Patterns of Nonresponse and Imputation in the IFO Business Survey

Statistical Consulting Program

Department of Statistics

Ludwig-Maximilians-Universität München

Chunyan Jiang

`Chunyan.Jiang@campus.lmu.de`

Qian Feng

`Qian.Feng@campus.lmu.de`

Munich, January 20, 2026

Supervisor:	PD Dr. Fabian Scheipl
Project partner:	Dr. Klaus Wohlrabe (ifo Institute)
	Mr. Stefan Sauer (ifo Institute)

Abstract

This study investigates patterns of nonresponse and evaluates alternative imputation methods in the ifo Business Survey, a large-scale monthly panel survey of German manufacturing firms. Focusing on the two core qualitative indicators Business Situation (BS) and Business Expectations (BE), we first conduct a detailed descriptive and quantitative analysis of unit nonresponse behavior. Using firm-level panel data from 1991 to 2024, we examine the determinants of unit nonresponse with a generalized additive model (GAM) featuring a logit link and firm-level cluster-robust inference. This framework allows for flexible nonlinear relationships while accounting for within-firm dependence. The results reveal pronounced and systematic nonlinear patterns in unit nonresponse. Calendar time exhibits strong non-monotonic dynamics, indicating distinct phases in the evolution of nonresponse behavior. Survey participation history emerges as a multidimensional construct: participation number and participation length capture different aspects of firms' engagement with the survey and display opposing nonlinear associations with unit nonresponse. In addition, differences are observed across states, sectors and months. Since the estimated spline shapes for Average Business Situation and Business Expectations are sensitive to the choice of the basis dimension k, they are treated primarily as flexible control components rather than as objects of direct interpretation.

To assess the implications of missing data for survey-based indicators, we compare three imputation strategies with increasing modeling flexibility: Last Observation Carried Forward (LOCF), homogeneous Markov Chain, and Random Forest. A simulation study based on artificially induced consecutive missingness that reflects more than 95% of observed missing patterns shows that the Random Forest consistently outperforms the alternative methods across gap lengths and evaluation metrics, while LOCF remains a strong benchmark. Imputation performance deteriorates with increasing gap length for all methods, and BS is imputed more accurately than BE under comparable settings. Applying the Random Forest imputation to the original survey data indicates that imputed BS balances are generally higher than the original series over most of the sample period, whereas imputed BE balances remain relatively close to the original balance prior to 2022.

Contents

1 Introduction

1.1 Research Background and Motivation

The mission of the ifo Institute is to shape economic policy debates in Germany and internationally by bridging academic excellence with policy relevance. Central to this mission is the ifo Business Survey, which has been conducted since 1949. Today, the importance of the results of this survey goes far beyond a mere statistical substitute. This importance stems from the central role of company-specific plans, expectations, and assessments of the current situation in economic theory, which the ifo Business Survey captures primarily through qualitative questions. Furthermore, the survey results offer prompt information on key economic variables that are typically published in official statistics only after considerable delays and revisions (Sauer et al., 2023).
Among the indicators obtained from the survey, two variables play a central role: the "Business Situation" (BS) and the "Business Expectations" (BE) for the next six months. Together, these form the basis of the ifo Business Climate Index, which is the most significant leading indicator for the German economy (Sauer et al., 2023).
As participation in the ifo Business Survey is voluntary, both unit and item nonresponse may arise in the data. Preliminary descriptive analysis reveals that the unit nonresponse rate in the manufacturing sector has exhibited a distinct trend over the study period. Furthermore, nonresponse rates exhibit notable differences across variables such as calendar month, region, and sector. If nonresponse depends on factors other than pure randomness(MCAR), it may affect the construction of the aggregated ifo Business Climate Index. Against this background, this paper investigates the determinants of nonresponse using cluster-robust Generalized Additive Models (GAM). It further evaluates the performance of three imputation methods, including Last Observation Carried Forward (LOCF), Markov Chains, and Random Forest, across missingness gaps of up to six months.

1.2 Conceptual Definitions and Data Context

To establish a rigorous framework for analyzing missing data within the ifo Business Survey, it is necessary to define the specific levels of nonresponse and the core variables under investigation.

Definitions of Nonresponse We distinguish between two types of nonresponse. *Unit nonresponse* occurs when a firm does not submit the questionnaire in a given survey wave, resulting in missing information for all survey items in that month. *Item nonresponse* arises when a firm participates in the survey and returns the questionnaire, but fails to respond to at least one survey item. Since unit nonresponse accounts for the majority of missing data, the quantitative analysis focuses on unit nonresponse, while the imputation analysis accounts for both unit and item nonresponse to restore the completeness of the panel structure.

Core Survey Outcomes The ifo Business Climate Index is based on two core variables. *Business Situation (BS)* asks firms to categorize their current business situation as "good," "satisfactory," or "poor." *Business Expectations (BE)* asks firms to forecast their business situation for the next six months as "become more favorable," "stay roughly the same," or "become less favorable."

Data The analysis is based on monthly firm-level panel data from the ifo Business Survey. The dataset covers the period from 1991 to June 2024 and includes more than 8,100 firms, resulting in over 1,054,000 firm–month observations. This panel structure allows for a detailed analysis of unit nonresponse and serves as the basis for the imputation analysis.

1.3 Structure of the Report

The remainder of this report is structured as follows. Chapter 2 describes the data used in the analysis, including the data source, the construction and cleaning of the dataset and the definitions of the main variables. Chapter 3 presents the ifo Business Survey data and a descriptive analysis of nonresponse patterns. Chapter 4 analyzes the determinants of unit nonresponse using cluster-robust Generalized Additive Models. Chapter 5 evaluates alternative imputation methods for missing observations and compares the imputed and original aggregated balance series. Chapter 6 concludes and discusses the limitations. Chapter 7 outlines the individual contributions of the authors. Chapter 8 provides the declaration regarding the use of AI tools.

2 Data

2.1 Data

2.1.1 Data source

We use firm-level data from the ifo Industrial Business Survey (IBS-IND), a monthly survey of German manufacturing firms conducted by the ifo Institute. The survey collects firms' assessments of their current business situation, short-term expectations, and related operational indicators. Participation in the survey is voluntary, which gives rise to both item nonresponse and unit nonresponse.

2.1.2 Sample construction and data cleaning

First, we restricted the sample to observations from 1991 onwards and retained only variables relevant for the analysis, including firm identifiers, time indicators, sector and regional information, and the main survey questions on business situation, orders, production, prices, employment, and expectations. Observations without a valid firm identifier were removed.

Second, due to changes in questionnaire design over time, several survey variables exist in different versions before and after specific reform dates. To ensure consistency over time, we harmonized these variables by constructing unified series. In particular, demand, orders, production, and price variables were combined by using the pre-reform versions for earlier periods and the post-reform versions thereafter. In addition, the indicator for online participation was recoded into a binary variable distinguishing online from non-online responses.

Third, we identified unit nonresponse at the firm-month level. An observation was classified as unit nonresponse if all core survey items were missing in a given month. Observations with at least one non-missing core item were classified as valid responses. Duplicate firm-month observations were removed by keeping a single record per firm and month. Firms with fewer than 12 total observations over the sample period were excluded to ensure a minimum participation history.

Fourth, the dataset was expanded to a balanced monthly panel at the firm level. For each firm, we generated a complete sequence of monthly observations between its first and last appearance in the survey. Newly created observations were flagged as non-original and treated as unit nonresponse. Static firm characteristics, such as sector, region, firm size, and survey weights, were carried forward across short gaps of nonresponse (up to 12 consecutive months) when the values before and after the gap were identical. This procedure avoids introducing spurious structural changes while preserving long-run firm characteristics.

Fifth, firms exhibiting long internal spells of unit nonresponse (at least 12 consecutive months occurring neither at the beginning nor at the end of the observation window) were removed from the sample, as such gaps typically reflect firms that dropped out of the survey and re-entered after several years rather than temporary nonresponse.

Sixth, we applied an end-of-sample trimming procedure. Since the data end in June 2024, firms that had not responded for at least 12 months prior to this date were treated as

having exited the survey. All artificial nonresponse observations beyond the firm's last actual response and exceeding this 12-month window were removed.
Seventh, December 2001 was excluded from the sample due to a structural break in the survey timing. To maintain a consistent monthly time index, we constructed a shifted year–month variable, in which all observations up to November 2001 were shifted forward by one month, while observations from December 2001 onwards remained unchanged.
Finally, additional variables required for the empirical analysis were constructed. These include calendar-time indices, month-of-year dummy variables, participation number, and participation length measured as the time since a firm's first observed response. For the Business Situation and Business Expectations variables, six-month backward-looking averages were computed, excluding the current month. Firms belonging to sectors with very few observations were excluded from the final sample.

2.1.3 Variable definitions

After data cleaning, three datasets were constructed. The first dataset is prepared for the analysis of item nonresponse and contains 1,062,925 observations from 8,176 firms. The second dataset is designed for the analysis of unit nonresponse and includes 1,054,749 observations from the same 8,176 firms. Compared to the item nonresponse dataset, this dataset excludes all observations with a participation length of zero. By definition, a participation length of zero implies a participation number of zero, indicating a firm's first survey participation, for which unit nonresponse is structurally impossible. These observations are therefore excluded from the regression analysis to avoid structural interference. The third dataset is prepared for imputation and contains 272,030 observations from 3,795 firms. Relative to the item nonresponse dataset, it is restricted to observations from January 2014 onward.
Given that more than 95% of missing observations in the data arise from unit nonresponse, the descriptive and quantitative analysis in the main text focuses on unit nonresponse behavior. Variable definitions and Descriptive results for item nonresponse are presented in the Appendix A.1.1 and A.2.2 for completeness.
Table 1 describes the construction of the key variables used in the quantitative analysis and imputation.

Table 1: Key Variables Description

Variable	**Description**
Unit Nonresponse	1: unit nonresponse; 0: not unit nonresponse
Company ID	company id
Calendar Time (Index)	1 for 02/1991, 2 for 03/1991, ..., 402 for 06/2024.
Month Indicators (is_jan–is_dec)	Month-of-year indicator variables (1 = corresponding month, 0 = otherwise; factors).
Month	1-12 (Jan-Dec)
Region	Western/Eastern.
Sector	EBDC sector classification.
Federal State	Federal state (ifo-code) .
Company Size	Size range.
Participation Number	The cumulative number of past months in which a firm responded to the survey (i.e., did not exhibit unit nonresponse), excluding the current month.
Participation Length	The number of months since a firm's first observed response, excluding the current month.
online	Online/Offline.
Avg. Business Situation (6-month)	Mean of the *Business Situation* responses over the past six months (not rounded).
Avg. Business Expectation (6-month)	Mean of the *Business Expectation* responses over the past six months (not rounded).
Lagged Business Situation	Business Situation of firm i at month $t-k$, where $k \in \{1,2,3,4\}$.
Lagged Business Expectations	Business Expectation of firm i at month $t-k$, where $k \in \{1,2,3,4\}$.

Notes: Observations with *Participation Number* equal to zero are excluded from the quantitative analysis, as unit nonresponse cannot occur prior to a firm's first response.
Observations with *Participation Length* equal to zero are excluded from the quantitative analysis, as unit nonresponse cannot occur prior to a firm's first response.

3 Descriptive Analysis

3.1 Unit Nonresponse

Due to substantial missing values or collinearity issues, the variables capturing Region(West/East), firm size, and survey response mode (online versus offline) are not included in the regression analysis. The specific reasons for their exclusion and the corresponding diagnostic analyses are documented in the Appendix A.2.1.

3.1.1 Analysis Based on Time-Related Variables

Trend in Time

Figure 1 plots the monthly unit nonresponse rate over time. The unit nonresponse rate exhibits substantial variation over the sample period. At the beginning of the survey, the nonresponse rate is approximately 5%, after which it gradually increases and reaches about 15% during 1993–1994. It then declines markedly and reaches a trough of approximately 2.5% in the early 2000s. Following this low point, the nonresponse rate begins to rise again from the mid-2000s onward and displays noticeable short-term fluctuations.
From the late 2010s onward, the upward trend becomes more pronounced and is accompanied by a certain degree of volatility, with the nonresponse rate reaching its highest level of around 24% toward the end of the observation period. Overall, the figure indicates that nonresponse behavior is not stable over time but instead evolves gradually, exhibiting both long-run trends and short-run fluctuations.
This pronounced and non-monotonic time pattern motivates the inclusion of flexible calendar-time controls in the quantitative analysis to capture potentially nonlinear temporal dynamics in unit nonresponse.

Figure 1: Unit Nonresponse Rate Over Time

Unit Nonresponse Rate by Month

Figure 2 illustrates how the average unit nonresponse rate varies across calendar months.

The unit nonresponse rate varies across months. At the beginning of the year, specifically from January to March, nonresponse rates are relatively lower and remain fairly stable, at around 9.6%. From April onward, the rate increases gradually, indicating a steady rise in nonresponse through April and May. In June and July, unit nonresponse rates decline, however, they remain higher than those observed at the beginning of the year. In August, the unit nonresponse rate increases sharply, followed by a gradual decline from September to November. The nonresponse rate reaches its highest level in December, at around 11.1%, suggesting markedly higher nonresponse in the final month of the calendar year.
Overall, the figure indicates that unit nonresponse is not evenly distributed across months. This monthly variation motivates the inclusion of month indicator variables in the quantitative analysis to account for differences in unit nonresponse across months.

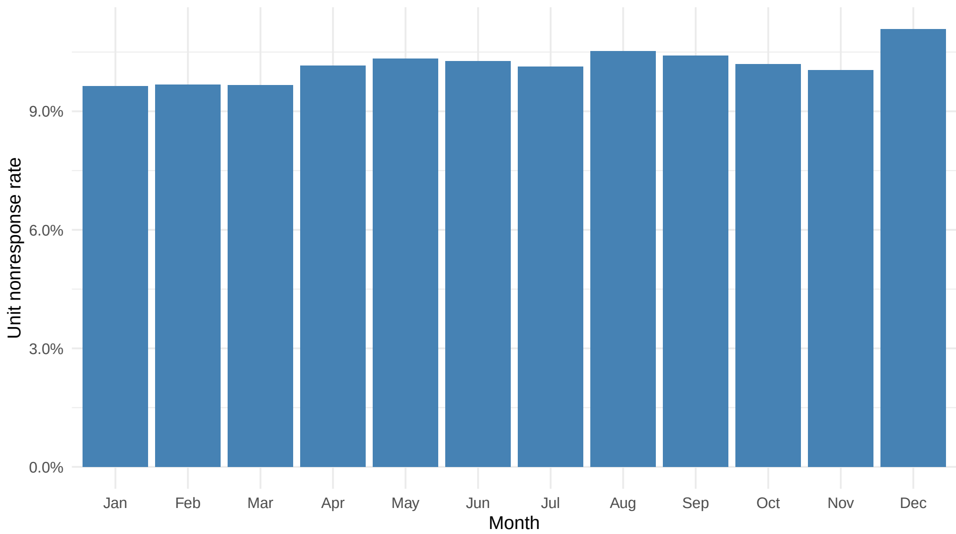

Figure 2: Unit Nonresponse Rate by Month

3.1.2 Analysis Based on Firm Characteristics Variables

Unit Nonresponse Rate by Federal State

Figure 3 shows the unit nonresponse rate by Federal State. Unit nonresponse rates vary considerably across regions, with Brandenburg, Mecklenburg-Vorpommern, Sachsen, Thueringen, and Sachsen-Anhalt exhibiting the highest rates (above 16%). In contrast, Bremen has the lowest rate at approximately 6.6%. Overall, states in East Germany tend to have higher unit nonresponse rates, while states in West Germany generally show lower rates.
Overall, the figure reveals substantial heterogeneity in unit nonresponse across federal states across firms. This pronounced state-level variation motivates the inclusion of state indicator variables in the quantitative analysis to account for differences in unit nonresponse across the 16 federal states.

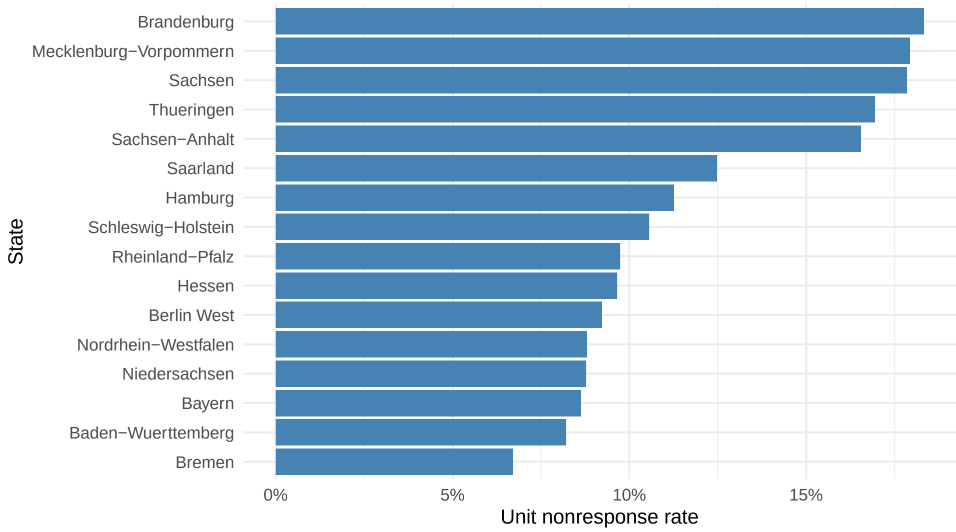

Figure 3: Unit Nonresponse Rate by Federal State

Unit Nonresponse Rate by Sector

Figure 4 shows the average unit nonresponse rate across manufacturing industries. The unit nonresponse rate varies across industries, indicating notable heterogeneity at the industry level. Food, Drinks and Tobacco exhibits the highest nonresponse rate, at approximately 13.7%, followed by the Chemical Industry and Textile, Clothing and Leather, with rates of around 12%.

A middle group of industries, including Electronics and Optical Products, Paper and Board Industry, Cokery and Oil Processing, Wood and Cork Products, and Glass, Ceramics and Stone Processing, shows nonresponse rates clustered around 10–11%. In contrast, relatively lower nonresponse rates are observed in the Metal Industry, Pharmaceutical Industry, Publishing and Printing, and Rubber and Plastics, where rates are closer to 9–10%.

Overall, the figure indicates that unit nonresponse differs across industries. This industry-level variation motivates the inclusion of industry indicator variables in the quantitative analysis to account for differences in unit nonresponse across industries.

