ORIGINAL ARTICLE

Analysing establishment survey non-response using administrative data and machine learning

Benjamin Küfner[1] | **Joseph W. Sakshaug[1,2,3]** | **Stefan Zins[1]**

[1]Institute for Employment Research (IAB), Nuremberg, Germany

[2]University of Warwick, Coventry, United Kingdom

[3]Ludwig Maximilian University of Munich, Munich, Germany

Correspondence
Benjamin Küfner, Institute for Employment Research (IAB), Nuremberg, Germany.
Email: benjamin.kuefner@iab.de

Funding information
Institute for Employment Research

Abstract

Declining participation in voluntary establishment surveys poses a risk of increasing non-response bias over time. In this paper, response rates and non-response bias are examined for the 2010–2019 IAB Job Vacancy Survey. Using comprehensive administrative data, we formulate and test several theory-driven hypotheses on survey participation and evaluate the potential of various machine learning algorithms for non-response bias adjustment. The analysis revealed that while the response rate decreased during the decade, no concomitant increase in aggregate non-response bias was observed. Several hypotheses of participation were at least partially supported. Lastly, the expanded use of administrative data reduced non-response bias over the standard weighting variables, but only limited evidence was found for further non-response bias reduction through the use of machine learning methods.

KEYWORDS

data quality, IAB Job Vacancy Survey, non-response bias, survey participation, weighting adjustment

1 | INTRODUCTION

Official statistics are based to a large extent on establishment surveys, which produce estimates that flow into price indices (e.g. the Producer Price Index Survey at the US Bureau of Labor

Statistics[1]), gross domestic products (e.g. the Purchase Survey at the UK Office for National Statistics[2]), and wage statistics (e.g. Survey of Employment, Payrolls and Hours at Statistics Canada[3]). Establishment surveys play an instrumental role in assessing the state of the economy and thus inform the development of economic policies. In addition, establishment survey data are widely used by researchers in various fields, including organisational behaviour (e.g. Bal & Dorenbosch, 2015), environmental studies (e.g. Ghisetti & Rennings, 2014), personnel economics (e.g. Houseman, 2001; White & Bryson, 2013), and labour economics (e.g. Bachmann et al., 2013). The demand for establishment data has increased in recent years as illustrated by a growing number of publications (e.g. Mercan & Schoefer, 2020), workshop series (e.g. ifo Conference on Macroeconomics and Survey Data), and new establishment surveys (e.g. the BeCovid study in Germany[4], the Decision Maker Panel in the UK[5]).

A prominent example of an official statistic captured by establishment surveys is the number and structure of job vacancies. Vacancy statistics describe the unfilled labour demand and are key economic indicators that provide insights into job matching efficiency and labour market tightness. In Germany, the Institute for Employment Research (IAB) collects vacancy data through the IAB Job Vacancy Survey (IAB-JVS), a voluntary survey which started in 1989. Since then, it has become one of the largest establishment surveys in Germany, annually collecting vacancy data from up to 15,000 establishments of all sizes and industry sectors. Its data are regularly used by Eurostat to compile European-wide vacancy statistics.

Similar to household surveys, voluntary establishment surveys face decreasing survey participation rates (Pielsticker & Hiebl, 2020). As stakeholders and researchers often see response rates as an important survey quality indicator, a decreasing response rate can undermine the reputation of a survey and increase costs as sample sizes must be increased to meet precision requirements. A comprehensive analysis of establishment survey non-response is essential not only for assessing survey quality, but also for understanding which subgroups are more prone to non-response and may benefit from interventions (e.g. tailored designs) or adjustment strategies (e.g. weighting) that mitigate the risk of non-response bias. The increasing use of large auxiliary data sources, such as administrative data, coupled with an expanding set of data-driven (e.g. machine learning) modelling tools offer a promising means of ascertaining mechanisms of establishment survey non-response, identifying subgroups that are most affected by it, and adjusting for potential bias.

While there is much literature documenting response rates, non-response bias, and adjustment strategies in household surveys (e.g. Brick & Williams, 2013; De Heer & De Leeuw, 2002; Groves & Peytcheva, 2008; Kreuter et al., 2010; Peytcheva & Groves, 2009; Williams & Brick, 2018), such analyses are not widespread in voluntary establishment surveys. This article contributes to the rather small strand of literature by analysing response rates and non-response bias over the last 10 years of the IAB-JVS. Using a comprehensive administrative database available for both respondents and non-respondents, we assess the magnitude of non-response bias and test several theory-driven hypotheses regarding survey participation. In addition, we evaluate the performance of using available administrative data and various modelling approaches, including machine learning methods, to adjust for non-response bias and improve on the traditional weighting strategy used in the IAB-JVS. Although some published studies have noted the

---

[1]www.bls.gov/respondents/ppi/

[2]www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/annualpurchasessurvey

[3]www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2612

[4]www.iab.de/de/befragungen/becovid.aspx

[5]www.decisionmakerpanel.co.uk/

potential advantages of machine learning methods for non-response bias adjustment, this is the first to evaluate a wide range of such methods for an establishment survey.

The remainder of this article is structured as follows. Section 2 briefly describes the theory of establishment survey participation and reviews the empirical evidence on response rates, non-response bias, and adjustment strategies. Section 3 presents the research objectives and survey participation hypotheses. Section 4 describes the data sources and analysis strategy. Section 5 presents the results and Section 6 concludes with a general discussion of the findings and their practical implications.

2 | THEORY AND BACKGROUND

2.1 | Establishment survey participation

Survey participation in an organisational context differs substantially from the household context. The professional goals of the establishment can shape the participation decision in a positive or negative way. Establishments conduct a rational cost-benefit analysis in which they weigh the costs of participation against the possible benefits in the context of their professional goals (Snijkers et al., 2007; Willimack et al., 2002; Willimack & Snijkers, 2013). Costs of participation include the perceived burden of allocating resources to the response task, searching for the requested information, and completing the questionnaire, which may be particularly burdensome for certain types of establishments. On the benefits side, while survey participation may not directly contribute to establishments' professional goals (e.g. making a profit), they may find other value in participating or use the data provided by the survey for planning purposes. In addition, some establishments perceive survey response as part of their corporate social responsibility and their contribution to a working society that informs the current political discussion (Willimack et al., 2002; Willimack & Snijkers, 2013). In the following we discuss specific factors that influence the participation decision based on the framework of Willimack and Snijkers (2013), which forms the basis for the forthcoming hypotheses.

Based on previous theoretical work (e.g. Tomaskovic-Devey et al., 1994; Willimack et al., 2002), Willimack and Snijkers (2013) distinguishes between participation factors under the control of the researcher, such as the sample design, and factors outside their control, namely the establishment's environment, the establishment itself, and the actual respondent within the establishment. The environmental factor includes all surrounding influences, including economic conditions, survey-taking climate and legal requirements or general norms. The establishment itself is characterised by the profile and organisation of the establishment including internal policies and resource availability. The last factor reflects the influences of the employee representative of the establishment who is assigned the response task, such as his or her experience level. These three factors are conceptualised hierarchically such that the environment shapes the establishment's decision, which in turn affects the responding employee (Willimack et al., 2002; Willimack & Snijkers, 2013). Within each hierarchical level, three further factors play a role: authority, capacity and motivation (Tomaskovic-Devey et al., 1994; Willimack & Snijkers, 2013). Authority reflects the formal and informal power to decide on the survey request. For example, organisational policies may shape the freedom of the representative to make the participation decision. Capacity is defined as the ability to comply with the survey request. This refers especially to the knowledge, time constraints and competence of the responding employee to gather the relevant information to complete the questionnaire. Lastly, motivation captures the

establishment's or individual respondent's attitude towards the survey and the drive to participate. Examples of the interrelationship between these higher- and lower-level factors are illustrated in the following subsections.

2.1.1 | Environmental factors

Laws or regulations that make a survey mandatory are one example of an environmental factor that affects the survey decision. Here the laws shape the authority as the establishment is obligated to respond or face a fine. Also, from an empirical perspective, there is clear evidence that mandatory participation leads to a higher likelihood of response (Snijkers, 2008; Snijkers et al., 2007; Willimack & Nichols, 2010; Willimack & Snijkers, 2013). The economic situation also influences the decision. Both a boom and a recession reduce the capacity of establishments to respond: either they have no time, because they are dealing with growing markets and influx of customers, or they reduce staff to stay solvent, which also reduces resources for survey participation (Davis & Pihama, 2009; Fisher et al., 2003; Seiler, 2014).

2.1.2 | Establishment factors

Establishments differ in various aspects that likely affect their cost-benefit analysis for participation. In addition to previously mentioned aspects, such as internal policies and corporate social responsibility, establishment size also plays a role. While owners of small businesses can usually handle the survey request themselves, more coordination between hierarchies and departments is needed for larger establishments (Willimack & Snijkers, 2013). In terms of capacity, easily derivable data from record systems, established response processes and clear organisational responsibilities for survey requests reduce response burden and facilitate participation (Willimack & Snijkers, 2013).

2.1.3 | Respondent factors

It is important to keep in mind that multiple persons could be involved in the response decision. For example, an owner or unit head may have the authority to comply with the survey request, while other employees have the capacity. Several employees from different units may be needed to provide the requested information in multi-topic surveys. Usually the researcher has only a minor influence on who responds within the establishment and can only address the survey request to the establishment as a whole or a specific role within (Willimack et al., 2002). The characteristics of all individuals involved (e.g. motivation, level of experience) also factors into the response decision.

2.2 | Response rates in establishment surveys

Participation rates vary strongly between establishment surveys. For instance, mandatory (mostly governmental) surveys such as the Survey of Occupational Injuries and Illnesses (SOII) in the United States (BLS, 2020) or the Survey on Investments in the Netherlands (Snijkers, 2018)

reach response rates of almost 90%. On the other extreme, voluntary multi-topic surveys conducted by private research organisations can have response rates below 5% (White & Luo, 2005). Further, there are indications that response rates are declining over time for some voluntary surveys. In 1990s Christianson and Tortora (2011) found that 30% of survey managers interviewed in 16 countries reported a declining response rate trend in their establishment surveys. In the early 2000s Petroni et al. (2004) found evidence of decreasing or stable response rates in both voluntary and mandatory surveys in the United States. Anseel et al. (2010) concluded from a meta-analysis of 2037 published studies that the increased use of response enhancement strategies prevented a strong decline in response rates. Most recently, a declining trend in response rates has been observed for voluntary surveys in the United States and Germany (BLS, 2020;Janik & Kohaut, 2012; König et al., 2021). An international meta-analysis of family firm surveys also confirms this trend (Pielsticker & Hiebl, 2020).

2.3 | Non-response bias in establishment surveys

Given the low and sometimes declining response rates in many establishment surveys, it is important to assess their potential for non-response bias. Most often non-response bias analyses are performed by comparing respondents and non-respondents based on the auxiliary information that is available for both groups, for example from the sampling frame (e.g. Earp et al., 2018; Lineback & Thompson, 2010). Alternatively, researchers compare early and late respondents, use previous census information (e.g. Earp et al., 2014) or conduct costly non-response surveys. With the rise of big data approaches, linking surveys with administrative data is gaining attention as a promising means to analyse non-response bias (Bavdaž et al., 2020). Administrative data offer potentially rich and up-to-date auxiliary information on responding and nonresponding establishments and are therefore uniquely suited for studying non-response. However, despite their high potential, administrative data are rarely used for non-response analysis (for exceptions, see Janik & Kohaut, 2012; König et al., 2021). The present study addresses this research gap by exploiting extensive administrative data to assess non-response bias in depth.

