

Do employees foresee the looming iceberg ahead? Empirical evidence on the predictability of corporate default events.

Jochen Lawrenz¹, Michael Strauß², Maximilian Strauß³, and Janette Walde⁴

¹*Department of Banking and Finance, University of Innsbruck, Austria*

²*Department of Banking and Finance, University of Innsbruck, Austria*

³*Faculty of Health Sciences, University of Copenhagen, Denmark*

⁴*Department of Statistics, University of Innsbruck, Austria*

January 2026

Abstract

We analyze the potential of crowd-sourced employee ratings for company default prediction. Beyond controlling for well-known financial predictors, we construct a peer-group-matched sample to address potential endogeneity and group-level effects. We find strong incremental predictive ability from both discriminatory and forecasting models, both in-sample and out-of-sample. In particular, we find that employee ratings are predictive up to twelve years before default with a minimum increase in balanced accuracy of three percentage points. In the arguably stricter setting of forecasting, we observe comparable effects up to three years forecast horizon. Our results provide strong evidence that employees possess relevant incremental information on the risk of company default that appears to go beyond the information encoded in classical financial and macroeconomic variables. (120 words)

JEL classification – G33, M50

Keywords – Bankruptcy Prediction, Credit Risk, Employee Ratings, Employee Satisfaction

1 Introduction

An increasing number of recent contributions analyze the information content of online employee ratings and their forecasting ability. A recent and growing literature relates employee ratings to several relevant company characteristics such as operating performance (Green et al., 2019; Huang et al., 2020; Symitsi et al., 2021), Tobin’s Q (Huang et al., 2015; Symitsi et al., 2018), stock returns and earnings announcement surprises (Sheng, 2025; Green et al., 2019; Chemmanur et al., 2019), post-IPO returns (Farhadi and Nanda, 2021), tax planning (Li, 2022), corporate misconduct (Campbell and Shang, 2022),

borrowing costs (Chi and Chen, 2021), CEO dismissal (Wang et al., 2023), and culture (Guiso et al., 2015). Li and Pinto (2025) and Gornall et al. (2025) investigate the extent to which employee satisfaction is affected by major ownership shifts, such as IPOs or LBOs. However, it is surprising that no study has yet analyzed employee ratings as a predictor of the arguably most existential risk to a company: business disruption and failure. This paper aims to fill this gap and provides evidence that employee ratings provide forward-looking incremental information about company default risk beyond financial and macroeconomic variables, which are routinely used in credit scoring methods.

The rationale for why employee ratings may be predictive is similar to that in previous contributions, which argue that employees have in-depth knowledge of their employers and possess valuable, often intangible inside information. Campbell and Shang (2022), for example, attribute their finding to inside information from employees about organizational cultures. Similarly, Farhadi and Nanda (2021) find that employees' ratings are informative about post-IPO excess returns due to employees' inside information about companies' future prospects, and Green et al. (2019) show that employees' rating changes more generally predict stock returns even after controlling for financial analysts' recommendations, suggesting that employees possess information beyond the one available to an informed specialist audience. This perspective is supported by findings from Wang et al. (2023), who show that employee approval of CEO leadership (as reflected in Glassdoor rating data) predicts CEO dismissal. Their study contributes to stakeholder theory and corporate governance literature by demonstrating how employee views provide valuable insights beyond financial and analyst assessments. Their result aligns with prior findings in Guiso et al. (2015), who use 679 companies taken from a dataset from the Great Place to Work Institute (GPTWI) that measures the relationship between culture and outcomes. Among others, they find that the high levels of perceived integrity (management) evaluated by employees are positively correlated with higher profitability. Finally, Naumzik et al. (2022) analyze online customer ratings – instead of employee ratings – of US restaurants as predictor variables. They find that their empirical model can predict business failures several months in advance. Building upon this evidence, we therefore expect that employees' value-relevant inside information is also predictive of companies' default or credit risk. We collect employee rating data from the online platform KUNUNU, which is the leading employee-based company rating platform in the German-speaking region. Our sample comprises approximately 25,000 employee ratings on 380 companies, located mainly in Germany, Austria, and Switzerland, covering the years 2007 to 2018. Employee ratings in KUNUNU can be reported in 13 categories that focus on different employee evaluation criteria, such as management behavior, company image, internal communication, and Employee Benefits. Based on results from the Principal Component Analysis,¹ we aggregate employee rating information into different variables: *Employee Rating*, *Latent Employee Rating*, and *Employee Benefit Score*. We include them in logistic regressions and test if they are significant explanatory variables for the default event indicator after controlling for financial variables and macroeconomic variables.

Testing employee ratings as explanatory variables faces several econometric challenges. First, it is reasonable to expect a significant endogeneity problem as the employee's assessment of company

¹ Using PCA to determine aggregate rating variables is in line with e.g., Farhadi and Nanda (2021).

characteristics is not independent of their prior experiences.² Second, as with any Likert-type scale assessment, substantial group-level effects can occur biasing inference (e.g., Beal and Dawson, 2007). Evidence, e.g., in Huang et al. (2015) shows a substantial variability in average rating scores across industries.³ It is expected that employees base their assessment on a relative comparison to relevant peer groups, which are substantially smaller than the industry, so it is likely to find group-level effects in even smaller subsamples. Third, in particular, for the identification of default risk, the problem of omitted variables is a relevant concern that may render employee ratings spurious.⁴