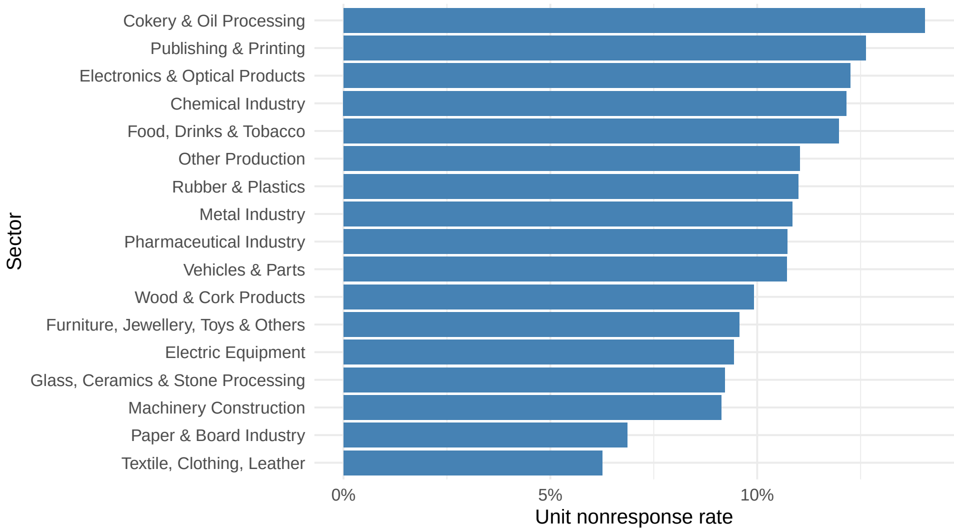

Figure 4: Unit Nonresponse Rate by Sector

3.1.3 Analysis Based on Survey-Related Variables

Unit Nonresponse Rate by Participation Number

Figure 5 illustrates the relationship between participation number, defined as the number of past survey participations up to the previous month, and the unit nonresponse rate, together with variability bands.

Figure 6 reports the number of available observations at each participation number.

The unit nonresponse rate is particularly high when participation number equals one, exceeding 25%. This reflects the fact that many firms respond only once and then experience extended periods of nonresponse before a second response occurs. Because firms may remain nonresponsive for several consecutive months after their first response, a single firm can contribute multiple observations with participation number equal to one.

As participation number increases, the unit nonresponse rate initially declines sharply and reaches a local minimum at moderate participation levels, at around 4%. The rate then increases at intermediate-to-high participation numbers before declining again at very high participation levels. Overall, the relationship between participation number and unit nonresponse is clearly non-monotonic, exhibiting a decline–increase–decline pattern rather than a simple monotonic trend.

At the same time, short-term fluctuations in the unit nonresponse rate become increasingly pronounced at higher participation numbers. As shown in Figure 6, the number of available observations decreases sharply with participation number, falling from more than 10,500 observations at participation number equal to one to only four observations at participation number equal to 400. This substantial reduction in the number of observations results in markedly greater variability in the unit nonresponse rate at high participation numbers. Consequently, the pronounced fluctuations observed in this range are likely to be driven primarily by data sparsity.

Overall, the descriptive evidence suggests a non-monotonic relationship between participation number and unit nonresponse. This pattern motivates the inclusion of participation number in a flexible, nonlinear form in the quantitative analysis.

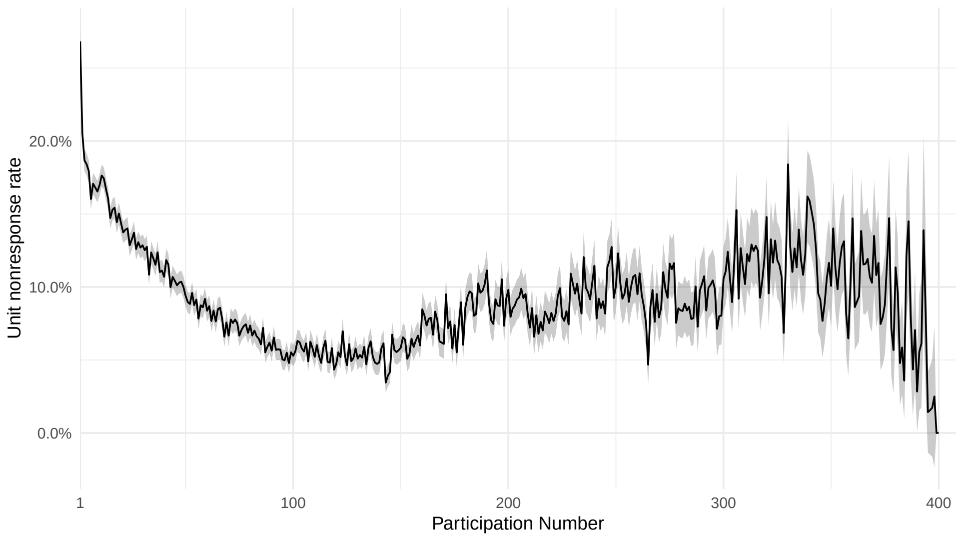

Figure 5: Unit Nonresponse Rate by Participation Number, with Variability Bands

Notes: Shaded areas indicate variability in the unit nonresponse rate arising from the finite number of observations at each participation number.

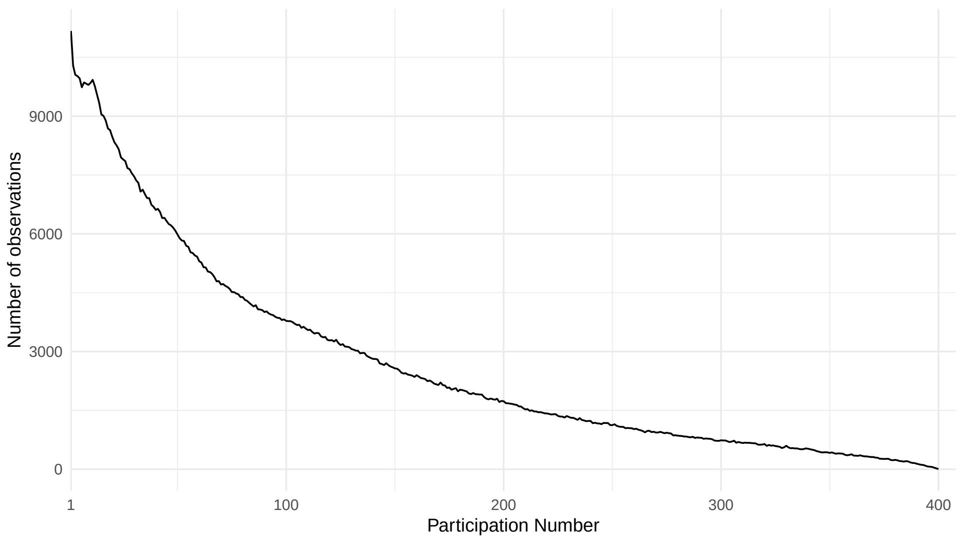

Figure 6: Number of Observations by Participation Number

Unit Nonresponse Rate by Participation Length

Figure 7 illustrates the relationship between participation length, defined as the number of months since a firm's first observed response up to the previous month, and the unit nonresponse rate, together with variability bands.

Figure 8 reports the number of available observations at each participation length.

The unit nonresponse rate is relatively high at very short participation lengths. As participation length increases, the unit nonresponse rate declines markedly and reaches a low level after firms have accumulated a moderate participation history.

Beyond this point, the unit nonresponse rate begins to increase gradually with participation length and displays pronounced short-term fluctuations. Overall, the relationship between participation length and unit nonresponse is clearly non-monotonic, exhibiting an initial decline followed by a gradual increase rather than a simple monotonic trend.
At the same time, short-term fluctuations in the unit nonresponse rate become increasingly pronounced at higher participation lengths. As shown in Figure 8, the number of available observations decreases steadily as participation length increases, falling from more than 8,000 observations at a participation length of one to around 320 observations at a participation length of 400. This reduction in the number of observations results in markedly greater variability in the unit nonresponse rate at longer participation lengths. Consequently, the increased volatility observed in this range is likely to be driven primarily by data sparsity.
Overall, the descriptive evidence suggests a non-monotonic relationship between participation length and unit nonresponse. This pattern motivates the inclusion of participation length in a flexible, nonlinear form in the quantitative analysis.

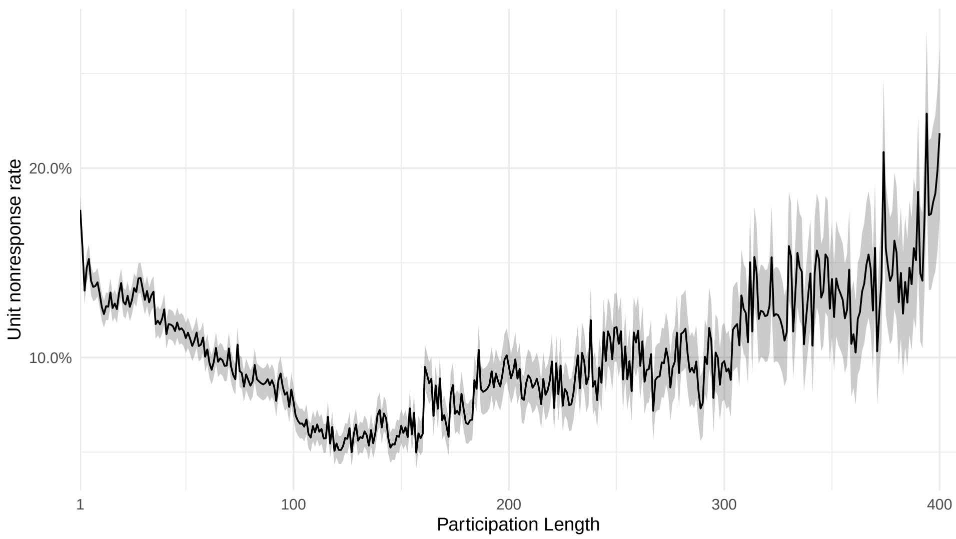

Figure 7: Unit Nonresponse Rate by Participation Length, with Variability Bands

Notes: Shaded areas indicate variability in the unit nonresponse rate arising from the finite number of observations at each participation length.

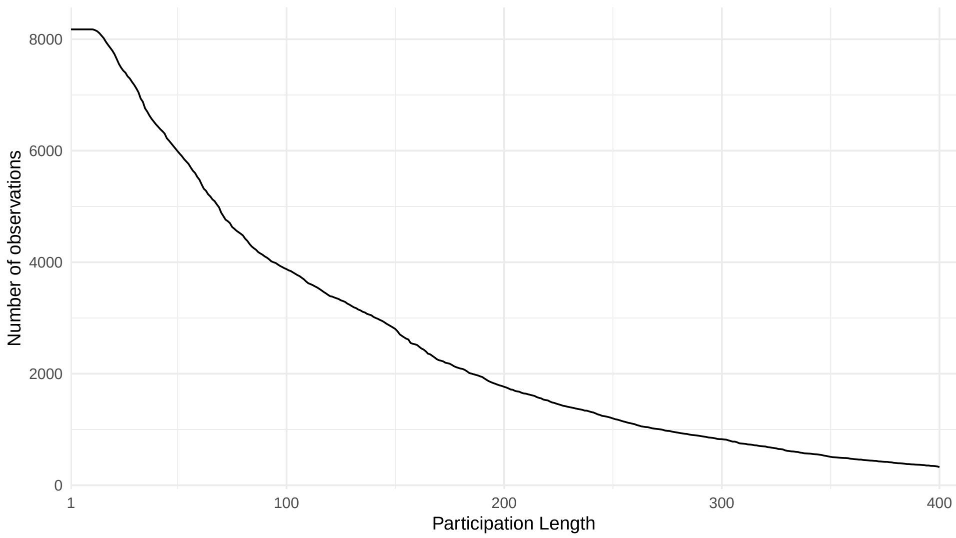

Figure 8: Number of Observations by Participation Length

Participation Number and Participation Length

Figure 9 illustrates how the relationship between participation number and participation length varies across firms' survey histories. For firms with short participation lengths, the two measures are highly correlated.

As participation length increases, however, the local correlation declines steadily, indicating that participation number and participation length gradually diverge.

This pattern suggests that, for long-standing firms, participation number and participation length capture distinct aspects of survey engagement rather than a single underlying dimension.

While participation length reflects cumulative exposure to the survey, participation number increasingly reflects selective response behavior.

This structural divergence provides descriptive evidence for treating the two measures separately in the subsequent nonlinear modeling framework.

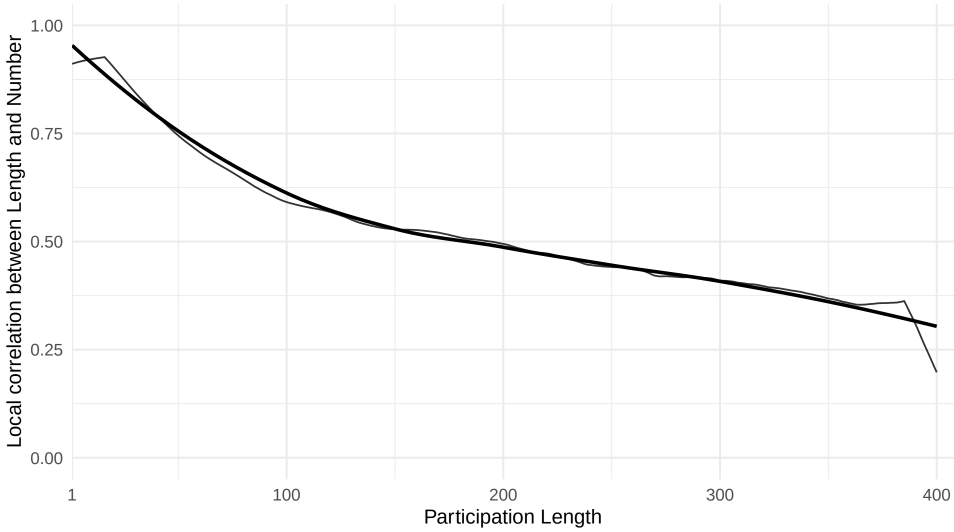

Figure 9: Local correlation between participation number and participation length. The figure shows the local correlation coefficient between participation number and participation length across the support of participation length. Correlation is computed within local neighborhoods of participation length, illustrating how the relationship between the two measures evolves over firms' survey histories.

3.1.4 Analysis Based on Two Main Questions

Unit Nonresponse Rate by Average Business Situation over the past 6 months

Figure 10 illustrates the relationship between the average Business Situation over the past six months and the unit nonresponse rate. Figure 11 reports the number of observations corresponding to each value of the average Business Situation over the past six months.
Across most values of the average Business Situation, the unit nonresponse rate lies within a relatively narrow range of approximately 7–14%. However, at several specific intermediate values, pronounced local spikes in the unit nonresponse rate are observed, with levels around 25% (for example, 1.25, 1.75, 2.25, and 2.75). In contrast, the unit nonresponse rate is relatively low at some adjacent values.
At the same time, the distribution of observations across values of the average Business Situation is highly uneven. As shown in Figure 11, observations are heavily concentrated at the value 2.0, which alone accounts for more than 300,000 observations. The next most frequent values are 1.0 and 3.0, each corresponding to more than 120,000 observations. By contrast, many other values are supported by far fewer observations; in particular, at values such as 1.25, 1.75, 2.25, and 2.75, the number of observations falls below 7,000.
This pronounced imbalance in the distribution of observations helps explain the irregular variation and sharp local fluctuations in the unit nonresponse rate across adjacent values of the average Business Situation. For values supported by a large number of observations (such as 1.0, 2.0, and 3.0), the unit nonresponse rate is relatively stable, whereas at sparsely populated values it is more prone to volatility and extreme outcomes. Consequently, part of the observed heterogeneity in unit nonresponse rates across values of the average Business Situation is likely attributable to differences in data density.

Overall, the relationship between the six-month average Business Situation and unit nonresponse is clearly non-monotonic. In light of this descriptive evidence, the relationship between average Business Situation and unit nonresponse cannot be appropriately characterized by a linear specification. Therefore, in the quantitative analysis, the six-month average Business Situation is treated as a pseudo-continuous variable and incorporated into the regression model using a flexible nonlinear form, which allows us to capture potential nonlinearities while avoiding unnecessary information loss that would arise from further discretization.

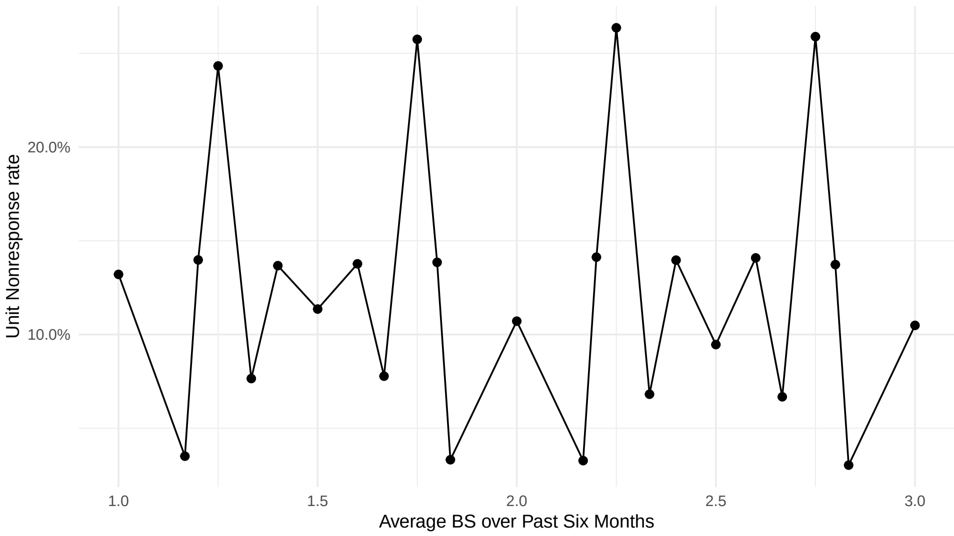

Figure 10: Unit Nonresponse Rate by Average Business Situation over the Past Six Months

Notes: The Business Situation variable is originally measured on an ordinal scale (1 = good, 2 = satisfactory, 3 = bad). The average Business Situation is computed as the mean of available monthly responses over the past six months (excluding the current month).

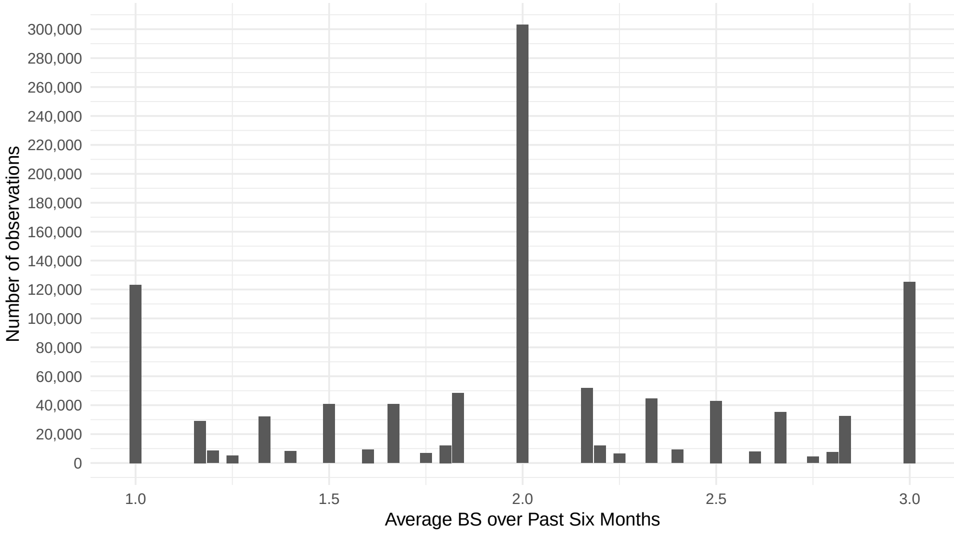

Figure 11: Number of Observations by Average Business Situation over the past 6 months

Notes: The Business Situation variable is originally measured on an ordinal scale (1 = good, 2 = satisfactory, 3 = bad). The average Business Situation is computed as the mean of available monthly responses over the past six months (excluding the current month).