Non-response bias can have large effects on establishment statistics when influential establishments do not respond. This is especially true for very large establishments that employ a considerable share of the workforce. Such establishments can have a substantial impact on key survey estimates (Lineback & Thompson, 2010; Riviére, 2002; Thompson & Oliver, 2012). With regard to job vacancy statistics, large establishments are especially critical as they typically contribute disproportionately to estimates of total vacancies. Thus, if larger establishments have lower response propensities than smaller establishments, then severe non-response bias in vacancy statistics could result. There is indeed evidence that larger establishments are less likely to participate in surveys (e.g. Earp et al., 2018; Janik & Kohaut, 2012; König et al., 2021; Phipps & Toth, 2012). Other correlates of establishment survey participation are also documented in the literature. For example, industry (Phipps & Jones, 2007; Tomaskovic-Devey, 1995), multi-unit establishments (Phipps & Toth, 2012), establishment age (Phipps & Jones, 2007), wages (Phipps & Toth, 2012) and region of the establishment (Janik & Kohaut, 2012; Phipps & Jones, 2007). The present study contributes to the existing literature by analysing additional hypothesised correlates of survey participation, including detailed workforce characteristics and diversity measures, employee demographic profile, and the development of the employment structure. To date, these workforce characteristics have not been explored in the establishment survey literature.

Additionally, most non-response bias studies present only measures of association (e.g. correlations, regression coefficients) between establishment characteristics and participation and do not evaluate the magnitude of non-response biases. Yet, magnitude is an important aspect of non response bias as it allows researchers to compare the sizes of the biases over time and assess their impact on substantive analyses. We address this research gap by presenting individual and aggregate bias estimates for important substantive variables derived from administrative data.

2.4 | Adjustment strategies and machine learning tools

To adjust for potential non-response bias in establishment surveys, sample-based weighting schemes, such as propensity score weighting, are often used (Valliant et al., 2013). Such weighting procedures rely on the availability and quality of auxiliary data. As illustrated by Little and Vartivarian (2005), weighting effects are optimised if the auxiliary data are correlated with both the response outcome and the target survey variables. In establishment surveys, auxiliary data are usually limited to available paradata or few sampling frame variables that may relate well to response propensity, but less so for the substantive survey variables. Administrative data offer a promising supplementary source, as they contain substantive attributes on the establishments (e.g. revenue) and their workforce (e.g. demographic composition) that likely have a stronger relationship with the key survey variables and the response propensity.

Propensity score weighting is traditionally performed by modelling the response outcome using a logistic (or probit) regression model conditional on the auxiliary variables and deriving response propensity scores to create weights for each sampled unit. However, within the last decade machine learning methods have become increasingly popular for modelling survey participation (e.g. Earp et al., 2014, 2018; Kern et al., 2019; Lohr et al., 2015; Phipps & Toth, 2012; Toth & Phipps, 2014; Zinn & Gnambs, 2022). A major advantage of these methods over traditional methods is that they can handle complex data structures with many variables and identify high-level interactions and other non-linear effects. As such, they offer the capability to identify intricate data-driven relationships between the auxiliary variables and the response outcome.

However, only a few studies have investigated the value of using machine learning algorithms for non-response weighting adjustments. In a simulation study, Buskirk and Kolenikov (2015) showed that logistic regression and Random Forests performed similarly well for inverse propensity weighting for a simple response pattern (only a few interactions), but logistic regression performed better for propensity score stratification. In a more complex response setting (with more interactions), Random Forest was superior for inverse propensity weighting and logistic regression was superior for propensity score stratification.

Lohr et al. (2015) conducted a similar simulation study comparing multiple tree-based methods, including Classification and Regression Trees (CART), Conditional Inference Trees (C-tree) and Random Forest, among others, to logistic regression. Response propensities were adequately estimated by logistic regression and C-Tree, if response was simulated linearly, with Random Forest and CART producing small deviations from the true response propensities. However, if interaction terms were used to simulate non-response, then logistic regression performed poorly compared to the tree-based methods, as expected. Using the estimated response propensities for weighting, the C-Tree algorithm performed rather well across different weighting schemes

and response models, outperforming CART. In the direct response propensity weighting scheme, Random Forest reduced the most bias, closely followed by C-Tree and logistic regression.

Earp et al. (2018) demonstrated the application of regression trees (recursive partitioning) to estimate response propensities in the BLS Job Opening Labor Turnover Survey, which is also a vacancy survey, but they did not compare it to other methods. Kern et al. (2019) showed that Extreme Gradient Boosting (XG-Boost) and Random Forest performed best for panel non-response prediction in the German Socio-Economic Panel, closely followed by Model-based Recursive Partitioning (MOB) and Bayesian Additive Regression Trees (BART). Logistic Regression, as the reference group, could not compete with the prediction accuracy of the tree-based methods.

3 | RESEARCH OBJECTIVES AND PARTICIPATION HYPOTHESES

3.1 | Research objectives

The present study has four research objectives derived from the aforementioned research gaps. The first research objective (RO1) investigates response rates in the 2010–2019 IAB-JVS to discern whether there is a noticeable trend over time. The second research objective (RO2) investigates the severity and trend (if any) of non-response bias in the IAB-JVS. Here, we utilise an extended set of administrative data containing detailed establishment and workforce characteristics to estimate non-response bias for each survey year. The third research objective (RO3) utilises the extended administrative data to test nine hypotheses of survey participation (described in detail below), including new hypotheses not yet considered in the literature. The last research objective (RO4) builds on the second by evaluating what we gain in non-response bias reduction by including the extended set of administrative variables in the IAB-JVS non-response weighting procedure, as compared to the smaller set of auxiliary variables used in the current weighting procedure.

Further, we compare the performance of several machine learning algorithms for reducing non-response bias in the IAB-JVS relative to a logistic regression-based weighting procedure. The evaluation includes various data-driven algorithms, including some not yet applied in an establishment survey context (e.g. C-Tree, XG-Boost, general additive models). We expect that including the extended set of administrative variables in the IAB-JVS weighting procedure will reduce non-response bias relative to the currently used weighting variables. Moreover, we expect that the machine learning algorithms will reduce non-response bias even further by accounting for complex, non-linear relationships between the response outcome and the administrative variables. The effectiveness of the methods is evaluated on the basis of non-response bias in the administrative variables and via a proxy measure of the key survey variable—*vacancies*.

3.2 | Hypotheses of survey participation

Pertinent to research objective 3, we use administrative data to test the following hypotheses: Relevance of the survey topic, establishment size, establishment age, average establishment wages, workforce diversity, interaction of establishment age and average employee age, response history and development of the employment structure. Each hypothesis is motivated and described in the following subsections.

3.2.1 | Relevance of the survey topic

One of the most frequently studied hypotheses is whether the topic of the survey influences the participation decision. Most of the literature, qualitative and quantitative, shows evidence that the topic of the survey matters (HMRC, 2010; Snijkers, 2018; Snijkers et al., 2013). That is, motivation to participate is higher if the survey topic is highly relevant to the establishment. Vacancy surveys and their statistics are especially relevant for establishments with many vacancies or those that employ many marginal employees, who are prone to change their jobs frequently. As the number of vacancies cannot be derived from administrative data, our analysis relies on new hires as a proxy measure for vacancies. This is reasonable given that vacancies are likely to be converted into new hires in the future. Thus, we hypothesise that establishments with a higher share of new hires and marginal employees (as proxies for survey topic relevance) are more likely to participate.

Hypothesis 1. *The likelihood of participation increases with a higher share of new hires and a higher share of marginal employees.*

3.2.2 | Establishment size

As previously stated, the size of establishments, measured by the number of employees, likely affects the participation outcome. However, empirical evidence on the direction of the effect is mixed. Two studies show that larger firms are more likely to participate than smaller ones (Davis & Pihama, 2009; Seiler, 2014). They argue that employees of large firms have specialised roles grouped into clear structures, leading to well-defined lines of authority and increased capacity to respond. However, both surveys have special procedures for handling large companies, which could have impacted the results. Other studies show that smaller establishments are more likely to respond, especially in voluntary surveys (Earp et al., 2018; Hecht et al., 2019; Janik & Kohaut, 2012; König et al., 2021). They argue that within smaller establishments less coordination is needed to organise the response task, it is easier to identify a capable employee to respond, and the same person can decide whether to participate and also complete the interview. Since the IAB-JVS is a voluntary survey similar in design to the cited studies, we hypothesise a negative effect of establishment size on participation.

Hypothesis 2. *The likelihood of participation decreases with establishment size.*

3.2.3 | Establishment age

We posit an establishment age effect given that older, more entrenched firms are likely to have more experience and better infrastructure for handling information requests. In contrast, younger establishments face additional challenges that have higher priority than survey response. Accordingly, younger establishments are expected to have less motivation and capacity to respond. Although Hecht et al. (2019) and Foo et al. (2019) find no age effect, Phipps and Jones (2007) find a positive age effect. Hence, we hypothesise that older establishments are more likely to respond than younger ones.

Hypothesis 3. *The likelihood of participation increases with establishment age.*

3.2.4 | Average establishment wages

According to rent-sharing theory, higher wages are associated with more profitable enterprises (Blanchflower et al., 1996), which in turn could be associated with more efficient organisation of enterprises and better (data) management (Ogbadu, 2009). More efficient organisation and data infrastructure should decrease response burden, as the required information can be gathered faster. On the other hand, the profitability of an establishment could lie in better prioritisation of revenue-generating tasks. As survey participation does not directly affect a firm's revenue, profitable establishments might be less motivated to take part and give it a low priority. Phipps and Toth (2012) find support for this claim as they showed responding establishments have lower (average) wages than non-responding establishments in the BLS Occupational Employment Statistics survey. Based on the empirical evidence, we hypothesise that establishments with higher wages are less likely to participate.

Hypothesis 4. *The likelihood of participation decreases with average establishment wages.*

Workforce diversity

The public expects that businesses are part of, and contribute to, a functioning society, but businesses follow this norm with different intensities. Participating in surveys is one way of engaging with the general public and expressing social responsibility. Diversity management is seen as a related aspect to social corporate responsibility (Colgan & McKearney, 2011; Hansen & Seierstad, 2017; Starostka-Patyk et al., 2015). We posit that an establishment's willingness to engage with society is related their hiring preferences with regard to nationality, sex and education. We expect that establishments with little demographic diversity are less interested in social responsibility, which should translate into a lower likelihood of survey participation and the opposite for establishments with higher levels of diversity.

Hypothesis 5. *The likelihood of participation increases with the diversity of the workforce.*

Interaction of establishment age and average employee age

We expect that younger establishments with a younger workforce (e.g. start-ups) differ from older establishments employing an older workforce with respect to survey participation. In particular, the first priority of the former group is to increase market share with less priority and capacity allocated to completing voluntary survey tasks. Thus, we hypothesise that younger establishments with a younger workforce are less likely to participate.

Hypothesis 6. *The likelihood of participation decreases for younger establishments with a younger workforce, compared to older establishments with an older workforce.*

3.2.5 | Response history

Although the IAB-JVS is a yearly cross-sectional survey, there are several establishments which are sampled at a higher rate than others due to their size or industry type. We posit that receiving more participation requests for the same survey has a negative effect on response.

Repeated requests could lead to suspicion regarding the random selection procedure or increase the perceived response burden, thus decreasing the response propensity.

Hypothesis 7. *The likelihood of participation decreases if an establishment was sampled in the previous year, compared to an establishment that was not sampled.*

Despite the expected negative effect of the previous-year's survey request, we anticipate that establishments whom already participated in the previous year are more likely to do so again. These establishments are familiar with the survey and its questionnaire and have already established a response process. Hence, the response task may be less burdensome compared to establishments who did not previously participate and must process the survey request from scratch (Earp et al., 2018; Janik & Kohaut, 2012; Smaill, 2012).