To address the three issues, our identification strategy starts with a particular sample construction, which we label peer-group matching. We intersect KUNUNU employee rating data with proprietary access to data from (legally imposed) restructuring reports on financially distressed companies. These mandatory reports, written by auditor experts, provide extensive information on the underlying reasons for financial distress and, most importantly for our purpose, provide detailed information on relevant peer companies that are comparable with respect to their business model in terms of industry, product and service portfolio, geographical footprint, and value chain depth. For our purposes, this allows to identify companies that are direct peers of any given distressed company. This approach shares similarities with a matched sample approach, such as propensity score matching (PSM), which is employed in Chemmanur et al. (2019), Lee et al. (2021), or Huang et al. (2015). However, PSM can only be done on the basis of measurable company variables. Instead, our peer-group approach can match based on difficult (or impossible) to measure aspects while at the same time retaining the possibility of using measurable variables as controls. Furthermore, our peer-group approach addresses the group-level effects problem by constructing counterfactuals from a subsample of companies with a similar rating distribution.

After constructing the sample, we estimate a logistic regression model using variables known to be potential predictors of default risk, and use it as the benchmark against which we evaluate the predictive power of employee rating information. We label the benchmark case as the traditional model $\mathcal{M}^{\text{Trad}}$. We then augment the regression model with our variables of interest, *Employee Rating*, *Latent Employee Rating*, and *Employee Benefit Score*. In either case, we employ several control variables as well as sector-fixed effects and cluster-robust standard errors. We label the suggested model as the employee model, abbreviated by $\mathcal{M}^{\text{Emp}}$.

Our empirical analysis proceeds in three key steps: First, we assess the discriminatory power of employee ratings using a logistic regression model, finding that companies with lower employee ratings exhibit a significantly higher likelihood of default. Second, we evaluate the predictive performance of this classification model, showing that predictive signals emerge up to twelve years before insolvency. The

² It is actually interesting to observe that while the majority of contributions take employee ratings as independent variables, studies such as Huang et al. (2015), Jing et al. (2019), Lee et al. (2021), or Dube and Zhu (2021) analyze the reverse causality and use employee ratings as dependent variables. Attempts to control for endogeneity include using two-stage least squares (2SLS) and / or GMM regressions such as in, e.g., Huang et al. (2015), Chi and Chen (2021), or Li (2022).

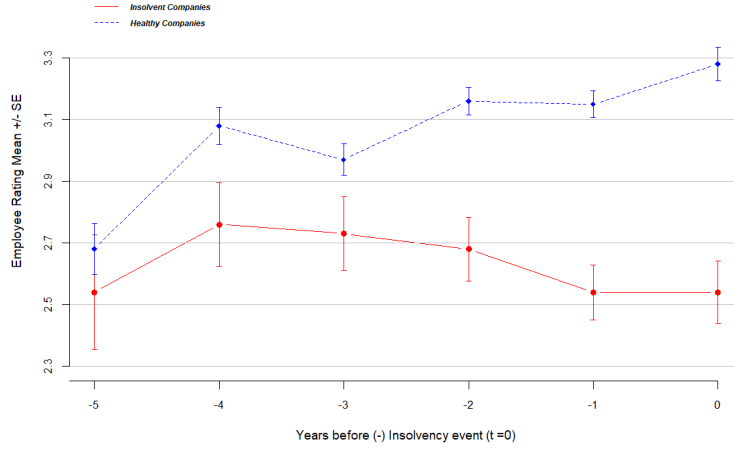
³ See Huang et al. (2015), Fig. 1.

⁴ See the large body of research aiming to find other variables for default prediction such as e.g., Altman et al. (2016) or Platt and Platt (1990).

inclusion of employee ratings substantially improves model fit, increasing Balanced Accuracy by at least three percentage points. Thirdly, we go beyond classification tests and implement a forecasting regression model, where current employee rating information is assumed to be a predictor for default events happening in the subsequent periods. We find strong support for the claim that employee information retains predictive power even in this arguably most challenging test setup. Thus, the ability of employee ratings to anticipate financial distress is reinforced.⁵

Figure 1 provides a tentative visualization of the primary driver of our results, showing the mean employee rating values for distressed companies (solid red line) and counterfactual companies (dashed blue line) before the default event (t_0). The graph indicates that up to four years before default, we can

Figure 1
Yearly Mean Employee Rating before Insolvency



This figure shows the yearly average *Employee Rating* variable for insolvent (red) and healthy (blue) companies from five years before (t_{-5}) the insolvency event until the insolvency event (t_0). A healthy company does not go insolvent during the sample period. Every insolvent company is matched to one or more healthy companies. The bars represent ± 1 standard error (*SE*) intervals. The *Employee Rating* variable is the mean of all employee rating variables as defined by KUNUNU. The horizontal axis represents event time, i.e. years before the insolvency event. The sample period ranges from January 2007 to December 2018.

observe statistically different levels for *Employee Rating* which tend to increase as default approaches. Thus, employee rating information is not only a substantial predictor of company default; we even find that employee ratings can forecast financial distress over a comparatively long time horizon. The logistic regression model estimates a coefficient of -0.47 , which translates into a 38% increase in the odds of being classified as insolvent for a unit change in the *Employee Rating* variable. The inclusion of employee rating information improves the model fit from 0.19 to 0.27, representing a significant improvement.⁶ By evaluating the prediction accuracy, we find that employee ratings increase Balanced Accuracy, and in particular, Sensitivity. We find that Balanced Accuracy is at least three and Sensitivity up to seventeen

⁵ See Table 7 for detailed results. In contrast to the majority of previous contributions, we do not only focus on the in-sample performance of employee rating information, but conduct extensive out-of-sample tests to assess the predictive ability. We find strong statistical and economic significance for the relevance of the employee rating variables in-sample as well as out-of-sample.