Unit Nonresponse Rate by Average Business Expectation over the past 6 months

Figure 12 illustrates the relationship between the average Business Expectation over the past six months and the unit nonresponse rate.

Figure 13 reports the number of observations corresponding to each value of the six-month average Business Expectation.

Across most values of the average Business Expectation, the unit nonresponse rate lies within a range of approximately 6–18%. However, at several specific intermediate values, pronounced local peaks in the unit nonresponse rate are observed, with levels approaching 25% (for example, at 1.25, 1.75, 2.25, and 2.75). By contrast, at other adjacent values, the unit nonresponse rate remains at comparatively lower levels.

At the same time, the distribution of observations across values of the average Business Expectation is highly uneven. As shown in Figure 13, observations are heavily concentrated at the value 2.0, which alone accounts for more than 340,000 observations. In contrast, many other values are supported by substantially fewer observations, with some values having fewer than 5,000 observations.

This pronounced imbalance in the distribution of observations helps explain the irregular variation and sharp local fluctuations in the unit nonresponse rate across adjacent values of the average Business Expectation. For values supported by a large number of observations, the unit nonresponse rate appears relatively stable, whereas at sparsely populated values it is more prone to large fluctuations and even extreme outcomes. Consequently, part of the observed differences in unit nonresponse rates across values of the average Business Expectation may be attributable to differences in data density.

Overall, the relationship between the six-month average Business Expectation and unit nonresponse is clearly non-monotonic. In light of this descriptive evidence, the effect of average Business Expectation on unit nonresponse cannot be adequately characterized by a simple linear specification. Therefore, in the quantitative analysis, the six-month average Business Expectation is treated as a pseudo-continuous variable and incorporated into the regression model using a flexible nonlinear form, which allows us to capture potential nonlinearities while avoiding unnecessary information loss that would arise from further discretization.

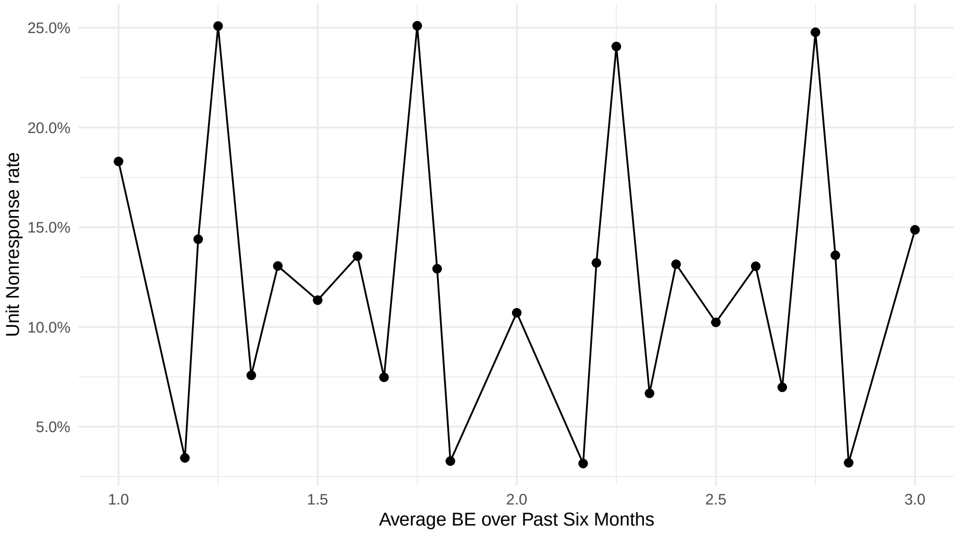

Figure 12: Unit Nonresponse Rate by Average Business Expectation over the past 6 months

Note: The Business Expectation variable is originally measured on an ordinal scale (1 = become more favorable, 2 = remain roughly the same, 3 = become less favorable.). The average Business Expectation is computed as the mean of available monthly responses over the past six months (excluding the current month).

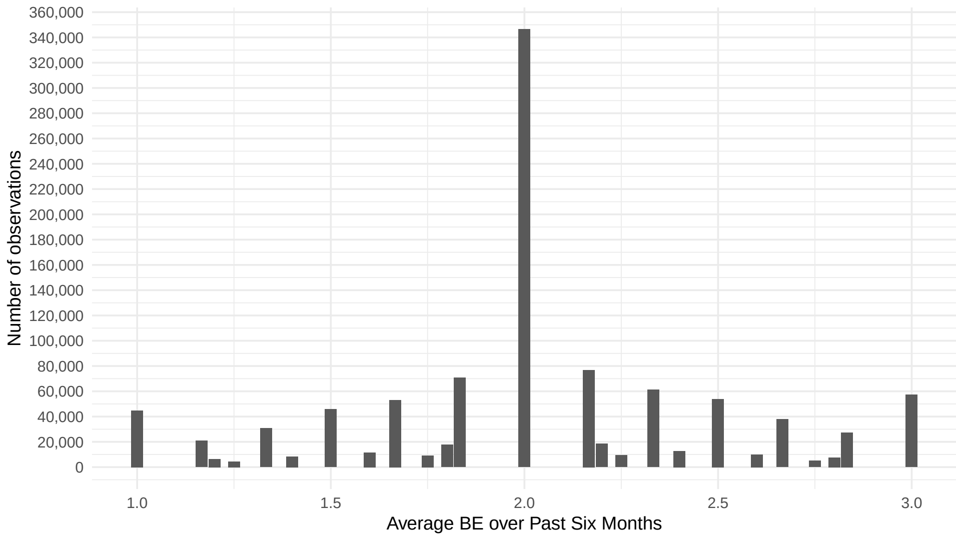

Figure 13: Number of Observations by Average Business Expectation over the past 6 months

Note: The Business Expectation variable is originally measured on an ordinal scale (1 = become more favorable, 2 = remain roughly the same, 3 = become less favorable.). The average Business Expectation is computed as the mean of available monthly responses over the past six months (excluding the current month).

4 Quantitative Analysis

4.1 Model Strategy

4.1.1 Model Overview

In this study, we model unit nonresponse at the firm–month level using a cluster-robust generalized additive model (GAM) with a logit link function and binomial response distribution. The outcome variable indicates whether a firm fails to provide any response in a given survey month, yielding a binary dependent variable that naturally motivates a nonlinear probability model. Our primary interest lies in population-average marginal effects on unit nonresponse.

The empirical setting is characterized by a large, unbalanced panel of manufacturing firms observed at monthly frequency over a long time horizon. Firms may enter and exit the survey at different points in time, and individual firms contribute repeated observations over successive months. This panel structure implies two key challenges for empirical modeling: first, the determinants of unit nonresponse are unlikely to follow simple linear relationships; second, observations from the same firm may exhibit serial dependence due to persistent firm-specific characteristics, survey fatigue, or unobserved organizational factors.

To address the first challenge, we adopt a generalized additive modeling framework that allows the conditional mean of unit nonresponse to depend on a set of covariates through flexible, data-driven smooth functions. In particular, smooth terms are used to capture potentially nonlinear effects of calendar time, firms' survey participation history, and average Business Situation/Busines Expectation over past six months. This flexibility is especially important in light of the descriptive evidence presented earlier, which reveals pronounced nonlinearity and non-monotonicity in several key relationships.

To address the second challenge, we conduct statistical inference using cluster-robust standard errors at the firm level, with the *Company ID* serving as the clustering variable. While the GAM is estimated under the assumption of independent observations, inference is adjusted ex post to allow for arbitrary correlation of errors within firms over time. This approach yields valid statistical inference in the presence of within-firm dependence, while preserving a population-average interpretation of the estimated marginal effects.

Overall, the modeling strategy combines the flexibility of generalized additive models with robust inference tailored to the panel structure of the survey data. This framework allows us to investigate how unit nonresponse varies with firm characteristics, calendar time, firms' survey participation history, and average Business Situation/Busines Expectation over past six months, while accounting for both nonlinearities and within-firm dependence.

4.1.2 Generalized Additive Model with Logit Link and Binomial Response Distribution

To model the relationship between unit nonresponse and its determinants flexibly, we employ a generalized additive model (GAM) with a logit link function and a binomial response distribution. Generalized additive models extend generalized linear models by allowing the systematic component to include smooth, data-driven functions of covariates,

thereby relaxing restrictive linearity assumptions while retaining a clear probabilistic interpretation (Hastie and Tibshirani, 1986, Wood, 2017).
Let y_{it} denote an indicator equal to one if firm i exhibits unit nonresponse in month t, and zero otherwise. Conditional on covariates X_{it}, the outcome follows a binomial distribution,

$$y_{it} \mid X_{it} \sim \text{Binomial}(1, p_{it}),$$

where $p_{it} = \Pr(y_{it} = 1 \mid X_{it})$. In the logistic regression framework, the model is specified as

$$\text{logit}(p_{it}) = \log\left(\frac{p_{it}}{1 - p_{it}}\right) = \alpha + \sum_k f_k(x_{kit}) + \sum_m \gamma_m d_{mit}.$$

Here, $f_k(\cdot)$ are unknown smooth functions of continuous or pseudo-continuous covariates, such as calendar time, measures of firms' survey participation history, and averages of Business Situation and Business Expectation over the past six months. The terms d_{mit} denote parametric indicator variables, including region indicators, sector effects, and month, with corresponding coefficients γ_m. The logit link ensures that predicted values lie in the unit interval and allows the model to be interpreted in terms of changes in the probability of unit nonresponse.

Nonlinear Modeling Framework

Each smooth function $f_k(\cdot)$ is approximated using penalized regression splines. This approach represents smooth effects as linear combinations of spline basis functions, while imposing a penalty on excessive curvature in order to avoid overfitting (Eilers and Marx, 1996, Wood, 2017). The degree of smoothness is determined endogenously from the data by minimizing a penalized likelihood criterion.
Penalized splines are particularly well suited for the present application, as several key covariates exhibit nonlinear and potentially non-monotonic relationships with unit nonresponse. These include calendar time, firms' survey participation history, and average business indicators over the past six months. As shown in the descriptive analysis, simple linear specifications would be inadequate to capture these patterns.

Smoothing and Regularization

We employ P-splines as the spline basis for all smooth terms. P-splines combine the flexibility of B-spline bases with difference penalties on adjacent coefficients, yielding smooth functions that are stable and interpretable even in large samples (Eilers and Marx, 1996, Wood, 2017). The number of basis functions is chosen sufficiently large to allow flexibility, while the effective smoothness of each term is automatically controlled through penalization.
Smoothing parameters are selected using restricted maximum likelihood. REML-based smoothing parameter selection has been shown to provide reliable performance and to avoid undersmoothing problems(Wood, 2011). This property is particularly important in the present study, where the primary objective is robust statistical inference.

In addition to penalizing smooth terms, we also apply parametric penalties to the linear component of the model. Specifically, coefficients associated with parametric covariates, including region indicators, sector, and month, are subject to an L_2 penalty implemented through a parametric penalty matrix. This regularization shrinks extreme coefficient estimates toward zero (Wood, 2017).

Suitability for the data

Within this framework, each smooth function captures the population-average marginal relationship between the corresponding covariate and the probability of unit nonresponse, holding other variables constant (Hastie and Tibshirani, 1986, Wood, 2017). For a binary outcome, the estimated smooth effects describe how the conditional probability of unit nonresponse varies across the support of a covariate, rather than imposing a constant marginal effect as in a linear specification.
The assumptions underlying the generalized additive model are well aligned with the characteristics of the data used in this study. The analysis is based on a large firm-month panel, which provides sufficient information to estimate smooth functions reliably. Several key covariates naturally lend themselves to a smooth representation, including calendar time, measures of survey participation history, and six-month averages of Business Situation and Business Expectation, which are continuous in practice.
At the same time, standard estimation and inference for generalized additive models are typically derived under an assumption of independent observations (Wood, 2017). In the present setting, however, firms contribute repeated observations over time, and outcomes within firms may be correlated. While this dependence does not invalidate estimation of the conditional mean, it may affect statistical inference. To address this issue, inference is adjusted using cluster-robust standard errors at the firm level. The cluster-robust inference procedure is discussed in detail in the following subsection.

4.1.3 Cluster-Robust Inference

The generalized additive model introduced in the previous section provides a flexible framework for modeling the conditional mean of unit nonresponse. However, standard inference procedures for GAMs rely on the assumption that observations are conditionally independent given the included covariates and smooth terms (Wood, 2017). In the empirical setting considered here, this assumption is unlikely to hold. Firms are observed repeatedly over time, and observations belonging to the same firm may be correlated. To obtain reliable statistical inference in this panel-data environment, we therefore employ cluster-robust inference at the firm level.

Motivation for Cluster-Robust Inference

The data used in this study consist of firm-level monthly observations spanning a long time horizon. As a result, each firm contributes multiple observations to the data, implying that observations within the same firm are likely to be correlated over time. Such within-firm dependence may arise from unobserved firm-specific characteristics, stable reporting

practices, or persistent response behavior, and may not be fully captured by the observed covariates or smooth functions included in the model.
Standard GAM inference relies on conditional independence of observations. When this assumption is violated, conventional standard errors may be biased, leading to distorted statistical inference (Cameron and Miller, 2015). This issue is particularly relevant in large panel datasets, where within-cluster dependence can affect uncertainty estimates if not accounted for.

Cluster-Robust Covariance Estimation

To account for potential within-firm correlation, we employ a cluster-robust covariance matrix estimator, commonly referred to as the sandwich estimator, using firm identifiers as the clustering variable. This approach adjusts the estimated covariance matrix of the model coefficients while leaving the point estimates unchanged. In doing so, it relaxes the assumption of conditional independence and allows observations within the same firm to exhibit arbitrary forms of correlation (Arellano, 1987).
In general form, the cluster-robust covariance matrix follows a sandwich structure,

$$\widehat{V}_{\text{CR}} = \text{Bread} \times \text{Meat} \times \text{Bread},$$

where the *Bread* matrix corresponds to the inverse of the penalized information matrix of the fitted model, and the *Meat* matrix captures the variability of the estimating functions aggregated at the cluster level.
In the clustered setting, the Meat matrix is constructed as

$$\widehat{B} = \sum_{g=1}^{G} \left(\sum_{i \in g} \hat{\psi}_i \right) \left(\sum_{i \in g} \hat{\psi}_i \right)',$$

where $g = 1, \ldots, G$ indexes clusters (firms in this study), and $\hat{\psi}_i$ denotes the estimated contribution of observation i to the model's estimating equations. Rather than summing over individual observations, residual contributions are first aggregated within each firm and then combined across firms to form the covariance estimator.
The key intuition underlying this estimator is that it captures arbitrary forms of dependence among observations within the same firm, including serial correlation and heteroskedasticity, without requiring explicit modeling of the within-firm correlation structure (White, 1980, Cameron and Miller, 2015). At the same time, the estimator does not alter the estimated smooth functions or parametric coefficients of the GAM. The model is estimated under a working assumption of independent observations, and the robustness adjustment is applied ex post to the covariance matrix.

Plausibility of the Assumptions in the Present Setting

The assumptions underlying cluster-robust inference are relatively mild and well suited to the data environment of this study.
First, the approach assumes independence across clusters. In the present context, firms constitute natural and well-defined clusters, and after controlling for common factors such

as calendar time, sector, and region, it is reasonable to assume that unit nonresponse behavior is independent across firms.
Second, cluster-robust inference relies on asymptotics in the number of clusters rather than the number of observations (Cameron and Miller, 2015, Wooldridge, 2010). The sample contains more than 8,000 firms, providing a sufficiently large number of clusters for reliable inference. With such a large number of clusters ($G > 8,000$), the asymptotic properties of the CRVE are well-assured. The sheer scale of the cluster dimension ensures that the estimator is consistent and that our inference is robust to any arbitrary within-firm correlation pattern without relying on small-sample corrections.
Furthermore, to assess whether cluster-robust inference may be unduly influenced by a small number of dominant clusters, we compute the effective number of clusters (G_{eff}) following MacKinnon et al. (2023). While the nominal number of clusters is $G = 8{,}176$, the estimated effective number is $G_{\text{eff}} \approx 4{,}895$. The fact that the effective number of clusters remains large and constitutes a substantial fraction of the nominal count suggests that cluster sizes are not excessively concentrated and that information is not dominated by a small subset of firms. This provides support for the use of cluster-robust inference and is consistent with the large-cluster asymptotic framework underlying the validity of the corresponding standard errors.
Finally, the cluster-robust approach does not require the within-firm dependence structure to be correctly specified. This feature is particularly important in the present setting, where within-firm error correlation is likely to be complex and difficult to model explicitly. By avoiding the need to impose a potentially misspecified parametric correlation structure, cluster-robust inference provides a flexible and robust way to account for within-firm dependence (White, 1980, Cameron and Miller, 2015).

4.2 Model Results

4.2.1 Model specification

$$
\begin{aligned}
\operatorname{logit}\big(P(\text{unit\_nonres} = 1)\big) = \beta_0 &+ \boldsymbol{\beta}_{\text{State}}^{\top}\,\text{State} + \boldsymbol{\beta}_{\text{Sector}}^{\top}\,\text{Sector} + \boldsymbol{\beta}_{\text{Month}}^{\top}\,\text{Month} \\
&+ s(\text{Calendar time}) \\
&+ s(\text{Participation number}) \\
&+ s(\text{Participation length}) \\
&+ s(\text{Avg. Business Situation over past 6 months}) \\
&+ s(\text{Avg. Business Expectations over past 6 months})
\end{aligned}
\tag{1}
$$

*Notes:*Firm–month observations with zero participation history are excluded, as unit nonresponse is structurally zero at first survey participation.

4.2.2 Overall model fit

The model explains approximately 16.2% of the deviance in unit nonresponse, which is substantial given the binary nature of the outcome and the large-scale panel structure of the data.

4.2.3 Parametric effects of categorical covariates

Table 14 presents the estimated effects of the parametric components of the model, including states, sector, and month. Coefficients are presented on odds ratios (exp(Estimate)) to facilitate interpretation. All results are obtained conditional on the included nonlinear smooth terms (calendar time, participation number, participation length, and average Business Situation and Business Expectations).

State indicators reveal clear and significant differences in corporate unit nonresponse across geographic regions. Specifically, firms located in the former East German states of Mecklenburg-Vorpommern, Brandenburg, and Sachsen exhibit odds ratios significantly greater than 1, indicating a substantially higher likelihood of nonresponse compared to Berlin West. Conversely, the majority of the former West German states, including Bremen, Niedersachsen, Nordrhein-Westfalen, Hessen, Baden-Wuerttemberg, and Bayern, show odds ratios significantly below 1. This suggests that enterprises in these Western regions exhibit a risk of unit nonresponse that is significantly lower than the level observed in Berlin West.