Hypothesis 8. *The likelihood of participation increases if an establishment participated in the previous year, compared to an establishment that did not participate.*

3.2.6 | Development of employment structure

We expect that changes in the establishment that occurred prior to the survey request affect survey participation. For example, if the share of women in the establishment moves closer to 50%, compared to the previous year, this would reflect a development towards greater diversity. In line with the aforementioned diversity Hypothesis 5, we would therefore expect this development to translate into a higher likelihood of participation. Similarly, we expect that a strong wage growth is a sign of a more profitable establishment which is expected to have a negative effect on participation, as previously suggested by Hypothesis 4. Lastly, in line with the survey topic relevance hypothesis (Hypothesis 1), an increasing proportion of marginal employees or new hires (as a proxy for vacancies) could translate into the survey topic becoming more relevant to the establishment, due to an increasing number of job recruiting processes. We also consider changes that occurred after the survey request, as they likely reflect procedures being implemented at the time of the survey.

Hypothesis 9. *The development of the employment structure affects the likelihood of participation.*

Hypothesis 9 consists of the following sub-hypotheses. The sub-indicator refers to the relevant main hypothesis:

- *Hypothesis 9.1a: The likelihood of participation increases if the share of new hires and the share of marginal employees (as proxies for survey topic relevance) increased from the year before the survey $(t-1)$ to the survey year (t), compared to no change or a decreasing share of new hires and marginal employees.*
- *Hypothesis 9.1b: The likelihood of participation increases if the share of new hires and the share of marginal employees (as proxies for survey topic relevance) increased from the survey year (t) to the year after the survey $(t+1)$, compared to no change or a decreasing share of new hires and marginal employees.*
- *Hypothesis 9.4a: The likelihood of participation decreases if the average establishment wage increased from the year before the survey $(t-1)$ to the survey year (t), compared to no change of the average establishment wage or a decreasing average establishment wage.*

- *Hypothesis 9.4b: The likelihood of participation decreases if the average establishment wage increased from the survey year (t) to the year after the survey ($t + 1$), compared to no change of average establishment wage or a decreasing average establishment wage.*

- *Hypothesis 9.5a: The likelihood of participation increases if the diversity of the workforce increased from the year before the survey ($t - 1$) to the survey year (t), compared to no change of diversity or a decreasing diversity.*

- *Hypothesis 9.5b: The likelihood of participation increases if the diversity of the workforce increased from the survey year (t) to the year after the survey ($t + 1$), compared to no change of diversity or a decreasing diversity.*

4 | DATA AND METHODS

4.1 | Data

4.1.1 | IAB Job vacancy survey

The IAB-JVS is a voluntary nationally-representative establishment survey that quantifies the size of the unfilled labour demand and other worker flows in Germany (Bossler et al., 2020). It is carried out annually as a repeated cross-sectional survey using a concurrent mixed-mode design, with establishments receiving paper questionnaires and the option of online completion. Random samples of about 110,000 establishments are drawn each year from the population of all establishments in Germany that have at least one regular employee liable for social security contributions. The sampling frame is the population on the 31 December in the previous year. Using an expert allocation, samples are disproportionately stratified by region, industry and establishment size, resulting in unequal inclusion probabilities. The IAB-JVS is fielded every fourth quarter (October–December) with short re-interviews conducted via telephone in the subsequent three quarters to update the number of vacancies. Since our focus is on cross-sectional non-response we do not consider the re-interviews. We analyse survey years 2010–2019 only as it is not possible to link the IAB-JVS to administrative data for years prior to 2010.The data used in this study are available from the Research Data Centre (RDC) of the Federal Employment Agency in Germany. Restrictions apply to the availability of these data, which are not publicly available. For more information on data access, see https://fdz.iab.de/en.aspx.

4.1.2 | Administrative data

To analyse non-response bias in the IAB-JVS, each yearly sample is linked to administrative data of the Establishment History Panel (BHP) of the Federal Employment Agency (Ganzer et al., 2020). The BHP is a longitudinal administrative database compiled by aggregating individual records of all employees to the establishment level. The reference date for the aggregation is the 30 June every year. This means there is one observation per year, which reflects the establishment profile in the quarter immediately prior to the survey. Since the IAB-JVS sampling frame and the BHP contain the same unique identifier it is straightforward to link them for almost every establishment. Exceptions are establishments that cease to exist between the reference dates of the sample selection and the BHP, which applies to 3.4% of all establishments.

In addition, we make use of the Administrative Wage and Labor Market Flow Panel (AWFP), which is an aggregation of employment biographies of individual employees and subsidy recipients to the establishment level (Seth & Stüber, 2018). It captures similar characteristics to the BHP and some additional aspects (e.g. mean employment tenure, standard deviation of wage) which we exploit to validate the results of the BHP through additional sensitivity analyses. However, the shares of some employee characteristics (e.g. males/females) are calculated differently by using only regular workers and excluding marginal employees. A key advantage of the AWFP over the BHP is that it is calculated quarterly and therefore the fourth quarter, which overlaps exactly with the quarter of data collection, can be used in the validation analysis. Hence, the validation analysis assesses non-response bias at the same time period of the survey. A major drawback of the AWFP is its availability only until 2014.

Table 1 provides an overview of the variables and data sources used for each research objective. For the non-response bias analysis, we categorise all variables of interest into approximately equal-sized categories. Descriptive statistics are presented in the Section A of Appendix S1.

We note that all administrative variables used in the analysis are treated as proxy variables for the actual IAB-JVS survey variables. This is reasonable considering that the administrative variables are likely correlated with the multiple topics covered in the survey questionnaire, including the variety of questions on vacancies and recruiting processes. To give a few examples, establishment size and the number of new hires is likely correlated with the number of reported vacancies; the share of fixed-term employees should be correlated with the reported number of fixed-term employees in the survey; and the administrative wage information is correlated with survey variables on hiring wages and wage negotiation.

4.2 | Methods

4.2.1 | Response rates

The first research objective investigates response rates in the IAB-JVS, which we define as the number of completed interviews divided by the sample size. This definition is equivalent to *Response Rate 1* as defined by the American Association for Public Opinion Research (AAPOR) (2016). As the definition is based on the full sample, it is a conservative calculation and can be considered the minimum response rate. The stratified sampling design of the IAB-JVS has unequal inclusion probabilities between strata, which we take into account when calculating the response rates. Thus, we report the population response rate.

A distinction is made between the *drawn* sample and the *fielded* sample, which depends on the particular year of analysis. Some establishments from the drawn sample could not be fielded (e.g. invalid addresses) and had no chance to participate in the survey. These non-fielded establishments can be identified only for years 2016–2019 and are excluded from the analysis for these years. For years 2010–2015, only the drawn sample can be used as the basis since it is not possible to identify the non-fielded cases. Design weights are based on the drawn sample between 2010 and 2015 and on the fielded sample between 2016 and 2019. We believe this distinction does not substantially affect the interpretation of the results, as the share of non-fielded establishments is small (below five percent for each year) and sensitivity checks for RO2, RO3, and RO4 showed no systematic differences between these two sample definitions, and no large differences between the bridge years 2015 and 2016. In the remainder, we use the term *analytic* sample to refer to the compilation of the *drawn* sample for years 2010–2015 and the *fielded* sample for years 2016–2019.

TABLE 1 Variable and data set overview

Variable	Bias measure	Bias measure validation	Hypothesis testing	Current response propensity estimation	Extended Response Propensity Estimation
Research objective	RO2, RO4	RO2, RO4	RO3	RO4	RO4
Data set	BHP	AWFP	BHP	BHP	BHP
Variables from the survey year					
East/West Germany[a]	Dum.	Dum.	Dum.	Dum.	Dum.
Number of employees[a]	Cat.	Cat.	Cat.	Cat.	Con.
Establishment age	—	—	Con.	—	Con.
Foundation year [a]	Cat.	Cat.	—	—	—
Industry[a]	Cat.	Cat.	Cat.	Cat.	Cat.
Average age of employees[b]	Cat.	Cat.	Cat.	Cat.	Con.
Share of female[b]	Cat.	Cat.	—	—	Con.
Sex diversity	—	—	Cat.	—	—
Share of fixed-term[b]	Cat.	—	—	—	Con.
Share of apprentices[b]	Cat.	Cat.	Cat.	—	Con.
Share of full-time[b]	Cat.	Cat.	Cat.	—	Con.
Share of part-time[b]	Cat.	Cat.	Cat.	—	Con.
Share of Germans[b]	Cat.	—	—	—	Con.
Nationality diversity	—	—	Cat.	—	—
Share of regular[b]	Cat.	Cat.	Cat.	—	Con.
Share of marginal[b]	Cat.	Cat.	Cat.	Cat.	Con.
Share of high-educated[b]	Cat.	Cat.	—	—	Con.
Share of mid-educated[b]	Cat.	Cat.	—	—	Con.
Share of low-educated[b]	Cat.	Cat.	—	—	Con.
Educational diversity	—	—	Cat.	—	—
Share of unknown educated[b]	Cat.	Cat.	Cat.	—	—
Quartile of wage distribution[b]	Cat.	Cat.	Cat.	—	Cat.
Mean tenure[b]	—	Cat.	—	—	—
SD of wages[b]	—	Cat.	—	—	—
Sampled in $t-1$	—	—	Dum.	—	Dum.
Participated in $t-1$	—	—	Dum.	—	Dum.
Establishment foundation in $t-1$[a]	Dum.	Dum.	—	—	Dum.
Establishment closure in $t+1$[a]	Dum.	Dum.	—	—	Dum.

(Continues)

T A B L E 1 (Continued)

Variable	Bias measure	Bias measure validation	Hypothesis testing	Current response propensity estimation	Extended Response Propensity Estimation
Development variables					
Change of sex diversity	—	—	Cat.	—	—
Change of apprentices	—	—	Cat.	—	—
Change of full-time	—	—	Cat.	—	—
Change of part-time	—	—	Cat.	—	—
Change of nationality diversity	—	—	Cat.	—	—
Change of regular	—	—	Cat.	—	—
Change of marginal	—	—	Cat.	—	—
Change of educational diversity	—	—	Cat.	—	—
Change of unknown educated	—	—	Cat.	—	—
Wage growth	—	—	Cat.	—	—
Change of average age of employees	—	—	Cat.	—	—
Share of new hirings	—	—	Cat.	—	—

Abbreviations: Cat., categorised variable; Con., continuous variable; Dum., dummy variable;
[a]Establishment characteristic.
[b]Employee characteristic.

4.2.2 | Non-response bias calculation

The second research objective pertains to non-response bias. Non-response bias is computed as the difference between the estimate of interest based on the set of respondents and the corresponding estimate based on the full sample:

$$\widehat{\text{NR bias}}_i = \hat{Y}_{i,r} - \hat{Y}_{i,n}, \tag{1}$$

where $\hat{Y}_{i,r}$ denotes the estimator for the ith statistic of interest based on the respondents and $\hat{Y}_{i,n}$ is the estimator based on the full sample.

Non-response bias is estimated for each category of each administrative variable shown in Table 1 (columns 2 and 3). As all biases are based on proportions, they can be compared on the same scale. Additionally, we construct and compare measures of absolute bias and average absolute bias, where absolute non-response bias is defined as:

$$\widehat{\text{Abs. NR bias}}_i = \left| \widehat{\text{NR bias}}_i \right|, \tag{2}$$

and average absolute non-response bias is defined as:

$$\widetilde{\text{Avg. abs. NR bias}} = \frac{\sum_{i=1}^{K} \widehat{\text{Abs. NR bias}_i}}{K}, \tag{3}$$

where K is the total number of variable categories for which non-response bias is estimated.

