Assessing Nonresponse in a Longitudinal Establishment Survey Using Regression Trees March 2018

Morgan Earp[1], Daniell Toth[1], Polly Phipps[1], and Charlotte Oslund[1]

This article introduces and discusses a method for conducting an analysis of nonresponse for a longitudinal establishment survey using regression trees. The methodology consists of three parts: analysis during the frame refinement and enrollment phases, common in longitudinal surveys; analysis of the effect of time on response rates during data collection; and analysis of the potential for nonresponse bias. For all three analyses, regression tree models are used to identify establishment characteristics and subgroups of establishments that represent vulnerabilities during the data collection process. This information could be used to direct additional resources to collecting data from identified establishments in order to improve the response rate.

Key words: Bureau of Labor Statistics; job openings and Labor Turnover Survey; response rates; panel survey; rpms.

1. Introduction

Government agencies administer many voluntary surveys in order to collect and disseminate official statistics that are important for informing policy and decision-making. Though the response rates for these surveys tend to be higher than surveys collected by other organizations (Heberlein and Baumgartner 1978; Paxson et al. 1995; Holbrook et al. 2007), it is important that the agencies maintain high response rates to help minimize the threat of nonresponse bias, loss of stakeholder confidence and the potential to inflate variance in survey estimates (Little and Vartivarian 2005; Office of Management and Budget 2006). Maintaining high response rates, especially in longitudinal surveys, requires substantial effort and resources (Freedman et al. 1980; Watson and Wooden 2009).

Traditionally, survey methodologists have used multiple approaches to increase response rates, such as offering incentives to respondents, tailoring notification letters, increasing contact attempts, or providing alternative data collection modes (Dillman 2000; Dillman et al. 2009; Groves et al. 2002). While considerable research has been conducted on the causes and effects of nonresponse in longitudinal surveys, the studies have focused on household surveys (Atrostic et al. 2001; De Heer and De Leeuw 2002; Groves 2006; Groves et al. 2002; Lepkowski and Couper 2002; Singer 2002; Watson and Wooden

[1] Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington DC, 20212-0001, U.S.A. Emails: earp.morgan@bls.gov, toth.daniell@bls.gov, phipps.polly@bls.gov, and oslund.charlotte@bls.gov

2009), and, in comparison, very few have focused on establishment surveys (Janik and Kohaut 2012; Seiler 2010; Smaill 2012).

Similar to the household survey longitudinal nonresponse literature, the existing studies on establishment surveys communicate the importance of understanding characteristics and patterns of nonresponse to improve survey quality, often through possible interventions with nonresponding groups. Thus, identifying the characteristics of nonrespondents is a key objective. The available characteristics used in the studies vary, but they all attempt to measure business characteristics, such as employment size, industry, and multi-establishment status, often utilizing auxiliary data available from the sampling frame or past panel participation. Also, each study analyzes the effect of response propensity over time differently, depending on the data availability and study definitions. For example, Seiler (2010) examines the long running effects of time over a 15-year period for the IFO Business panel survey in Germany, finding that the risk of nonresponse decreases over time, and that employment size and industry predict response, as well as economic conditions. Janik and Kohaut (2012) investigate nonresponse in the Institute for Employment Research establishment panel survey in Germany. The authors note that willingness to participate in the panel survey for the first time is much lower than response for establishments that have participated at least once. They analyze time by investigating the response behavior of establishments in the 2006 wave that previously participated in the survey. Their findings show the importance of establishment characteristics in predicting refusal to participate, including employment size and independent company status. They also find several characteristics associated with past panel experience important: duration of past participation, recent refusal, and a change in interviewer from the previous wave. Smaill (2012), analyzing the New Zealand Retail Trade Survey, uses mixture modeling to investigate how the initial response in the first six months of the survey predicts the probability of response over the subsequent 36 months, finding that previous response behavior is the most consistent predictor of future response.

Response propensity is most often modeled using a logistic regression model. There are some disadvantages with using this approach (Phipps and Toth 2012). First, the response propensity curve may not be logistic, which could lead to inaccurate conclusions. Second, selecting the variables to use in the logistic model can be challenging, especially if interactions between available model variables are to be considered. Additionally, the inclusion of interactions in the model can make it hard to interpret the effect of a variable on the response rate.

Indeed, the existing longitudinal establishment survey studies rarely consider interaction effects when analyzing response propensity. Yet, interactions between characteristics that define businesses have been found to be important in understanding nonresponse in establishment surveys. For example, in their analysis of the Occupational Employment Survey of the U.S. Bureau of Labor Statistics (BLS), Phipps and Toth (2012) find that small, independent establishments have a high response propensity, while large establishments are less likely to respond, particularly in certain industries and larger metropolitan areas. As they note, variable and parametric model selection and interpretation in nonresponse analyses using many continuous and categorical variables with interactions is difficult. As an alternative, regression tree methods can be utilized to easily identify and interpret associations between establishment characteristics and response propensity.

Morgan and Sonquist (1963) proposed the idea of recursive partitioning in order to detect the presence of interaction effects between variables in survey data. Several methods have used this idea to build tree models that automatically produce a partitioning of the data into groups based on characteristics of the observations. The resulting model is a linear function of characteristics, where characteristics are defined by successive splits on the included variables. Thus, any interactions among the variables are automatically detected and included in the model and their effects on the response rate are easily interpreted.

The regression tree models used in this analysis are created using the R Package "Recursive Partitioning for Modeling Survey Data" (rpms) (Toth 2017). The rpms function builds a tree recursive partitioning using a linear model fit on time in survey for each node of the tree like those proposed in Su et al. (2004), using an unbiased variable selection procedure adapted from a method proposed by Hothorn et al. (2006) to account for sample design. Of the methods examined by Lohr et al. (2015), the method proposed by Hothorn and colleagues was found to be the best for modeling response propensity. Since the rpms algorithm uses a stopping rule based on a design consistent hypothesis test to decide on the number of splits, no pruning is required. All the trees built for the application used in our analysis use a *p*-value of .005.

Since surveys used to collect information for official statistics usually attempt to collect data from a wide variety of establishments, it is important to be able to identify establishment characteristics associated with and assess the effect of lower response rates. In this article, we propose a method for carrying out nonresponse analysis of a longitudinal establishment survey using tree models.