The sectoral heterogeneity is a relevant factor in survey participation behavior, with the majority of industrial sectors exhibiting a significantly lower risk of unit nonresponse compared to the reference category (Food, Drinks and Tobacco). Specifically, the Textile, Clothing, Leather sector and the Paper & Board Industry display the strongest negative effects, with odds ratios significantly below 0.9. Conversely, the Publishing & Printing sector represents a notable outlier, showing a statistically significant positive effect (OR > 1) on the probability of unit nonresponse, suggesting a higher likelihood of non-participation relative to the baseline.

The month indicators reveal clear and differences in unit nonresponse across calendar months. Holding all other covariates constant, the odds of unit nonresponse are higher in several months relative to the reference month (January). In particular, the months of August and December exhibit the highest risks, with odds ratios of approximately 1.15. This indicates that, ceteris paribus, the odds of unit nonresponse in August and December are roughly 15% higher than in January.

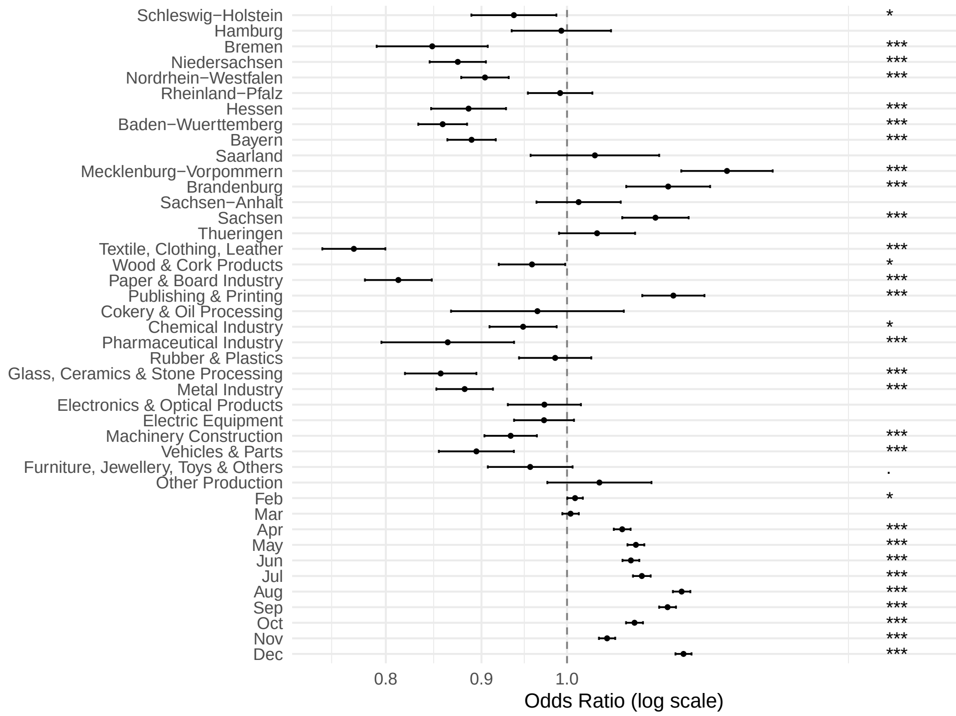

Figure 14: Cluster-robust odds ratios for parametric covariates from the GAM model. Dots indicate point estimates, and horizontal bars denote 95% confidence intervals based on firm-level clustered standard errors. The vertical dashed line marks an odds ratio of one.

Notes: Estimates are reported on the log-odds scale. Odds ratios are obtained by exponentiating the corresponding coefficients. Berlin West, and the sector “Food, Drinks & Tobacco” and January serve as reference categories.

4.2.4 Nonlinear effects of key covariates

To formally assess the statistical relevance of the nonlinear components, Table 2 summarizes the estimated smooth terms and their corresponding significance tests. All smooth terms included in the model exhibit highly significant nonlinear effects on the probability of unit nonresponse.

Table 2: Approximate Significance of Smooth Terms

Smooth term	**edf**	**Ref.df**	**Chi.sq**	**p-value**
s(Calendar time)	44.73	49	7,005	$< 2 \times 10^{-16}$ ***
s(Participation number)	47.10	49	63,080	$< 2 \times 10^{-16}$ ***
s(Participation length)	47.00	49	140,266	$< 2 \times 10^{-16}$ ***
s(Average BS over six months)	17.90	19	32,695,050	$< 2 \times 10^{-16}$ ***
s(Average BE over six months)	16.96	19	109,473	$< 2 \times 10^{-16}$ ***

The Spline figures below report the partial effects of each covariate on the log-odds of unit nonresponse, holding all other variables constant. Shaded areas indicate pointwise

95% confidence intervals based on cluster-robust standard errors.

Time-Related Variables
Figure 15 illustrates the estimated smooth effect of calendar time on the log-odds of unit nonresponse. The results reveal a highly non-linear temporal trend.
A positive value of the calendar-time smooth indicates that calendar time contributes positively to the log-odds of unit nonresponse, corresponding to a higher level of nonresponse risk relative to its overall average level implied by the model over the observed data. Conversely, a negative value indicates a lower level of nonresponse risk relative to this average, holding other covariates constant.
In the early 1990s, the calendar time contribution is initially negative, suggesting relatively low nonresponse risk during this period.
This is followed by a rapid increase in nonresponse risk and then a clear decline, reaching a minimum in the late 1990s to early 2000s.
After the early 2000s, the effect increases markedly.
From around the mid-2000s onward, the calendar time smooth becomes positive and remains relatively stable, with only limited short-term fluctuations. This pattern indicates that, from around the mid-2000s onward, calendar time generally contributes positively to the log-odds of unit nonresponse, corresponding to a higher level of nonresponse risk relative to the overall average.
Overall, the estimated smooth highlights several distinct phases in the long-run evolution of unit nonresponse behavior.

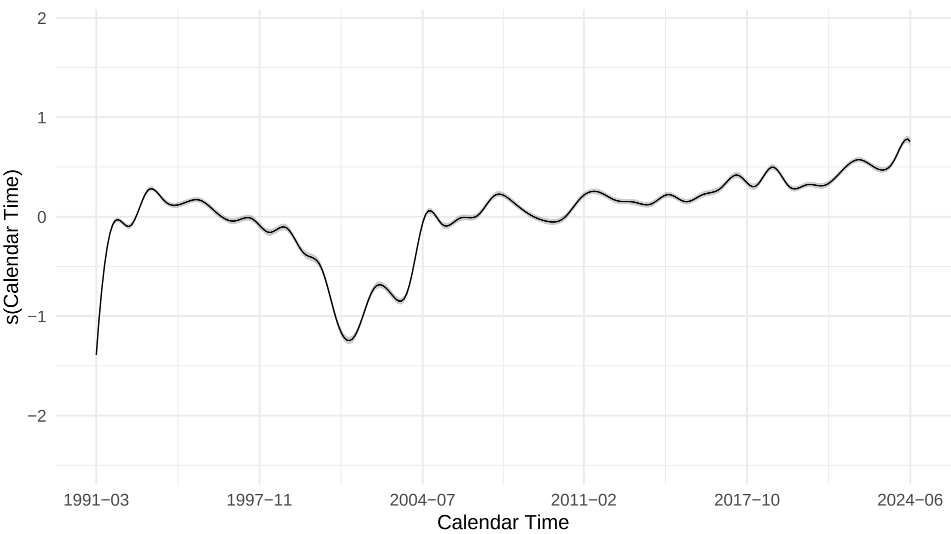

Figure 15: Nonlinear effect of calendar time on unit nonresponse

Survey-Related Variables
Figure 16 illustrates the estimated smooth effect of participation number on the log-odds of unit nonresponse. The results indicate a nonlinear relationship between the cumulative number of completed survey responses and nonresponse risk.

A positive value of the participation-number smooth indicates that participation number contributes positively to the log-odds of unit nonresponse, corresponding to a higher level of nonresponse risk relative to the overall average implied by the model over the observed data. Conversely, a negative value indicates a lower level of nonresponse risk relative to this average, holding other covariates constant.
At low levels of participation number, the smooth effect is positive, indicating relatively high nonresponse risk among firms with few completed responses. As participation number increases, the smooth function declines and becomes negative, suggesting that firms with a growing history of completed responses are less likely to exhibit unit nonresponse.

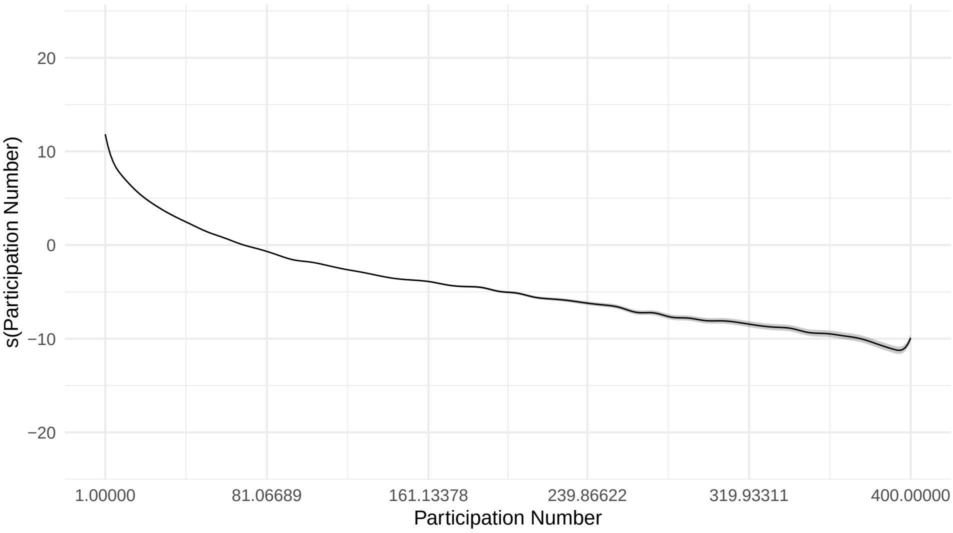

Figure 16: Nonlinear effect of participation number on unit nonresponse

Figure 17 illustrates the estimated smooth effect of participation length on the log-odds of unit nonresponse. The results reveal a pronounced and highly nonlinear relationship between participation length and nonresponse risk.
A positive value of the participation-length smooth indicates that participation length contributes positively to the log-odds of unit nonresponse, corresponding to a higher level of nonresponse risk relative to the overall average implied by the model over the observed data. Conversely, a negative value indicates a lower level of nonresponse risk relative to this average, holding other covariates constant.
At short participation lengths, the smooth effect is negative, suggesting that firms in the early stages of survey participation exhibit relatively low nonresponse risk. As participation length increases further, the smooth effect rises steadily and eventually turns positive at higher participation lengths, indicating that firms with long participation histories face a higher level of unit nonresponse risk relative to the model-implied average.

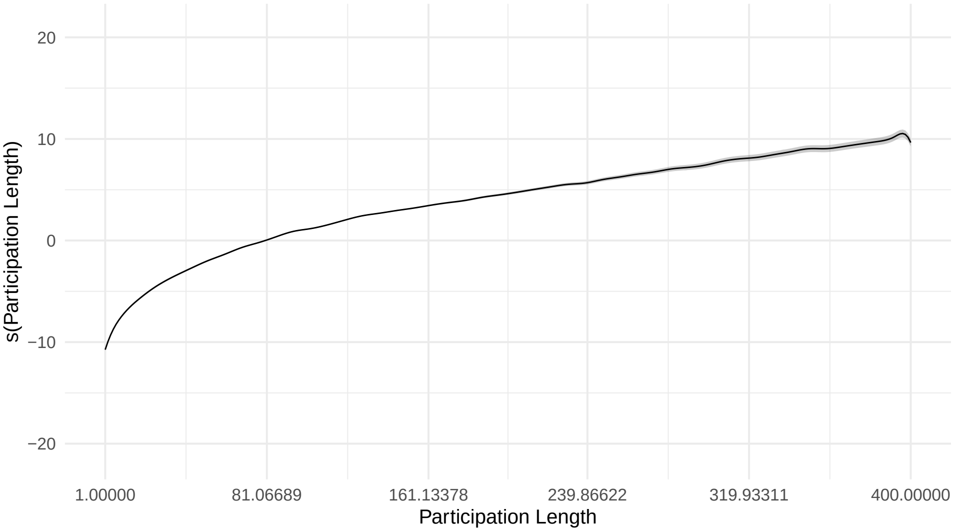

Figure 17: Nonlinear effect of participation length on unit nonresponse

Although participation number and participation length are often highly correlated empirically, they differ in their conceptual construction and capture distinct aspects of survey participation behavior rather than representing alternative measures of the same latent concept.

Participation length measures the total duration of a firm's presence in the survey since its first appearance in the sample, reflecting cumulative exposure to the survey system and not requiring a response in every survey wave. In contrast, participation number records the cumulative count of survey waves in which the firm actually provided a response, thereby more directly capturing realized response behavior.

As shown in the Figure 9, a high correlation between participation number and participation length at beginning. As participation length increases, however, this correlation declines substantially, indicating that response behavior among long-standing firms becomes increasingly selective. In other words, as the survey history unfolds, whether a firm responds in a given wave is no longer closely aligned with the length of time it has remained.

Against this background, the distinct and even opposing nonlinear patterns observed in the smooth estimates of the two variables admit a clear structural interpretation. Higher participation number reflects accumulated response activity and familiarity with the survey process and is associated with lower unit nonresponse risk. In contrast, longer participation length captures prolonged exposure to the survey and, at long durations, is associated with higher unit nonresponse risk.

Therefore, the opposing nonlinear patterns of the smooth functions for participation number and participation length do not constitute a contradiction but rather highlight the multidimensional nature of survey participation history. Treating the two measures as separate covariates allows the model to distinguish between response incidence and long-term survey exposure, distinctions that would be obscured in a simple linear specification or by relying on a single participation measure.

Two Main Questions

Table 2 shows that the smooth terms for average Business Situation (BS) and Business Expectations (BE) are statistically significant, indicating that these variables enter the model in a nonlinear fashion. However, both variables take values on a discrete and unevenly distributed support, with only a limited number of distinct realizations. As a result, the estimated spline shapes for BS and BE are sensitive to the choice of the basis dimension k.

When the shapes of smooth functions vary substantially with alternative specifications of k, their detailed functional forms do not admit a robust substantive interpretation. Accordingly, the smooth effects of BS and BE are treated primarily as flexible control components rather than as objects of direct interpretation in the main text.

For completeness, the estimated spline functions for BS and BE based on the baseline specification ($k = 20$) are reported in the Appendix A.3.1. The Appendix A.3.1 further presents corresponding spline estimates obtained under alternative choices of the basis dimension ($k = 12$ and $k = 8$), illustrating the sensitivity of the BS and BE smooths to the choice of k.

4.3 Summary

This section examines the determinants of unit nonresponse in the ifo Business Survey using a cluster-robust generalized additive model at the firm–month level. The modeling framework combines flexible nonlinear smooth terms with robust inference to account for both complex covariate effects and within-firm dependence in a large panel of manufacturing firms.

The results reveal substantial nonlinearities in several key dimensions. Calendar time exhibits a pronounced and highly non-monotonic pattern, indicating that the underlying level of unit nonresponse evolves in distinct phases over the observation period. Survey participation history also plays a central role. Participation number and participation length display opposing nonlinear effects: firms with a higher cumulative number of completed responses tend to exhibit lower nonresponse risk, while firms with long participation histories face elevated nonresponse risk at higher durations. These patterns underscore the importance of distinguishing between different dimensions of survey participation history.

Among the parametric covariates, clear differences emerge across states, sectors and months. State indicators show that firms in some East Germany states have significantly higher odds of unit nonresponse compared to Berlin West. In contrast, most West German states exhibit a significantly lower risk of nonresponse. The majority of industrial sectors show a lower risk of unit nonresponse than the reference group, while Publishing & Printing is a rare exception with a significantly higher nonresponse risk. Month indicators reveal variation in nonresponse throughout the year. Relative to January, the likelihood of non-participation is approximately 15% higher in August and December.

Finally, average Business Situation and Business Expectations over the past six months enter the model in a statistically significant but highly nonlinear manner. However, due to their discrete and uneven support, the detailed shapes of their smooth functions are sensitive to the choice of spline basis dimension and are therefore treated primarily as flexible control components rather than objects of substantive interpretation.

5 Imputation

5.1 Missing-data setting and notation

We study a firm-level monthly panel indexed by firms $i = 1, \ldots, N$ and calendar months $t = 1, \ldots, T$, based on observed responses from a repeated business survey. In each survey wave, firms answer a set of questions about their current situation, recent developments, and expectations. For the imputation analysis, we focus on two core outcomes central to the survey and its derived indicators: Business Situation (BS) and Business Expectations (BE).

Let Y_{it}^{BS} denote the reported *Business Situation* of firm i in month t, and let Y_{it}^{BE} denote the firm's reported *Business Expectations*. Both variables take values in the finite set $\{1, 2, 3\}$ and are ordinal in nature. For Business Situation, the categories correspond to "good," "satisfactory," and "poor," while for Business Expectations they correspond to "become more favorable," "remain roughly the same," and "become less favorable."

For notational convenience, we use Y_{it} to generically denote either outcome variable, with $Y_{it} \in \mathcal{S}$ and $\mathcal{S} = \{1, 2, 3\}$.

Under unit nonresponse, the questionnaire is not returned, implying that survey outcomes such as Y_{it}^{BS} and Y_{it}^{BE} are not observed for that firm-month. In addition, missing values may arise from *item nonresponse*, where a firm participates in month t but leaves one or more questions unanswered. To capture missingness at the item level, let R_{it}^{BS} and R_{it}^{BE} denote indicators for whether BS and BE are observed, respectively, with

$$R_{it}^{m} = \begin{cases} 1, & \text{if } Y_{it}^{m} \text{ is observed,} \\ 0, & \text{if } Y_{it}^{m} \text{ is missing,} \end{cases} \qquad m \in \{BS, BE\}.$$

In the remainder of this section and the subsequent imputation analysis, we treat missing values in Y_{it}^{BS} and Y_{it}^{BE} arising from either unit or item nonresponse within a unified framework.

A key feature of the missing-data structure is the presence of *consecutive missing observations*. For a given firm i, we define a missing gap of length k as a sequence of months $t, t+1, \ldots, t+k-1$ for which $R_{i\tau} = 0$ for all $\tau \in \{t, \ldots, t+k-1\}$. The gaps of length $k = 1, \ldots, 6$ account for more than 95% of all missing observations, and the subsequent simulation study concentrates on this range of gap lengths accordingly.

Let X_{it} denote the vector of covariates available for imputation, including firm characteristics, calendar-time indicators, survey-related variables, and lagged values of Y_{it}^{BS} and Y_{it}^{BE} when observed. These covariates are incorporated exclusively in the Random Forest imputation approach, reflecting its ability to accommodate heterogeneity, whereas LOCF and homogeneous Markov Chain methods rely solely on past observations of the outcome variable.