Average absolute non-response bias is calculated across all variables and separately for two variable groups: establishment characteristics and employee characteristics (see Table 1). Separating these variable groups sheds light on which one is most impacted by non-response bias.

These three measures are used to assess non-response bias and corresponding non-response bias trends in the IAB-JVS (RO2) and examine the performance of the various non-response adjustment models (RO4). As some variables are not available in 2010 and 2019 (e.g. establishment closure in $t + 1$, share of fixed-term employees), the non-response bias analysis is restricted to years 2011–2018. As a robustness check, we also estimate absolute relative non-response biases and report them in the Appendix S1 (see Section F.5). All non-response bias estimates are design weighted to account for unequal inclusion probabilities.

4.2.3 | Modelling survey participation

RO3 tests hypotheses of survey participation using a series of logistic regressions modelling response ($1 = response$, $0 = non\text{-}response$) for each survey year. Each model specification builds on the previous one by cumulatively adding more explanatory variables. Model 1 consists of the current set of IAB-JVS weighting variables. Model 2 adds static variables which are measured during the survey year (t). Model 3 adds the development variables which reflect changes in the establishment since the previous year ($t - 1$). This is followed by Model 4, which includes development variables reflecting subsequent changes in the establishment from the survey year until the following year ($t + 1$). Additional control variables about the establishment (e.g. industry, region, share of full-time employees) are included in all four models presented below.

Model 1: The current IAB-JVS logistic regression model for estimating response propensities:

$$\text{logit}(R_{k,t}) = \alpha + \boldsymbol{\beta}_1 \mathbf{x}_{1,k,t}^\top + \boldsymbol{\gamma}_1 \mathbf{z}_{1,k,t}^\top, \tag{4}$$

where $R_{k,t}$ is the response indicator for the kth establishment ($R_k = 1 = response$, $R_k = 0 = non\text{-}response$) in survey year t, $\mathbf{x}_1$ is a vector of current IAB-JVS weighting variables, and $\mathbf{z}_1$ is a set of additional control variables.

Model 2: Extended response model with static variables:

$$\text{logit}(R_{k,t}) = \alpha + \boldsymbol{\beta}_1 \mathbf{x}_{1,k,t}^\top + \boldsymbol{\beta}_2 \mathbf{x}_{2,k,t}^\top + \boldsymbol{\gamma}_1 \mathbf{z}_{1,k,t}^\top, \tag{5}$$

where $\mathbf{x}_2$ includes the extended set of administrative (static) variables.

Model 3: Extended response model with static variables and previous-year change variables:

$$\text{logit}(R_{k,t}) = \alpha + \boldsymbol{\beta}_1 \mathbf{x}_{1,k,t}^\top + \boldsymbol{\beta}_2 \mathbf{x}_{2,k,t}^\top + \boldsymbol{\beta}_3 \mathbf{x}_{3,k,t-1}^\top + \boldsymbol{\gamma}_1 \mathbf{z}_{1,k,t}^\top + \boldsymbol{\gamma}_2 \mathbf{z}_{2,k,t-1}^\top, \tag{6}$$

where $\mathbf{x}_3$ includes administrative change variables from $t - 1$ and $\mathbf{z}_2$ is a set of control variables reflecting change from $t - 1$.

Model 4: Extended response model with static variables and previous- and subsequent-year change variables:

$$
\begin{aligned}
\text{logit}(R_{k,t}) = \alpha &+ \beta_1 \mathbf{x}_{1,k,t}^\top + \beta_2 \mathbf{x}_{2,k,t}^\top + \beta_3 \mathbf{x}_{3,k,t-1}^\top + \beta_4 \mathbf{x}_{4,k,t+1}^\top \\
&+ \gamma_1 \mathbf{z}_{1,k,t}^\top + \gamma_2 \mathbf{z}_{2,k,t-1}^\top + \gamma_3 \mathbf{z}_{3,k,t+1}^\top,
\end{aligned}
\tag{7}
$$

where $\mathbf{x_4}$ includes administrative change variables from $t + 1$ and $\mathbf{z_3}$ is a set of control variables reflecting change from $t + 1$.

Design weights (i.e. inverse inclusion probabilities) and strata are accounted for in all model estimations. As the estimated model coefficients and test statistics were found to be stable over the years, we also report the pooled-data results. By using pooled data and controlling for year effects in the logistic regression, we assume stable effects of our variables of interest within the observation period. With more observation years available, one could consider fitting a multi-level model to account for year-specific effects. As a robustness check, we also estimated a random intercept model with respondents clustered within years. The results of the random intercept model supported the results of the logistic regression model using pooled data.

To facilitate comparisons between the different model specifications, the analytic sample is restricted to all establishments with observed variables for the survey year, the year before the survey, and the year after the survey. Thus, the number of observations is held constant for every model specification. Sensitivity checks incorporating establishments with missing variable information at $t - 1$ and/or $t + 1$ did not affect the study conclusions (results not shown).

4.2.4 | Response propensity models and adjustment weights

The fourth research objective (RO4) investigates whether including the extended set of administrative variables in the response propensity estimation improves non-response bias reduction relative to the current set of IAB-JVS auxiliary variables. To do this, two separate logistic regression models are fitted: one using only the current IAB-JVS weighting variables and the other adding the extended administrative variables. The resulting weights derived from both models are then compared in terms of their bias-reducing performance.

More complex models are also evaluated, including several data-driven modelling methods, such as Lasso regression, generalised additive models and supervised machine learning (ML) algorithms. All of these methods are applied to estimate response propensities based on the full set of current and extended administrative auxiliary variables. In sum, the following modelling approaches are evaluated:

- Logistic regression (with and without extended administrative variables) (Cox, 1958)
- Lasso regression (with second order polynomials) (Lasso) (Tibshirani, 1996)
- Ridge regression (with second order polynomials) (Ridge) (Hoerl & Kennard, 1970)
- General additive model (GAM) (Hastie & Tibshirani, 1990)
- Generalised Additive Model Selection (GAMSEL) (Chouldechova & Hastie, 2015)
- Decision tree using the CART algorithm (CART) (Breiman et al., 1984)
- Decision tree using the C-Tree algorithm (C-Tree) (Hothorn et al., 2006)
- Model-based recursive partitioning (MOB) (Zeileis et al., 2008)

- Random Forest (Breiman, 2001)
- Extreme Gradient Boosting (XG-Boost) (Chen & Guestrin, 2016)
- Bayesian additive regression trees (BART) (Chipman et al., 2010)

As the goal is not to predict out-of-sample non-response, but to estimate response probabilities based on the explanatory variables, the data are not split into test and training sets. That is, the complete data are used both for training the models and estimating the response propensities. To optimise the CART, C-Tree, XG-Boost algorithms a hyper-parameter tuning is performed by conducting a grid search on various parameter settings with fivefold cross validation. The BART algorithm is applied with the default setup and Random Forest with specific selected parameters to avoid overfitting. Table E.1 in Appendix S1 provides an overview of the parameters used for the machine learning models. We follow Lohr et al. (2015) and estimate response propensities without using design weights, knowing that this implicitly assumes that our sampling design is non-informative for the response indicator (i.e. inclusion probabilities are unrelated to the response indicator, Pfeffermann, 2011). Since we control for the variables used to create the sampling strata the effect of a possibly informative design is mitigated. The full analysis is implemented in Stata (StataCorp, 2019) and in R (R Core Team, 2019) using the packages *glmnet* (Friedman et al., 2010), *gam* (Hastie, 2019), *gamsel* (Chouldechova & Hastie, 2015), *rpart* (Therneau & Atkinson, 2019), *partykit* (Hothorn & Zeileis, 2015), *randomForest* (Liaw & Wiener, 2002), *xgboost* (Chen et al., 2019), *bartMachine* (Kapelner & Bleich, 2016), and *caret* (Kuhn, 2020). The code that was used to analyse the data can be obtained from Appendix S1.

To avoid overfitting, each target administrative variable for which non-response bias is assessed is left out of the corresponding set of explanatory variables for the response propensity estimation. This "leave-one-out" approach results in different sets of response propensities estimated for each establishment corresponding to each target variable of interest. As the proportion of unknown educated employees is co-linear with the proportions of low-educated, middle-educated and high-educated employees, this variable is left out of the explanatory set for all response propensity estimations. The adjustment weight for this outcome variable is based on the full set of explanatory variables. The inverse of these propensities are the raw non-response weights. To reduce the variance of the raw weights, they are trimmed at the 99th percentile. The final adjustment weights are constructed by multiplying the non-response weight with the design weight. The adjustment weights are then used to compute weighted estimates of the corresponding target administrative variables. Non-response bias is assessed by comparing the non-response-adjusted weighted estimates under each modelling approach against the design-weighted benchmark values. This comparison provides information about which modelling approaches perform best in terms of reducing non-response bias.

The same set of explanatory variables are used in all modelling approaches. In contrast to the models used to test the survey participation hypotheses (RO3), the continuous variables are not categorised to allow the machine learning algorithms to make use of the full depth of information. The traditional response propensity estimation implemented in the IAB-JVS is based on categorised variables. In order to ensure a fair comparison with the machine learning methods, we use the continuous versions of these variables for all modelling approaches. For the Lasso and Ridge regressions second-order interactions and quadratic terms are included in the set of explanatory variables. To control for outliers, establishment size is top-coded at 20,000 employees.

5 | RESULTS

5.1 | Response rates in the IAB-JVS between 2010 and 2019

Figure 1 shows the design-weighted response rates of the IAB-JVS for years 2010–2019. The corresponding table can be found in Section B of Appendix S1. One can see that the yearly response rates have always been below 21% since 2010. Over these years, the response rate has dropped from 18.87% (2010) to 14.65% (2019), representing an average design-weighted response rate of 16.40% and an average decline of 0.4 percentage points per year. A stabilising trend is observed since 2016, which is the first year the fielded sample is analysed (as opposed to the drawn sample). However, there are signs that this trend is not purely driven by the change in sample type, as the field reports also indicate a stabilising trend with less decline in recent years (see Section B of Appendix S1). This decline is even more evident when looking at the response rates based on the field reports since 1989 (see Figure B.1 of Appendix S1). The unweighted response rates declined from 40.1% in 1989 to 20.4% in 2009. In sum, the response rates of the IAB-JVS can be considered low compared to other establishment surveys worldwide. Moreover, the general decreasing trend in the IAB-JVS is consistent with indications of declining participation in other establishment surveys (see Section 2.2).