As a motivating example, we analyze the response rate of the 2012 Job Openings and Labor Turnover Survey (JOLTS). This longitudinal establishment survey conducted by the BLS collects data every month from a panel of establishments to provide national estimates of job openings (jobs not filled on the last business day of the month), hires (additions to payroll during the month), and total separations (from the payroll during the month; the sum of quits, layoffs, and discharges) in the United States. The JOLTS surveys approximately 16,000 establishments per month from all 50 states and includes both the government and private sectors. Once selected, an establishment remains in the survey for 24 months. A new panel of establishments is rotated in and an old panel is rotated out every month. The JOLTS selects establishments for each panel using a sample stratified by ownership (private or public), region, industry sector, and employment size class.

Our analysis focuses on three aspects of survey nonresponse, covering time in two ways: the time periods associated with identification and recruitment phases that are early in the survey process, as well as a 24-month longitudinal data collection phase. Using regression trees, we identify establishment characteristics associated with relatively high failure rates for each of these two phases separately. First, we consider the loss of selected units as they are being identified and contacted for enrollment as potential survey participants. These administrative processes, which take place before data collection, are often present in longitudinal surveys and could cause selected units to be removed before data collection due to failure to locate or successfully enroll the unit for survey participation.

Second, the analysis focuses on the effect of time on response rates during the data collection phase. For longitudinal surveys, selected establishments may be more or less

likely to become nonresponding units the longer they are in the survey. This part of the analysis models the response rate as a linear function of time-in-survey and uses regression trees to group establishments based on establishment characteristics with similar relationships between response rate and time.

The third part of the analysis focuses on the potential for nonresponse bias. If establishments that do not respond and those that do respond are systematically different on key outcome variables, nonresponse can raise concerns about the bias and precision of survey estimates (Kreuter et al. 2010). Thus, an assessment of possible nonresponse bias is important when carrying out a comprehensive nonresponse analysis for longitudinal surveys. The subdomains defined by terminal nodes created in the second part of our analysis could be used to define adjustment cells for nonresponse weight adjustments. In this part of the analysis, we study the effectiveness of such adjustment cell weighting on reducing nonresponse bias by systematically comparing differences between respondents and nonrespondents conditioning on the establishment characteristics identified in the regression tree analysis.

Section 2 describes the data set and results of modeling the response rates separately for the address refinement and enrollment phases of the survey. In Section 3, we analyze the effect of time-in-survey on response rates during the data collection phase. In Section 4, we present an analysis of potential nonresponse bias for two key survey outcomes (hires and separations) using proxy variables from the BLS Quarterly Census of Employment and Wages (QCEW). Section 5 contains a discussion of the main results of the analysis.

2.   Identifying Characteristics of Nonresponse at Each Phase of Data Collection

During longitudinal data collection, there are different phases in the collection process when nonresponse can occur. Identifying the group characteristics of nonresponding establishments during these phases is important and can lead to the development of targeted strategies to address nonresponse. The JOLTS data collection process is broken up into three phases. Each sampled establishment starts with address refinement, where the BLS data collection center attempts to verify the address and contact information for the establishment. After the establishment's address is verified, they are moved into enrollment, where the establishment is recruited to participate in the survey. Establishments that have agreed to participate in the survey begin data collection (see Figure 1). An establishment may be classified as a nonrespondent during any one of these
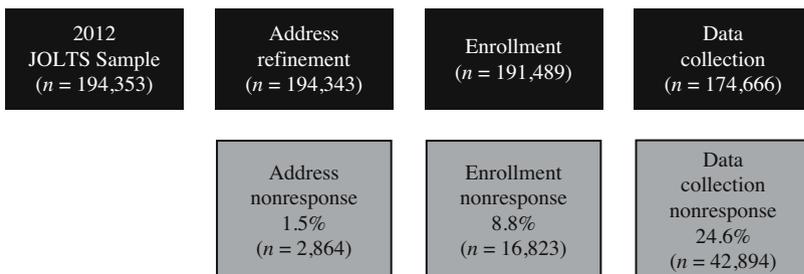

Fig. 1.   JOLTS response process from the time of sampling through data collection, displaying all potential opportunities for nonresponse.

Table 1.   Establishment characteristics.

Variables	Description
Ownership	Federal, State, Local government, or Private
Private	Government or Private
Sector	Industry Sectors (20)
Super Sector	Groupings of Industry Sectors (11)
White-Collar Services	Information, Professional & Business Services, and Financial Activities vs. All Other Super Sectors
Certainty Unit	Certainty Unit or Not Certainty Unit
Multi-Establishment	Multi-Establishment business or a Non Multi-Establishment business
Size Class	1: 1-9, 2: 10-49, 3: 50-249, 4: 250-999, 5: 1,000-4,999, or 6: 5,000 + Employees
Time in Survey	Number of Months in Survey (0 = First Month in Survey)

phases and any given month of data collection. By modeling the phases separately, we can identify and compare establishment characteristics associated with low response rates during initial contact, recruitment, and data collection; which in turn, allows BLS to make the best use of resources throughout the data collection process.

We use regression tree models to identify establishment characteristics associated with low response for each phase of JOLTS. We utilize available characteristics that have been found to be associated with response propensity, such as establishment size, industry, and whether the establishment is independent or is part of a larger firm (Tomaskovic-Devey et al. 1994; Davis and Pihama 2009; Seiler 2010; Janik and Kohaut 2012; Phipps and Toth 2012). These types of establishment characteristic variables are usually contained in the sampling frames of government agencies and used for sample selection; for this study we use auxiliary data from QCEW, which is the sample frame for JOLTS. The variables we consider for the analyses are shown in Table 1.

The data for our study was originally made up of 207,567 JOLTS observations (from 24,158 establishments across 36 panels (eight percent of which were large, influential units selected into the sample with certainty) collected from January through December of 2012. An observation in the dataset is a record of the establishment and its reported values for each month it was in the survey during our twelve-month study period. We excluded observations from establishments that were out of business ($n = 12,734$) and/or post offices ($n = 480$). Post offices were excluded since the postal service provides data to the JOLTS as a census by state. We excluded a small number of observations with no record of any contact or collection attempt, since we are interested in classifying establishments that do not respond given the opportunity. After removing these records, the final dataset used for analysis consisted of 194,353 observations (from 22,776 establishments).

2.1.   Address Refinement

During address refinement, BLS obtains and verifies the contact information of sampled establishments by telephone. Establishment contact information is provided by states and is included as part of the sample frame. By the time the frame is used to draw the sample

for the JOLTS, this information is at least twelve months old. Most sampled establishments have some known contact information, but there are a few with little or no information available. Even in the case where contact information is provided, the quality and extent vary. A street address is provided for most establishments, and in some cases a telephone number, but for the majority there is no contact name. Also, when contact information is available, it may be out of date, given the twelve-month lag time. If the contact information for an establishment cannot be verified or found by the BLS during the address refinement phase, these establishments are considered nonrespondents.