⁶ See Table 3. For comparison, the recent contribution of Campbell and Shang (2022) report an increase in R^2 – also from logistic regression – from 27.3% to only 29.2%.

percentage points better. Finally, we demonstrate that forecasting the insolvency event is possible up to three years ahead with an out-of-sample AUC of 0.69 (see Table 7). Again, these increases are both statistically and economically large.

We perform several robustness checks to corroborate our findings. First, explanatory power is validated through different considerations of the rating variables (for instance, including a latent variable computed from the rating variables). Second, we perform different in-sample robustness tests. Therefore, we evaluate the effects of imputing missing data, excluding the one-year lag for *FIN* variables, and limiting the analysis to one sector. We also employ two validation strategies to support our forecasting results. We run Monte Carlo Simulations for all forecasting models to capture the observed variance in the forecasting metrics across different sample splits. In addition, we perform out-of-sample tests. Our conclusions remain stable, as none of our main results is substantially affected.

Our paper contributes to different streams of the literature. First, we add additional evidence to the growing literature on the relevance of employee rating information. Investigating employee ratings has thereby attracted significant interest in many different management areas: Within a financial context, Green et al. (2019) and Sheng (2025) focus on the possibility to predict stock returns from employee information, while recent contributions by Li and Pinto (2025) and Gornall et al. (2025) take the opposite perspective and use employee ratings as dependent variable and investigate to what extent they are affected by major corporate financial events, namely after IPOs or LBOs. Besides finance, other disciplines such as Organization studies (Abdulsalam et al., 2025), strategic management (Wang et al., 2023), or marketing (Naumzik et al., 2022) are interested in employee rating information. We distinguish ourselves as the first study to relate employee ratings to a company’s default risk. To address endogeneity issues, omitted variables, and group-level effects, we propose a peer-group comparison approach that uses selected information from legally mandated restructuring reports to identify counterfactual companies. In line with previous studies, our results provide strong evidence that employees possess relevant inside information about companies’ prospects, which appears to go beyond the information encoded in classical financial and macroeconomic variables

Although our focus is on company default risk, complementary literature has emerged that examines the credit scoring process for *retail* customers against the backdrop of the growing availability of alternative data, particularly from online shopping behavior and peer-to-peer lending platforms. In this regard, our contribution relates to, for example, Berg et al. (2020), who consider the information content of a simple ‘digital footprint’ for retail customers’ default risk discrimination.⁷ They show that the digital footprint in isolation provides the same discriminatory power as a credit bureau scoring, and improves accuracy when combined with the classical credit score. Similarly, Jagtiani and Lemieux (2019) analyze LendingClub credit scores and compare them with traditional FICO scores.⁸ They summarize their

⁷ As ‘digital footprint’, Berg et al. (2020) define a number of generic pieces of information such as the operating system, device, e-mail host, etc. which a customer routinely leaves when visiting an online store.

⁸ LendingClub is one of the leading peer-to-peer online lending platforms. FICO scores are among the most widespread credit assessment tool which are based on traditional information such as payment behavior,

findings by stating that “the rating grades assigned by LendingClub are more powerful in predicting a borrower’s default probability than a set of FICO scores, other traditional risk variables, and economic factors combined”.⁹ Jagtiani and Lemieux (2019) also show that LendingClub scores are only weakly correlated with FICO scores; a result which appears also in Gambacorta et al. (2023) whose results support the view that alternative data captures different aspects than traditional variables. Furthermore, Gambacorta et al. (2023) show that these alternative data appear to have the ability to replace and crowd out the usage of collateral. Our contribution is also related to the literature on online customer ratings, such as in e.g. Naumzik et al. (2022), who show that financial indicators (e.g., revenue, profit) often lag behind customer sentiment, making ratings a more timely early warning signal for business distress. Other contributions in this direction include Li et al. (2023), or Tao and Zhou (2023).

Finally, we contribute to the classical credit risk literature which starts with seminal contributions such as Ohlson (1980) and others, who tried to identify and introduce various non-financial variables in credit risk models, such as management turnover, delayed accounts, risk compliance signs (Keasey and Watson, 1987), industry-relative ratios (Platt and Platt, 1990), the evaluation of management capabilities (Lussier, 1995), payment behavior (Kallberg and Udell, 2003), macroeconomic variables (Bonfim, 2009) or a combination of different factors related to the type of company, sector, size, age, reporting, compliance, and operational risk (Altman et al., 2016).

The remainder of the paper is organized as follows. Section 2 proceeds with the description of our sample, the related literature, and our modeling methodology. In Section 3, we illustrate the performance of our models that include financial and / or employee variables. We validate the robustness of our findings in Section 4. The paper concludes in Section 5.

2 Data and methodology

2.1 Peer-group matching approach

The key concerns in the empirical identification strategy stem from the special characteristics of the employee ratings data: First, it is clear that employees’ assessments of a company are subject to endogeneity. An imminent threat of financial distress is likely to negatively affect employees’ perceptions of the company. Vice versa, it is easily conceivable that employees’ negative mood can deteriorate operating performance and cause financial distress. The a priori unclear direction of causality is reflected in the literature’s ambiguity regarding whether to treat employee ratings as dependent or independent variables. Attempts to address the issue include using two-stage least squares estimation or propensity score matching.