The goal of this chapter is to evaluate and compare different imputation methods for missing BS and BE values, focusing on predictive performance, calibration, and the resulting aggregate balance measures.

5.2 Imputation methods

We compare three imputation strategies that differ in their assumptions and degree of flexibility. The Last Observation Carried Forward (LOCF) approach assumes that a missing observation takes the same value as the most recently observed outcome for the same firm. The homogeneous Markov Chain introduces dynamic transitions between outcome categories, but imposes the restriction that transition probabilities are identical across firms and over time. In contrast, the Random Forest approach accounts for heterogeneity through covariates.

5.2.1 Last Observation Carried Forward (LOCF)

The Last Observation Carried Forward (LOCF) method is a simple rule-based imputation approach commonly used in longitudinal data settings. Under LOCF, a missing observation for firm i at time t is imputed using the most recent non-missing value observed for the same firm. For a missing observation at time t, the imputed value is given by

$$\hat{Y}_{it} = Y_{it^*}, \qquad t^* = \max\{s < t : R_{is} = 1\}.$$

In the case of consecutive missing observations, the last observed value is carried forward repeatedly until a new observation becomes available.
The LOCF approach implicitly assumes that the outcome remains constant over the missing period. As a deterministic procedure, LOCF assigns a single fixed value to each missing observation and therefore does not account for uncertainty in the imputed values. Moreover, the method relies exclusively on a firm's own last observed value and does not incorporate additional information. As a result, LOCF does not allow for heterogeneity. In this study, LOCF is included as a transparent and easily interpretable benchmark for comparing alternative imputation strategies.

5.2.2 Homogeneous Markov Chain (Homogeneous MC)

We model the dynamics of the categorical survey outcomes as a homogeneous Markov Chain with a finite state space(Norris, 1998). The imputation is conducted under the missing-data setting and notation introduced in Section 5.1, where missing values in the outcome variables may arise from both unit and item nonresponse. A Markov Chain of order p assumes that the conditional distribution of Y_{it} depends only on the previous p realized states, that is,

$$\mathbb{P}(Y_{it} = b \,|\, Y_{i,t-1} = a_1, \ldots, Y_{i,t-p} = a_p, \mathcal{H}_{i,t-p-1}) = \mathbb{P}(Y_{it} = b \,|\, Y_{i,t-1} = a_1, \ldots, Y_{i,t-p} = a_p)$$

for all $b \in \mathcal{S}$, where $\mathcal{H}_{i,t-p-1}$ denotes the history prior to $t-p$.
The homogeneity assumption imposes that the transition probabilities do not vary across firms i or months t. Hence, the transition mechanism is fully characterized by the order-p transition probabilities

$$P^{(p)}_{a_1,\ldots,a_p \to b} := \mathbb{P}(Y_{it} = b \,|\, Y_{i,t-1} = a_1, \ldots, Y_{i,t-p} = a_p), \qquad a_1, \ldots, a_p, b \in \mathcal{S}.$$

In our application, the Markov model is estimated separately for Business Situation (BS) and Business Expectations (BE). We experimented with higher-order homogeneous

Markov Chains, but focus on the first-order specification for simplicity, as higher-order variants did not yield meaningful performance gains.

Estimation. Transition probabilities are estimated from observed transitions by pooling all firms and months. For each observed history $(a_1, \dots, a_p)$, we compute the share of times the outcome moves to state b in the next month. Formally,

$$\widehat{P}^{(p)}_{a_1,\dots,a_p \to b} = \frac{N^{(p)}_{a_1,\dots,a_p \to b}}{\sum_{b' \in \mathcal{S}} N^{(p)}_{a_1,\dots,a_p \to b'}},$$

whenever the denominator is positive. If a particular history $(a_1, \dots, a_p)$ is not observed in the data (i.e., the denominator is zero), we fall back to a global distribution over states:

$$\widehat{\pi}_b = \frac{\sum_{i,t} \mathbf{1}\{Y_{it} = b \text{ is observed}\}}{\sum_{i,t} \mathbf{1}\{Y_{it} \text{ is observed}\}}, \qquad b \in \mathcal{S}.$$

Imputation of consecutive missing observations. Consider a missing gap of length k starting at month t, i.e., $Y_{i,t}, \dots, Y_{i,t+k-1}$ are missing. Imputation proceeds sequentially. Given the most recent available history of length p prior to the gap, we generate an imputed value for $Y_{i,t}$ based on the estimated transition probabilities $\widehat{P}^{(p)}$ and then update the history by appending the imputed state. The same procedure is repeated for $Y_{i,t+1}, \dots, Y_{i,t+k-1}$.
At each step, the imputed outcome is chosen as the modal category[1] implied by the conditional transition distribution, that is,

$$\hat{Y}_{it} = \arg\max_{b \in \mathcal{S}} \widehat{P}^{(p)}_{a_1,\dots,a_p \to b},$$

where $(a_1, \dots, a_p)$ denotes the current history, consisting of observed and/or previously imputed outcomes.

Discussion of assumptions. The homogeneous Markov Chain captures time dynamics in the survey outcomes via state transitions, but it imposes that the transition mechanism is identical across firms and months and does not incorporate covariates. Consequently, any heterogeneity related to firm characteristics, calendar time, or survey conditions is not modelled within this approach.

5.2.3 Random Forest (RF)

We impute missing Business Situation (BS) and Business Expectations (BE) using a Random Forest (RF) classifier (Breiman, 2001), which estimates the conditional distribution

[1] Two approaches are commonly used to generate imputations from a Markov Chain: sampling from the conditional transition distribution and selecting the modal category. We experimented with both approaches under identical simulation settings and Markov Chain orders. In our simulations, selecting the modal category consistently outperformed stochastic sampling. We therefore drop the stochastic sampling approach throughout the analysis.

of the categorical outcome $Y_{it} \in \mathcal{S} = \{1, 2, 3\}$ given a set of covariates X_{it}. LOCF and the homogeneous MC do not incorporate heterogeneity, whereas the RF approach introduces heterogeneity through covariate-dependent predictions and does not impose a parametric functional form.
As a predictive imputation tool, the Random Forest relies on the informativeness of the included covariates and is most naturally aligned with a missing-at-random (MAR) perspective conditional on X_{it}.

Model and prediction target. Let X_{it} denote the covariate vector available at firm i and month t. The Random Forest estimates the conditional class probabilities

$$\widehat{p}_{it}(b) \;:=\; \widehat{\mathbb{P}}(Y_{it} = b \,|\, X_{it})\,, \qquad b \in \mathcal{S},$$

where $\widehat{p}_{it}(b)$ is given by the proportion of trees in the forest that predict class b at (i, t). In our application, separate RF models are trained for Business Situation (BS) and Business Expectations (BE), with the same set of covariates.

Covariates. The feature set X_{it} used in the Random Forest imputation is summarized in Table 3.[2]

Table 3: Covariates used in the Random Forest imputation

Category	**Variables**
Firm identifier	Company ID
Firm characteristics	Region (West/East)
	Sector
Survey mode	Online vs. offline questionnaire
Time-related variables	Calendar time
	Indicator for August
	Indicator for December
Lagged survey outcomes	Lagged BS values (lags 1–4)
	Lagged BE values (lags 1–4)

With this covariate specification, the Random Forest accounts for heterogeneity in the imputation process. In particular, firm-level and time-related heterogeneity are captured through the inclusion of a firm identifier and calendar time. Additional covariates further account for heterogeneity related to firm characteristics, survey mode, and recent response history, while lagged outcomes reflect short-run persistence in survey responses.
The choice of including lags 1–4 of BS and BE is motivated by comparability considerations. In the homogeneous Markov Chain approach, we evaluate models of order $p = 1, \ldots, 4$ (MC1–MC4). Using the same lag length in the Random Forest ensures that both approaches are based on an equivalent amount of historical outcome information.

[2] To ensure a complete covariate matrix, we impute missing values in all covariates using a firm-level LOCF procedure prior to Random Forest training and imputation.

Imputation of consecutive missing observations. Consider a missing gap of length k starting at month t, i.e., $Y_{i,t}, \ldots, Y_{i,t+k-1}$ are missing. Imputation is carried out sequentially over the gap. For lagged outcome covariates, observed lag values are used whenever available. Only when a required lag value is missing is it replaced by the corresponding imputed outcome from previous steps.
Given $X_{i\tau}$, we compute the predicted class probabilities $\widehat{p}_{i\tau}(b)$ and impute the outcome as the modal class

$$\hat{Y}_{i\tau} \;=\; \arg\max_{b \in \mathcal{S}} \widehat{p}_{i\tau}(b).$$

Hyperparameter tuning. We tune the RF via a grid search over three hyperparameters: (i) *mtry* (number of variables sampled at each split), (ii) *ntree* (number of trees), and (iii) *nodesize* (minimum terminal node size). The optimal combination is selected based on the out-of-bag (OOB) prediction error, which provides an internal cross-validation mechanism without an explicit train–test split.

5.3 Simulation

Objective and general design. In the observed survey data, missing values in Business Situation (BS) and Business Expectations (BE) do not have an observable ground truth, which precludes a direct evaluation of imputation performance. To assess and compare the performance of different imputation methods, we conduct a controlled simulation study. The key idea is to randomly mask observed outcome values and to compare the resulting imputations with the original observed responses. To ensure comparability, all imputation methods are compared using identical simulated datasets for each consecutive-missingness scenario.

Construction of missingness patterns and repetition. In the dataset, more than 95% of all missing observations in both Business Situation (BS) and Business Expectations (BE) occur in consecutive gaps of length 1 to 6. The simulation study therefore focuses on these consecutive-missingness patterns. To reduce computational burden while preserving the empirical panel structure relevant for the simulation, we restrict the analysis to the period from 2014 to 2024. This subsample retains a large number of firms and observations and continues to exhibit the dominant consecutive-missingness patterns of the full dataset. For each gap length k, we randomly select 10% of observed outcomes and mask the corresponding BS or BE values for k consecutive months. For each consecutive-missingness setting, we repeat the simulation 30 times with independent random draws. Repeating the simulation reduces the influence of random variation and allows us to assess the stability of imputation performance across methods.

Simulation procedure. For each gap length $k = 1, \ldots, 6$ and each simulation repetition $r = 1, \ldots, 30$:

1. Randomly sample blocks of k consecutive observed outcomes within firms such that the total number of sampled observations is as close as possible to 10% of the observed outcomes.

2. Mask the outcomes within the sampled k-period blocks.
3. Apply each imputation method to the resulting masked dataset.
4. Store the resulting imputed dataset for evaluation.

The procedure is carried out separately for Business Situation (BS) and Business Expectations (BE).

5.4 Evaluation Metrics

In this section, we compare the performance of different imputation methods based on the simulation results. Three complementary evaluation metrics are used to assess different aspects of performance: Accuracy, Cohen's κ, and Spearman's ρ. In addition, we evaluate the calibration of predicted class probabilities for the best-performing method, namely the Random Forest approach. LOCF and MC are deterministic and therefore excluded from calibration analysis.

5.4.1 Point prediction metrics

Accuracy. Accuracy measures the proportion of masked observations that are correctly imputed and is defined as

$$\text{Accuracy} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(\hat{y}_i = y_i),$$

where $\hat{y}_i$ and y_i denote the imputed and true class labels for observation i, respectively, and n is the total number of masked observations.
Figure 18 and Figure 19 report the distribution of accuracy across simulation repetitions for BS and BE, respectively. Across simulation repetitions, the Random Forest approach exhibits higher accuracy on average than the alternative methods across gap lengths. LOCF serves as a strong baseline and yields competitive accuracy. MC1 exhibits lower accuracy than the other imputation methods across the considered settings. Results for MC2–MC4 are omitted, as MC1 shows better performance than these variants under identical configurations. For Cohen's κ and Spearman's rank correlation ρ, results for MC2–MC4 are also omitted for the same reason. As expected, imputation performance deteriorates as the length of consecutive missing periods increases, while BS generally exhibits higher accuracy than BE under the same imputation method and gap-length setting.

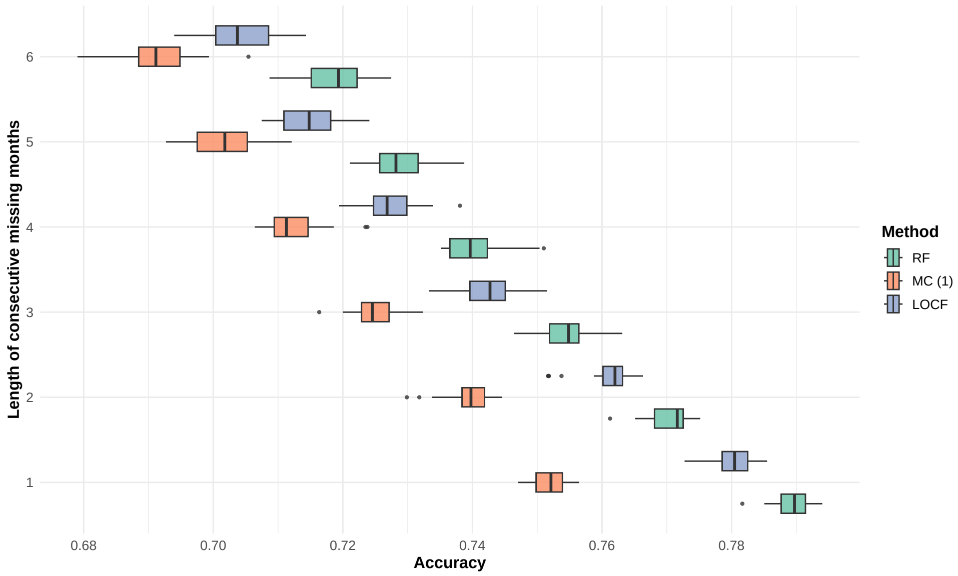

Figure 18: Distribution of accuracy across simulation repetitions for BS.

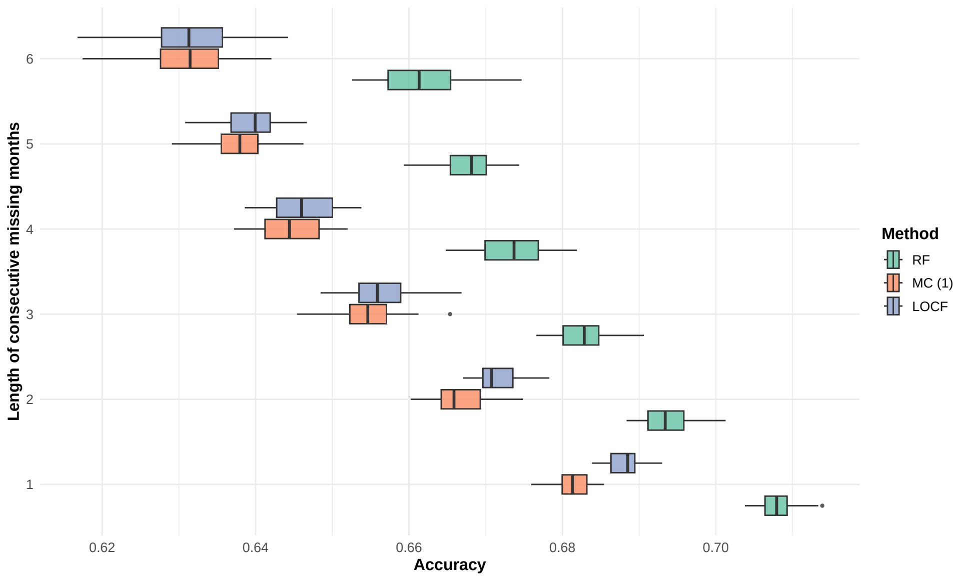

Figure 19: Distribution of accuracy across simulation repetitions for BE.

Cohen's κ. Cohen's κ measures the level of agreement between the imputed and true class labels beyond what would be expected by chance. It is defined as

$$\kappa = \frac{p_o - p_e}{1 - p_e},$$

where p_o denotes the observed agreement and p_e denotes the expected agreement under random guessing. The observed agreement p_o is computed as

$$p_o = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(\hat{y}_i = y_i).$$

The expected agreement p_e is calculated from the class frequencies of the imputed and true labels,

$$p_e = \sum_{c=1}^{C} p_c^{(\mathrm{imp})}\, p_c^{(\mathrm{true})},$$

with

$$p_c^{(\mathrm{imp})} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\hat{y}_i = c), \qquad p_c^{(\mathrm{true})} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(y_i = c),$$

where C is the total number of classes. By accounting for agreement under the empirical class frequencies, Cohen's κ complements accuracy by distinguishing genuine predictive agreement from agreement that may arise due to the class distribution. This distinction is particularly important in our setting, as both BS and BE exhibit class imbalance with class 2 being the majority class. Higher values of κ indicate stronger agreement between the imputed and true labels beyond what would be expected by chance.
Figure 20 and Figure 21 report the distribution of Cohen's κ across simulation repetitions for BS and BE, respectively. The performance comparison based on Cohen's κ leads to conclusions consistent with those obtained using accuracy. κ values are generally lower than the corresponding accuracy values across methods, reflecting the adjustment for agreement arising from the majority class.

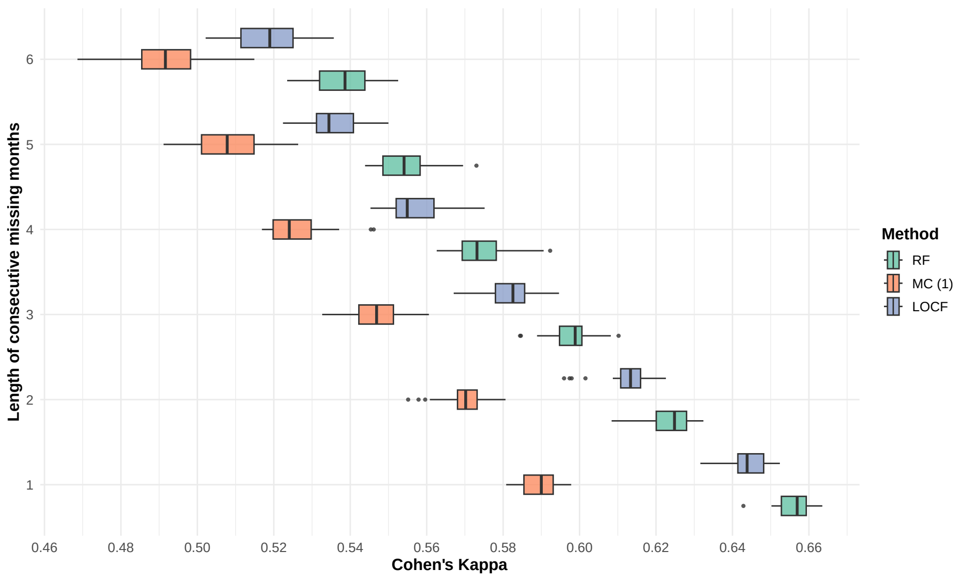

Figure 20: Distribution of Cohen's κ across simulation repetitions for BS.