While the overall response rate for address refinement is high at 98.5 percent, the regression tree model, shown in Figure 2, identifies three subgroups with significantly lower response rates, which are likely candidates for additional effort or prioritization during this phase. Note that the tree graphs used in this article use a common numbering convention for the nodes. Starting with the root node at the top (Node 1), given node numbered n, the left-child is $2n$ and the right child is $2n + 1$. This ensures that every node of a tree has a unique number label.

Figure 2 identifies the first two groups with significantly lower response rates as federal government multi-establishments (94% of federal government establishments are classified as multi-establishments). In Node 20, federal establishments with 250–999 employees have an average response rate of 80 percent during address refinement, well below the mean of 98.5 percent. In addition, smaller establishments with 1–249 employees, as well as extremely large establishments with 5,000 or more employees, have a low average response rate of 90 percent (Node 21). The JOLTS program indicates this is likely due to two factors: 1) it can be hard to find the correct building for large federal agencies, and 2) the sampling frame is often missing the addresses and specific locations of the federal establishments. The third group with lower response rates during address refinement includes multi-establishments of 50–999 employees in the retail trade sector with an extremely low response rate of 58 percent (Node 8).

2.2. *Enrollment*

Once an establishment's address is verified, it moves into the enrollment phase; otherwise it does not continue into the subsequent phases (see Figure 1). Enrollment is the second phase of JOLTS, when interviewers contact establishments and solicit their participation in the survey. The goal of the enrollment phase is to gain consent from the establishment to participate in the JOLTS program, which involves providing monthly employment and turnover data. During enrollment, each establishment is mailed an "introductory packet" explaining the survey and the importance of their participation; these packets include a customized cover letter, the JOLTS brochure, a business information guide, a fact sheet explaining how the data are used, and the JOLTS survey form. About three to five days after the introductory packet is mailed out, interviewers follow-up by calling the establishment to solicit participation (Bureau of Labor Statistics 2015).

The response rates modeled during the enrollment phase are conditional on establishments responding during the address refinement phase. The conditional response rate during enrollment at 91.2 percent is still quite high, but lower than that of address refinement. The model shown in Figure 3 identifies two groups with significantly lower

Address refinement

$multi \in \{1\}$ — $multi \in \{0\}$

$ret \in \{retail\}$ — $ret \in \{other\}$ — node 3 / value 1

$size \in \{4,5\}$ — node 8 / value 0.58

$size \in \{1,2,3,6\}$

$size \in \{3\}$ — node18 / value 0.96

$size \in \{1,2,6\}$ — node19 / value 1

$fed \in \{federal\}$ — $fed \in \{other\}$ — node 11 / value 0.99

$size \in \{4,5\}$ — node 20 / value 0.8

$size \in \{1,2,3,6\}$ — node 21 / value 0.9

Legend:

Multi:
0: Non-Multi-Establishment, 1: Multi-Establishment

Cert:
0: Non-Certainty Unit, 1: Certainty Unit

Size:
1: 1–9, 2: 10–49, 3: 50–249, 4: 250–999, 5: 1,000–4,999, or 6: 5,000+

Establishment Employees

Fig. 2. Address refinement response rate tree model.

Enrollment

size ∈ {3, 4, 5}

private ≤ 0

own ∈ {F, S}

node 8 | value 0.93

own ∈ {L}

size ∈ {4, 5}

node 18 | value 0.96

size ∈ {3}

node 19 | value 0.97

multi ∈ {1}

wc ∈ {1}

node 20 | value 0.81

wc ∈ {0}

node 21 | value 0.85

private > 0

multi ∈ {0}

size ∈ {4, 5}

wc ∈ {1}

node 44 | value 0.77

wc ∈ {0}

size ∈ {5}

node 90 | value 0.87

size ∈ {4}

node 91 | value 0.9

size ∈ {3}

wc ∈ {1}

node 46 | value 0.91

wc ∈ {0}

node 47 | value 0.93

size ∈ {1, 2, 6}

multi ∈ {1}

private ≤ 0

node 12 | value 0.96

private > 0

wc ∈ {1}

node 26 | value 0.88

wc ∈ {0}

node 27 | value 0.91

multi ∈ {0}

wc ∈ {1}

node 14 | value 0.94

wc ∈ {0}

node 15 | value 0.96

Legend:

Multi:
0: Non-Multi-Establishment, 1: Multi-Establishment

Size:
1: 1–9, 2: 10–49, 3: 50–249, 4: 250–999, 5: 1,000–4,999, or 6: 5,000+ Employees

Private:
0: Not-Privately Owned, 1: Privately Owned

WC:
0: Not White Collar Service Sector, 1: White

Own:
F: Federal, S: State, L: Local

Fig. 3. *Enrollment response tree model. Response rates during enrollment are conditional on address refinement response.*
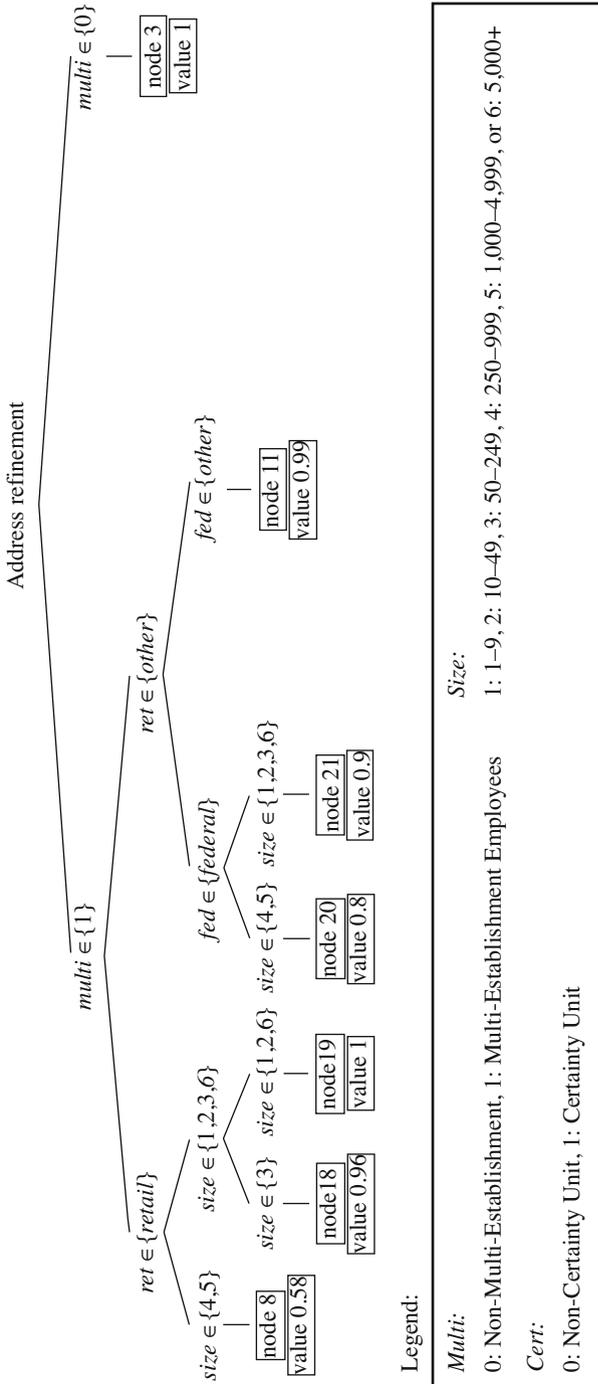
enrollment response rates that may be candidates for resource allocation during this phase. The first group includes establishments that are part of multi-establishment firms, privately owned, and in the white-collar service industries (Node 20), with a conditional response rate of 81.4 percent – nearly ten percentage points below the average response rate at this phase. The second group includes establishments that are not part of a multi-establishment firm, with 250 employees or more, and in white-collar services, with a conditional response rate of 77.9 percent (Node 44)– 13.3 percentage points below the average response rate at this phase.

