches conducted for the Consumer Expenditure (CE) Surveys redesign and subsequent implementation of redesign. The principal objective of the redesign is enhancing the data quality of CE Surveys, with decreased and verified measurement error. Reducing measurement error may need to ask more questions to increase data quality, but the trade off is increasing respondents burden. Therefore, it is important to be able to measure respondent burden which also contributes to data quality. We would like to evaluate respondents' perceived level of burden over time.

The data of U.S. consumer's purchasing activities is collected by CE: a consumer unit (CU: family or single consumer) household spending, family income, demographic and social-economic characteristics, etc. CE is conducted by the U.S. Census Bureau. There are two components of CE: Quarterly Interview Survey (CEQ which is the focus of this study) and the Diary Survey (CED). The current CEQ consists of four waves, the first wave is where a census field representative will collect a participating CU's demographics, social-economic characteristics and baseline spending figures of previous month. In the second, the third and the final wave, a selected CU or household will be followed-up with multiple panel survey questionnaires to compile spending data.

CE has unique values for economics researches: first, CE is the only nationwide survey to collect a wide range from U.S. household's expenditure, income to demographic and social-economic characteristics measurements; second, CE provides a comprehensive observation of economic aspects from U.S. consumers. CE has been used in a broad range of economic researches: to evaluate the effects of economic policy modifications on subpopulation, to serve as inputs for the U.S. Census Supplemental Poverty Measure, to examine U.S. household consumption pattern and tendency by scholars and institutions. Perhaps,

*U.S. Bureau of Labor Statistics Office of Survey Methods Research, 2 Massachusetts Avenue Suite 5930, NE Washington, DC 20212

†Disclaimer: Any opinions expressed in this paper are those of the author(s) and do not constitute policy of the U.S. Bureau of Labor Statistics.

the most important aspect of CE is to periodically update the Consumer Price Index (CPI) market basket of goods and services and their relative importance by providing expenditure and demographic inputs of U.S. household.

For other Federal agencies, the CE quarterly expenditure estimates have been used to update the cost of living. CE data have been used to gauge National Income and Product Accounts (NIPA), annual growth rates and benchmarking by the Bureau of Economic Analysis (BEA). Researchers have been conducting numerous statistical models on CE data to assess U.S. household spending behavior to address respective research questions or economic hypothesis (Yang [2013]).

CE is a burdensome survey which takes almost an hour long, with a lot of non-trivial questions. To assess burden, after completing the standard expenditure questionnaire at the end of the final interview wave, respondents are asked a series of research questions, including ten questions that ask respondents for their assessment of burden from participating in the survey.

In survey practice, psychologists deploy multivariate questions (a.k.a. items) to measure different behavior aspects or dimensions of respondents, e.g., burden questions (or items) in the CE. Psychologists will often develop an index based on these multivariate questions (items) to reflect latent constructs of a respondents behavior or perception. The burden questions collected in the CE research section provide data to implement multivariate techniques and to develop composite burden measures (Yang [2015]), as an extension of prior research, e.g., Fricker et al. [2012] and Kopp et al. [2013]. Respondents perception of burden has been shown to have a negative association with response propensity in the survey panel (Fricker et al. [2012]), which may be a contributing factor to data collection strategy adjustment.

In this study, we developed (composite) burden index scores to track perceived respondent burden over time. Yan [2015] also applied a single burden question (or item) to examine the impact of the respondents perceived burden on data quality. Those respondent burden measures would allow CE to detect and understand changes in burden following modifications to the survey, to evaluate the association between the burden measures and other variables of interest, and to develop interventions that reduce respondents perception of burden. For example, suppose burden index scores and single burden question are used to collect the perception of burden from each respondent, then it could be integrated as an element of an intermediate summary or report to inform survey management, which in turn would help to mitigate the bias maybe introduced by burdened-out respondents.

Currently, burden questions are collected in the final interview wave, the burden index scores and single burden question could also be used to improve the data collection process in the next round of the survey to offset measurement error. If burden questions were asked in earlier interview waves, then potentially the respondent burden measures could be used to identify respondents that are at the greatest risk of dropping from the panel or providing low quality data. This information could then be used to change some aspect of the survey (e.g., incentive levels, interviewer engagement) in order to retain respondents or improve reporting.

In this paper, the burden questions and corresponding CE data in 2012 and 2017 will be described in Section 2. We will introduce a composite burden index score using Polychoric correlation PCA and a simple summation score for respondents in Section 3. Section 4 uses the nonparametric recursive partitioning model to analyze respondent burden proxy indicators under a complex survey design for both the 2012 CE Quarterly Interview Survey (CEQ) and the 2017 CEQ newly revised respondent burden questions. The results of this study is then summarized in the Section 5.

2. CE Respondent Burden Data

2.1 Burden Questions (or Items)

Between October 2012 and September 2013, a series of questions were asked (to respondents) in the interview survey at the end of the final wave, including 10 questions assessing respondents perceived burden (CEQ [2013]), e.g.

How burdensome was this survey to you (bbur)?

1 = Not at all burdensome

2 = A little burdensome

3 = Somewhat burdensome

4 = Very burdensome

How sensitive did you feel the questions I asked today were?

1 = Not at all sensitive

2 = A little sensitive

3 = Somewhat sensitive

4 = Very sensitive

Do you feel that the length of today's interview was too long, too short, or about right?

1 = too short

2 = about right

3 = too long

How interesting was this survey to you?

1 = Very interesting

2 = Somewhat interesting

3 = A little interesting

4 = Not at all interesting

How difficult or easy was it for you to answer the questions in this survey?

1 = Very easy

2 = Somewhat easy

3 = Somewhat difficult

4 = Very difficult

You were asked to participate in five interviews. Would you say that this was too many interviews or did it seem like a reasonable number?

1 = A reasonable number

2 = Too many interviews

How agreeable would you be to take another survey like this in the future?

1 = Very agreeable

2 = Somewhat agreeable

3 = Not at all agreeable

If we had to extend this survey for another 15 minutes, how willing would you have been to continue this interview?