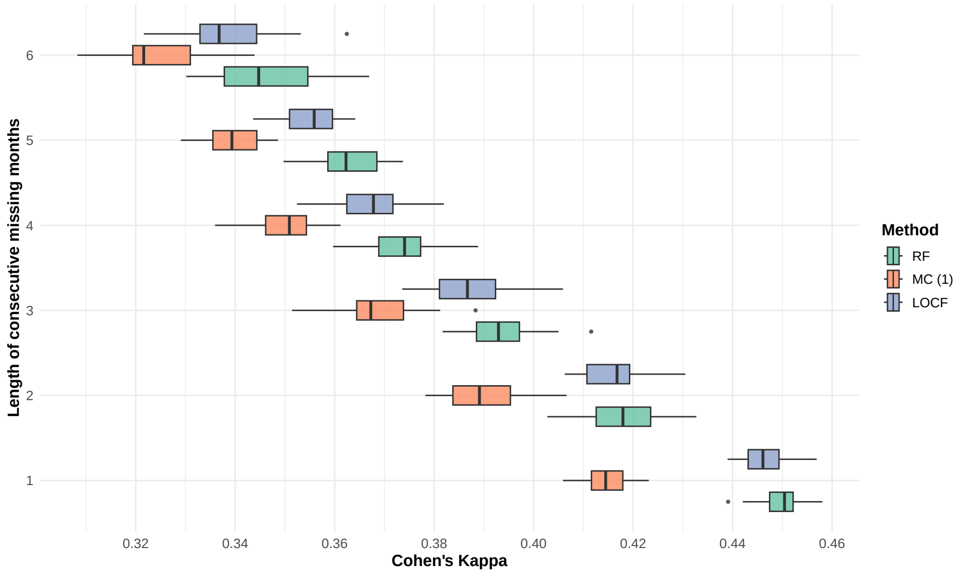

Figure 21: Distribution of Cohen's κ across simulation repetitions for BE.

Spearman's ρ. Spearman's ρ measures the consistency between the rankings of samples derived from the true labels and those derived from the imputed labels. Formally, Spearman's ρ is defined as the Pearson correlation between the ranked imputed and true labels,

$$\rho = \operatorname{corr}\big(\operatorname{rank}(\hat{y}), \operatorname{rank}(y)\big),$$

where $\operatorname{rank}(\cdot)$ denotes the rank transformation.

Unlike accuracy and Cohen's κ, which treat class labels as nominal categories, Spearman's ρ accounts for ordinal structure and can capture agreement in relative ordering even when exact label matches are not achieved.

Figure 22 and Figure 23 report the distribution of Spearman's ρ across simulation repetitions for BS and BE, respectively. The performance comparison based on Spearman's ρ leads to conclusions that are consistent with those obtained using accuracy and Cohen's κ.

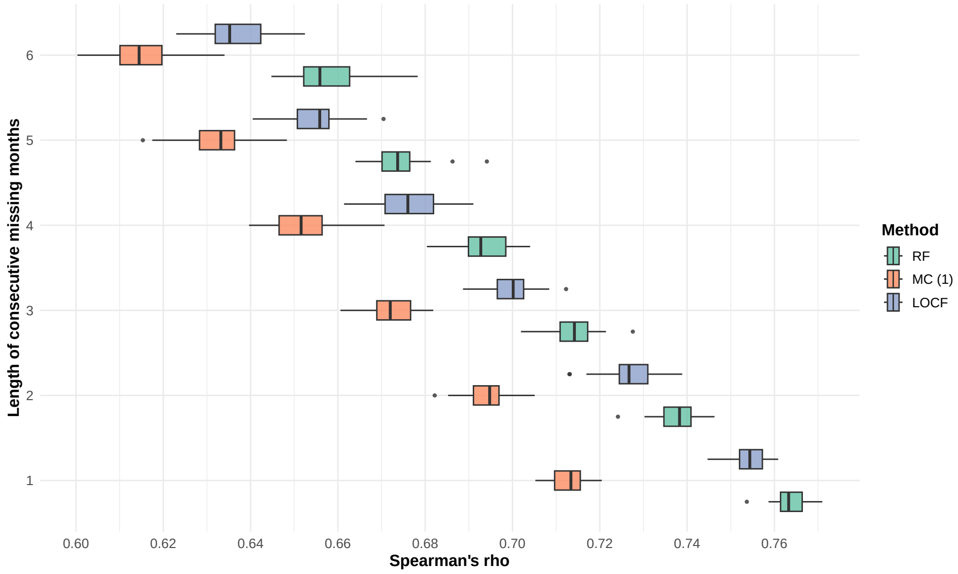

Figure 22: Distribution of Spearman's ρ across simulation repetitions for BS.

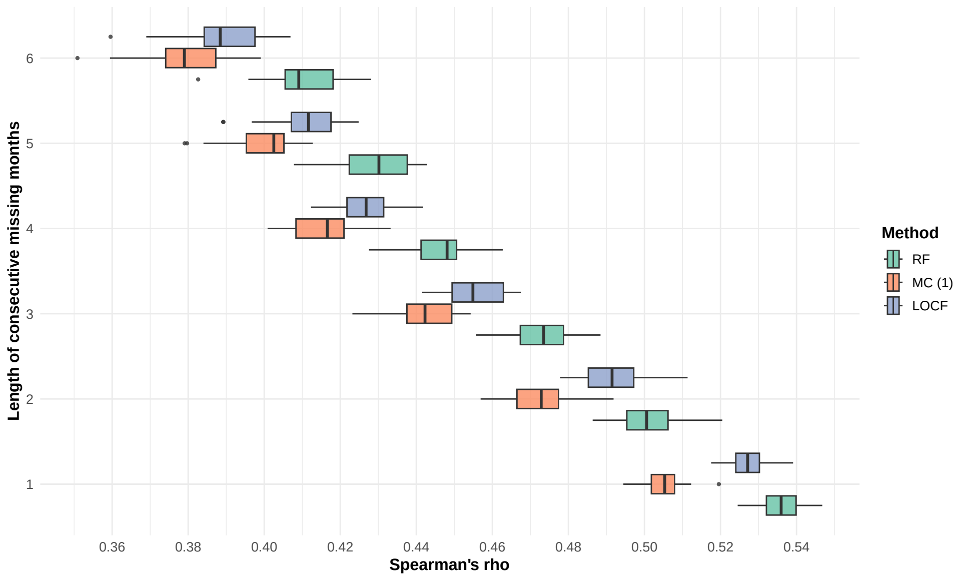

Figure 23: Distribution of Spearman's ρ across simulation repetitions for BE.

5.4.2 Calibration (Random Forest)

While point prediction metrics such as accuracy, Cohen's κ, and Spearman's ρ evaluate different aspects of agreement between imputed and observed outcomes, they do not reveal whether an imputation method systematically favors certain outcome categories over others.

Consequently, if the predicted probabilities for a given category are systematically overestimated (or underestimated) relative to the corresponding empirical frequencies, the argmax rule will tend to over-assign (or under-assign) missing observations to that category. Since LOCF and the mode-based Markov Chain imputations are deterministic, the calibration analysis is restricted to the Random Forest approach.

Short missing gaps ($k = 1$). Figures 24 and 25 show that, for short missing gaps ($k = 1$), the calibration curves for all outcome categories lie close to the diagonal, with no clear evidence of systematic over- or under-assignment to any particular class. This indicates that under short-term missingness the Random Forest imputation is well calibrated across the three categories for both BS and BE.

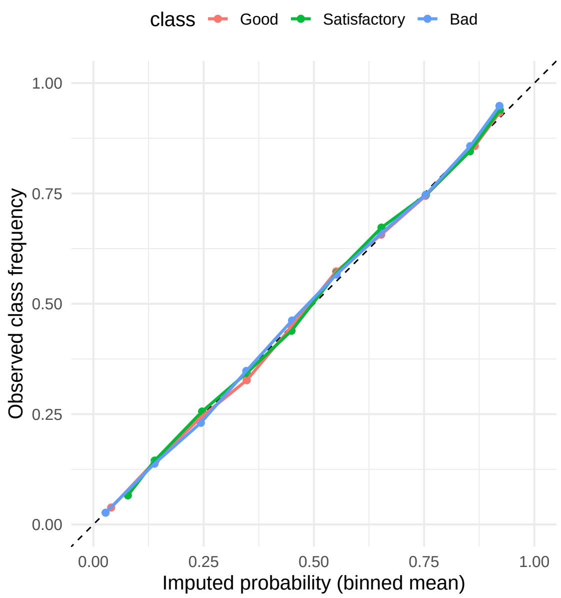

Figure 24: Calibration plot for Random Forest imputation of Business Situation with short missing gaps ($k = 1$).

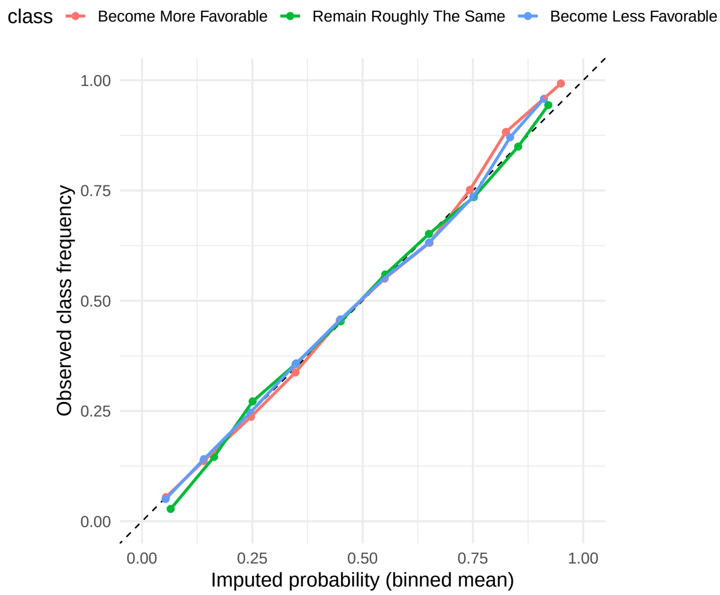

Figure 25: Calibration plot for Random Forest imputation of Business Expectations with short missing gaps ($k = 1$).

Long missing gaps ($k = 6$). For long missing gaps ($k = 6$), Figures 26 and 27 display clearer deviations from the diagonal. In the low-frequency region, predicted probabilities tend to be underestimated, whereas in the high-frequency region predicted probabilities tend to be overestimated. For BS, the category "Satisfactory" is best calibrated in the high-frequency region, while "Bad" is best calibrated in the low-frequency region. A similar pattern holds for BE, where "Remain roughly the same" is best calibrated at high frequencies and "Become less favorable" at low frequencies.

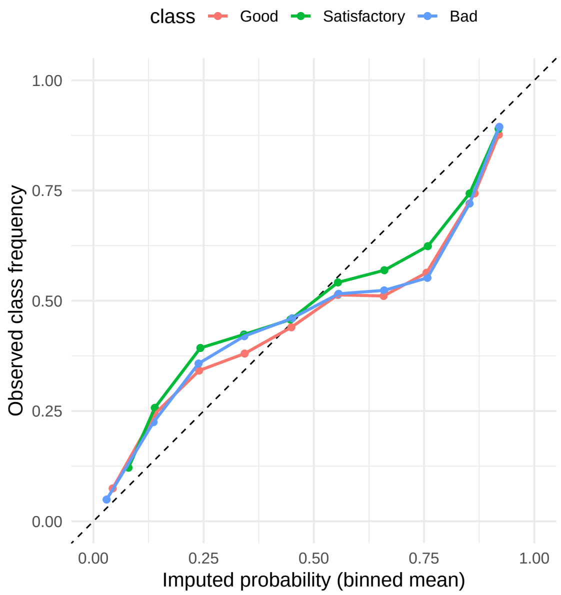

Figure 26: Calibration plot for Random Forest imputation of Business Situation with long missing gaps ($k = 6$).

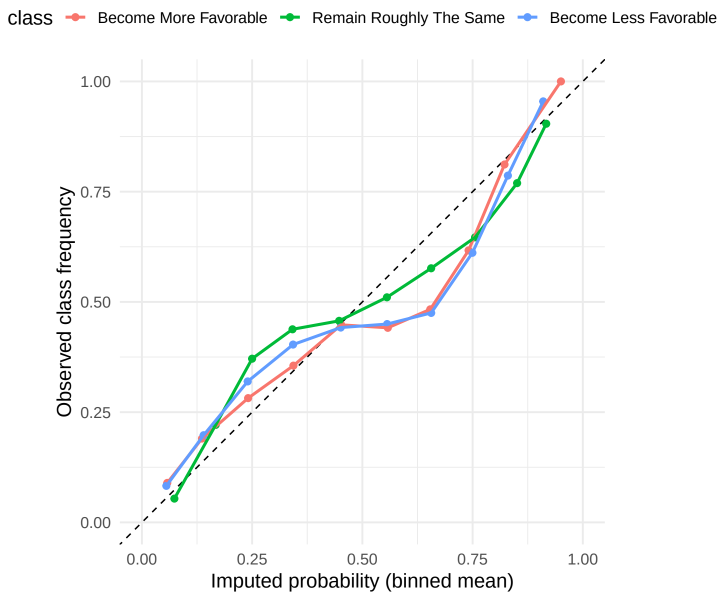

Figure 27: Calibration plot for Random Forest imputation of Business Expectations with long missing gaps ($k = 6$).

5.5 Balance

Motivation and definition. The evaluation metrics in Section 5.4 quantify imputation performance at the point level. Both the evaluation metrics and the calibration analysis are based on simulated missingness scenarios, in which the true underlying val-

ues are observed and methods can be assessed under controlled conditions. In contrast, balance measures are aggregate indicators evaluated using the original survey data, where imputation is applied to real missingness rather than to simulated missingness.
For both Business Situation and Business Expectations, balance is the difference between positive and negative assessments (categories 1 and 3), divided by the total number of observations in a given month. Formally, the balance measures are defined as

$$\text{Balance}_t^{BS,\text{real}} = \frac{\sum_i \mathbb{I}\big(Y_{it}^{BS} = 1\big) - \sum_i \mathbb{I}\big(Y_{it}^{BS} = 3\big)}{\sum_i \mathbb{I}(Y_{it}^{BS} \in \{1,2,3\})},$$

$$\text{Balance}_t^{BS,\text{imp}} = \frac{\sum_i \mathbb{I}\Big(\tilde{Y}_{it}^{BS} = 1\Big) - \sum_i \mathbb{I}\Big(\tilde{Y}_{it}^{BS} = 3\Big)}{\sum_i \mathbb{I}\Big(\tilde{Y}_{it}^{BS} \in \{1,2,3\}\Big)},$$

$$\text{Balance}_t^{BE,\text{real}} = \frac{\sum_i \mathbb{I}\big(Y_{it}^{BE} = 1\big) - \sum_i \mathbb{I}\big(Y_{it}^{BE} = 3\big)}{\sum_i \mathbb{I}(Y_{it}^{BE} \in \{1,2,3\})},$$

$$\text{Balance}_t^{BE,\text{imp}} = \frac{\sum_i \mathbb{I}\Big(\tilde{Y}_{it}^{BE} = 1\Big) - \sum_i \mathbb{I}\Big(\tilde{Y}_{it}^{BE} = 3\Big)}{\sum_i \mathbb{I}\Big(\tilde{Y}_{it}^{BE} \in \{1,2,3\}\Big)}.$$

For completeness, we also computed balance measures based on the LOCF-imputed data. The resulting balance series are very similar to those obtained under the RF approach. To avoid redundancy and to maintain clarity of presentation, the LOCF-based balance results are therefore relegated to the Appendix A.4.

Business Situation. Figure 28 compares the time series of BS balances computed from the original data and from the RF-imputed dataset. The imputed balance closely follows the overall dynamics of the original series over time. At the same time, the imputed balance is higher than the original balance in most months. This difference is particularly pronounced during the periods 2017–2019 and from mid-2021 to mid-2022.

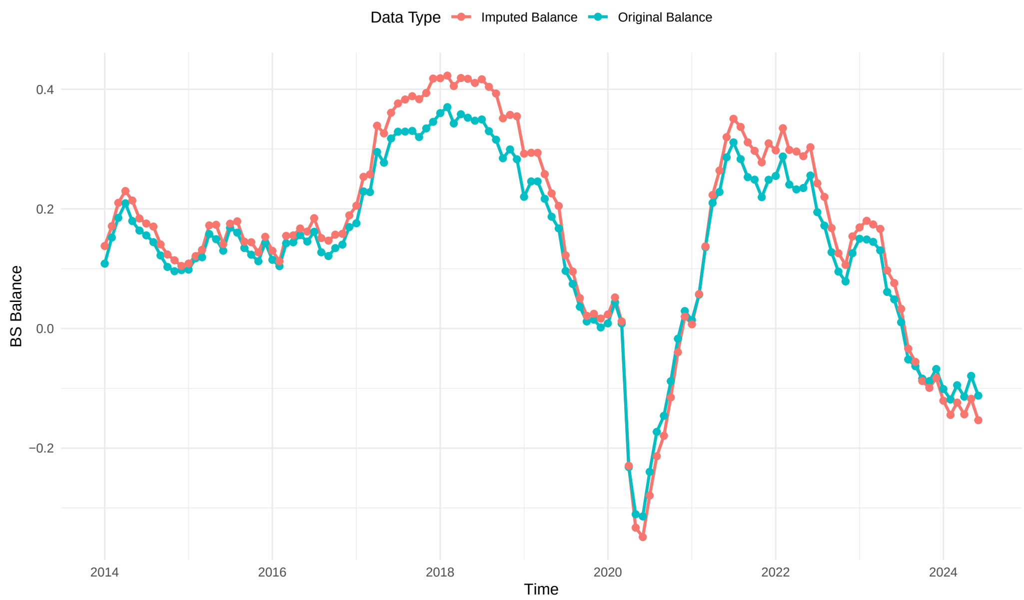

Figure 28: BS balance based on RF imputation.

Business Expectations. Figure 29 reports the comparison of Business Expectations balances computed from the original data and from the RF-imputed dataset. For most of the period, the imputed balance is very close to the original balance. From 2022 onwards, the imputed balance tends to be lower than the original balance.