Similar to address refinement, the enrollment model exhibits a significant relationship between establishment ownership type and industry; in this case, private ownership and white-collar service sectors have lower response rates. Also similar to address refinement, the enrollment model shows that being part of a multi-establishment firm is linked to lower response rates. Both models also exhibit a relationship between larger establishment sizes (≥ 250 employees) and lower response rates. The association between large and multi-establishments and lower response rates is typical in BLS establishment surveys, and establishments in the white-collar service industry sectors have been shown to have low response rates in other BLS surveys (Phipps and Toth 2012).

3. Data Collection

After an establishment is successfully enrolled in the survey, the interviewer schedules an appointment and moves the unit into the data collection phase, at which point, the interviewer attempts to collect the requested data. Establishments are asked every month to report the number of employees, hires, total separations, quits, layoffs, discharges, and job openings over a 24-month period, except for certainty units that remain in the survey indefinitely. For the first three months of data collection, most establishments complete the survey via computer-assisted telephone interviewing (CATI); after that time the data collection is carried out via the web. As of 2012, about 60 percent of responding establishments supplied their data through the web, but if the establishment insists on another mode of data delivery, they are usually accommodated. Offering a variety of collection methods helps accommodate respondent preferences, which is important, since the JOLTS is a voluntary survey program.

The next phase of the analysis considers establishment response rates conditional on having agreed to participate in the survey during the enrollment phase. In order for an establishment to be counted as a respondent in this phase, we required them to have provided data that was used for the survey estimates, so at least one of hires, total separations, or job openings was requisite. Almost every establishment surveyed for JOLTS either provided both hires and total separations or neither. Also establishments that provided job openings almost always provide hires and total separations. Therefore, the tree model for response rates in data collection presented in this section is the same tree we get if we model response rates of hires or separations separately.

Since the data collection phase typically lasts 24 months, we consider the effect of time in survey on the response rate. In this section we model the relationship between establishment characteristics and the effect of time on the response rate. This allows us to investigate establishment characteristics associated with nonresponse over time, and

identify respondents that could be subject to attrition over time (Seiler 2010).Though the rpms package as of this writing only allows linear models in the end node, there could be potential for further analysis if different functions were possible, such as logistic and proportional hazards models.

Figure 4 shows the response rates for certainty and non-certainty units over the amount of time they have been in the survey. While the response rate does not have an obvious linear trend for either certainty or non-certainty units in aggregate, we can use rpms to assess if the relationship between time in survey and response rates varies by establishment characteristics. The model partitions the establishments based on the fit of the linear model, relating response rate to time in survey. This allows us to create homogeneous groups of establishments not only in terms of response rates, but also with regard to the effect of time, which helps in understanding longitudinal survey response.

We modeled the response rates over time in survey for certainty and non-certainty units separately. The algorithm did not find any linear effect for time in survey on the response rate for certainty units, resulting in a tree with no splits. However, for non-certainty units, differences in the effect of time on the response rate became clear through the regression tree model. The effect of time on the response rate varied depending on establishment characteristics. Figure 5 shows the model obtained from the rpms algorithm with the linear equation $r(t)$, the response rate at time t, as well as the initial response rate $r(0)$, and the predicted response rate for their last month in the survey $r(23)$ given at the bottom of each node.

Starting at the root node of the tree, we see that the algorithm determined that size class of the establishment had the biggest impact on the relationship between response rates and time, separating out establishments in the two smallest size classes, which had the highest
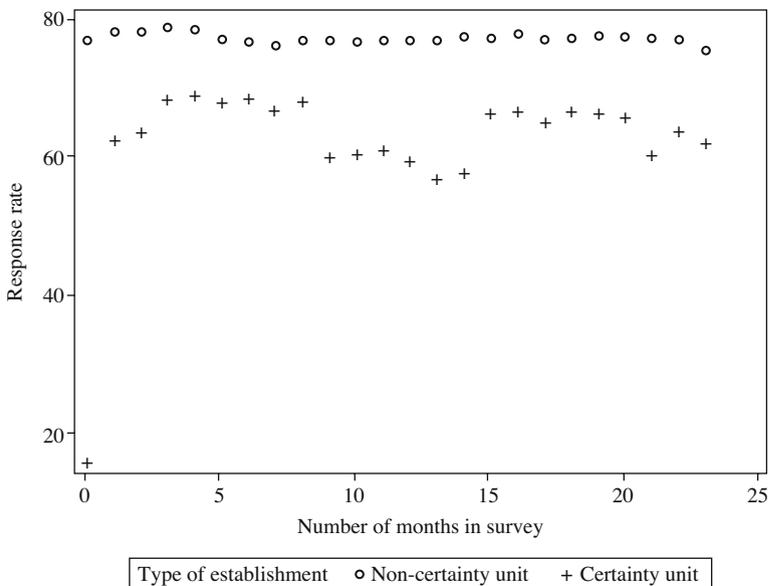