A second concern arises from the fact that the ratings are a summary of various hard-to-define aspects measured on a Likert-like scale. Ratings of, e.g., a company’s work atmosphere, company image, or management behavior are obviously strongly subjective, not only in the sense of differences in personal

accounts owned and credit structure of retail customers.

⁹ Jagtiani and Lemieux (2019), p. 1023.

assessments of these aspects, but also in the absence of an objective anchor. Since an objective reference point is missing, employees most likely base their individual assessment on a relative judgment against the background of her previous work experience and peer group opinions. Thus, employee ratings show significant level effects across different subsamples, which is e.g., accompanied in prior studies such as Huang et al. (2015), who find substantial variation in mean ratings across industries.¹⁰ However, considering a whole industry as the relevant benchmark appears still too coarse, as employees' personal experiences are limited to a much smaller sample. Ideally, such a sample may be constructed from the set of companies that are former or prospective employers, as they are active in the same region and similar product lines.

Guided by these concerns, our analysis is based on a unique data sample drawn from legally mandated restructuring reports. According to German legal requirements,¹¹ financial institutions must assign a recognized auditor to provide an extensive, external restructuring report for financially distressed companies that are not yet in default.¹² The structure, content, and method of these reports follow strict guidelines issued by the Institute of Public Auditors in Germany and represent the leading standard for restructuring concepts in German-speaking countries.¹³ They provide detailed information not only on the financial situation and the causes of distress of the individual company, but also about its market environment, its competitors, its product and service portfolio, its value chain, legal information, as well as the planned restructuring strategy and a business plan.

We have access to a proprietary database of one of the Big-4 auditors which initially includes 1,100 reports from 2004 to 2018, of which 230 have sufficient information for our purpose.¹⁴ Due to the confidential nature of the data, we had to ask for individual approval for data usage, to which 87 companies eventually gave their permission.

The unique opportunity of this data set is that it allows us to construct a sample in which the set of defaulted companies can be matched to a set of counterfactuals that are identified as their peers based

¹⁰ Huang et al. (2015) analyze 102,888 individual ratings from Glassdoor for 993 U.S. companies between 2008-12 classified in the 49 Fama-French industries and find “[...] *industry averages ranging from 3.51 for food products to 2.65 for personal services.*”

¹¹ Paragraph 18 Abs. 1 Kreditwesengesetz (KWG) obligates companies to make their financial situation transparent to lenders. The jurisdiction requires a restructuring plan that shows that a company is in financial trouble. Otherwise lenders are held liable due to transactions (loans) contrary to public policy (see for instance BGH v. 9.7.1953 — IV ZR 242/52, BGH v. 26.1.2017 — IX ZR 285/14, BGH v. 12.5.2016 — IX ZR 65/14).

¹² A financial institution usually recommends at least three advisers to the distressed companies, where the company eventually determines one to develop the restructuring concept. Events that trigger the assignment of a report includes e.g., covenant breaches, delayed annual report provision, additional financing, indications of insolvency such as illiquidity.

¹³ The standard is known under ‘IDW Standard 6’. IDW being the acronym of Institut der Wirtschaftsprüfer (Institute of Public Auditors).

¹⁴ The chosen 230 projects represented a full restructuring concept including all the detailed information mention above. The remaining 870 reports offered only fractions of the required information, e.g., a financial quick analysis of a selected company in distress

on different company-level criteria. In this sense, we label our approach peer group matching.¹⁵ More specifically, we use information from the restructuring reports to identify the main competitors based on four criteria: industry, product and service portfolio, geographical footprint, and depth of the value chain. First, we identify potential matches for the company in distress by selecting the same industry (e.g., automotive), then compare the product and service portfolios of companies in each industry to group them into sub-industries where applicable (e.g., manufacturing of car seats). Third, we create groups based on either the geographical footprint (e.g., the number of company locations per country/continent) and/or the depth of the value chain (e.g., whether the company has an in-house research and development department or uses contract manufacturers). Based on this peer-group approach, we identify 293 companies as close matches for the 87 distressed companies. Thus, our sample consists of $n = 380$ entities. Based on the peer-group matching approach, a comparison of variables makes little sense, since it is not the goal to construct a sample that maximizes the similarity of selected ratios. However, we document comparative statistics for selected key finance and control variables before and after peer-group matching, as shown in Panel B of Table 1. Interestingly, we observe a lower variability (standard deviation) after the peer group matching for three variables *Debt to Equity*, *Sales to Assets*, and *Average Assets*. Although not explicitly intended by the matching process, the decline in variability supports the validity of the matching procedure, as it indicates a more homogeneous sample among matched entities.

2.2 Data sources

We use data from different sources. First, we complement (and cross-validate) the data from the restructuring reports, with company-level data from S&P’s Capital IQ, the Markus database of Bureau van Dijk,¹⁶ and annual reports at the national register courts. Second, macroeconomic data are obtained via various publicly available sources. Consumer price index,¹⁷ insolvency data,¹⁸ business climate index,¹⁹ unemployment data,²⁰ and EURIBOR interest rates.²¹ Third, we collect employee rating data from KUNUNU, the largest database of employee-based company ratings in the German-speaking area.²² Via web-scraping, we collected 24,679 individual employee ratings within the period from 01/2007 to 12/2018 for all 380 companies in our sample. In summary, our data-set consists of four types of variables: employee ratings, finance, controls, and macro.