5.2 | Non-response bias

Figure 2 illustrates the average absolute non-response bias, estimated using only design weights, between 2011 and 2018 for all administrative variables, and separately for establishment and employee characteristics. Across all variables, the average absolute non-response bias lies between 1.37% (2012) and 1.74% (2015) across the 8 years without any noticeable trend over time. These aggregate values are considered rather small. Given the low response rates reported earlier, it is reassuring that the aggregate bias is not particularly high. The subset of employee and establishment characteristics range between 1.18 (2011) and 1.63 (2017)% and 1.76 (2012) and 2.08 (2015)%, respectively. Thus, the establishment characteristics tend to be more impacted by non-response bias than the employee characteristics.
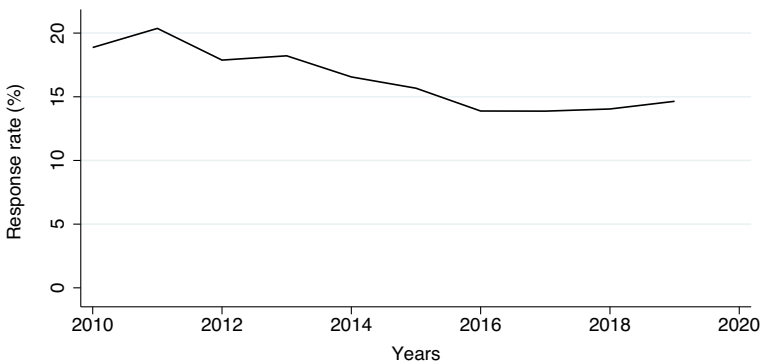

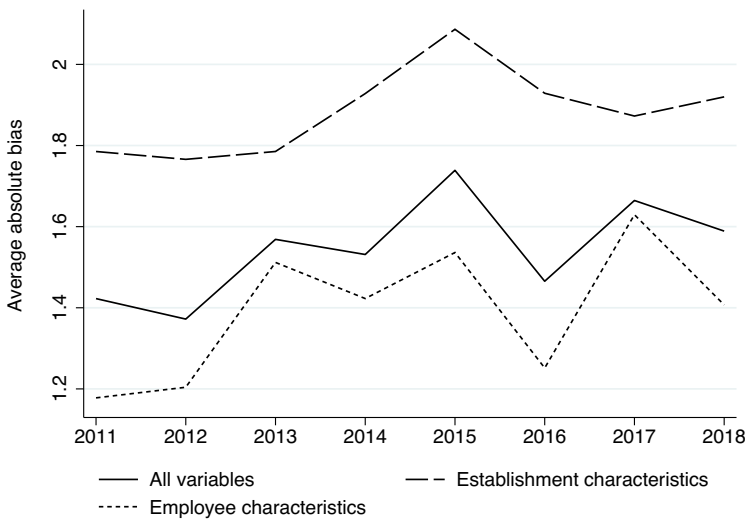**FIGURE 1**   Design weighted response rates, 2010–2019 IAB-JVS

F I G U R E 2   Design weighted average absolute bias by all administrative variables, establishment characteristics and employee characteristics, BHP 2011–2018

With respect to the 56 individual bias estimates (see Section C.1 of Appendix S1), the number of those that exceeded an arbitrary threshold of 2% ranged from 13 (2012) to 21 (2015) across the years, with more such bias estimates occurring in the later years than in the earlier years. There are particularly large biases for industry groups, establishment foundation year, and indicators of establishment closure in $t + 1$, reaching up to 6.5%. Other large biases are observed for the mean age of employees, the share of high-educated employees, and the share of German employees with values up to 5.5% (see Section E.4 of Appendix S1 for detailed information on individual bias estimates). Similar patterns of bias are also observed for the absolute relative non-response bias and additional validation data (see Section C.2 of Appendix S1).

5.3 | Hypotheses of survey participation

Table 2 presents the results of the four response models used to test the survey participation hypotheses. In addition to odds-ratios, average response propensities are shown to allow readers to assess the effect size of the predictor variables. As the results do not differ systematically between years, only the pooled results (i.e. across all years) are shown. The random intercept model (robustness check), the separate yearly regression results, and a yearly summary are displayed in the Section D of Appendix S1. Compared to the current IAB-JVS response model (Model 1), there are improvements in model fit when the extended set of (static) administrative variables are added to the model (Model 2). However, the additional effects of the developmental variables on model fit (Models 3 and 4) are negligible. The full model (Model 4) explains only little variation in the response outcome (Pseudo - R^2 of 0.025).

Next, we turn to the hypothesis testing results. Table 3 provides a short summary of the hypothesis testing results based on the pooled-data analysis and a significance level of 5%. An extended summary table, including operationalisation, sub-hypotheses, and potential effect sizes are shown in the Table D.2 of Appendix S1. We expected that establishments with a higher share of new hires

TABLE 2 Logistic regression models of survey participation (pooled results)

DV: Response	Model 1 OR	Model 1 (SE)	Model 2 OR	Model 2 (SE)	Model 3 OR	Model 3 (SE)	Model 4 OR	Model 4 (SE)	RPr
Number of employees (Ref.: 1–9)									18.14
10–19	0.914***	(0.022)	0.924**	(0.024)	0.968	(0.027)	0.964	(0.028)	17.47
20–49	0.854***	(0.018)	0.875***	(0.023)	0.917**	(0.026)	0.913**	(0.027)	16.95
50–249	0.733***	(0.016)	0.766***	(0.025)	0.787***	(0.028)	0.778***	(0.029)	15.14
250–499	0.563***	(0.018)	0.612***	(0.028)	0.600***	(0.029)	0.586***	(0.031)	11.57
500–999	0.487***	(0.019)	0.563***	(0.030)	0.542***	(0.031)	0.527***	(0.032)	10.39
≥1000	0.523***	(0.024)	0.624***	(0.038)	0.587***	(0.038)	0.572***	(0.039)	11.35
Average employees age (Ref.: 0–39)									13.73
39.00–43.49	1.184***	(0.029)	1.166***	(0.052)	1.153**	(0.052)	1.145**	(0.051)	16.19
43.50–47.99	1.211***	(0.030)	1.130**	(0.051)	1.101*	(0.051)	1.085	(0.050)	17.32
≥48.00	1.271***	(0.031)	1.285***	(0.056)	1.229***	(0.054)	1.207***	(0.054)	18.76
Share of marginal employees (Ref.: 0)									15.90
0.00–14.99	1.196***	(0.030)	1.226***	(0.041)	1.222***	(0.048)	1.189***	(0.051)	16.88
≥15	1.053*	(0.022)	1.210***	(0.049)	1.217***	(0.054)	1.192***	(0.056)	16.61
Wage distribution (Ref.: first quartile)									15.43
Second quartile	1.066*	(0.030)	1.038	(0.030)	1.043	(0.030)	1.037	(0.030)	16.68
Third quartile	1.156***	(0.035)	1.098**	(0.034)	1.099**	(0.034)	1.089**	(0.034)	17.72
Fourth quartile	1.060	(0.033)	0.998	(0.032)	0.992	(0.032)	0.980	(0.032)	15.88
Missings	1.032	(0.031)	1.114**	(0.039)	1.019	(0.074)	1.142	(0.103)	17.24
Establishment age			1.009***	(0.002)	1.009***	(0.002)	1.008***	(0.002)	16.34
Employees age × establishment age									[a]
39.00–43.49 × establishment age (continuous)			0.999	(0.002)	0.999	(0.002)	0.999	(0.002)	[a]
43.50–47.99 × establishment age (continuous)			1.001	(0.002)	1.001	(0.002)	1.001	(0.002)	[a]
≥48.00 × Establishment age (continuous)			0.996	(0.002)	0.996	(0.002)	0.996	(0.002)	[a]

(Continues)

TABLE 2 Continued

	Model 1		Model 2		Model 3		Model 4		
	OR	(SE)	OR	(SE)	OR	(SE)	OR	(SE)	RPr
Sex diversity (Ref.: Non-diverse)									
Slightly diverse establishments			1.021	(0.021)	1.040	(0.024)	1.040	(0.025)	16.82
Diverse establishment			1.068*	(0.030)	1.082**	(0.033)	1.086**	(0.034)	16.34
Nationality diversity (Ref.: Non-diverse)									16.07
Diverse			0.868***	(0.020)	0.920**	(0.025)	0.936*	(0.028)	18.75
Educational diversity (Ref.: low level of diversity)									14.01
Medium level			1.017	(0.023)	1.052*	(0.026)	1.059*	(0.028)	17.56
High level			1.052	(0.027)	1.086**	(0.031)	1.104***	(0.033)	16.82
Sampled in $t-1$ (Ref.: not sampled)									15.25
Sampled			0.450***	(0.012)	0.450***	(0.012)	0.450***	(0.012)	17.15
Participated in $t-1$ (Ref.: not participated)									14.63
Participated			10.454***	(0.412)	10.428***	(0.412)	10.418***	(0.414)	14.90
Share of new hirings in t (Ref.: 0–6.25)									49.75
6.26–15.37					0.922**	(0.027)	0.919**	(0.027)	19.17
15.38–29.51					0.900***	(0.026)	0.908***	(0.026)	16.99
≥29.52					0.868***	(0.025)	0.882***	(0.025)	15.74
Δ_{t-1} Sex diversity (Ref.: no change)									13.98
Decrease					0.986	(0.026)	0.985	(0.027)	17.71
Increase					0.973	(0.027)	0.977	(0.028)	15.93
Δ_{t-1} Nationality diversity (Ref.: no change)									15.78
Decrease					0.919**	(0.029)	0.929*	(0.030)	17.77
Increase					0.940	(0.031)	0.955	(0.032)	13.84
Δ_{t-1} Education diversity (Ref.: no change)									13.52
Decrease					0.980	(0.028)	0.986	(0.029)	17.91
Increase					0.951	(0.026)	0.953	(0.026)	15.97
									15.64

(Continues)

TABLE 2 Continued

	Model 1		Model 2		Model 3		Model 4		RPr
	OR	(SE)	OR	(SE)	OR	(SE)	OR	(SE)	
Δ_{t-1} Share of marginal employees (Ref.: no change)									16.66
Decrease					1.037	(0.035)	1.030	(0.035)	16.24
Increase					1.013	(0.037)	1.012	(0.037)	16.40
Δ_{t-1} Employees age (Ref.: stable age)									15.59
More than a half year older					0.968	(0.025)	0.970	(0.025)	16.82
More than a half year younger					1.027	(0.029)	1.026	(0.029)	16.62
Δ_{t-1} Wage growth (Ref.: no change)									17.13
Negative wage growth					0.957	(0.029)	0.960	(0.030)	15.85
Positive wage growth of up to 5%					0.998	(0.029)	0.997	(0.029)	16.69
Positive wage growth of more than 5%					0.978	(0.028)	0.983	(0.029)	16.11
Wage missing value					1.025	(0.071)	1.051	(0.073)	17.02
Share of new hirings in $t+1$ (Ref.: 0–5.95)									18.64
5.96–14.99							1.035	(0.031)	17.54
15.00–27.99							0.962	(0.029)	15.65
≥28.00							0.920**	(0.027)	14.00
Δ_{t+1} Sex diversity (Ref.: no change)									17.47
Decrease							0.995	(0.028)	15.90
Increase							1.027	(0.028)	16.04
Δ_{t+1} Nationality diversity (Ref.: no change)									17.75
Decrease							0.960	(0.034)	13.69
Increase							0.987	(0.029)	13.92
Δ_{t+1} Education diversity (Ref.: no change)									17.71
Decrease							0.971	(0.028)	15.83
Increase							1.022	(0.028)	15.93

(Continues)

TABLE 2 Continued

	Model 1		Model 2		Model 3		Model 4		
	OR	(SE)	OR	(SE)	OR	(SE)	OR	(SE)	RPr
Δ_{t+1} Share of marginal employees (Ref.: no change)									16.52
Decrease							1.037	(0.036)	16.29
Increase							1.052	(0.035)	16.54
Δ_{t+1} Employees age (Ref.: stable age)									15.52
More than a half year older							1.008	(0.027)	16.94
More than a half year younger							1.005	(0.029)	16.45
Δ_{t+1} Wage growth (Ref.: no change)									16.93
Negative wage growth							0.992	(0.031)	16.03
Positive wage growth of up to 5%							1.014	(0.030)	16.68
Positive wage growth of more than 5%							1.020	(0.030)	16.17
Wage missing value							0.839**	(0.054)	16.75
Constant	0.363***	(0.017)	0.249***	(0.018)	0.293***	(0.023)	0.316***	(0.027)	
Observations	636105		636105		636105		636105		
Pseudo R^2	0.013		0.023		0.024		0.025		
AIC	12,916,940		12,788,558		12,771,784		12,761,739		
BIC	12,917,475		12,789,330		12,772,863		12,763,125		

Notes: Control variables: West/East Germany, industry, year, share of unknown qualification, Δ_{t-1} share of part-time, share of full-time, share of part-time, share of apprentices, share of regular, share of fixed-term, Δ_{t-1} share of unknown qualification, Δ_{t-1} share of part-time, Δ_{t-1} share of full-time, Δ_{t-1} share of apprentices, Δ_{t-1} share of regular, Δ_{t-1} share of fixed-term, Δ_{t+1} share of unknown educated, Δ_{t+1} share of part-time, Δ_{t+1} share of full-time, Δ_{t+1} share of apprentices, Δ_{t+1} share of regular, Δ_{t+1} share of fixed-term.