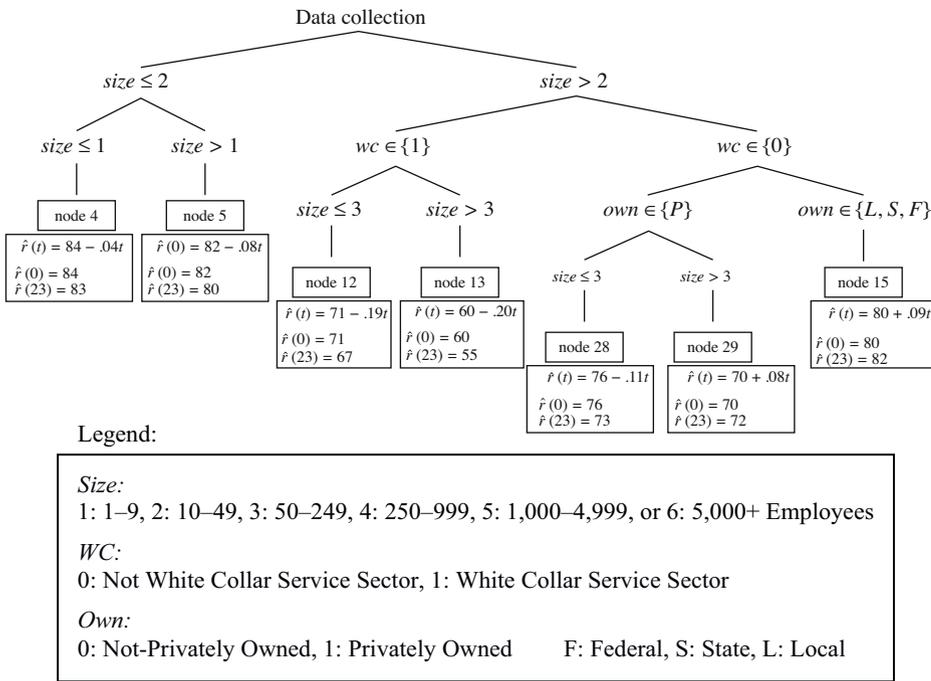*Fig. 4. JOLTS response rate by time in survey for certainty and for non-certainty units.*

Data collection

$size \leq 2$ | $size > 2$

$size \leq 1$ | $size > 1$ | $wc \in \{1\}$ | $wc \in \{0\}$

$size \leq 3$ | $size > 3$ | $own \in \{P\}$ | $own \in \{L, S, F\}$

node 4
$\hat{r}(t) = 84 - .04t$
$\hat{r}(0) = 84$
$\hat{r}(23) = 83$

node 5
$\hat{r}(0) = 82 - .08t$
$\hat{r}(0) = 82$
$\hat{r}(23) = 80$

node 12
$\hat{r}(t) = 71 - .19t$
$\hat{r}(0) = 71$
$\hat{r}(23) = 67$

node 13
$\hat{r}(t) = 60 - .20t$
$\hat{r}(0) = 60$
$\hat{r}(23) = 55$

$size \leq 3$ | $size > 3$

node 15
$\hat{r}(t) = 80 + .09t$
$\hat{r}(0) = 80$
$\hat{r}(23) = 82$

node 28
$\hat{r}(t) = 76 - .11t$
$\hat{r}(0) = 76$
$\hat{r}(23) = 73$

node 29
$\hat{r}(t) = 70 + .08t$
$\hat{r}(0) = 70$
$\hat{r}(23) = 72$

Legend:

Size:
1: 1–9, 2: 10–49, 3: 50–249, 4: 250–999, 5: 1,000–4,999, or 6: 5,000+ Employees

WC:
0: Not White Collar Service Sector, 1: White Collar Service Sector

Own:
0: Not-Privately Owned, 1: Privately Owned   F: Federal, S: State, L: Local

Fig. 5.   Regression tree model of establishment responserate by time (during data collection). Note that since we assess the linear relationship between establishment size and nonresponse bias in Section 4 of the paper, establishment size is treated as a continuous variable in the preceding Data Collection tree model.

initial response rates, from larger establishments. The two smallest size classes were then split again (into nodes 4 and 5), where the smallest establishments had slightly higher initial response rates and only half the rate of attrition of establishments with 10–49 employees.

For establishments in the larger size classes, the algorithm then separated out white-collar service sector establishments (Nodes 12 and 13), which were identified in Phipps and Toth (2012) as being particularly poor responders as part of a nonresponse analysis for another government survey. Establishments in these industries have a low initial response rate and the highest rate of attrition, where we expect the response rate to decrease by over four percentage points over the 24 months. The algorithm further identified the largest establishments in this industry (over 250 employees) as having a much lower initial response rate of 60 percent compared to 71 percent for these establishments with between 50 and 250 employees.

The larger establishments that were not in the white-collar service sector industries split into private versus government (Local, State, or Federal) establishments. The government establishments (Node 15) have a high initial response rate for large establishments and their response rate actually improves over the 24 months in the survey. The privately held establishments split into medium (between 50 and 249 employees) and large (250 or more employees) establishments. While the medium establishments have a higher initial response rate than the larger ones (76% compared to 70%), their rate of attrition leads to a reduction of three percentage points in response rate, while the response rate of the larger

establishments of this type improves by two percentage points over the 24 months in the survey.

4.   Assessing Nonresponse Bias in the Overall Sample

Each of the models in the previous sections provide useful information that can be used to allocate resources and collection efforts to lower response propensity groups at each stage of the JOLTS survey process. However, what they do not tell us is where we are likely to see error in the estimates due to nonresponse. An estimate that has been adjusted for heterogeneous response rates, using establishment characteristics, could still be biased if, conditioned on the establishment characteristics, the response rates of the establishments are related to the variable of interest. For example, the JOLTS estimate of hires would be biased downward if establishments with the most hires, conditioned on the establishment's characteristic variables, respond at a lower response rate than establishments with fewer hires.

In this section, we assess the possible bias for two of the JOLTS estimates (hires and separations) conditioned on the establishment characteristics associated with response rate. This is done by considering the difference in the mean of the variable of interest between responding and nonresponding units for each size class within each end node of the tree in Figure 5. If we find evidence of a difference in the mean of the variable of interest between respondents and nonrespondents within cells of homogeneous response rates, this could indicate a potential for nonresponse bias to remain in the estimate of total hires (or separations) even after adjusting for nonresponse using the observed data.