1 = Very willing

2 = Somewhat willing

3 = Somewhat unwilling

4 = Very unwilling

Thinking about the amount of effort that you put forth into answering today's survey, would you say that you put forth:

1 = A little effort

2 = A moderate amount of effort

3 = A lot of effort

Please tell me if you agree or disagree with the following statement: I trust the U.S. Census Bureau to safeguard the information that I have provided them.

1 = Strongly agree

2 = Agree

3 = Neither agree or disagree

4 = Disagree

5 = Strongly disagree

We used data sets in (Yang [2015]) for demonstration purposes. The data set includes household units who participated in 5 waves of the CE survey panel which covered the implementation of the 2012-2013 CE Research Section Questions of interest. The actual burden questions data were collected between October, 2012 (the 4^{th} quarter or Q4) and September, 2013 (the 3^{rd} quarter or Q3) in the CE Research Section, the corresponding 1^{st} wave interview was conducted in October, 2011 and September, 2012, respectively.

Between April 2017 (the 2^{nd} quarter or Q2) and March 2018 (the 1^{st} quarter or Q1), an updated version of 4 questions assessing respondents perceived burden were asked in the interview survey at the end of the final wave with an extra level of category added in CEQ [2017], e.g.

How burdensome was this survey to you (bbur)?

1 = Not at all burdensome

2 = A little burdensome

3 = Somewhat burdensome

4 = Very burdensome

5 = Extremely burdensome

How difficult was it for you to answer the questions in this survey?

1 = Not at all difficult

2 = A little difficult

3 = Somewhat difficult

4 = Very difficult

5 = Extremely difficult

How sensitive did you feel the questions I asked today were?

1 = Not at all sensitive

2 = A little sensitive

3 = Somewhat sensitive

4 = Very sensitive

5 = Extremely sensitive

In thinking about the length of today's survey, would you say it was?

1 = Very short

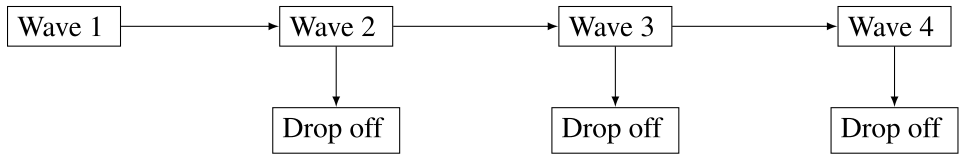
2 = Somewhat short

3 = Neither short nor long

4 = Somewhat long

5 = Very long

The 2017 CE Data has 4 waves, burden questions were only collected from participants in their final wave. Here is an illustration of attrition by the final wave:



CE also takes into account replacement households for people that move. It is true that people drop off, but there are also people that move and are replaced by new people at the address (though likely at a lower rate than they are dropping off). We excluded households with missing values in any of the burden questions, final samples had 6,369 households (in 2012 with 10 burden questions) and 6,067 (in 2017 with 4 burden questions).

2.2 Burden Proxy Indicators

CE is entirely rely on respondent's answer of burden, however, there are other objective indicators, e.g. other sets of variables people used to indicate burden, such as "burden proxy indicators." Some Census researchers use heavily of those objective indicators, in this study, the burden proxy indicators of interest for CE includes: household income before tax ($FINCBTAX$), interview length ($TOTTIMEmin$, minutes), number of expenditures ($NUMEXPN$, unedited), mortgage indicator ($mortgageind$, 0 or 1), converted refusal indicator ($CONVREF1$, Yes or No). In addition, we have interview mode ($telph$) and category variables of information booklet usage ($INFOBOOK2$) and records usage ($RECORDS2$):

Interview Mode:

0 = Visit (personal visit)

1 = Phone (telephone)

Information Booklet:

5 = Almost (Almost always, 90% of the time or more)

4 = Most (Most of the time, 50% to 89% of the time)

3 = Occas. (Occasionally, 10% to 49% of the time)

2 = Never (Never or almost never, less than 10% of the time)

1 = No I.B. (The respondent did not have access to the information booklet, reference level)

Record Usage:

4 = Almost (Almost always, 90% of the time or more)

3 = Most (Most of the time, 50% to 89% of the time)

2 = Occas. (Occasionally, 10% to 49% of the time)

1 = Never (Never or almost never, less than 10% of the time, reference level)

We also have door step concerns ($ICONCnew$), a variable from Contact History Instrument (CHI). We did not simply dichotomize door step concerns because respondents burden perception may be different among different subgroups.

Door Step Concerns:

0 = No concerns

1 = Privacy/gov. (government concerns)

2 = Busy/logist. (logistics)

3 = Other

3. Develop Burden Index Scores from Polychoric Correlation Principal Component Analysis (PCA)

Those burden questions in 2012 and updated version in 2017 provided data to implement multivariate techniques and to develop composite burden index scores (based on 10 items and 4 items, respectively). The socio-economic status (SES) of a household or an individual covers multiple dimensions of characteristics. The most common method to aggregate those multivariate questions (items) is to allocate weights to those questions, then summarize those weighted (values) into a composite measure (Kolenikov and Angeles [2004]). [Bollen et al., 2001, 2002] suggested that principal components provide one of the best performances under a regression model framework. Principal Component Analysis (PCA) works when the variables under consideration are continuous and have an approximate Normal distribution.

According to [Kolenikov and Angeles, 2004, 2009], directly applying PCA on ordinal data violates the Normal assumption. In addition, since PCA focuses on the second moment (covariance or correlation) of the data, the estimated covariance or correlation matrix of categorical variables does not reflect the actual covariance or correlation matrix of the latent (continuous) variables which had not been observed. Furthermore, those estimates are more likely to be biased toward 0. PCA estimates of factor loadings (or component weights) on categorical variables will be biased or inconsistent, and estimates of the proportion of variance explained by the first few components will be downward biased. Lastly, the order of categorical variables is not likely to be preserved by PCA. The violation of the Normal assumption, inaccurate covariance or correlation estimation of non-continuous variables, and inconsistent proportion of variance explained by components will make classical PCA a less desirable option for this type of burden data analysis. For categorical data, Kolenikov and Angeles [2004] found that applying PCA on a Polychoric correlation matrix, which had been introduced by Pearson and Pearson [1922], and Olsson [1979], provides a desirable approach.