Abbreviations: OR, Odds Ratio; RPr, Average Response Propensity by Category; SE, Standard Error.

[a]Average response propensity for interaction terms does not provide useful insights.

Source: BHP & IAB-IVS 2011–2018 (pooled).

TABLE 3 Hypotheses and findings

Hypothesis	Pooled Result
Hypothesis 1: The likelihood of participation increases with a higher share of new hires and a higher share of marginal employees.	*mix.*
Hypothesis 2: The likelihood of participation decreases with establishment size.	✓
Hypothesis 3: The likelihood of participation increases with establishment age.	✓
Hypothesis 4: The likelihood of participation decreases with establishment wages.	✗
Hypothesis 5: The likelihood of participation increases with the diversity of the workforce.	*mix.*
Hypothesis 6: The likelihood of participation decreases for younger establishments with a younger workforce, compared to older establishments with an older workforce.	✗
Hypothesis 7: The likelihood of participation decreases if an establishment was sampled in the previous year, compared to an establishment that was not sampled.	✓
Hypothesis 8: The likelihood of participation increases if an establishment participated in the previous year, compared to an establishment that did not participate.	✓
Hypothesis 9: The development of the employment structure affects the likelihood of participation.	(✓)

Notes: Significance Level (α) = 5%.

Abbreviations: ✗, Rejected; ✓, Supported; (✓), Partially supported (*Partially supported* refers to the situation where only one of the hypothesised variables yields a statistically significant result.); *mix.*, Mixed results (*Mixed results* refers to the case where multiple hypothesised variables yield statistically significant results in both directions.).

(a proxy for job vacancies), an indication of greater topic relevance, would be positively associated with response (Hypothesis 1). The results do not confirm this relationship. Even more, we find that establishments with a higher share of hirings are less likely to respond. However, the second operationalisation, which is based on the share of marginal employees, supports the posited hypothesis. Compared to establishments without any marginal employees, those with are more likely to participate. In line with the results of Janik and Kohaut (2012), Earp et al. (2018) and König et al. (2021) establishments with more employees are less likely to participate than those with fewer employees, which supports Hypothesis 2. Older establishments are more likely to respond than younger establishments, supporting hypothesis (Hypothesis 3). Relatedly, the interaction effect of establishment age and the average age of employees is not statistically significant, yielding no support for Hypothesis 6.

Regarding the relationship between survey participation and wages, the results do not support Hypothesis 4. Establishments whose median wages belong to the third quartile of the wage distribution are significantly more likely to participate compared to the first quartile, but the fourth quartile is not significantly different from the first quarter. Thus, there is no support for Hypothesis 4.

The three diversity measures, which capture the corporate social responsibility of the establishment, indicate different relationships. While education and sex diversity support the hypothesis that the social responsibility of an establishment is positively associated with response (Hypothesis 5), establishments that are diverse with respect to the nationality of their workforce are less likely to respond. The effect size of these associations is rather small and partly insignificant for some years. Nonetheless, these findings are the first (partial) evidence of a positive effect of corporate social responsibility on establishment survey participation.

The response history variables clearly confirm the posited relationships. There is a strong negative effect of the sampling indicator (Hypothesis 7) on participation, suggesting that establishments who received a survey request in the previous year are less likely to participate in the current year, supporting Hypothesis 7. In addition, there is strong evidence that participation in the previous year is positively associated with response in the current year, lending support to Hypothesis 8.

Turning to the development of the employment structure, the majority of coefficients show no significant association with response, indicating that changes in the establishment within the preceding or subsequent year are unrelated to response (Hypothesis 9). Only the development of nationality diversity in $t-1$ shows a significant effect, suggesting that development towards lesser diversity is associated with a lower likelihood of response. Thus, there is support for sub-hypothesis Hypothesis 9.5a. Overall, there is only partial support for the global development hypothesis (Hypothesis 9).

5.4 | Evaluation of non-response bias adjustments

Lastly, we evaluate the potential of using extensive administrative data and machine learning algorithms to adjust for non-response bias. Four bias measures are computed before and after the adjustments: average absolute bias, the number of individual significant biases, the mean squared error and the magnitude of bias in the mean number of new hires in $t+1$ (a key proxy measure for vacancies in the current survey year).

Figure 3 shows the average absolute bias for each year between 2011 and 2018 and for each modelling approach used to estimate propensity score weights. The unadjusted bias value, which is measured with design weights only, is also shown as a reference (Bar 1). The corresponding tables for average and individual biases are displayed in Sections E.3 and E.4 of Appendix S1, respectively.

As expected, the inclusion of the extended set of administrative variables (Bar 3) in the traditional logistic regression model improves non-response bias reduction in each survey year relative to the current IAB-JVS auxiliary variables (Bar 2). In general, all modelling approaches reduce non-response bias for each year. With regard to the machine learning algorithms, random forest (Bar 11), XG-Boost (Bar 12) and the BART (Bar 13) algorithms compete well with traditional logistic regression (Bar 3), with no clear-cut winner among these tree ensemble methods. Lasso (Bar 4), Ridge (Bar 5) and GAM (Bar 6) all perform similar to logistic regression (Bar 3) in all years. Gamsel (Bar 7) performs less well than the other regression approaches. Out of the three single tree methods CART (Bar 8), C-Tree (Bar 9) and MOB (Bar 10), CART model performs worst in terms of bias reduction and C-Tree slightly outperforms MOB. These conclusions hold when comparing absolute relative non-response biases and analysing the validation data set (see Section F of Appendix S1).

The same patterns are present when analysing average bias separately for the establishment and employee characteristics (see Section E.3 of Appendix S1). That is, the extended set of administrative variables perform better than the current weighting variables in the traditional logistic regression model, and the regression and ensemble tree methods perform better than the single tree methods. These patterns are generally similar for the individual bias estimates (see Section E.4 of Appendix S1). However, there are some methods that perform better than others for some variables and years. Such an example is the Random Forest algorithm, which reduces bias in
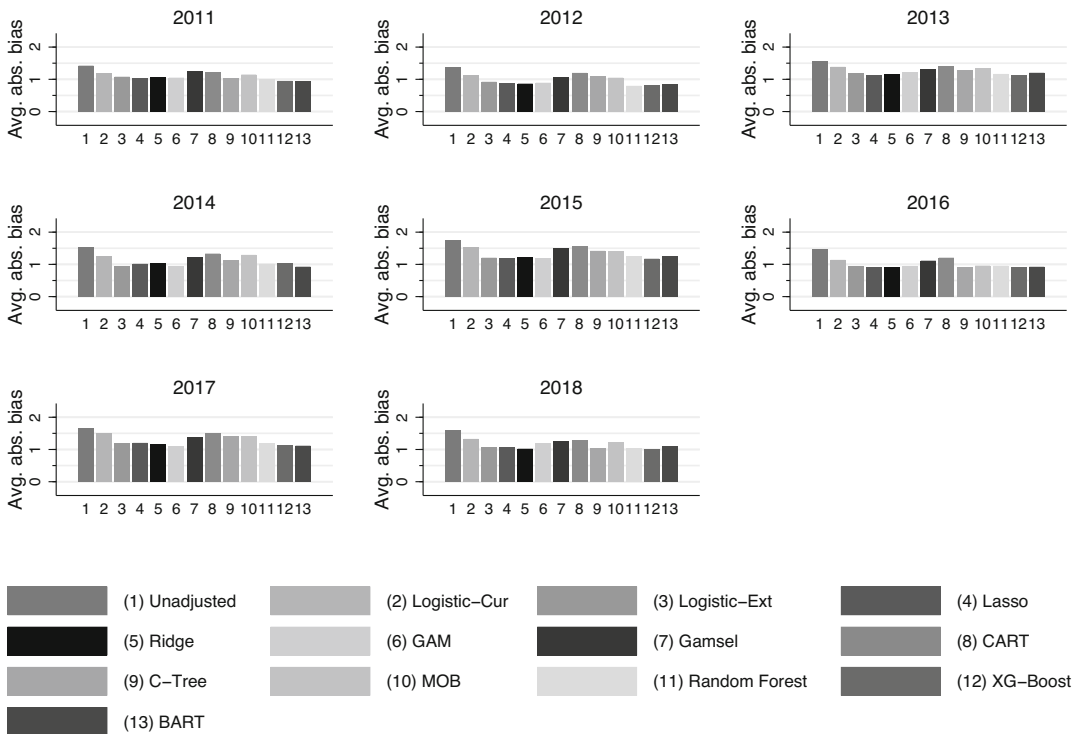
FIGURE 3 Average absolute bias by modelling approach (see p. 17) and set of explanatory variables, BHP 2011–2018

estimates of the number of employees to a greater extent than all other methods (except in year 2017).

Next, the performance of the weighting strategies are compared in terms of their ability to reduce the number of individual statistically significant non-response biases. A non-response bias is statistically significant if the full sample estimate of the target variable lies outside the confidence interval of the weighted respondent estimate. Standard Errors are derived using a linearisation-based variance estimator and Wald confidence intervals are used. Stratification effects are accounted for in the variance estimation. Section E.5 of Appendix S1 shows the number of significant bias estimates by year and by weighting strategy. In the unadjusted scenario, which again serves as the benchmark, between 16 (2012) and 31 (2015) out of 56 bias estimates are significant in each year, resulting in an average of 22.38 across the years. Including only the current set of IAB-JVS auxiliary variables in the standard logistic regression weighting procedure reduces the average number of statistically significant non-response biases to 16.38, a reduction of six estimates across the years. Including the extended set of administrative variables in the response propensity estimation further reduces this number to 9.38, a reduction of 7 estimates across the years compared to the model with current weighting variables. The XG-Boost algorithm performs best in terms of reducing the average number of significant non-response biases (5.88). Lasso (6.00), Ridge (6.63), Random Forest (7.38) and BART (7.75) are the runners-up, followed by standard logistic regression (9.38), C-Tree (9.50) and GAM (10.50). These methods perform better than the other single-tree methods—CART (17.50) and MOB (14.00)—and Gamsel (14.0), with CART being the least performing method.

For a combined assessment of the weighting schemes on non-response bias and variance, we also analyse the mean squared error (MSE). The MSE is estimated as the sum of the variance and the squared non-response bias estimated under each weighting approach. For a more detailed description and the corresponding tables and figures see Section F.6 of Appendix S1. The results do not differ from the aforementioned metrics and are consistent with the conclusions previously drawn. The extended use of the administrative data leads to a lower MSE, and regression and ensemble-tree methods outperform single-tree methods in reducing MSE.

Lastly, we compare the methods in terms of reducing non-response bias in the mean number of establishment new hires in $t + 1$, which is a key proxy for the number of job vacancies in the survey year (see also Section E.2 of Appendix S1). As this target variable is not part of the response propensity estimation, the weights are based on all available explanatory variables, without any "leave-one-out" procedure. Figure 4 shows the (unadjusted) mean number of new hires in the full sample and the weighted mean new hires for all models used to create the response propensity weights. The tabular values are provided in Section E.6 of Appendix S1. The horizontal reference line represents the full sample estimate. Values below the reference line indicate an underestimation of mean new hires, while values above the reference line indicate an overestimation of mean new hires.