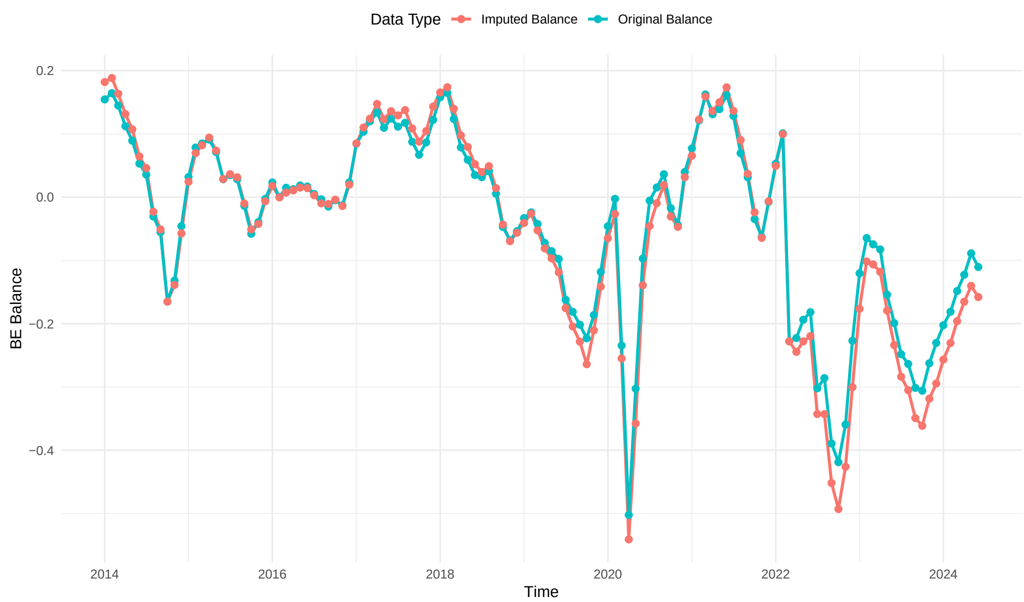

Figure 29: BE balance based on RF imputation.

5.6 Summary

This chapter evaluates LOCF, homogeneous Markov Chain, and Random Forest across multiple point-level performance metrics under simulated missingness scenarios. Results show that the Random Forest approach consistently outperforms the alternatives across all evaluation metrics and gap lengths, while LOCF remains a strong benchmark.
Imputation performance deteriorates for all methods as the length of consecutive missingness increases. Across identical settings and imputation methods, BS consistently exhibits better imputation performance than BE.
Calibration analysis of the Random Forest model shows good agreement between predicted probabilities and observed frequencies for short missing gaps, while calibration quality deteriorates as gap length increases.
Random Forest is used to impute missing values in the original survey data. Balance measures are then calculated separately for the original and imputed datasets. The comparison shows that imputed BS balance is generally higher than the original balance over most of the sample period, whereas for BE the two balance series are relatively close prior to 2022.

6 Conclusion

The descriptive and quantitative analyses provide a comprehensive empirical assessment of unit nonresponse behavior in a long-running firm-level survey. By combining generalized additive models with cluster-robust inference, the analysis accommodates both complex nonlinear relationships and the panel structure of the data, yielding population-average insights that remain robust to within-firm dependence.

The findings demonstrate that unit nonresponse is governed by systematic and highly nonlinear patterns that cannot be captured by simple linear specifications. In particular, calendar time exhibits pronounced non-monotonic dynamics, indicating that the underlying level of nonresponse evolves in distinct phases over the observation period rather than following a single long-run trend. Moreover, survey participation history emerges as a multidimensional construct: participation number and participation length capture distinct aspects of firms' engagement with the survey and display opposing nonlinear associations with unit nonresponse. Treating these measures separately is therefore essential for an adequate characterization of nonresponse behavior.

In addition, the results reveal differences in unit nonresponse across firm characteristics and months. State indicators show that several states belonging to the former East German region exhibit significantly higher odds of unit nonresponse relative to the reference category, while many former West German states display lower nonresponse risks. The majority of industrial sectors show a lower risk of unit nonresponse than the reference group, with Publishing & Printing being a notable exception characterized by a significantly higher nonresponse risk. Month indicators reveal pronounced variation over the year, with August and December exhibiting the highest risks of unit nonresponse relative to January.

Average Business Situation and Business Expectations over the past six months also enter the model in a statistically significant nonlinear manner. However, due to the discrete and unevenly distributed support of these variables, the detailed shapes of their smooth functions are sensitive to the choice of spline basis dimension. As a result, these variables are treated primarily as flexible control components rather than as objects of direct interpretation.

In the imputation analysis, we compare three methods with increasing structure: Last Observation Carried Forward (LOCF) as a benchmark, a homogeneous Markov Chain capturing outcome dynamics, and a Random Forest model allowing for heterogeneity through covariates. Across all gap lengths and evaluation metrics, RF consistently outperforms the alternative approaches, while LOCF remains a strong benchmark. Imputation performance deteriorates for all methods as the length of consecutive missingness increases and BS is imputed more accurately than BE under identical settings for any given method among the three. Applying RF to the original survey data shows that imputed BS balances are generally higher than the original series over most of the period, whereas BE balances remain relatively close prior to 2022.

Limitations should be noted. First, due to substantial missing values in key firm characteristics, such as firm size and survey mode (e.g., paper-based versus online questionnaires), these factors could not be explicitly incorporated into the empirical model. While the inclusion of flexible smooth terms and a rich set of covariates helps account for systematic

heterogeneity in nonresponse behavior, the absence of these variables implies that certain sources of variation, such as differences in organizational capacity or changes related to survey mode, are not directly captured in the analysis.
Second, the smooth effects of average Business Situation and Business Expectations are sensitive to the choice of spline basis dimension due to the discrete and unevenly distributed support of these variables. Although their nonlinear significance is robust, the precise functional forms are not, which limits the interpretability of their detailed shapes.
Third, the Markov Chain imputation approach relies on a homogeneity assumption, under which transition probabilities are assumed to be identical across firms and constant over time. In contrast, the Random Forest imputation explicitly allows for covariate-dependent conditional distributions and therefore accommodates heterogeneity. This difference in modeling flexibility limits the comparability between the Markov Chain approach and the Random Forest. A natural extension of the Markov Chain framework is to relax the homogeneity assumption by allowing transition probabilities to vary across firms, over time, or as functions of observed covariates.
Forth, the simulation-based evaluation of the imputation methods relies on a MAR assumption. However, the quantitative analysis indicates that nonresponse behavior may exhibit MNAR features, which could affect the generalizability of the simulation results. An important direction for future research is therefore to relax the MAR assumption underlying the simulation design.

7 Contributions

Chunyan Jiang was primarily responsible for developing the research plan for each stage of the analysis and carrying out the coding. Specifically, Chunyan wrote the code for the preliminary descriptive analysis, including the descriptive analysis of unit and item nonresponse, and implemented the code for the quantitative analysis and the imputation analysis. In the report, Chunyan wrote the Introduction, Imputation, and Conclusion sections.

Qian Feng participated in the development of the research design and the implementation of the empirical analysis code. Specifically, Qian Feng was primarily responsible for data cleaning, and Qian Feng contributed to parts of the descriptive analysis, quantitative analysis, and imputation analysis. In terms of report writing, Qian Feng was responsible for the Data, Descriptive Analysis, Quantitative Analysis, and Conclusion sections.

8 AI Tools

We hereby confirm that we have written all parts of this work ourselves.
For the preparation of the written report, we adopted suggestions concerning grammar and writing style only after a manual review. In the Simulation section, ChatGPT-5 was employed to assist with visualization and formula translation. Additionally, we used ChatGPT-5 to summarize key ideas for the Conclusion section, which we subsequently revised and independently implemented.
Regarding the technical implementation, specifically within the imputation part, we employed Gemini for debugging and error correction. Gemini was also used to refactor the code structure, such as converting code blocks for different simulation settings into iterative loops. For the visualization of the simulation results, the initial version of the code for the boxplots comparing different methods was generated via Prompting, which we subsequently modified and refined to fit our specific requirements. In addition, ChatGPT-5 Debug was used during the data cleaning, descriptive analysis, and quantitative analysis stages, with particular emphasis on the visualization of model results.

References

Arellano, M. (1987). Computing robust standard errors for within-groups estimators., *Oxford Bulletin of Economics & Statistics* **49**(4).

Breiman, L. (2001). Random forests, *Machine learning* **45**(1): 5–32.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference, *Journal of human resources* **50**(2): 317–372.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties, *Statistical Science* **11**(2): 89–121.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models, *Statistical Science* **1**(3): 297–310.

MacKinnon, J. G., Nielsen, M. Ø. and Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice, *Journal of Econometrics* **232**(2): 272–299.

Norris, J. R. (1998). *Markov chains*, number 2, Cambridge university press.

Sauer, S., Schasching, M. and Wohlrabe, K. (2023). *Handbook of ifo Surveys*, number 100, ifo Beiträge zur Wirtschaftsforschung.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica: journal of the Econometric Society* pp. 817–838.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, *Journal of the Royal Statistical Society: Series B* **73**(1): 3–36.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*, 2 edn, CRC Press.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*, MIT press.

List of Figures

List of Tables

A Appendix

A.1 Data

A.1.1 Item nonresponse

Table 4: Variable Descriptions of Standard Questions (Current Situation)

Variable	Question
Business Situation	1. We characterize our current business situation as good / satisfactory / poor
Inventories	2. We characterize our inventories of unsold manufactured goods as too low / sufficient / too high / warehousing not customary
Orders	3a. We characterize our overall order backlog as comparatively large / sufficient / too low
Foreign Orders	3b. We characterize our order backlog for export as comparatively large / sufficient / too low / we don't export

Table 5: Variable Descriptions of Standard Questions (Review)

Variable	Question
Demand (vs. Past)	4. We compare the demand situation in October to September as improved / not changed / worsened
Orders (vs. Past)	5. We compare our order backlog in October to September as increased / remained roughly the same / decreased
Production (vs. Past)	6. We compare our production activities in October to September as increased / remained roughly the same / decreased / no significant domestic production
Prices (vs. Past)	7. We compare our prices in October to September as risen / not changed / fallen
Employment (vs. Past) (removed)	8. We compare our workforce in October to September as increased / remained roughly the same / decreased

Table 6: Variable Descriptions of Standard Questions (Expectation)

Variable	Question
Production Expectation	9. We expect our production activity in the next 3 months to increase / remain roughly the same / decrease / no significant domestic production
Price Expectation	10. We expect our prices in the next 3 months to rise / remain roughly the same / fall
Export Expectation	11. We expect the scope of our export business to widen / remain roughly the same / decrease / we don't export
Employees Expectation(removed)	12. We expect our workforce in the next 3 months to increase / remain roughly the same / decrease
Business Expectation	13. We expect our business situation in the next 6 months to become more favorable / remain roughly the same / become less favorable

A.2 Descriptive Analysis

A.2.1 Unit nonresponse

Unit Nonresponse Rate by Region

Due to strong collinearity between state-level location and the regional indicator (West/East), the Region (West/East) variable is excluded from the regression, as state-level indicators provide a more fine-grained measure of geographic location.

Figure 30 compares the average unit nonresponse rate between firms located in Western and Eastern Germany. The unit nonresponse rate also varies across regions. Firms located in Western Germany exhibit a relatively lower unit nonresponse rate, at around 8.5%. In contrast, firms in Eastern Germany display a substantially higher nonresponse rate, reaching approximately 17.5%, which is more than double the level observed in the West.

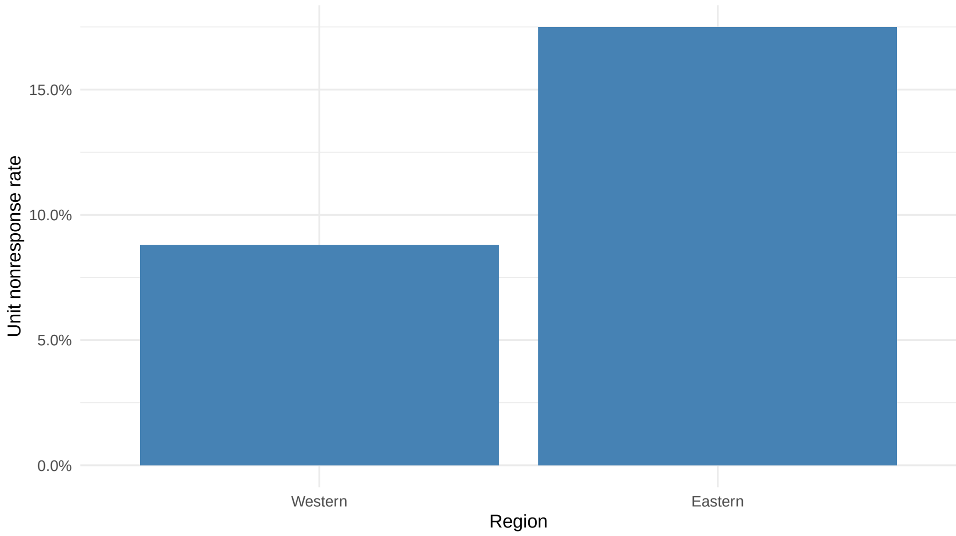

Figure 30: Unit Nonresponse Rate by Region

Unit Nonresponse Rate by Company Size

The dataset covers the period from January 1991 to June 2024. However, Company Size has only been collected since April 2018, which means that about 90 percent of observations in the full dataset have missing values for this variable.

Focusing on the period from April 2018 onward, we find that approximately 22 percent of firms experience changes in company size, while the remaining 78 percent remain stable. If company size is backfilled for these stable firms to their earlier observations, the share of missing values is still around 60 percent.

Alternatively, if the first observed company size after April 2018 is backfilled for all firms, regardless of whether their size changes later, the proportion of missing observations is reduced to about 50 percent but remains substantial. Therefore, Company Size is excluded from the regression analysis.

Figure 31 shows the unit nonresponse rate by company size. The "Solo self-employed" group has the highest rate (above 27%), followed by "Small businesses" (around 21.5%).

Companies with 1–999 and more than 1000 employees exhibit similar rates of about 15–17%.

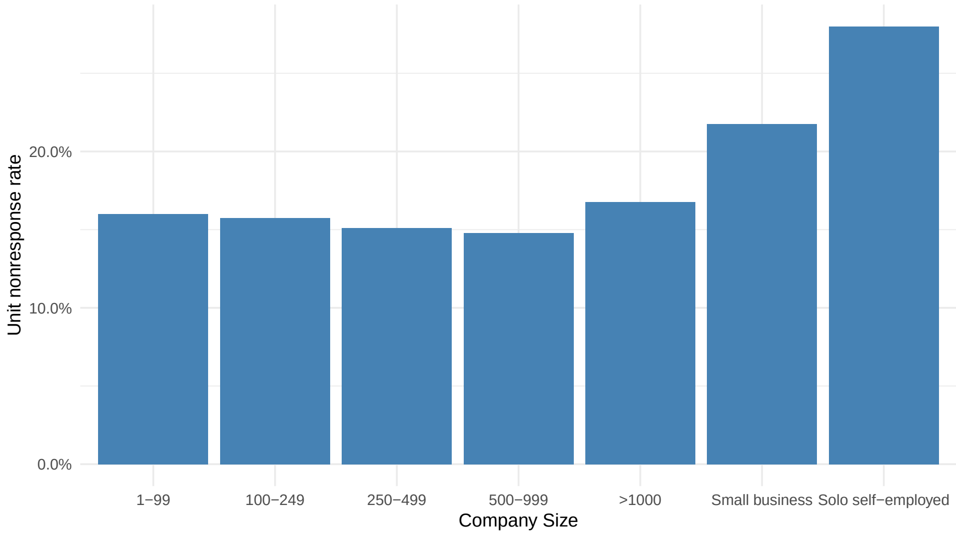

Figure 31: Unit Nonresponse Rate by Company Size

Unit Nonresponse Rate by Online Status

For the majority of unit nonresponse observations, the online status is simultaneously missing. As a result, the online-response indicator cannot be included in the regression analysis.

Figure 32 shows the unit nonresponse rate by online status. The absolute levels for both groups are very low, with missing rates below 0.03%. In most cases, when there is unit nonresponse, the online variable is NA.

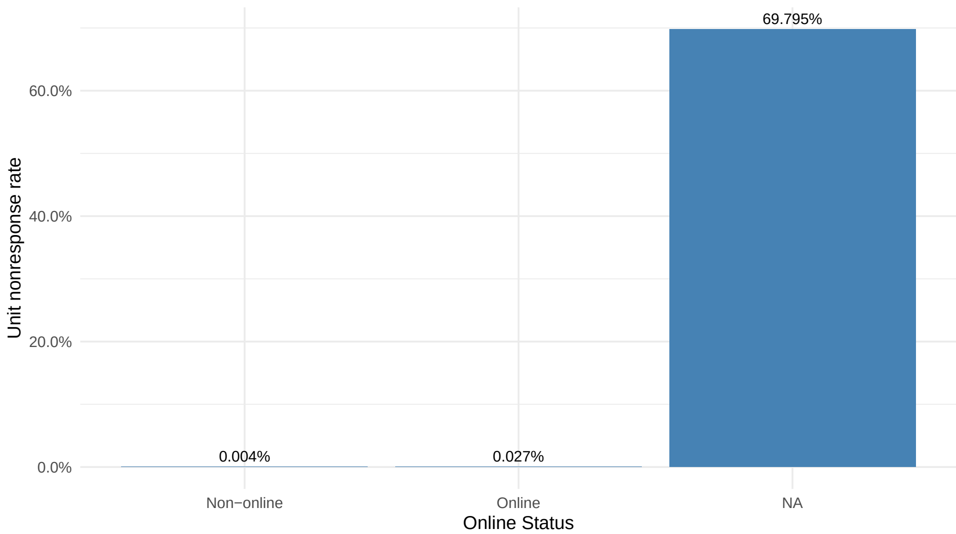

Figure 32: Unit Nonresponse Rate by Online Status

A.2.2 Item nonresponse

Question 8 (Employment (vs. Past)) has been asked since 2018-07. Question 12 (employees Expectation) was a special question before 1997-07. Therefore we exclude these two questions in the analysis. We analyzed item nonresponse of Inventories, Business Situation, Production Expectation, Price Expectation, Orders, Foreign Orders, Export Expectation, Business Expectation, Production (vs. Past), Prices (vs. Past), Orders (vs. Past) and Demand (vs. Past)

Item nonresponse rate by time

Figure 33 shows that overall, the missing rates of most variables exhibit a long-term upward trend. Notably, the missing rate of Foreign Orders surged sharply around 2016.

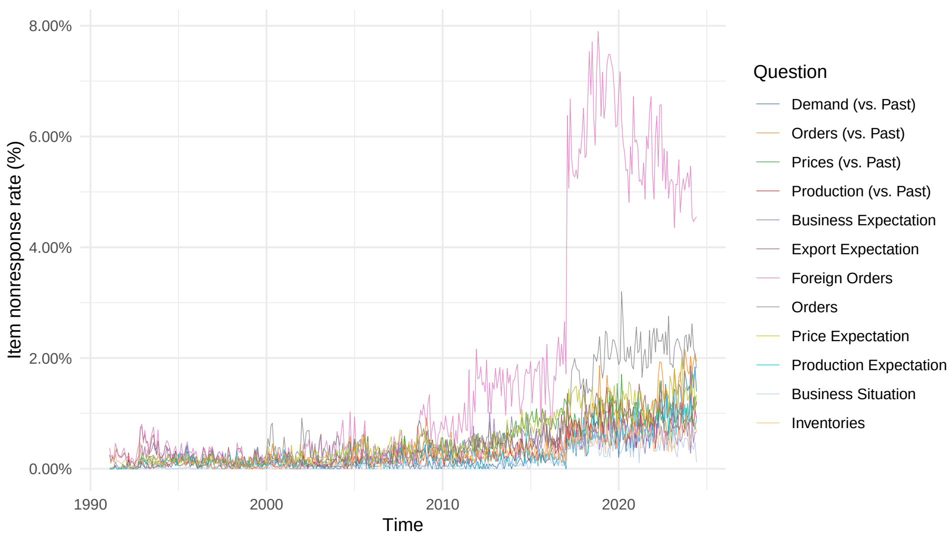

Figure 33: Item nonresponse rate over time

Item nonresponse rate by month

Figure 34 illustrates monthly item nonresponse rates for different survey items over the calendar year. The horizontal axis shows the months (January–December), while the vertical axis reports item-specific nonresponse rates. Each colored line corresponds to a distinct survey item.