The steps of constructing the composite burden index scores are summarized in the following:

1. Compute a Polychoric correlation matrix of the burden data,

2. Compute PCA on this Polychoric correlation matrix,

3. Determine the number of principal components to be selected:

   (a) apply the Broken-stick method and the proportion of the total variance explained criteria,

(b) identify the largest loading magnitude of selected principal components,

4. Propose ways to compute the overall composite burden index scores based on Polychoric correlation PCA: a simple summation and a proportional weighted summary.

3.1 Compute PCA on the Polychoric Correlation Matrix of the Burden Data

We applied R$\{polycor\}$ package by using the maximum likelihood (ML) estimation to compute a Polychoric correlation matrix on burden data. We used the R $princomp()$ function by specifying the covariance/correlation matrix input to compute PCA on a Polychoric correlation matrix. Because in this computation, the input is a Polychoric correlation matrix but not the original data set, the scores (principal components) need to be manually calculated (see Subcestion 3.3).

3.2 Determine the Number of Principal Components to be Selected

The number of selected principal components needs to be determined to represent the data or Polychoric correlation matrix. If the number of selection is too large, then we miss the main purpose of dimension reduction of PCA. If the number of selection is too small, then we do not have principal components to account for a sufficient amount of variation. Here, we introduce a combination of guidelines to select the number of principal components consisting of a simple and easy to implement Broken-stick method, and two criteria of proportion of the total variance explained.

1. Broken-Stick Method: Suppose the total variance is partitioned randomly among the principal components, then the eigenvalues follow a Broken-stick distribution (Peres-Neto et al. [2005]). Jackson [1993] illustrated in a simulation study that the Broken-stick method provided robust performance. Let p be the total number of principal components, so for the k^{th} component, the critical value is

   $$b_k = \sum_{i=k}^{p} \frac{1}{i}.$$

   An equivalent critical value to account for percentage of total variance is

   $$b_k^{'} = \frac{1}{p} \sum_{i=k}^{p} \frac{1}{i}.$$

   However, only one principle component may not be able to account for sufficient proportion of total variation explained. Thus, the criteria of selected principle components number needs to be established as in the following:

2. A Proportion $1/p$ of the Total Variance Explained: Johnson and Wichern [2007] suggested that selecting the principal components which account for the minimum of a proportion $1/p$ of the total variance explained (or alternatively, estimated variances > 1). Hence, for the CE burden data, the first and the second principal components are selected as minimum.

3. Cumulative Proportion of the Total Variance Explained: Rencher [2003] suggested that selecting the principal components to account for a pre-specified proportion of

the total variation. For simplicity, we chose 50% as a pre-specified level, in other words, the cutoff threshold to accept principal components is explaining at least a cumulative 50% of total variation. For example, The CE 2012 burden data shows that the first and the second principal components account for a cumulative 57.96% of the total variance explained, hence, those two principal components should be selected. The Broken-stick method selected the first principal component, however, a proportion $1/p$ of the total variance explained method and cumulative proportion of the total variance explained method selected the first and the second principal components. Therefore, considering both the Broken-stick method and the sufficient proportion of the total variance explained criteria, we chose to select the first and the second principal components.

4. Identify the Largest Loading Magnitude of Selected Principal Components:

   The CE 2012 burden data showed that the burden questions with relatively high loading magnitude (absolute value) in the first principal component are:

   (a) Whether the five waves of interview is too many,

   (b) Respondents feeling about the length of the interview,

   (c) How burdensome did the respondent feel about this survey,

   (d) How likely would a respondent agree to take another survey like this in the future.

   Therefore, the first principal component may be considered as a respondent's burden assessment component to indicate its impact on perceived burden. In the second principal component, burden question which asked about respondents level of effort in responding to the questionnaire, has the largest loading magnitude (absolute value). Hence, the second principal component may be regarded as a "respondent's effort" component to indicate its impact on perceived burden.

3.3 Compute the Overall Composite Burden Index Scores

We used the results of Polychoric correlation matrix PCA to compute respondents overall composite burden index scores (in R). The loadings of PCA on a Polychoric correlation matrix were not produced from the original burden item data set. Therefore, the intermediate scores should be created based on the selected principal components and Polychoric correlation matrix.

1. First, the intermediate scores were obtained from selected principal components and Polychoric correlation matrix (i.e. $\boldsymbol{S_i} = \boldsymbol{\rho PC}$).

2. Second, we computed composite burden index scores from intermediate scores and the original burden item data ($\boldsymbol{X}$), e.g. in the CE 2012 burden data, we selected the first and second principal components, each respondent will have two separate composite burden scores from the first and second principal components, respectively.

3. Finally, we produced the overall summation of composite burden scores by applying the proportional weighted summation which used the proportion of total variation ($\boldsymbol{p_i}$) explained in the Polychoric correlation matrix PCA as weights to compute a weighted summary composite burden score (i.e. $\boldsymbol{p_i X S_i}$).

3.4 Compute the Likert Scale Summation Scores (A Standardized Shifted Simple Summation of Burden Questions)

A simple summation was calculated to serve as a comparison for the Polychoric correlation PCA proportional weighted summation overall composite burden index scores. If the two scores have similar properties and similar predictive power, than the simpler, more interpretable Likert Scale Summation Scores could be easily adopted. If the composite burden index scores provide more predictive power, are more stable, or have other statistical properties, than it could be used. Since the burden questions had different numbers of response ranges (from 2 to 5), we first standardized each burden question Likert Scale (e.g. X_i) using the formula $Z_i = \frac{X_i - \bar{X}}{SD(X)}$. We then shifted each distribution by the minimum of all standardized burden items so that the standardized shifted burden questions are all positive. We did this because having negative levels of burden is not meaningful. Then compute the sum of all standardized shifted burden question Likert Scale values into a measurement of Likert Scale Summation Scores for each respondent.

4. Nonparametric Recursive Partitioning Model of CE Burden Proxy Indicators under a Complex Survey Design

4.1 Nonparametric Recursive Partitioning Model under a Complex Survey Design

There were 10 questions collected in 2012 and 4 questions collected in 2017 on respondent burden perception, respectively. It is natural to consider combining those questions into a single measure (by weighting and other functions), therefore, we now have three burden measures:

1. Single Burden Question.
2. Likert Scales Summation Scores: for simplicity, the Likert Scales summation scores were computed by summarizing all 10 or 4 burden questions (of Likert Scales responses), respectively.
3. Composite Burden Index Scores: proportional weighted summation of overall burden index scores using Polychoric correlation PCA from burden questions.

