Analyzing Nonresponse Bias in the 2010-2018 IAB Job Vacancy Survey Using Administrative Data and Machine Learning

**Keywords:** Nonresponse, Nonresponse Bias, Establishment Survey, Machine Learning

# Introduction

While there is much literature documenting the declining trend of household response rates and resulting effects on nonresponse bias over time, analyses of nonresponse rate and nonresponse bias trends in voluntary establishment surveys are not so widespread. This research project contributes to this rather small strand of literature by analyzing trends in response rates and their relationship with nonresponse bias over the last 8 years of the IAB-Job Vacancy Survey. The JVS is a nationally-representative and voluntary establishment survey, which mainly aims to quantify the size of the unfilled labor demand in Germany and other kinds of worker flows. It is conducted yearly as a repeated cross-sectional survey carried out using a concurrent mixed-mode design, where paper questionnaires are sent to all establishments with the option of online completion.

In addition, we evaluate the performance of various machine-learning approaches to further reduce nonresponse bias relative to the conventional weighting approach. While some studies have already pointed out the potential of machine learning methods in non-response settings, this is the first to investigate the usefulness of a wide range of different machine learning methods to reduce non-response bias for an establishment survey. For all analyses we exploit comprehensive administrative data sources available for both respondents and nonrespondents, which includes extensive information on establishment characteristics as well as on the composition of the workforce.

# Methods

## Response Rate

Referring to the definitions of the American Association of Public Opinion Research (2016), we compute the response rate based on Response Rate 1, which is the number of completed interviews divided by the sample size. As this definition includes the full sample in the reference group, it is a conservative calculation of response rates and can be interpreted as a minimum response rate.

## Nonresponse Bias

Nonresponse bias is directly measured as the difference between the target estimate based on the respondents and the corresponding estimate based on the full sample. Nonresponse bias is calculated for all available administrative establishment characteristics. In addition, we construct and compare absolute relative bias and average absolute relative bias measures [1].

$NR Bias= Y\_{i,r}-Y\_{i,n}$ (1)

$Abs.rel. NR Bias= \left|\frac{Y\_{i,r}-Y\_{i,n}}{Y\_{i,n}}\right|$ (2)

$Avg. abs. rel. NR bias= \frac{\sum\_{i=1}^{K}\left|\frac{Y\_{i,r}-Y\_{i,n}}{Y\_{i,n}}\right|}{K}$ (3)

where i denotes the estimable category of interest (i=1, 2,...,K), r denotes the respondents, and n denotes the full sample

As formula 2 expresses, absolute relative nonresponse bias is calculated as the difference between the ith estimated percentage based on the respondent sample and the corresponding estimated percentage based on the full sample relative to the estimated percentage based on the full sample. Thereby, we categorised all variables of interest into approximately equal-sized categories and estimated bias by comparing estimates of each category of the variables of interest. As a summary measure for multiple biases, we average the absolute relative biases for all establishment information and two subgroups: establishment characteristics and employee characteristics. These measures enable us to investigate the relationship of the response rate and the corresponding unadjusted nonresponse bias.

## Response Propensity Models and Weights

To investigate the usefulness of various modelling approaches in a non-response reduction scheme, we estimate response propensities from traditional logistic regression models and supervised machine learning algorithms. In line with previous literature on predicting nonresponse using a machine-learning approach, we concentrate on tree-based methods ([2], [3], [4]). As highlighted by Kern, Weiß and Kolb (2019) these methods are well-suited for tasks involving many variables as well as non-additive and/or non-linear relationships. Furthermore, they are able to handle complex interactions between the explanatory variables. Additionally we use a penalized general additive regression model (gamsel). To summarize, we evaluate the performance of the following modelling approaches: Logistic regression, General additive model, Decision tree using the CART-algorithms, Decision tree using the C-Tree-algorithms, Model-based recursive partitioning (MOB), Random Forest, Extreme Gradient Boosting and Bayesian Additive Regression Trees. Based on these models we estimate response propensities and use the inverse of these propensities to construct adjustment weights.

# Results

## Response Rate

Figure 1 (left panel) shows the response rate of the IAB-JVS for years 2010-2018. One can see that the response rate of IAB-JVS has always been below 20 percent since 2010 and the average response rate for the observation period is 16.32 percent. Compared to other establishment survey response rates worldwide it is rather low. Already starting from this comparably low response rate of 19.69 percent in 2010 the response rate has decreased further to 13.20 percent in 2018. This represents an average decline of 0.8 percentage points per year.