¹⁵ Note, that contrary to the procedure that is usually termed as matched-sample approach in the literature, our matching process does not consist in using statistical methods to maximize some (metric) similarity score. We rather define similarity in the sense of being member of the same peer group.

¹⁶ We obtained permission to use the database for research purposes from the audit company.

¹⁷ Federal Statistical Office Germany (Genesis Database)

¹⁸ Statista (Creditreform database)

¹⁹ ifo Institute

²⁰ Federal Employment Agency

²¹ European Central Bank (Statistical Data Warehouse)

²² As of June 2021, they cover approximately 978,000 companies.

2.3 Employee rating data

The KUNUNU system allows for feedback in 14 categories, such as *Management Behavior*, *Work Atmosphere*, and *Company Image*. Table A.2 in the Appendix provides a complete list and an explanation of the variables. Employees can comment on each aspect and assign a 5-point Likert-like rating to 13 of the criteria. The last variable is defined by the number of benefits (*Employee Benefit Score*). We refrain from conducting textual analysis and rely solely on ratings.²³ We aggregate all individual assessments into a particular category by forming monthly or yearly averages.²⁴ Reflecting the general trend toward platform-based business models, we find a strong time trend in the number of ratings during our sample period. Although in 2007, 411 employee ratings were assigned (no monthly or yearly averages - real-time data), that number increased 13 times to 5,218 in 2017.

The correlation plot (see Figure S.3 in the Supplementary Material) indicates a high correlation among the employee rating variables (and between the macro variables and to a lesser extent between the finance variables). In line with the previous literature, for example Farhadi and Nanda (2021), we therefore construct aggregate variables. We aim to better understand the data and perform a principal component analysis (PCA) of the variables, which yields a couple of interesting findings. First, the types of variables capture different dimensions of the data, as shown in Table 2. The 14 rating variables are highly loaded on the first principal component. Only one criterion – namely the *Employee Benefit Score* – appears somewhat distinct from the other 13 categories within the first principal component. However, macro variables mainly explain the second dimension, while financial and control variables explain the third, fourth, and fifth dimensions.

Guided by these results, we draw the following conclusion for variable selection: Regarding the employee rating, we define three variables for subsequent analysis. First, we calculate the mean of the first 13 variables, which serve as our primary explanatory variable, labeled *Employee Rating* (a 5-point Likert-like scale). The second variable is the *Employee Benefit Score*, defined as the absolute number of benefits. Finally, we define a latent variable, *Latent Employee Rating*, based solely on all available employee rating variables. We test the explanatory power of all three variables *Employee Rating*, *Latent Employee Rating*, and *Employee Benefit Score* later within our models. We choose the first principal component, which captures 70% of the total variance of the employee rating variables (the second component explains only 5.1%). We use the variables to defer the full explanatory power of our newly introduced variables later in Section 3.

We report descriptive statistics for *Employee Rating*, *Employee Benefit Score*, and the other *EMP* variables in Table 1. Across the entire sample, the mean rating is 3 (i.e., at the center of the rating scale). However, we observe substantial variation across time and in the cross-section. As illustrated in

²³ See e.g., Campbell and Shang (2022) or Symitsi et al. (2021) for textual analysis of employee assessments.

²⁴ We calculate yearly averages for the employee rating data due to two reasons: First, we match employee rating data with financial data which is only available at an annual frequency, and second, we average the larger number of ratings of large companies. Usually, bigger companies have more employees, which on average produce more ratings than smaller companies. For specific descriptive analyzes, we will also use monthly averages.

Table 1
Descriptive Statistics

Panel A: Univariate Summary Statistics					
	Healthy		Insolvent		
	Mean	SD	Mean	SD	
Employee Rating					
Employee Rating ^(a)	3.10	0.873	2.69	0.963	
Employee Benefit Score ^(b)	5.52	3.329	3.83	2.648	
Older Colleagues	3.42	1.042	3.13	1.040	
Management Behaviour	2.88	1.077	2.42	1.191	
Career Development	2.79	1.081	2.36	1.124	
Challenging Work	3.36	0.969	3.02	1.080	
Company Image	3.13	1.077	2.58	1.168	
Environmental F.	3.18	1.034	2.73	1.145	
Gender Equality	3.24	1.022	2.95	1.132	
Internal Communication	2.76	1.034	2.36	1.132	
Work Atmosphere	3.01	1.085	2.59	1.179	
Office/Work Env.	3.23	1.012	2.87	1.077	
Overall Compensation	3.06	1.011	2.56	1.057	
Teamwork	3.58	0.884	3.46	0.951	
Work Life Balance	2.99	1.043	2.55	1.120	
Finance					
EBITDA to Assets	0.10	0.117	0.03	0.162	
EBIT to Assets	0.06	0.116	-0.03	0.189	
EBIT to Sales	0.04	0.126	-0.10	0.669	
Net Income to Assets	0.03	0.108	-0.15	0.837	
Sales to Assets	1.61	1.249	1.40	1.032	
Working Capital to A.	0.23	0.189	0.21	0.203	
Debt to Assets	0.64	0.274	0.92	0.792	
Equity to Assets	0.36	0.272	0.05	0.786	
Debt to Equity	3.24	25.181	2.73	77.568	
Cash to Assets	0.11	0.1264	0.06	0.0834	
Stock to Working Capital	0.94	13.642	1.28	6.677	
Creditors to Liabilities	0.14	0.128	0.13	0.134	
Debtors to Assets	0.14	0.121	0.12	0.121	
Firm Characteristics					
Age	80.61	56.085	64.25	53.568	
Average Assets	206.55	232.995	140.85	193.083	
Macro					
Euribor	0.91	1.36	0.91	1.36	
CPI Change p.a.	1.4	0.82	1.4	0.82	
IFO GER	12.15	9.99	12.15	9.99	
IFO EUR	-3.54	21.56	-3.54	21.56	
Unemployment Rate	7.17	1.25	7.17	1.25	
Insolvency Rate Change	-0.05	0.05	-0.05	0.05	
Panel B: Descriptive Statistics on the Matching Procedure ^(c)					
Variables	Mean		SD		Test
	Before	After	Before	After	Wilcoxon p-value
Net Income to Assets	0.10	0.10	0.14	0.12	0.80
Debt to Equity	4.13	3.84	48.08	24.68	0.01**
Working Capital to Assets	0.24	0.23	0.18	0.14	0.20
Sales to Assets	1.59	1.58	1.17	0.83	0.08*
Age	77.27	77.27	54.62	40.07	0.13
Average Assets	186.03	180.91	219.91	143.91	0.01**