Multivariate Analysis is powerful for revealing the connections and relative significance between factors, variables that tends to be large in numbers and complex in structure. PCA has particular advantages:

1. Robust due to its nonparametric nature.
2. Diagnoses important variables with minimal information loss.
3. Produces independent principal components.
4. "Requires no sparsity constraint" (Hu [2018]) [1].

While looking at those 3 various measures which will be used interchangeably throughout this paper, a natural extension is to see which one is better associated with, e.g. some kind of data quality measurements, hence, we have done some numerical analyses on these 3 measures. Here are some of our findings in the 2012 data where all 10 burden questions were used:

[1] We also applied Sparse PCA (with sparseness constraint) as suggested in Hu [2018] and the results are identical to regular PCA.

1. There is no conclusive evidence of differences in correlations in data quality measures with burden measurements.

2. From the data we have, for both the single burden question and burden scores, excluding most-burdened respondents does not appear to have much of an effect on selected expenditure variable mean estimates.

We wanted to see which one is better to be associated with some kind of data quality measurements, but the results are not very "predicative", therefore, we need to look at alternatives, like a nonparametric approach, to see how burden measures are associated with variables we think are related to burden.

In addition, we had experienced questions such that, could we "predict" perceived burden index scores? For example, in an incentives field test study, burden questions were not asked, however, it did have variables that considered related to burden (such as regarding records usage, information booklet usage, survey mode, length of interview, respondent efforts and doorstep concerns).

Therefore, if one's only interest is measuring respondent's burden, then a single burden question is sufficient. However, the caveat here is if one is interested in the dimensions of burden which may lead to improvement or intervention strategy, then burden questions cover multiple aspects need to be asked. Now we try to address the question of "how to extrapolate among burden proxy indicators, after taking into account the complex design?" It is natural to consider a nonparametric approach which does not require assumptions, e.g. like in linear model, and more robust, such as recursive partitioning to model single burden question (categorical) and/or burden index scores (continuous).

Recursive partitioning creates a decision tree that strives to correctly classify members of the population by splitting it into sub-populations based on several independent variables. 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 (Loh [2011]).

Recursive partitioning can be regarded as "predictive" by using input variables ("branch") to predict a target variable ("leaf"). The goal is to create a model that predicts the value of a target variable based on several input variables. 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 leaf. For simplicity, if the target variable takes a discrete set of values then it is called a classification tree; while if the target variable takes continuous values (usually in real numbers), then it is called a regression tree. Typically, in the decision tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

4.2 Nonparametric Recursive Partitioning Models for CE 2012 Burden Proxy Indicators with 10 Burden Questions

Respondents perception of burden could be very different for different sub-populations. The Recursive Partitioning for Modeling Survey Data R$\{rpms\}$ package (2018 release, Toth [2019]) had been implemented on the 2012 Q4 - 2013 Q3 CE burden data for all 10 burden questions.

We tried to look into how burden measures are related to those burden proxy indicators in the CE 2012 data with 10 burden questions. First, we used recursive partitioning to split for the single burden question among the burden proxy indicators after taking into account the complex survey design (Figure 1). For the single burden question, we saw that those respondents with no doorstep concerns reported the lowest levels of burden. Households who

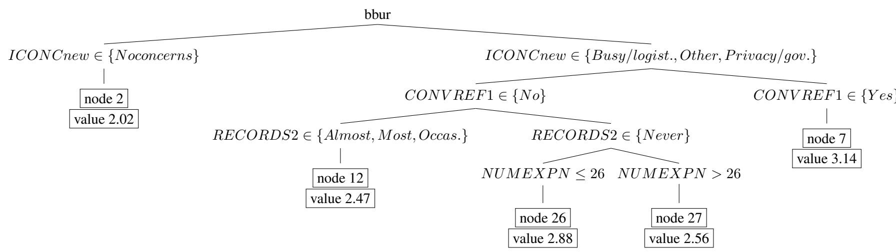


Figure 1: 2012 Q4 - 2013 Q3 Single Burden Question (bbur)

expressed concerns and had to be further convinced to participate in the survey, that were converted refusals, expressed the highest levels of burden. Respondents who expressed concerns originally but used records during the interview reported a moderate amount of burden. For households who expressed door step concerns but never used record during the interview, the model further split at whether the number of reported expenditures is > 26, those who reported > 26 expenditures had a lower reported burden level than those who reported ≤ 26.

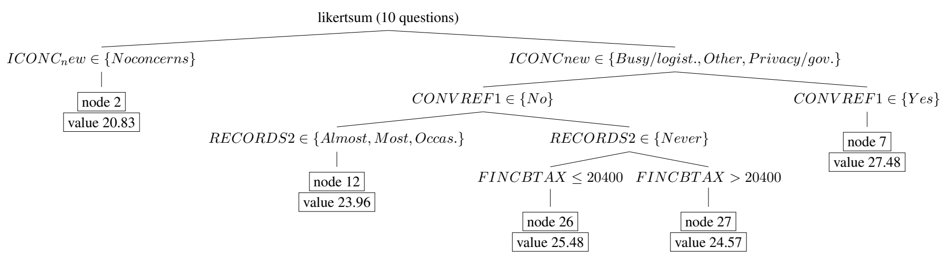


Figure 2: 2012 Q4 - 2013 Q3 Likert Scales Summation Scores (likertsum, 10 Questions)

Second, we used recursive partitioning to split for the Likert Scale summation scores on the burden proxy indicators after taking into account the complex survey design (Figure 2). Once again, those households with no doorstep concerns reported the lowest level of burden, converted refusals who had expressed concerns reported the highest level of burden, and respondents who used records expressed moderate level of burden. For respondents who had doorstep concerns, who were not a converted refusal, did not use records, and whose household income $>$ \$20,400 (e.g. which may be close to some kind of lower income threshold) reported the higher burden than those who reported income $\leq \$20,400$.