All weighted values underestimate mean hirings in $t + 1$. Although there is some variation in the performance of the weighting strategies from year to year, the pattern is fairly consistent and resembles the pattern observed for the previous three bias measures. In particular, the positive impact of including the additional administrative variables in the traditional logistic
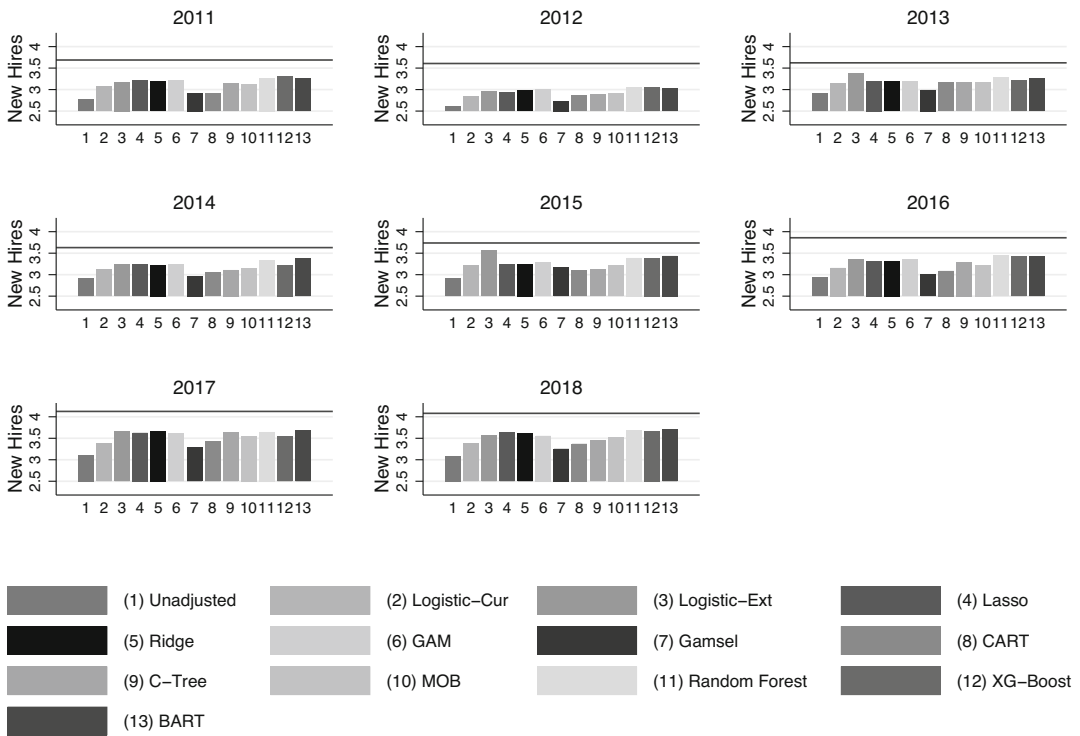

FIGURE 4 Mean number of new hires in $t + 1$ by year and modelling approach, BHP 2011–2018. Note: The horizontal line represents the number of new hires in the full sample.

regression weighting procedure (Bar 3) persists when compared to the using only the current IAB-JVS weighting variables (Bar 2). Turning to the comparison of machine learning methods, logistic regression (Bar 3), Lasso (Bar 4), Ridge (Bar 5), GAM (Bar 6), Random Forest (Bar 11), XG-Boost (Bar 12) and BART (Bar 13) all perform very well in reducing the discrepancy between the weighted and full sample means. In 2015 logistic regression does a remarkably good job and reduces the non-response bias almost entirely. The next best performing group of algorithms consists of C-Tree (Bar 9) and MOB (Bar 10). The CART (Bar 8) and Gamsel (Bar 7) algorithms perform the worst, on average. However, all methods reduce the non-response bias at least somewhat.

To conclude, the ensemble tree methods (Random Forest, BART, XG-Boost) slightly outperform the traditional logistic regression and general additive regression weighting procedures in some years and for some bias measures. However, logistic regression and the other regression approaches (Lasso, Ridge, GAM) perform remarkably well and even better than some machine learning algorithms (CART, C-Tree, MOB, GAMSEL).

6 | DISCUSSION

This article evaluated the use of extensive administrative data and machine learning techniques for analysing and adjusting for the effects of non-response in a large-scale job vacancy survey, that is, the IAB-JVS. The response rate of the IAB-JVS has been declining by about a half percentage point per year since 2010, which is indicative of similar declines in many establishment surveys worldwide. Despite the high level of non-response, the average non-response bias, calculated across 56 estimates from administrative data, was found to be reassuringly low. However, biases for individual estimates, such as industry or establishment closure in $t + 1$, were more severe.

Exploiting the large administrative data source also permitted testing several hypotheses regarding survey participation and identified many establishment characteristics associated with the response outcome. As expected, smaller and older establishments were more likely to participate than their larger and younger counterparts. Consistent with the literature, the previous-year response history of the establishment explained a lot of the variation in current-year participation. The analysis found only limited support for the notion that year-to-year changes in the employment structure are associated with participation. However, the notion that higher levels of corporate social responsibility, expressed through greater workforce diversity, is positively associated with survey participation was supported, providing the first evidence of such a correlation. There was mixed evidence regarding the relevance of the survey topic for establishments that handle many recruiting processes. While having a greater share of marginal employees was positively associated with participation, having a higher share of new hires had a negative association. This negative effect could be due to HR departments being too occupied with filling vacancies that they cannot afford to allocate time or resources to completing the voluntary survey task, even if the topic is particularly relevant at the time.

To adjust for the aforementioned non-response biases in the IAB-JVS, the performance of several machine learning algorithms was compared for generating response propensity weights using the extended administrative data as auxiliary information. Even without using sophisticated data-driven approaches, utilising the additional administrative data was an improvement over the current weighting variables used in the IAB-JVS standard logistic regression weighting procedure.

Further reductions in non-response bias were observed in some years for some machine learning methods, namely, Random Forest, BART and XG-Boost. GAM, Lasso and Ridge performed similarly well to the standard logistic regression approach, while all other machine-learning methods (Gamsel, CART, MOB, C-Tree) were inferior to the standard modelling approach. The good performance of the traditional logistic regression approach relative to the majority of the machine learning algorithms might be explained by this particular case study, as there did not appear to be high-level interactions or higher polynomial functions that explained participation in the IAB-JVS, which may not be the case for other establishment surveys. Additionally this analysis showed that the selection of auxiliary variables seems to be more important than the modelling approach used for creating response propensities, because several approaches produced comparable results. Similar conclusions were also drawn by Rizzo et al. (1996), Brick (2013) and Mercer et al. (2018).

If survey organisations or sponsors are able to access and link large auxiliary data (e.g. administrative data) to their surveys, the present study can serve as a blueprint for utilising such data for the purposes of analysing response patterns and estimating and adjusting for non-response bias. The information gleaned from these analyses could be used to develop adaptive designs and contact strategies that are tailored towards important subgroups most susceptible to non-response (e.g. large establishments) with the goal of reducing non-response bias at the design stage. Furthermore, incorporating additional auxiliary data into non-response adjustment procedures could improve the effectiveness of non-response weights, even without the use of data-driven, machine learning methods. However, in order to take advantage of the full potential of the auxiliary data, we recommend evaluating machine learning methods to optimise bias adjustment. Survey organisations would be best served by evaluating several algorithms and comparing their performance before deciding on a single approach. Further research could assist this decision by analysing a wide range of methods and comparing their performance under multiple realistic settings, including the setting where only limited auxiliary data are available or when non-response bias is very large.

Although we made use of detailed administrative information on both establishment and employee characteristics to analyse non-response, these data do not provide information on the internal structure and the internal policies of the establishment. Theoretical and qualitative research suggests that internal factors, such as establishments' data sharing policies and the personal attitudes of the employees involved in the response decision predict survey participation much better than high-level establishment characteristics (e.g. Bavdaž, 2010; Snijkers et al., 2013; Willimack et al., 2002). Therefore, future research would benefit from identifying ways in which data describing these internal factors could be generated and made available for non-response analyses.

In conclusion, this study demonstrated the important roles that large-scale administrative data and data-driven approaches can play in understanding the response behaviour of establishments, identifying specific mechanisms of participation, and reducing non-response bias in establishment surveys. These tools are especially important at a time when response rates in voluntary surveys are very low and the risk of non-response bias is very high. Such tools may also prove useful in identifying subgroups most prone to non-response and informing tailored survey designs aimed at increasing their likelihood of participation.

AAPOR Conference, the 2021 Joint Sociology Conference of the German and Austrian Sociological Association, and an internal IAB seminar. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data used in this study are available from the Research Data Centre (RDC) of the Federal Employment Agency in Germany. Restrictions apply to the availability of these data, which are not publicly available. For more information on data access, see https://fdz.iab.de/en.aspx.

ORCID

Benjamin Küfner https://orcid.org/0000-0003-3367-8494

REFERENCES

American Association for Public Opinion Research (AAPOR). (2016) *The American association for public opinion research. Standard definitions: final dispositions of case codes and outcome rates for surveys*, 9th edition. AAPOR.

Anseel, F., Lievens, F., Schollaert, E. & Choragwicka, B. (2010) Response rates in organizational science, 1995–2008: a meta-analytic review and guidelines for survey researchers. *Journal of Business and Psychology*, 25, 335–349.

Bachmann, R., Elstner, S. & Sims, E.R. (2013) Uncertainty and economic activity: evidence from business survey data. *American Economic Journal: Macroeconomics*, 5, 217–249.

Bal, P.M. & Dorenbosch, L. (2015) Age-related differences in the relations between individualised HRM and organisational performance: a large-scale employer survey. *Human Resource Management Journal*, 25, 41–61.

Bavdaž, M. (2010) The multidimensional integral business survey response model. *Survey Methodology*, 36, 81–93.

Bavdaž, M., Snijkers, G., Sakshaug, J.W., Brand, T., Haraldsen, G., Kurban, B. et al. (2020) Business data collection methodology: current state and future outlook. *Statistical Journal of the IAOS*, 36, 1–16.

Blanchflower, D.G., Oswald, A.J. & Sanfey, P. (1996) Wages, profits, and rent-sharing. *The Quarterly Journal of Economics*, 111, 227–251.

BLS. (2020) *Response rates*. Washington, DC: U.S. Bureau of Labor Statistics. Available at: https://www.bls.gov/osmr/response-rates/. [Accessed 25th February 2021].

Bossler, M., Gürtzgen, N., Kubis, A., Küfner, B. & Lochner, B. (2020) The IAB job vacancy survey: design and research potential. *Journal for Labour Market Research*, 54, 1–12.

Breiman, L. (2001) Random forests. *Machine Learning*, 45, 5–32.

Breiman, L., Friedman, J., Stone, J. & Olshen, R.A. (1984) *Classification and regression trees*. Belmont: Wadsworth.

Brick, J.M. (2013) Unit nonresponse and weighting adjustments: a critical review. *Journal of Official Statistics*, 29, 329.

Brick, J.M. & Williams, D. (2013) Explaining rising nonresponse rates in cross-sectional surveys. *The Annals of the American Academy of Political and Social Science*, 645, 36–59.

Buskirk, T.D. & Kolenikov, S. (2015) Finding respondents in the forest: a comparison of logistic regression and random forest models for response propensity weighting and stratification. *Survey Methods: Insights from the Field*, 1–17.