The table illustrates the mean, standard deviation ('SD') of sample variables based on yearly averages (to balance companies with many and/or few ratings) and match financial (yearly) variables. The table shows data only six years before the insolvency event (t_0) to avoid diluting insolvency firms that may recover. The values shown not imputed for missing values. The employee rating variables range from 1 (very bad) to 5 (very good), except the *Employee Benefit Score*, which shows the number of benefits. Finance variables are ratio values. *Age* shows years. *Average Assets* show million euros. Macro variables are percentages - only *IFO GER* and *IFO EUR* are absolute values. The sample period ranges from January 2007 to December 2018. The variable definition is explained in Table A.1 in the Appendix. ^(a) *Employee Rating* is the mean of the 13 rating variables. ^(b) *Employee Benefit Score* is measured by the number of benefits offered to employees.

B^(c) *Note Panel* The table illustrates the effect of the 4-step matching process by comparing the standard deviation (=SD) of selected variables before and after the matching process. The Wilcoxon test statistic determines whether the matching process leads to a significant reduction in SD, an indicator of a successful matching process.

Figure S.1 in the Supplementary Material, the mean employee rating shows a drop (presumably after the 2007 financial crisis) around 2008/09, followed by a steady increase from 2010 onward. The mean

Table 2
Principal Component Analysis (PCA).

	Loadings				
	DIM.1	DIM.2	DIM.3	DIM.4	DIM.5
Rating					
Employee Rating	0.99	0.10	-0.03	0.05	-0.02
E. Benefit Score	0.63	0.06	0.07	-0.04	0.02
Older Colleagues	0.82	0.10	-0.05	0.03	-0.02
Management Behaviour	0.88	0.06	-0.04	0.03	-0.01
Career Development	0.88	0.09	-0.01	0.05	0.02
Challenging Work	0.82	0.10	-0.05	0.01	-0.03
Company Image	0.87	0.14	0.04	0.06	-0.03
Environmental F.	0.86	0.11	0.00	0.07	-0.08
Gender Equality	0.82	0.08	-0.06	0.07	-0.01
Internal Communication	0.88	0.12	-0.02	0.09	0.02
Work Atmosphere	0.92	0.07	-0.03	0.07	-0.02
Office Work Environment	0.83	0.16	-0.04	-0.01	-0.04
Overall Compensation	0.82	0.07	0.02	0.01	-0.01
Teamwork	0.74	0.14	-0.06	0.04	-0.03
Work Life Balance	0.86	0.07	-0.02	-0.01	-0.02
Finance					
EBITDA to Assets	0.08	-0.10	0.79	0.16	0.41
EBIT to Assets	0.13	-0.12	0.82	0.17	0.39
EBIT to Sales	0.15	-0.07	0.61	0.15	0.50
Net Income to Assets	0.07	-0.12	0.81	-0.09	-0.27
Sales to Assets	-0.16	-0.09	-0.02	0.91	-0.03
Working Capital to A.	-0.09	-0.08	0.07	0.28	-0.57
Debt to Assets	-0.06	0.14	-0.79	0.22	0.44
Equity to Assets	0.07	-0.14	0.80	-0.21	-0.43
Debt to Equity	0.05	0.02	0.01	-0.01	-0.01
Cash to Assets	0.02	-0.12	0.10	0.18	0.42
Stock to Working Capital	-0.01	0.02	-0.02	-0.04	0.03
Creditors to Liabilities	-0.12	-0.10	0.12	0.35	-0.05
Debtors-Assets	-0.04	0.03	0.06	0.58	-0.33
Controls					
Age	0.02	-0.09	0.14	-0.15	-0.14
Average Assets	0.23	0.18	0.02	-0.43	0.17
Macro					
Euribor	-0.23	0.92	0.15	0.03	0.00
CPI Change p.a.	-0.22	0.85	0.14	0.03	0.02
IFO GER	0.27	-0.95	-0.10	-0.04	0.02
IFO EUR	0.27	-0.88	-0.06	-0.04	0.03
Unemployment Rate	-0.24	0.94	0.12	0.03	-0.01
Insolvency Rate Change	-0.28	0.93	0.11	0.04	-0.01
Eigenvalues					
	32.2	14.0	9.9	6.9	4.5

The table illustrates the first five dimensions of the Principal Component Analysis (PCA), which explain 67.5 percent of the variance in the data. The objective of PCA is to summarize the information in a multivariate data-set by reducing the dimensionality. These new dimensions are a linear combination of the original variables. The eigenvalues represent the percentage of variation explained by each principal component. The gray-shaded values highlight correlations greater than 0.55 or less than -0.55. The sample period ranges from January 2007 to December 2018. Variable definition is explained in Table A.1 in the Appendix.