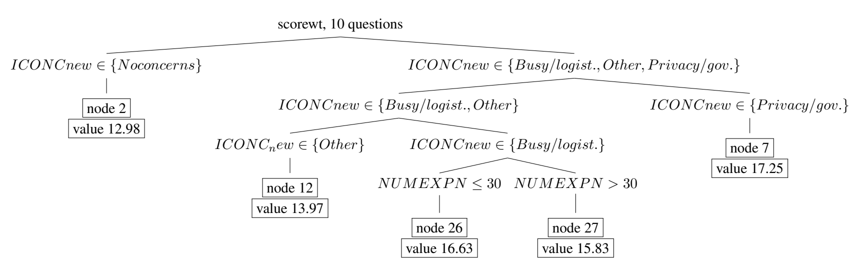


Figure 3: 2012 Q4 - 2013 Q3 Composite Burden Index Scores (scorewt, 10 Questions)

Finally, we implemented another recursive partitioning to model relationships between the burden proxy indicators and the composite burden index scores after taking into account the complex survey design (Figure 3). The decision tree looked a bit different here. The first three splits were all related to doorstep concerns. In this tree structure, the actual type of concern, whether it was about privacy or the government compared to concerns about being to busy, logistics or other concerns, was related to burden. This tree ended with a split on number of expenditure reported, respondents who reported > 30 expenditures had lower burden than those who reported ≤ 30.

When we converted information booklet and records usage into indicators, the recursive partitioning analysis produced the same classification trees in all the above models.

2012 Q4 – 2013 Q3 (10 burden questions)	Single Burden Question	Likert Scales Summation Scores	Composite Burden Index Scores
First Split	Door Step Concerns vs. Not	Door Step Concerns vs. Not	Door Step Concerns vs. Not
Second Split	Converted Refusal vs. Not	Converted Refusal vs. Not	Door Step: Privacy, or Gov. vs. Busy, Logistic, Other
Third Split	Record Usage vs. Not	Record Usage vs. Not	Door Step: Busy or Logistic vs. other
Fourth Split	Number of Expenditures > 26 vs. ≤ 26	Income $> \$20,400$ vs. Income $\leq \$20,400$	Number of Expenditures > 30 vs. ≤ 30

Table 1: CE 2012 Recursive Partitioning for Modeling Survey Respondent's Burden Perception Decision Tree Comparison - 1

Looking across all three burden measures in 2012 models where all 10 burden questions were used (Table 1), we saw that there were commonalities. Whether a respondent had door step concerns or not had the strongest association with the burden measures. Respondents with door step concerns consistently reported higher levels of burden. Within the composite burden index scores, specific door step concerns drove the second and third split, suggesting that the topic of the concern, e.g. if the respondent was worried about privacy, was related to the level of burden they would report. In the other two measures, the Likert Scales summation scores and the single burden question, whether a respondent was a converted refusal had the next strongest association. Use of records was found to be associated with burden in those two measures as well. As for the burden measures themselves, the single burden question model shared a similar decision tree structure to the summation score model, all the way to the very end node. In addition, household income and number of expenditures reported were found to be related to respondent burden. These burden proxy indicators, as well as record usage, are interesting to be pay attention when researchers try to model burden.

4.3 Nonparametric Recursive Partitioning Models for CE 2012 Burden Proxy Indicators with 4 Burden Questions

The latest version of R$\{rpms\}$ package (2019 release) had been implemented on the 2012 Q4 - 2013 Q3 CE burden data with 4 burden questions.

We also tried to look into how burden measures are related to those burden proxy indicators in the CE 2012 data with only 4 burden questions similar to 2017. First, we used recursive partitioning to split for the single item burden question among the burden proxy indicators after taking into account the complex survey design (Figure 4). For the single item burden question, we saw that respondents with busy/logistics, privacy/government doorstep concerns reported the highest levels of burden. Those who expressed no or other concerns originally but conditioning on interview mode, mortgage and income at \$54,600

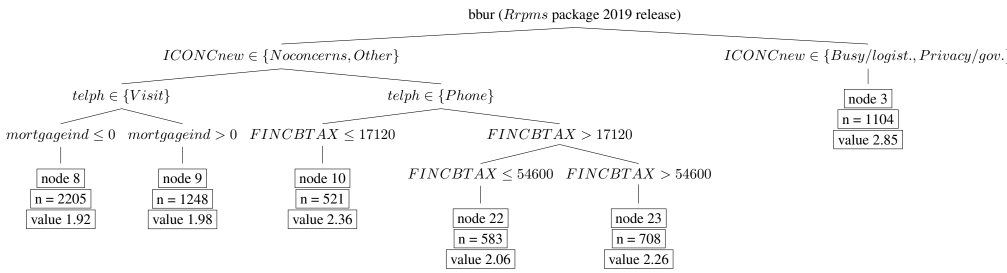


Figure 4: 2012 Q4 - 2013 Q3 Single Burden Question (bbur, R$\{rpms\}$ package 2019 release)

(e.g. perhaps close to some kind of median household income threshold) reported a moderate amount of burden (where the end node average of perceived burden question ranges 1.92 – 2.26). For households who expressed no or other concerns, were visited by person and income $\leq \$17,120$ reported the lowest levels of burden. Those who reported income $> \$17,120$ had a higher burden level than those who reported income is $\leq \$17,120$. The same classification tree was produced by the recursive partitioning model when we converted information booklet and records usage to indicators.