Since we do not observe the variable of interest for the nonresponding units, this analysis cannot be done using only the JOLTS data. The Quarterly Census of Employment and Wages (QCEW) data, which is used by JOLTS for a sampling frame, includes monthly employment counts for every establishment. The QCEW employment dataset is available about six months after JOLTS produces its monthly estimates, so it cannot be used to improve JOLTS estimates, but may be used to assess their quality.

Using the change in the QCEW employment count data, we compute the net gain (or loss) of employees for every establishment over a given month. Reasoning that establishments with a large net increase in employees probably have more hires than establishments with no increase in employment on average (after all, they must have had at least as many hires as the net increase), we use the computed change in employment level from the QCEW data as a proxy for hires (and separations) in our nonresponse analysis.

There are difficulties with using this difference directly as a proxy. First, the JOLTS collects the number of hires (and separations) for a given calendar month, and the QCEW data has employment counts for the pay period covering the twelfth of the month, so the timeframes for which hires and separations are being measured are different. Additionally, a small increase or even a decrease in employment does not imply few hires because there could have been a large number of separations as well. However, empirically, there is a relationship between change in employment and hires, indicating that the variables are predictive of survey outcome variables of interest (Kreuter et al. 2010).

Every establishment with a net increase of employment that is in the highest quartile in its size class, we label as in the high-hire (H) group. Using the JOLTS data collected

from the responding units, we calculate the mean number of hires for establishments in the high-hire group, and the mean number of hires for establishments not in that group (we label this Group L) by size class. Figure 6 shows the difference in the mean hires for establishments in the high-hire group (Group H) compared to the mean for the rest of the establishments by month and size class. The difference varies by month, but the average number of hires is larger in Group H than Group L for every month and in every size class. The difference in average total separations is also variable, and except for the $1,000-4,999$ size class in the month of December, is all positive. This gives us sufficient assurance that, for a class of establishments, if the proportion of responding establishments in the high-hire group is differs greatly from the proportion in the high-hire group overall, then we could have the potential for nonresponse bias.

Assuming this observed relationship between net employment calculated from the administrative data and hires calculated from the JOLTS data holds for nonresponding establishments as well, we can use this relationship to estimate the potential contributed nonresponse bias at each adjustment cell (defined as each end node by size class). Since our target estimate (hires or separations) is a function of the estimated conditional mean number of hires per establishment (Y'), we can assess the possibility of bias due to nonresponse by considering the difference $Y'_o - Y'$, between the conditional mean using the observed data Y'_o and the conditional mean for all the data Y'. Since

$$Y' = Y'_H p + Y'_L (1 - p),$$

where p is the proportion of establishments in the (high-hires) Group H, Y'_H represents the mean of establishments in Group H and Y'_L is the mean of establishments not in Group H.

If we assume that the difference in average number of hires for establishments in the high-hire group versus that of establishments not in Group H is the same for responding and nonresponding units, then

$$Y'_o - Y' = Y'_H p_o + Y'_L (1 - p_o) - Y'_H p - Y'_L (1 - p)$$
$$= (Y'_H - Y'_L)(p_o - p),$$

where p_o is the proportion of observed (responding) units that are in Group H. Then our estimated bias for the total number of hires based on the observed data is

$$\theta = N(Y'_H - Y'_L)(p_o - p).$$

In order to evaluate the significance of these estimates of expected bias, we need a standard error for our estimator θ. For fixed proportions p_o and p and population total N, the standard error of θ is given by

$$SE(\theta) = N_o |p_o - p| \sqrt{\frac{S_H^2}{n_H} + \frac{S_L^2}{n_L}},$$

where S_H^2, n_H, S_L^2, and n_L are the sample variance of the mean hires and number of responding establishments in the high-hire and low-hire groups respectively. These standard errors account for the JOLTS complex sample design and weighting when the proportions are considered fixed.
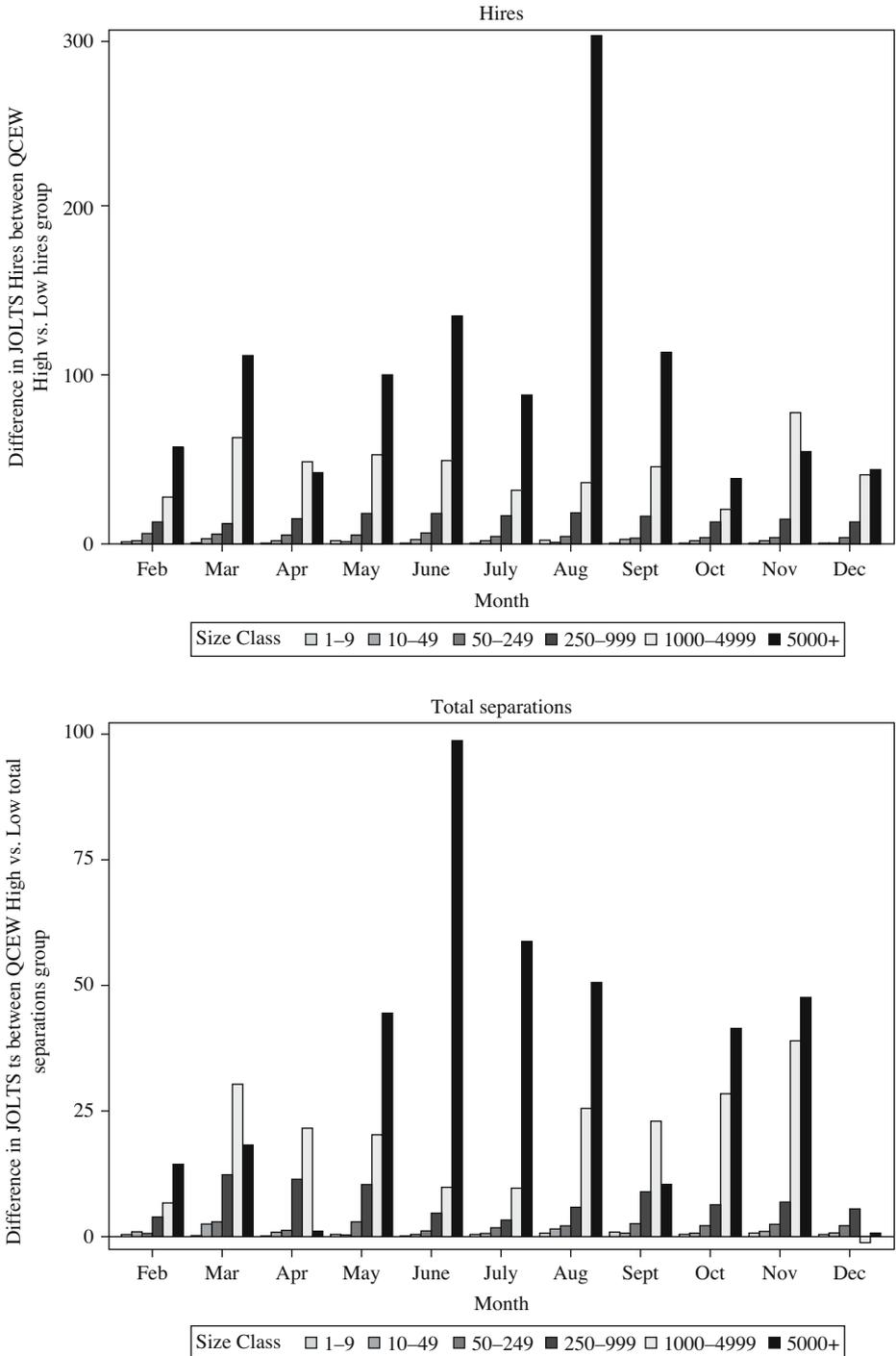
We calculated the statistic (θ) for every size class in every end node of the tree in Figure 5 by month. Based on the 95 percent confidence intervals of these statistics, we saw no real risk of bias due to nonresponse. While this is good news for the JOLTS survey for two of the key estimates, this analysis should be conducted regularly to ensure no pattern of consistently over-representing either high- or low-hire (or separation) group develops in the future for a class of establishments. Also, note that we are not able to carry out a similar nonresponse bias analysis for a third key estimate, job openings, due to the lack of a proxy variable, so the possibility of bias remains for that important outcome.