Employee Benefit Score, however, is fluctuating around the mean of 5.13. By aggregating into industries, we find substantial variation (see, for example, Table S.2 in the Supplementary Material) that is in line with previous literature such, as e.g., Huang et al. (2015).

More interestingly, for our purposes is the distinction between distressed and non-distressed companies. As reported in Table 1, we find a substantial difference in *Employee Rating* between the two sub-samples, with non-distressed companies reporting a mean *Employee Rating* of 3.10 (standard deviation (SD) of 0.87) compared to 2.69 (SD of 0.96) for distressed companies. Although this is only descriptive evidence, we can get a first hint of the results of our inferential analysis by comparing the development of mean ratings prior to the default event for the sub-sample of distressed and non-distressed companies (Figure 1). We find a clear difference in levels, distressed companies show lower *Employee Rating* values. And the spread between distressed and non-distressed companies widens as early as 4 years prior to default, providing a first indication that employee ratings may signal poor company performance ahead of the distress event itself. We elaborate on the role as an early warning indicator in more detail in Section 3.

2.4 Company-level and macroeconomic data

To control for various company-specific and macroeconomic effects, we retrieve an extensive set of variables that are summarized in Table 1. For financial variables, we follow well-established results from the credit risk literature, such as e.g., Altman et al. (2010), Altman et al. (2016), Ohlson (1980), Zmijewski (1984), or Grunert et al. (2005), and select key ratios that are known to be default predictors. Furthermore, we generally include sector-fixed effects to estimate sector-specific, time-invariant effects. Therefore, we add a sector variable²⁵ and a second variable to account for industry-specific risk. We could have added industry-fixed effects (each company in the data set is associated with a sector and an industry). However, due to the low sample size (and consequently low degrees of freedom), we refrained from adding 28 industry-specific dummy variables each year. In addition, this also emphasizes the well-known incidental variable problem, which we aim to minimize here. Therefore, we develop a proxy variable to represent industry-specific risk. We define this variable as the *Industry Risk* variable and set a five-point scale with the levels ‘average risk’ (1), ‘increased industry risk’ (2), ‘temporal industry decline’ (3), ‘shrinking industry’ (4), and ‘industry disruption’ (5). The risk of insolvency is the highest in the level of ‘industry disruption’ and subsequently decreases to an average risk in the level of ‘average risk’. The five-scale variable thus represents the industry risk for that industry. We have inferred the methods based on the market section of the restructuring reports. See Table A.3 in the Appendix for a detailed definition and the methodology of this variable.

Several studies, such as, e.g., Bonfim (2009) stress the impact of the macroeconomic environment. We therefore retrieve data on unemployment, inflation, interest rates, and insolvency quota (Hillegeist et al.

²⁵ We have tested our models with two different levels of cluster-fixed-effects: Sector and industry level. We apply sector controls in order to deal with a low sample size. The sector variable comprises of ‘Manufacturing’, ‘Energy’, ‘Logistics’, ‘Trade’, ‘Service’, and ‘Construction’. For an overview of industries and sectors, see Table S.4 in the Appendix.

(2004) or Altman et al. (2010)). As our sample consists of 93% German companies, there is not much variation in the macro variables between companies just across years. Thus, the macro variable models relatively a time-fixed effects, and using one macro variable is sufficient. We use the yearly percentage change in the (detrended) German unemployment rate. Finally, in line with, e.g., Altman et al. (2016) we include age as a control variable. As previous studies have detected a non-monotonic relation between age and the probability of insolvency, we follow the literature and take the log of the variable. Table 1 reports descriptive statistics; the distressed-company sample shows lower operating performance, higher leverage, and lower cash levels.

2.5 Timing of financial variables and data selection

We intend to assess the predictive ability of employee ratings relative to the information content of (traditional) accounting-based models. A relevant issue concerns the real-world timing of data availability. Although employee rating data are available immediately, it is well known that annual reports are only issued with substantial delays. Regular publication delays are likely to be even significantly longer for distressed companies, as has been shown, e.g., Keasey and Watson (1991), Whittred and Zimmer (1984), or Lawrence (1983). For example, these studies state that nearly 50% of distressed companies experience a reporting delay of four up to thirteen months after the end of the financial year. Based on empirical findings, we have checked the availability of financial information for insolvent companies in our sample at the registration court, and in 45% of the cases, no annual reports were available one year before the default event. We also found submission lags for the available financial reports up to two years. Restructuring reports provide reasons for these delays. In fact, more than 90% of reports indicate weaknesses in the financial organization and processes of distressed companies. Therefore, we follow other studies and include a one-year time lag for all financial variables to account for reporting delays (e.g., Shumway, 2001) unless stated otherwise.