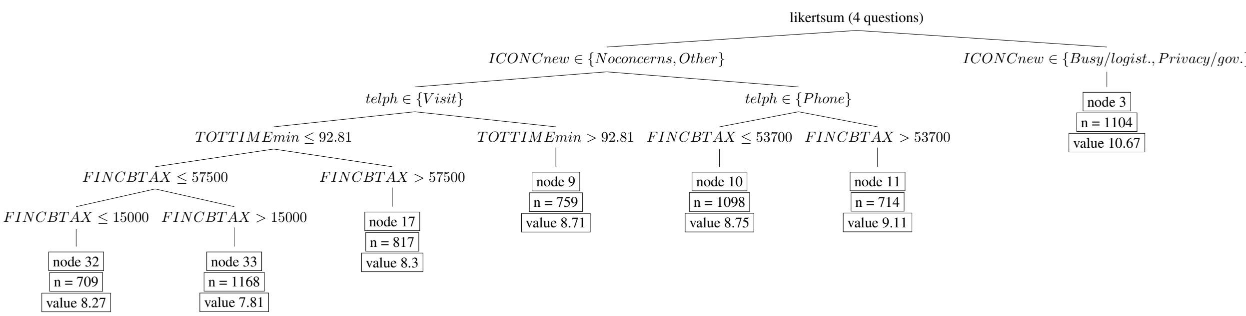


Figure 5: 2012 Q4 - 2013 Q3 Likert Scales Summation Scores (likertsum, 4 Questions)

Second, we implemented recursive partitioning to split for the Likert Scale summation scores on the burden proxy indicators after taking into account the complex survey design (Figure 5). The tree looks a little bit different here. Those respondents with busy/logistics, privacy/government doorstep concerns reported the highest levels of burden, respondents who expressed no or other concerns, were visited by person, interviewed ≤ 92.81 minutes and income $> \$15,000$ reported the least level of burden, others expressed moderate level of burden (where the Likert Scale summation scores end node average ranges 8.27 – 9.11). Once again, when we converted information booklet and records usage to indicators, the recursive partitioning model produced the same classification tree.

Finally, we implemented another recursive partitioning to model relationships between the burden proxy indicators and the composite burden index scores after taking into account the complex survey design (Figure 6). The exactly same decision tree structure was produced as the Likert Scale summation scores model except the end node averages were different. And the same classification tree was produced by the recursive partitioning model when we converted information booklet and records usage to indicators.

Among all three burden measures in 2012 models where ONLY 4 burden questions similar to 2017 were used (Table 2), we too saw commonalities. Once again, whether a respondent had door step concerns or not had the strongest association with the burden measures. Respondents with busy/logistics, privacy/government door step concerns con-

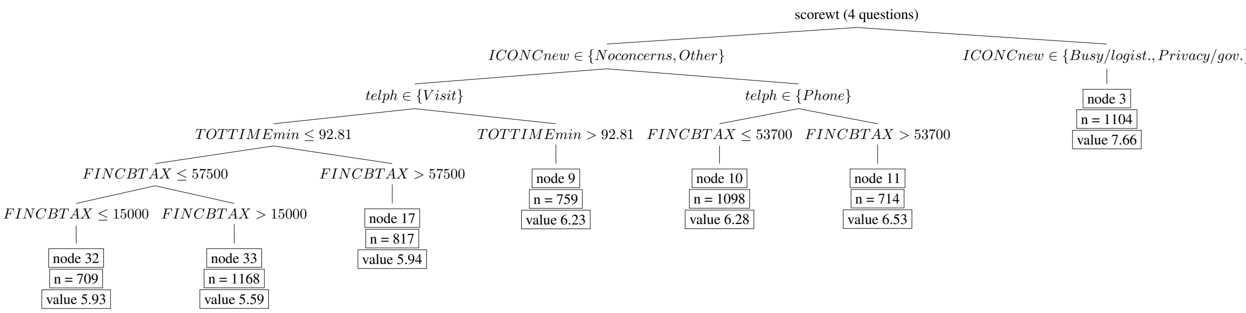


Figure 6: 2012 Q4 - 2013 Q3 Composite Burden Index Scores (scorewt, 4 Questions)

2012 Q4 – 2013 Q3 (4 burden questions)	Single Burden Question	Likert Scales Summation Scores	Composite Burden Index Scores
First Split	Door Step Concerns vs. Not, Other	Door Step Concerns vs. Not, Other	Door Step Concerns vs. Not, Other
Second Split	Personal Visit vs. Telephone	Personal Visit vs. Telephone	Personal Visit vs. Telephone
Third Split	Mortgage vs. Not, Income $> \$17,120$ vs. $\leq \$17,120$	Interview Length $>$ 92.81 min. vs. $\leq$ 92.81 min., Income $> \$53,700$ vs. $\leq \$53,700$	Interview Length $>$ 92.81 min. vs. $\leq$ 92.81 min., Income $> \$53,700$ vs. $\leq \$53,700$
Fourth Split	Income $> \$54,600$ vs. $\leq \$54,600$	Income $> \$57,500$ vs. $\leq \$57,500$	Income $> \$57,500$ vs. $\leq \$57,500$
Fifth Split		Income $> \$15,000$ vs. $\leq \$15,000$	Income $> \$15,000$ vs. $\leq \$15,000$

Table 2: CE 2012 Recursive Partitioning for Modeling Survey Respondent's Burden Perception Decision Tree Comparison - 2

sistently reported higher levels of burden. Whether the interview mode is personal visit or telephone had the second strongest association with the burden measures. Within the single burden question, whether a respondent had a mortgage or not drove the third split, and the household income was found to be related to respondent burden at the fourth split. In the other two measures, the Likert Scales summation scores and the composite burden index scores, interview length had the next strongest association, and household income was found to be associated with burden in those two measures as well. The Likert Scales summation score model shared an identical structure to the composite burden index scores model, and the single burden question model is not that far away. Also, mortgage indicator, interview length and household income were found to be related to respondent burden.

4.4 Nonparametric Recursive Partitioning Models for CE 2017 Burden Proxy Indicators

We used the latest version of Recursive Partitioning for Modeling Survey Data R$\{rpms\}$ package (2019 release) to analyze the 2017 Q2 - 2018 Q1 CE burden data (with 4 burden questions).