5. Discussion

This study provides a comprehensive framework for assessing establishment nonresponse in a longitudinal survey. We identify and compare the characteristics of nonresponding establishments across the various phases of data collection – both before and during data collection. At all phases, we find that employment size, structure (being part of a multi-establishment firm), ownership and industry are associated with nonresponse. While these establishment characteristics have been identified as important in understanding nonresponse in other longitudinal studies (Janik and Kohaut 2012; Seiler 2010; Smaill 2012), the regression tree models provide us with a tool to identify and interpret critical interactions between these business characteristics, and response propensity. The groups formed by regression trees are highly interpretable, allowing for allotment of resources to identified subgroups that have low response rates.

By looking at nonresponse separately for each collection phase, we observe that the specific characteristics of nonrespondents vary, which helps us to better understand when and for what type of establishment nonresponse is an issue. For example, the results for JOLTS suggest that directing efforts toward federal government establishments during data collection would not be nearly as effective as doing so during address refinement. Also, waiting to target white-collar service sector establishments until enrollment is a potential strategy, since there is less difficulty locating and verifying addresses and contact information compared to gaining their participation in the survey. It is also helpful to know, prior to sampling and initial contacts, that employment size and structure are negatively associated with low response at each phase.

Using the rpms tree model, we model the effect of time in survey on response rates during data collection, so that we can determine which establishments require more follow-up at the beginning and over the course of the survey to ensure that they respond. For non-certainty units, the tree algorithm identifies important relationships between response rate and time, involving establishment size, industry and ownership. Small establishments have high initial response rates and low attrition, large establishments in white collar services have low initial response rates and a high rate of attrition, and government establishments have high initial response rates and actually increase response over time. These patterns are distinct and can be used to help the survey program determine when they need to focus their calling efforts.

To assess nonresponse bias, we are able to link the results from the data collection regression tree model to QCEW proxy data. This enables us to determine whether there are groups of establishments with similar response and time effects for which the

distributions of hires and separations vary for respondents versus nonrespondents indicating potential bias. While we find little evidence of nonresponse bias in two key estimates, we are unable to assess a third outcome variable. Since establishment surveys differ on the key estimates, they produce potential proxy variables that are not always easily identified and developed from existing sampling or alternative frames. Another approach is to utilize model-based methods, such as the proxy pattern-mixture model approach presented in Andridge and Little (2011) and Andridge and Thompson (2015). Since the predictor variables for the response rates are available for all sampled units and if the available frame data covariates are correlated with the variables of interest, then the potential effects of nonresponse bias could be assessed using the fraction of missing information (FMI) obtained from maximum likelihood or multiple imputation. Likewise, Wagner (2010) and Wagner (2012) contain alternative approaches to estimating the FMI from survey data for the similar nonresponse bias analysis purposes.

This type of retrospective analysis is important. However, there are ways of using these same identified characteristics on an ongoing basis to monitor the potential for nonresponse bias. Examples include the R-indicator and partial R-indicators (Schouten et al. 2009, Schouten, Shlomo, and Skinner 2011), which measure the deviation of the sample from missing-completely-at-random using modeled response propensities, the balance and distance indicators proposed by Särndal (2011) and Särndal and Lundquist (2014). Here, the distance measure is analogous to the comparison provided in Section 4, and the Fraction of Missing Information as proposed in Wagner (2012), if some or all of the identified characteristics are also predictive of outcome.

Since minimizing bias is critical to the quality of survey estimates and response rates remain important in maintaining stakeholder confidence, the framework we have set out to examine establishment survey longitudinal nonresponse has the potential to be a useful tool in analyzing and improving the efficiency of many longitudinal surveys. Most establishment sample frames have rich auxiliary variables to assess the relationships between business characteristics and nonresponse, including bias, allowing the surveys to benefit from this type of analysis. In an era of declining resources, determining the timing of nonresponse for establishment subgroups allows survey programs to experiment with and tailor interventions with the goal of most efficiently allocating resources or dynamically shifting resources as part of an adaptive data collection plan.

6.   References

Andridge, R.R. and R.J.A. Little. 2011. "Proxy Pattern-Mixture Analysis for Survey Nonresponse." *Journal of Official Statistics* 27: 153–180.

Andridge, R.R. and K.J. Thompson. 2015. "Assessing Nonresponse Bias in a Business Survey: Proxy Pattern-Mixture Analysis for Skewed Data." *Annals of Applied Statistics* 9(4): 2237–2265. Doi: http://dx.doi.org/10.1214/15-AOAS878.

Atrostic, B.K., N. Bates, and A. Silberstein. 2001. "Nonresponse in US Government Household Surveys: Consistent Measures, Recent Trends, and New Insights." *Journal of Official Statistics* 17: 209–226.

Bureau of Labor Statistics. 2015. *Handbook of Methods*. "Job Openings and Labor Turnover Survey" Chapter 18. Available at: https://www.bls.gov/opub/hom/pdf/homch18.pdf (accessed October 2017).