Figure 1. Response Rate between 2010 and 2018 (left panel) and Average absolute relative bias between 2011 and 2017 (right panel)

## Nonresponse Bias

Figure 1 (right panel) illustrates the average absolute relative nonresponse bias between 2011 and 2017. The average relative bias measure lies in all observation years around 7 percent. Hence, the relative nonresponse bias is on average not very large. Relating that to a comparably high nonresponse rate of more than 80 percent, it is reassuring that the average nonresponse bias is not particularly high. Between 2011 and 2017 we see a slight increase of the average absolute relative bias from 6.38 percent in 2011 to 7.14 percent in 2017. Corresponding to the response rate the strongest increase of average absolute relative bias occurs between 2013 and 2015.

A deeper look shows that many of the individual relative biases exceed a threshold of 10 percent. Already when differentiating between the establishment and employee characteristics, it is remarkable that the average absolute relative bias of the establishment characteristics shows values around 10 percent. In total, we find between 15 individual bias estimates in 2011 and 20 in 2017 with values greater than 10 percent. Thus, the number of relative biases exceeding 10 percent seems to be increasing over time. Since we analyze 70 single bias estimates, the share of bias estimates that exceeds 10 percent amounts to 21 percent in 2012 to 29 percent in 2017 of all biases. Thus, the share of bias estimates above the threshold of 10 percent increased by 8 percentage points during the observation period.

## Response Propensity Models and Weights

As one measure of nonresponse bias reduction, we compare the mean number of hirings per establishment, which is not only an important measure by itself but also correlated to the number of vacancies in the previous year. We use the weights generated by including all explanatory variables into the modeling approaches to compute weighted estimates of this target variable. By comparing the weighted hirings and the unweighted full sample we provide evidence on which model performs best in terms of reducing non-response bias out of sample. Figure 3 relates the number of hirings in the full sample to the weighted hirings with each machine learning algorithm. The reference line represents the perfect fit of the weighted estimates and the full sample and thus, the complete reduction of non-response bias in regard to the number of hirings. All weighted hirings underestimate the number of hirings in the full sample. Although the performance of the weighting strategies varies from year to year the pattern is consistent over the years. Interestingly, the traditional logistic regression with a small set of categorized variables is just slightly outperformed by the Bart-Algorithm. Both approaches can reduce the nonresponse bias with regard to hirings to a minimum. All other machine-learning approaches underestimate the number of hirings even more with XG Boost, MOB and logistic regression with the large set of explanatory variables as the second best group. The random forest algorithm did not work very well in our setting and could not reduce the nonresponse bias at all.

# Conclusions

Our analyses showed that the response rate decreased between 2010 and 2018, while the corresponding nonresponse bias indicated a slight increase. The level of non-response bias was moderate overall, despite high non-response rates. Although machine-learning applications offer new possibilities to model response behaviour and use these models in weighting schemes, they do not automatically outperform traditional logistic regression with a well-chosen set of explanatory variables. In our application, just the BART-algorithm reduced non-response bias more than the standard logistic regression approach. Even though the machine learning methods in our application did not show any fundamental improvement over previous approaches, further research is necessary to better understand the interaction between different weighting strategies and machine learning methods on nonresponse bias reduction.

Figure 4. Comparison of weighting strategies with regard to hirings

# References

1. Groves, R. M. and Heeringa, S. G., Responsive design for household surveys: Tools for actively controlling survey errors and costs, Journal of the Royal Statistical Society. Series A (Statistics in Society), 169 (2006), 439–457.
2. Kern, C., Klausch, T. and Kreuter, F., Tree-based machine learning methods for survey research, Survey Research Methods, 13 (2019a), 73–93.
3. Earp, M., Mitchell, M., McCarthy, J. and Kreuter, F., Modeling nonresponse in establishment surveys: Using an ensemble tree model to create nonresponse propensity scores and detect potential bias in an agricultural survey, Journal of Official Statistics, 30 (2014), 701–719.
4. Earp, M., Toth, D., Phipps, P. and Oslund, C., Assessing nonresponse in a longitudinal establishment survey using regression trees, Journal of Official Statistics, 34 (2018), 463–481