Chen, T. & Guestrin, C. (2016) *Xgboost: a scalable tree boosting system*. arXiv preprint arXiv:1603.02754.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. (2019) *xgboost: extreme gradient boosting*. R package version 0.90.0.2. Available at: https://CRAN.R-project.org/package=xgboost.

Chipman, H.A., George, E.I. & McCulloch, R.E. (2010) BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4, 266–298.

Chouldechova, A. & Hastie, T. (2015) *Generalized additive model selection*. arXiv preprint arXiv:1506.03850.

Christianson, A. & Tortora, R.D. (2011) Ch. 14. Issues in surveying businesses: An international survey. In: Cox, B.G., Binder, D.A., Chinnappa, B.N., Christianson, A., Colledge, M.J. & Kott, P.S. (Eds.) *Business survey methods*. Hoboken, NJ: John Wiley & Sons, Ltd., pp. 235–256.

Colgan, F. & McKearney, A. (2011) Equality, diversity and corporate responsibility. *Equality, Diversity and Inclusion: An International Journal*, 30, 719–734.

Cox, D.R. (1958) The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B*, 20, 215–232.

Davis, W. R. & Pihama, N. (2009) Survey response as organisational behaviour: an analysis of the annual enterprise survey, 2003-2007. *Paper presented at New Zealand Association of Economists Conference*, Wellington.

De Heer, W. & De Leeuw, E. (2002) Trends in household survey nonresponse: a longitudinal and international comparison. In: Groves, R.M., Dillman, D.A., Eltinge, J.L. & Little, R.J. (Eds.) *Survey nonresponse*, Vol. 41. New York: John Wiley, pp. 41–54.

Earp, M., Mitchell, M., McCarthy, J. & Kreuter, F. (2014) 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, 701–719.

Earp, M., Toth, D., Phipps, P. & Oslund, C. (2018) Assessing nonresponse in a longitudinal establishment survey using regression trees. *Journal of Official Statistics*, 34, 463–481.

Fisher, S., Bosley, J., Goldenberg, K., Mockovak, W. & Tucker, C. (2003) A qualitative study of nonresponse factors affecting BLS establishment surveys: results. *Paper Presented at joint statistical meetings*. San Francisco.

Foo, C.Y., Reidpath, D.D. & Sivasampu, S. (2019) The association between hospital characteristics and nonresponse in an organization survey: an analysis of the national healthcare establishment and workforce survey in Malaysia. *Evaluation & the Health Professions*, 42, 3–23.

Friedman, J., Hastie, T. & Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33, 1.

Ganzer, A., Schmidtlein, L., Stegmaier, J. & Wolter, S. (2020) Establishment History Panel 1975-2019 (No. 202016 (en)). Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany]. FDZ-Methodenreport 16.

Ghisetti, C. & Rennings, K. (2014) Environmental innovations and profitability: how does it pay to be green? An empirical analysis on the German innovation survey. *Journal of Cleaner Production*, 75, 106–117.

Groves, R.M. & Peytcheva, E. (2008) The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public Opinion Quarterly*, 72, 167–189.

Hansen, K. & Seierstad, C. (2017) *Corporate social responsibility and diversity management*. Switzerland: Springer International Publishing.

Hastie, T. (2019) *gam: generalized additive models. R package version 1.16.1*. Available at: https://CRAN.R-project.org/package=gam.

Hastie, T.J. & Tibshirani, R.J. (1990) *Generalized additive models*, Vol. 43. London: Chapman & Hall.

Hecht, V., Litzel, N. & Schäffler, J. (2019) Unit nonresponse at the firm level: a cross-border analysis using the IAB-RELOC data. *Journal for Labour Market Research*, 53, 1–18.

HMRC. (2010) Large business methodology review. *Her Majesty's Revenue and Customs Research Report*, 98, 1–94.

Hoerl, A.E. & Kennard, R.W. (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55–67.

Hothorn, T., Hornik, K. & Zeileis, A. (2006) Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics*, 15, 651–674.

Hothorn, T. & Zeileis, A. (2015) partykit: a modular toolkit for recursive partytioning in R. *The Journal of Machine Learning Research*, 16, 3905–3909.

Houseman, S.N. (2001) Why employers use flexible staffing arrangements: evidence from an establishment survey. *ILR Review*, 55, 149–170.

Janik, F. & Kohaut, S. (2012) Why don't they answer? Unit non-response in the IAB establishment panel. *Quality & Quantity*, 46, 917–934.

Kapelner, A. & Bleich, J. (2016) bartMachine: machine learning with Bayesian additive regression trees. *Journal of Statistical Software*, 70, 1–40.

Kern, C., Klausch, T. & Kreuter, F. (2019) Tree-based machine learning methods for survey research. *Survey Research Methods*, 13, 73–93.

König, C., Sakshaug, J.W., Stegmaier, J. & Kohaut, S. (2021) Trends in nonresponse rates and nonresponse bias in establishment surveys: evidence from the 2001-2017 IAB establishment panel. *Journal of Official Statistics*, 37, 931–953.

Kreuter, F., Müller, G. & Trappmann, M. (2010) Nonresponse and measurement error in employment research: making use of administrative data. *Public Opinion Quarterly*, 74, 880–906.

Kuhn, M. (2020) *caret: classification and regression training*. R package version 6.0-85. Available at: https://CRAN.R-project.org/package=caret.

Liaw, A. & Wiener, M. (2002) Classification and regression by randomforest. *R News*, 2, 18–22. Available at. https://CRAN.R-project.org/doc/Rnews/

Lineback, J. F. & Thompson, K. J. (2010) Conducting nonresponse bias analysis for business surveys. *Proceedings of the American statistical association, government statistics section*, pp. 317–331.

Little, R.J. & Vartivarian, S. (2005) Does weighting for nonresponse increase the variance of survey means? *Survey Methodology*, 31, 161–168.

Lohr, S., Hsu, V. & Montaquila, J. (2015) Using classification and regression trees to model survey nonresponse. *Paper presented at joint statistical meetings*. Seattle.

Mercan, Y. & Schoefer, B. (2020) Jobs and matches: quits, replacement hiring, and vacancy chains. *American Economic Review: Insights*, 2, 101–124.

Mercer, A., Lau, A. & Kennedy, C. (2018) *For weighting online opt-in samples, what matters most*. Pew Research Center.

Ogbadu, E.E. (2009) Profitability through effective management of materials. *Journal of Economics and International Finance*, 1, 99–105.

Petroni, R., Sigman, R., Willimack, D., Cohen, S. & Tucker, C. (2004) Response rates and nonresponse in BLS and census bureau establishment surveys. *Proceedings of the American statistical association, survey research methods section*. pp. 4159–4166.

Peytcheva, E. & Groves, R.M. (2009) Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. *Journal of Official Statistics*, 25, 193–201.

Pfeffermann, D. (2011) Modelling of complex survey data: why is it a problem? How should we approach it? *Survey Methodology*, 37, 115–136.

Phipps, P. & Jones, C. (2007) Factors affecting response to the occupational employment statistics survey. *Proceedings of the 2007 federal committee on statistical methodology research conference*.

Phipps, P. & Toth, D. (2012) Analyzing establishment nonresponse using an interpretable regression tree model with linked administrative data. *The Annals of Applied Statistics*, 6, 772–794.

Pielsticker, D.I. & Hiebl, M.R.W. (2020) Survey response rates in family business research. *European Management Review*, 17, 327–346.

R Core Team. (2019) *R: a language and environment for statistical computing*. Vienna, Austria: Available at: R Foundation for Statistical Computing. https://www.R-project.org/

Riviére, P. (2002) What makes business statistics special? *International Statistical Review*, 70, 145–159.

Rizzo, L., Kalton, G. & Brick, J.M. (1996) A comparison of some weighting adjustment methods for panel nonresponse. *Survey Methodology*, 22, 43–53.

Seiler, C. (2014) The determinants of unit non-response in the IFO business survey. *AStA Wirtschafts-und Sozialstatistisches Archiv*, 8, 161–177.

Seth, S. & Stüber, H. (2018) The administrative wage and labor market flow panel. *FAU Discussion Papers in Economics*, 1, 1–47.

Smaill, K. (2012) Trajectory modelling of longitudinal non-response in business surveys. *Statistical Journal of the IAOS*, 28, 137–144.

Snijkers, G. (2008) Getting data for business statistics: a response model. *Proceedings of Q2008 - 4th European conference on quality in official statistics*.

Snijkers, G. (2018) The effect of response measures in business surveys. *Paper presented at conference of European statisticians*, Geneva.

Snijkers, G., Berkenbosch, B., Luppes, M. & Netherlands, S. (2007) Understanding the decision to participate in a business survey. *Proceedings of the 3rd International Conference on Establishment Surveys*. pp. 1048–1059.

Snijkers, G., Haraldsen, G., Jones, J. & Willimack, D. (2013) *Designing and conducting business surveys*. Hoboken, NJ: John Wiley & Sons.

Starostka-Patyk, M., Tomski, P. & Zawada, M. (2015) Diversity management as a part of corporate social responsibility policy. *Procedia Computer Science*, 65, 1038–1045.

StataCorp. (2019) *Stata statistical software: release*, Vol. 16. College Station, TX: College Station.

Therneau, T. & Atkinson, B. (2019) *rpart: recursive partitioning and regression trees*. R package version 4.1-15. Available at: https://CRAN.R-project.org/package=rpart.

Thompson, K.J. & Oliver, B.E. (2012) Response rates in business surveys: going beyond the usual performance measure. *Journal of Official Statistics*, 28, 221–237.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267–288.

Tomaskovic-Devey, D. (1995) Item nonresponse in organizational surveys. *Sociological Methodology*, 25, 77–110.

Tomaskovic-Devey, D., Leiter, J. & Thompson, S. (1994) Organizational survey nonresponse. *Administrative Science Quarterly*, 39, 439–457.

Toth, D. & Phipps, P. (2014) Regression tree models for analyzing survey response. *Proceedings of the American statistical association, government statistics section*. pp. 339–351.

Valliant, R., Dever, J.A. & Kreuter, F. (2013) *Practical tools for designing and weighting survey samples*. New York: Springer.

White, G. D. & Luo, A. (2005) Business survey response rates - can they be improved. *Proceedings of the American statistical association, survey research methods section*. pp. 3666–3668.

White, M. & Bryson, A. (2013) Positive employee attitudes: how much human resource management do you need? *Human Relations*, 66, 385–406.

Williams, D. & Brick, J.M. (2018) Trends in US face-to-face household survey nonresponse and level of effort. *Journal of Survey Statistics and Methodology*, 6, 186–211.

Willimack, D.K. & Nichols, E. (2010) A hybrid response process model for business surveys. *Journal of Official Statistics*, 26, 3–24.

Willimack, D.K., Nichols, E. & Sudman, S. (2002) Understanding unit and item nonresponse in business surveys. In: Groves, R.M., Dillman, D.A., Eltinge, J.L. & Little, R.J.A. (Eds.) *Survey nonresponse*. New York: Wiley, pp. 213–227.

Willimack, D.K. & Snijkers, G. (2013) The business context and its implications for the survey response process. In: Snijkers, G., Haraldsen, G., Jones, J. & Willimack, D. (Eds.) *Designing and conducting business surveys*. Hoboken, NJ: John Wiley & Sons, pp. 39–82.

Zeileis, A., Hothorn, T. & Hornik, K. (2008) Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, 17, 492–514.

Zinn, S. & Gnambs, T. (2022) Analyzing nonresponse in longitudinal surveys using Bayesian additive regression trees: a nonparametric event history analysis. *Social Science Computer Review*, 40(3), 678–699.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Küfner, B., Sakshaug, J.W. & Zins, S. (2022) Analysing establishment survey non-response using administrative data and machine learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 1–33. Available from: https://doi.org/10.1111/rssa.12942