We looked into how burden measures are related to those burden proxy indicators in the CE 2017 burden data. First, we used recursive partitioning to split for the single item burden question among the burden proxy indicators after taking into account the complex

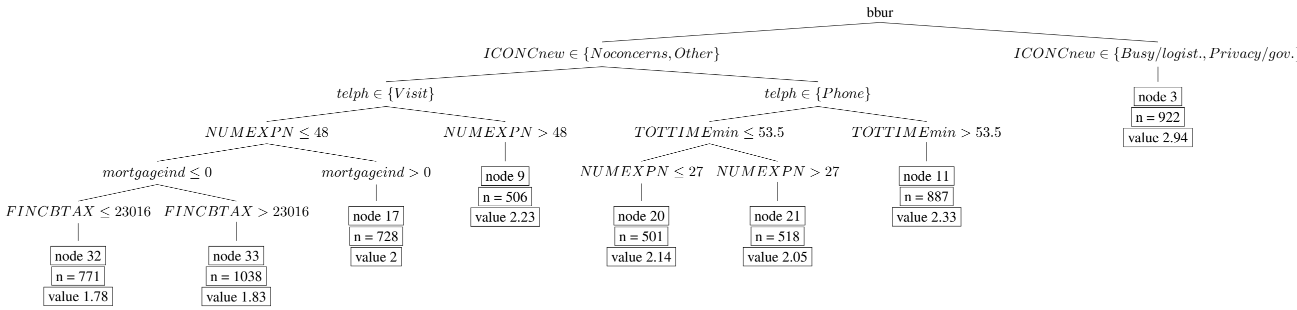


Figure 7: 2017 Q2 - 2018 Q1 Single Burden Question (bbur)

survey design (Figure 7). For the single item burden question, we saw that respondents with busy/logistics, privacy/government doorstep concerns reported the highest levels of burden. Those who expressed no or other concerns originally but conditioning on interview mode, number of expenditures, interview length and mortgage reported a moderate amount of burden (where the end node average of perceived burden question ranges 2 – 2.33). For households who expressed no or other concerns, were visited by person, had number of expenditures ≤ 48, without mortgage and income $\leq \$23,016$ reported the lowest levels of burden. Those who reported income $> \$23,016$ had a slightly higher burden level than those who reported income $\leq \$23,016$. We repeated the same recursive partitioning model analysis with information booklet and records usage were converted to indicators, the resulting classification tree was identical, except for households who expressed no or other concerns, were visited by person, had number of expenditures ≤ 48 and without mortgage reported the lowest levels of burden.

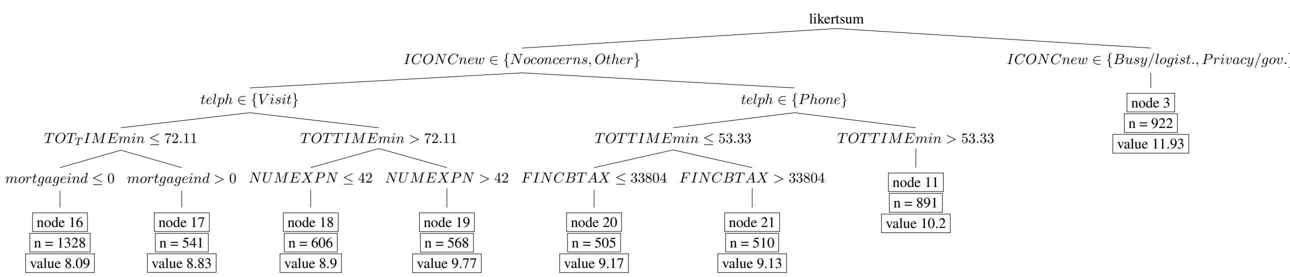


Figure 8: 2017 Q2 - 2018 Q1 Likert Scales Summation Scores (likertsum)

Second, we implemented recursive partitioning to split for the Likert Scale summation scores on the burden proxy indicators after taking into account the complex survey design (Figure 8). Once again, those households with busy/logistics, privacy/government doorstep concerns reported the highest levels of burden, respondents who expressed no or other concerns, were visited by person, interviewed < 72.11 minutes and without mortgage reported the least burden, others expressed moderate level of burden (where the Likert Scale summation scores end node average ranges 8.83 – 10.2). We repeated the same recursive partitioning model analysis with information booklet and records usage were converted to indicators, those with busy/logistics, privacy/government doorstep concerns reported the highest levels of burden, respondents who expressed no or other concerns, were visited by person, interviewed < 72.11 minutes reported the least burden, others expressed moderate level of burden (where the Likert Scale summation scores end node average ranges 8.9 – 10.2).

Finally, we implemented another recursive partitioning to model relationships between

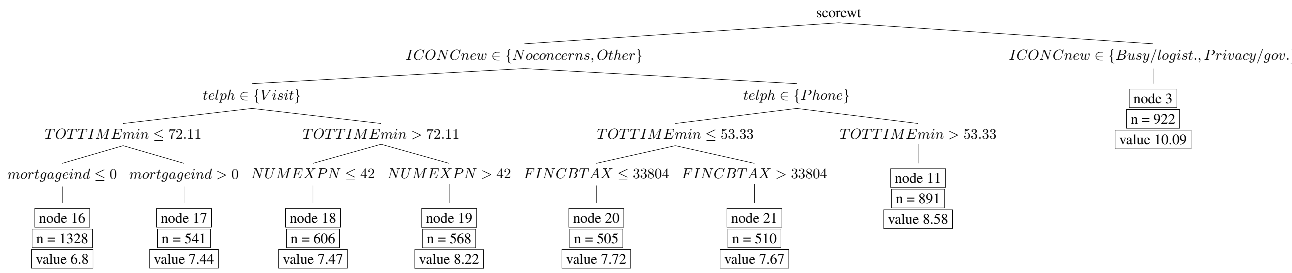


Figure 9: 2017 Q2 - 2018 Q1 Composite Burden Index Scores (scorewt)

the burden proxy indicators and the composite burden index scores after taking into account the complex survey design (Figure 9). The exactly same decision tree structure was produced as the Likert Scale summation scores model except the end node averages are different. We repeated the same recursive partitioning model analysis with information booklet and records usage were converted to indicators, we obtained the exactly same decision tree structure as the Likert Scale summation scores model except the end node averages are different.