Davis, W.R. and N. Pihama. 2009. "Survey Response as Organisational Behaviour: an Analysis of the Annual Enterprise Survey, 2003–2007." New Zealand Association of Economists Conference. New Zealand: New Zealand Association of Economists. Available at: http://ro.uow.edu.au/eispapers/826/ (accessed October 2017).

De Heer, W. and E. De Leeuw. 2002. "Trends in Household Survey Nonresponse: a Longitudinal and International Comparison." In *Survey Nonresponse*, edited by R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, 41–54. New York: John Wiley and Sons.

Dillman, D.A. 2000. *Mail and Internet Surveys: The Tailored Design Method* (second edition). New York: John Wiley & Sons.

Dillman, D.A., J.D. Smyth, and L.M. Christian. 2009. *Internet, Phone, Mail, and Mixed-Mode Surveys: the Tailored Design Method* (third edition). Hoboken, NJ: John Wiley & Sons.

Freedman, D.S., A. Thornton, and D. Camburn. 1980. "Maintaining Response Rates in Longitudinal Studies." *Sociological Methods & Research* 9: 87–98.

Groves, R.M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70: 646–675.

Groves, R.M., D.A. Dillman, J.L. Eltinge, and R.J.A. Little. 2002. *Survey Nonresponse*. New York: John Wiley and Sons.

Heberlein, T.A. and R. Baumgartner. 1978. "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature." *American Sociological Review* 43: 447–462.

Holbrook, A., J.A. Krosnick, and A. Pfent. 2007. "The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms." In *Advances in Telephone Survey Methodology*, edited by J.M. Lepkowski, C. Tucker, J.M. Brick, E. De Leeuw, L. Japec, P.J. Lavrakas, M.W. Link and R.L. Sangster, 499–528. Hoboken, NJ: Wiley.

Hothorn, T., K. Hornik, and A. Zeileis. 2006. "Unbiased Recursive Partitioning: a Conditional Inference Framework." *Journal of Computational and Graphical Statistics* 15: 651–674.

Janik, F. and S. Kohaut. 2012. "Why Don't They Answer? Unit Nonresponse in the IAB Establishment Panel." *Quality & Quantity* 46: 917–934.

Kreuter, F., K. Olson, J. Wagner, T. Yan, T.M. Ezzati-Rice, C. Casas-Cordero, M. Lemay, A. Peytchev, R.M. Groves, and T.E. Raghunathan. 2010. "Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Non-Response: Examples from Multiple Surveys." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 173: 389–407.

Lepkowski, J.M. and M.P. Couper. 2002. "Nonresponse in the Second Wave of Longitudinal Household Surveys." In *Survey Nonresponse*, edited by R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, 259–272. New York: John Wiley and Sons.

Little, R.J.A. and S. Vartivarian. 2005. "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology* 31: 161–168.

Lohr, S., V. Hsu, and J. Montaquila. 2015. "Using Classification and Regression Trees to Model Survey Nonresponse." In *Joint Statistical Meetings, Proceedings of the Survey Research Methods Section: American Statistical Association*. 2071–2085. Alexandria, VA: American Statistical Association. Available at: http://ww2.amstat.org/sections/srms/Proceedings/y2015f.html (accessed October 2017).

Morgan, J.N. and J.A. Sonquist. 1963. "Problems in the Analysis of Survey Data, and a Proposal." *Journal of the American Statistical Association* 58: 415–434.

Office of Management and Budget. 2006. *Statistical Directive No. 2, Standards and Guidelines for Statistical Surveys 2006*. Washington DC: OMB. Available at: https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf (accessed October 2017).

Paxson, M.C., D.A. Dillman, and J. Tarnai. 1995. "Improving Response to Business Mail Surveys." In *Business Survey Methods*, edited by B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge, and P.S. Kott, 303–316. New York: John Wiley and Sons.

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: 772–794.

Särndal, C. 2011. "The 2010 Morris Hansen Lecture. Dealing with Survey Nonresponse in Data Collection, in Estimation." *Journal of Official Statistics* 27(1): 1–21.

Särndal, C. and P. Lundquist. 2014. "Accuracy in Estimation with Nonresponse: A Function of Degree of Imbalance and Degree of Explanation." *Journal of Survey Statistics and Methodology* 2(4): 361–387. Doi: https://doi.org/10.1093/jssam/smu014.

Schouten, B., F. Cobben, and J. Bethlehem. 2009. "Indicators for the Representativeness of Survey Response." *Survey Methodology* 35(1): 101–113.

Schouten, B., N. Shlomo, and C. Skinner. 2011. "Indicators for Monitoring and Improving Representativeness of Response." *Journal of Official Statistics* 27(2): 231–253.

Seiler, C. 2010. *Dynamic Modelling of Nonresponse in Business Surveys*. No. 93. Ifo Working Paper. Available at: https://econpapers.repec.org/paper/cesifowps/_5f93.htm (accessed October 2017).

Singer, E. 2002. "The Use of Incentives to Reduce Nonresponse in Household Surveys." In *Survey Nonresponse*, edited by R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, 163–177. New York: John Wiley and Sons.

Smaill, K. 2012. "Trajectory Modelling of Longitudinal Non-Response in Business Surveys." *Statistical Journal of the International Association of Official Statistics* 28: 137–144.

Su, X., M. Wang, and J. Fan. 2004. "Maximum Likelihood Regression Trees." *Journal of Computational and Graphical Statistics* 13: 586–598.

Tomaskovic-Devey, D., J. Leiter, and S. Thompson. 1994. "Organizational Survey Nonresponse." *Administrative Science Quarterly* 39: 439–457.

Toth, D. 2017. *rpms: Recursive Partitioning for Modeling Survey Data. R package version 0.2.0*. Available at: https://CRAN.R-project.org/package=rpms (accessed October 2017).

Wagner, J. 2010. "The Fraction of Missing Information as a Tool for Monitoring the Quality of Survey Data." *Public Opinion Quarterly* 74: 223–243. Doi: http://dx.doi.org/10.1093/poq/nfq007.

Wagner, J. 2012. "A Comparison of Alternative Indicators for the Risk of Nonresponse Bias." *Public Opinion Quarterly* 76: 555–575. Doi: http://dx.doi.org/10.1093/poq/nfs032.

Watson, N. and M. Wooden. 2009. "Identifying Factors Affecting Longitudinal Survey Response." In *Methodology of Longitudinal Surveys*, edited by Peter Lynn, 151–181. Hoboken, NJ: John Wiley and Sons.