Overall, substantial heterogeneity is observed across items. Some items exhibit consistently higher nonresponse rates throughout the year, such as Foreign Orders, whereas others remain at comparatively low levels, including Business Situation. Along the time dimension, month-to-month variation is generally modest, with most items displaying mild fluctuations.

For several items, nonresponse rates increase slightly toward the end of the year, although this pattern is neither uniform across items nor monotonic. This suggests that monthly variation in item nonresponse is largely item-specific rather than driven by a common time-related factor. Business Expectations belongs to the group of items with relatively larger fluctuations, exhibiting notably lower nonresponse rates between June and August.

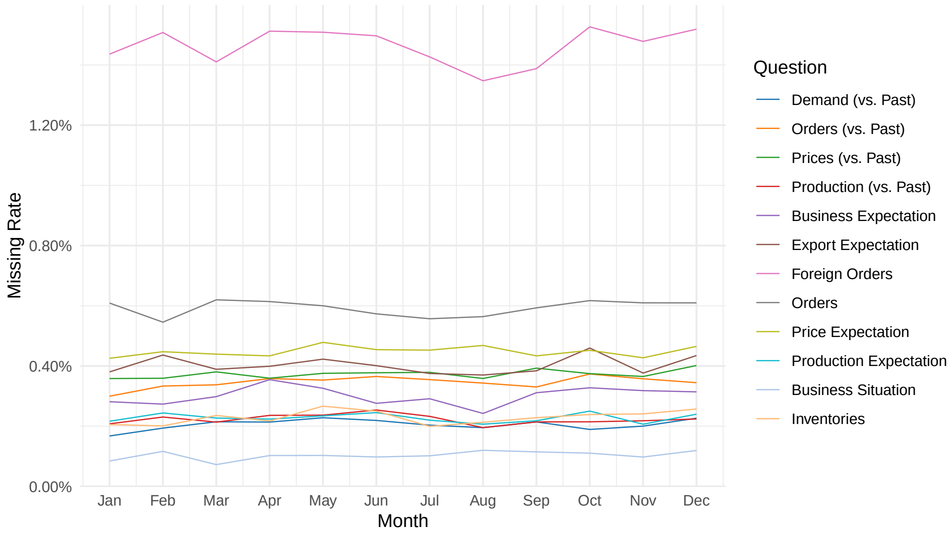

Figure 34: Item nonresponse rate by month

A.3 Quantitative Analysis

A.3.1 Sensitivity of BS and BE smooths to the choice of basis dimension

Figures 35–40 present the estimated smooth functions for average Business Situation (BS) and Business Expectations (BE) under alternative choices of the spline basis dimension k. While the nonlinear effects of both variables remain statistically significant across specifications, the detailed shapes of the estimated smooths vary substantially with the choice of k.
In particular, increasing the basis dimension alters the curvature, local extrema, and apparent turning points of the smooth functions. These changes indicate that such features are not robustly identified from the data and are sensitive to modeling choices rather than being uniquely determined by the underlying empirical relationship.
This sensitivity reflects the fact that BS and BE take values on a discrete and unevenly distributed support, providing limited information to pin down a unique smooth functional form. As a result, although the presence of nonlinearity is robust, the precise shape of the spline functions does not admit a stable substantive interpretation.
Accordingly, in the main specification, BS and BE are treated as flexible control variables. Their smooth effects are included to account for nonlinear economic conditions, while detailed interpretations of their functional forms are deliberately avoided. The figures reported here are provided for completeness and to illustrate the sensitivity of the estimated smooths to the choice of k.

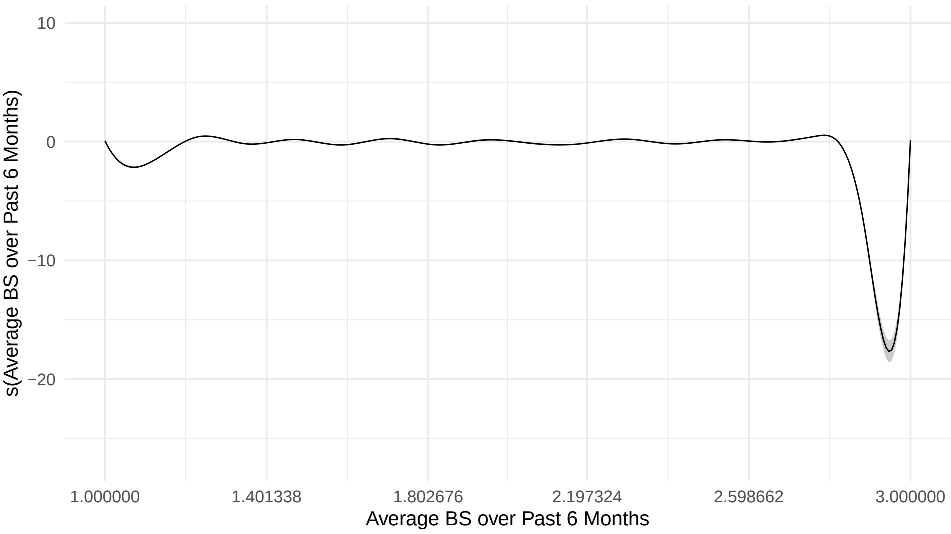

Figure 35: Nonlinear effect of average Business Situation over the past six months on unit nonresponse ($k_{\text{BS}} = k_{\text{BE}} = 20$)

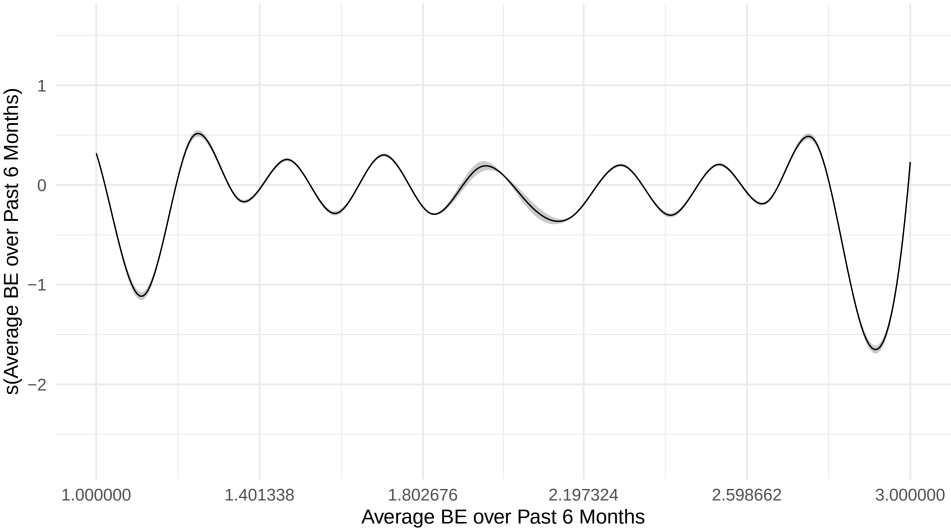

Figure 36: Nonlinear effect of average Business Expectation over past six months on unit nonresponse ($k_{\mathrm{BS}} = k_{\mathrm{BE}} = 20$)

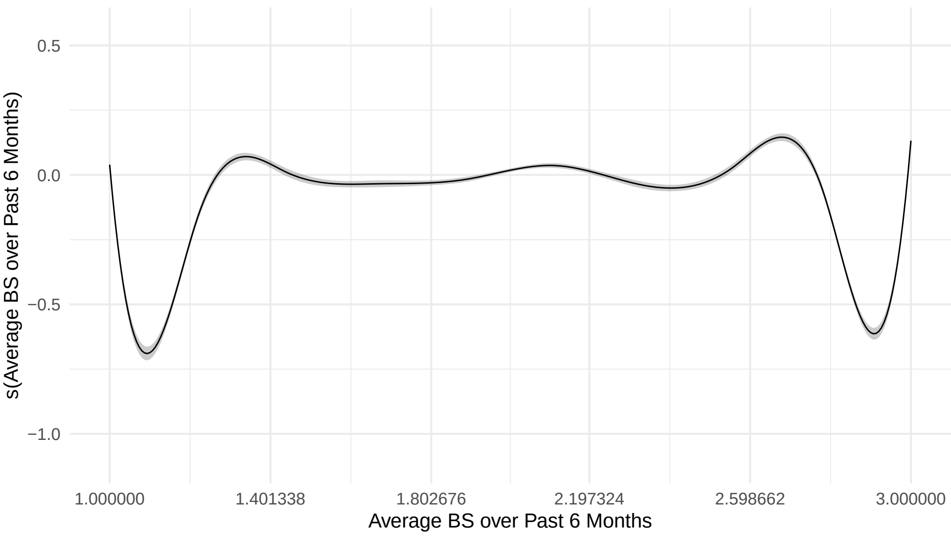

Figure 37: Nonlinear effect of average Business Situation over the past six months on unit nonresponse ($k_{\mathrm{BS}} = k_{\mathrm{BE}} = 12$)

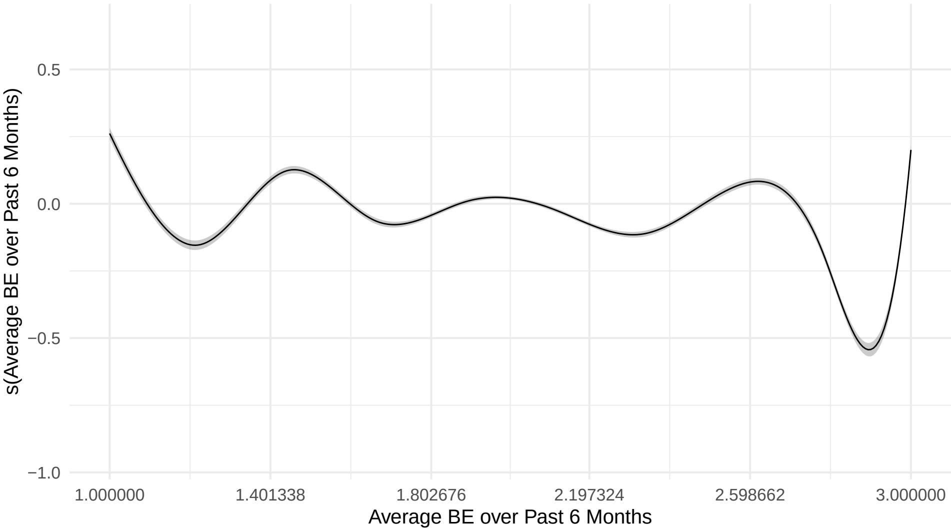

Figure 38: Nonlinear effect of average Business Expectation over past six months on unit nonresponse ($k_{\mathrm{BS}} = k_{\mathrm{BE}} = 12$)

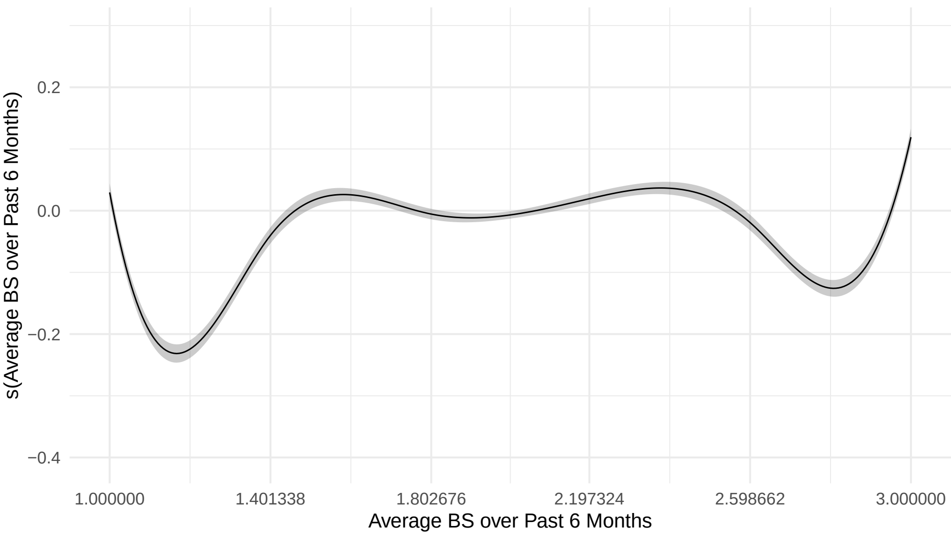

Figure 39: Nonlinear effect of average Business Situation over the past six months on unit nonresponse ($k_{\mathrm{BS}} = k_{\mathrm{BE}} = 8$)

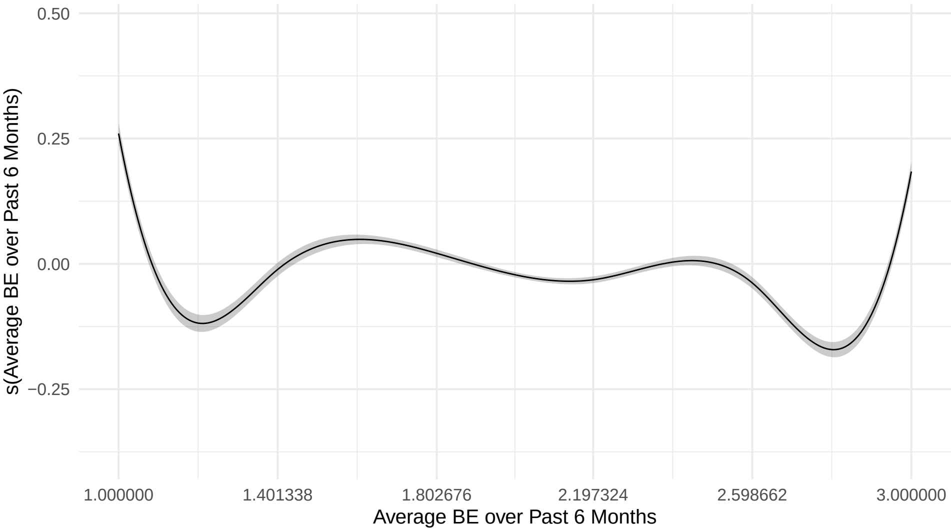

Figure 40: Nonlinear effect of average Business Expectation over past six months on unit nonresponse ($k_{\mathrm{BS}} = k_{\mathrm{BE}} = 8$)

A.4 LOCF Balance

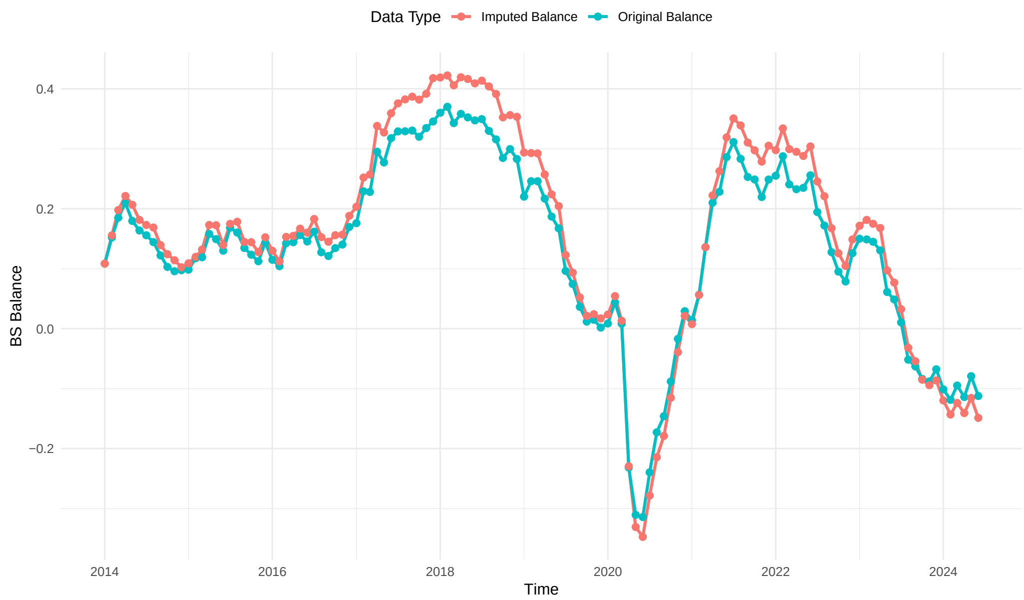

Figure 41: BS balance based on LOCF imputation.

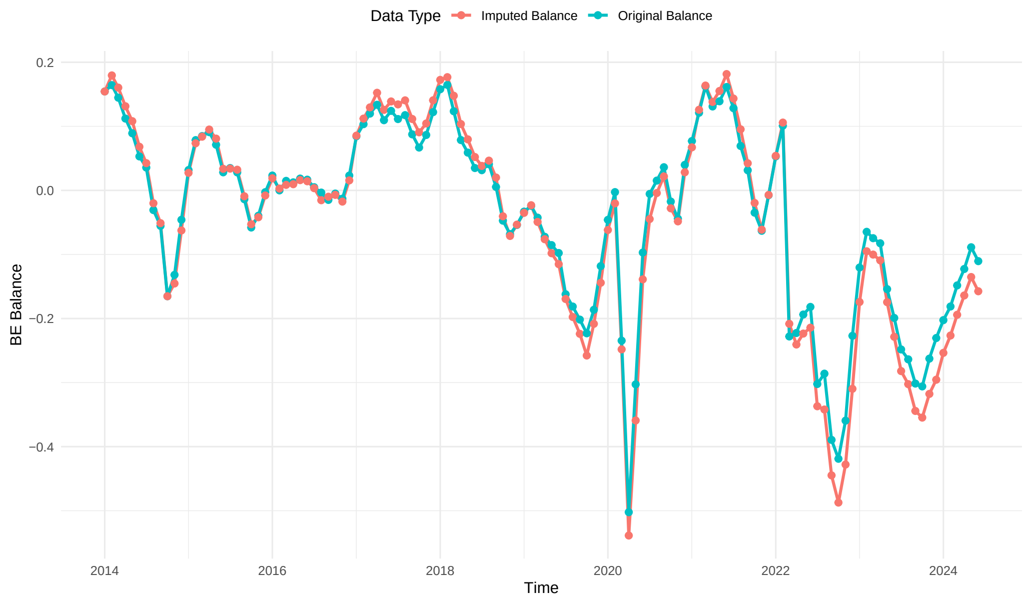

Figure 42: BE balance based on LOCF imputation.

B Electronic appendix

This repository provides the program code, test data, and outputs generated from the test data: `https://github.com/Chunyan-Jiang/nonres_analysis`.
A readme file in the repository documents the folder structure and provides instructions for reproducing the results.