2017 Q2 – 2018 Q1	Single Burden Question	Likert Scales Summation Scores	Composite Burden Index Scores
First Split	Door Step Concerns vs. Not, Other	Door Step Concerns vs. Not, Other	Door Step Concerns vs. Not, Other
Second Split	Personal Visit vs. Telephone	Personal Visit vs. Telephone	Personal Visit vs. Telephone
Third Split	Number of Expenditures > 48 vs. ≤ 48, Interview Length > 53.5 min. vs. ≤ 53.5 min.	Interview Length: $>$ 72.11 min. vs. ≤ 72.11 min., > 53.33 min. vs. ≤ 53.33 min.	Interview Length: $>$ 72.11 min. vs. ≤ 72.11 min., > 53.33 min. vs. ≤ 53.33 min.
Fourth Split	Mortgage vs. Not, Number of Expenditures > 27 vs. ≤ 27	Mortgage vs. Not, Number of Expenditures > 42 vs. ≤ 42, Income $> \$33,800$ vs. $\leq \$33,800$	Mortgage vs. Not, Number of Expenditures > 42 vs. ≤ 42, Income $> \$33,800$ vs. $\leq \$33,800$
Fifth Split	Income $> \$23,020$ vs. $\leq \$23,020$		

Table 3: CE 2017 Recursive Partitioning for Modeling Survey Respondent's Burden Perception Decision Tree Comparison

Looking across all three burden measures in 2017 models (Table 3), we saw that there are commonalities. Whether a respondent had door step concerns or not had the strongest association with the burden measures. Respondents with busy/logistics, privacy/government door step concerns consistently reported higher levels of burden, which suggesting that the topic of the concern, e.g. if the respondent was worried about privacy or logistics, was related to the level of burden they would report. Whether the interview mode is personal visit or telephone had the second strongest association with the burden measures. Within the single burden question, reported number of expenditures and interview length drove the third split, whether a respondent had a mortgage or not and number of expenditures reported drove the fourth split. And household income was found to be related to respondent burden at the fifth split. In the other two measures, the Likert Scales summation scores and the composite burden index scores, interview length had the next strongest association. Mortgage indicator, reported number of expenditures and household income was found to be associated with burden in those two measures as well. As for the burden measures themselves, the Likert Scales summation score model shares an identical structure to the composite burden index scores model, and the single burden question model is not so far away. Mortgage indicator, number of expenditures reported and household income were found to be related to respondent burden. These burden proxy indicators, as well as information booklet and record usage, are interesting to pay attention when modeling burden.

5. Summary and Future Steps

For all three burden measures, there were a few burden proxy indicators were repeatedly identified to be associated with burden. A tiered burden proxy indicators can be listed as the following:

1. Tier 1: Door step concerns.
2. Tier 2: Interview mode, converted refusal.
3. Tier 3: Number of expenditures, interview length, mortgage indicator, records usage.
4. Tier 4: Household income before tax.

In future studies, we would recommend that these indicators should be further explored as they may be useful in understanding respondent behaviors that could be caused by burden (e.g., attrition, data quality). Interestingly in the 2012 CE burden data, we see that in the case of the single burden question and the composite index scores that reporting more expenditures was associated with lower burden. This is the opposite of what one might expect and points to a difficulty in understanding the relationship between burden and these proxy indicators. One may ask that do variables like survey length and number of questions lead to greater burden or do respondents who feel burdened engage in behaviors like satisficing or under reporting to speed the interview along? These lead to very different conclusions about how we interpret these proxy indicators. In particular interview mode, do respondents feel burdened by a personal visit and opt for telephone to alleviate the burden, or do they feel burden because they opted for the telephone? And for some of these decision tree splits, it will be hard to identify the direction of the association. Following this study, we would like to extend this study to seek:

1. Whether new burden proxy indicators could be included in the recursive partitioning for modeling survey data?
2. Whether the recursive partitioning model could be further explored to prediction error? (This would require inputs from CE program experts).
3. Explore the possibility of extending to regression trees and random forest modeling, e.g. predicting a respondents anticipated burden perception (in terms of a single burden response or burden scores) from burden proxy indicators and etc.

References

Consumer expenditure survey interview dictionary phase 3 2013. Technical report, 2013.

Consumer expenditure survey interview questionnaire (ceq) - 2017. Technical report, 2017.

K. A. Bollen, J. L. Glanville, and G. Stecklov. Socioeconomic status and class in studies of fertility and health in developing countries. *Annual Review of Sociology*, 27:153185, 2001.

K. A. Bollen, J. L. Glanville, and G. Stecklov. Economic status proxies in studies of fertility in developing countries: Does the measure matter? *Population Studies*, 56(1):81–96, 2002.

S. Fricker, C. Kreisler, and L. Tan. Exploratory research on the construction of a summary index for respondent burden. Technical report, Bureau of Labor Statistics, 2012.

Mingzhao Hu. Income and expenditure in us households: a multivariate analysis of consumer expenditure fmli161 dataset. 2018.

Donald A. Jackson. Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology*, 74(8):22042214, 1993.

R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis, Sixth Edition*. Pearson Education Inc., 6 edition, 2007.

Stanislav Kolenikov and Gustavo Angeles. The use of discrete data in pca: Theory, simulations, and applications to socioeconomic indices. page 59, 2004.

Stanislav Kolenikov and Gustavo Angeles. Socioeconomic ocioeconomic status measurement with discreteproxy variables: Is principal component analysis a reliable answer? *Review of Income and Wealth*, (1):128–165, 2009.

B. Kopp, B. McBride, and L. Tan. An exploratory study on the association of doorstep concerns with three survey quality measures for the ce interview survey. Technical report, Bureau of Labor Statistics, 2013.

Wei-Yin Loh. Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1(1):14–23, 2011.

Ulf Olsson. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4):443460, 1979.

K. Pearson and E. S. Pearson. On polychoric coefficients of correlation. *Biometrika*, 14 (1/2):127156, 1922.

Pedro R. Peres-Neto, Donald A. Jackson, and Keith M. Somers. How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics Data Analysis*, 49(4):974–997, 2005.

A. C. Rencher. *Methods of Multivariate Analysis, Second Edition*. Wiley-Interscience, 2 edition, 2003.

Daniell Toth. Recursive partitioning for modeling survey data. page 13, 2019.

Ting Yan. Response burden: What predicts it and what is its impact on response quality? Technical report, Bureau of Labor Statistics, 2015.

Daniel K Yang. Numerical impact of a simple random subsample on consumer spending for children. *Proceedings of the Federal Committee on Statistical Methodology*, 2013.

Daniel K Yang. Consumer expenditure interview survey composite respondent burden index final report. Technical report, Bureau of Labor Statistics, 2015.