2.6 Data imputation

Data availability poses several challenges regarding sample size for model estimates. Reduction of sample size results from (i) the restriction to use only rating data before an insolvency event in t_0 , (ii) the yearly availability of financial data and thus the calculation of annual averages for employee ratings, and (iii) the fact that complete data rows for estimation is necessary. Peduzzi et al. (1996), for example, determine the threshold for the optimal sample size (Subsection 4.1). To check whether sample size restrictions are a problem, we also use data imputation via the k -nearest neighbors algorithm to predict the outcome of a new observation as the average of the five most similar observations. We apply this algorithm to employee rating variables (e.g., *Employee Rating*, *Employee Benefit Score*) and to the possible control variable *Average Assets*. We investigate the results of the models before and after imputation in Subsection 4.1.²⁶

²⁶ The data sample shows 1,999 rating values; after imputation we receive 5,320 values.

2.7 Model specification

Our focus is on assessing the incremental value of employee rating information for predicting default beyond traditional credit risk indicators. Therefore, we estimate a benchmark model comprising well-established explanatory variables and label it generically as the traditional model, $\mathcal{M}^{\text{Trad}}$. When adding our variables of interest, namely the employee rating information, we refer to the augmented model as the employee model, $\mathcal{M}^{\text{Emp}}$.

A logistic regression model is used to estimate the probability of insolvency. We apply a penalized log-likelihood method²⁷ to account for the rare-event structure and to address perfect separation in our sample, where maximum-likelihood estimates tend to infinity.²⁸ Finally, we apply cluster-robust standard errors (heteroskedasticity and autocorrelation-consistent (HAC)), which are clustered at sector and year(s) of observation before the insolvency event. Thus, our workhorse estimation model is

$$\text{logit } Y_{i,t} = \beta_0 + \beta_1 EMP_{i,t} + \beta_2 FIN_{i,t-1} + \beta_3 CON_{i,t} + \delta_{i,j} + \gamma_{i,k} + \varepsilon_{i,t}, \quad (1)$$

where $Y_{i,t}$ is the default indicator variable for company i at time t . EMP , FIN , and CON are vectors summarizing our employee rating variables, financial accounting data, and additional control variables, respectively. We include sector fixed effects $\delta_{i,j}$ for sector j and an industry risk variable $\gamma_{i,k}$ that captures the market riskiness k of the industry to which company i belongs. As mentioned above, CON can include macroeconomic factors (e.g., the unemployment rate) that are constant across companies for a given period. Since this essentially controls for the different time periods, we do not add a separate time-fixed effects dummy.

We use different definitions for the default indicator $Y_{i,t}$ in subsequent analyzes. In Section 3.2 on forecasting, we use the usual definition that $Y_{i,t}$ equals 1 in time t at which company i filed for insolvency and zero otherwise. In contrast, in (the next) Section 3.1, we employ the definition that $Y_{i,t}$ equals 1 for *all* t if company i defaulted within the sample period. The latter analysis is effectively akin to a discriminant analysis and tests whether employee information can successfully categorize defaulting versus non-defaulting companies.

For the employee model, $\mathcal{M}^{\text{Emp}}$, EMP may include *Employee Rating*, *Latent Employee Rating*, and *Employee Benefit Score*, while EMP is omitted to estimate $\mathcal{M}^{\text{Trad}}$. We report goodness-of-fit criteria, such as Nagelkerke's R^2 (see Nagelkerke, 1991). In evaluating the model's predictive capability, we focus on three key aspects. First, we compare Balanced Accuracy²⁹ of the traditional $\mathcal{M}^{\text{Trad}}$ and employee model $\mathcal{M}^{\text{Emp}}$ by evaluating whether our new variables of interest improve the metric. Second, we consider

²⁷ Technically speaking, penalization is done using Jeffrey's invariant prior or using the bias-reducing modified scores. We apply bias-reducing modified scores, as in the logistic regression, maximum penalized likelihood as well as the modified-scores approach coincide for natural exponential families.

²⁸ The average insolvency rate in Germany between 2006 – 2018 was 0.82 percent. However, the share of insolvent companies in our sample for the main analysis is approximately 20%, mainly to avoid negative effects from the composition of the sample. See Firth (1993) for details.

²⁹ Due to the imbalanced data-set, Balanced Accuracy, defined as $Balanced\ Accuracy = (Sensitivity + Specificity)/2$ is a more suitable metric compared to Accuracy (see for instance Thölke et al., 2023).

AUC and the improvement in the ROC-curve. AUC can be interpreted as the average power of the model across all possible cut-off values for healthy and insolvent companies. Finally, we inspect Sensitivity and Specificity. Sensitivity is the ratio of correctly identified insolvent companies, which is of the highest economic value in such models. In contrast, specificity measures the ratio of correctly identified healthy companies. We standardize all numeric variables to have a standard deviation of one for comparable interpretation of effect sizes. All statistical analyzes are performed with the software R.

3 Results

Our analysis proceeds in three steps. First, in the following subsection, the potential of our newly introduced *EMP* variables is investigated to improve the discrimination between defaulting and non-defaulting companies. Second, we analyze whether the fitted model has predictive ability for identifying defaulting companies in future time periods $t + q$ conditional on the information set at t . Third, in Section 3.2, we re-specify our estimation model (1) such that it is a proper forecasting regression. We evaluate the fitted models in- and out-of-sample.

3.1 Discriminatory power of employee rating information

This section tests the in-sample ability of employee ratings to discriminate between insolvent and non-insolvent companies. For this purpose, the default indicator $Y_{i,t}$ (Eq. (1)) defined as one for all t for companies i that defaulted at any time during the sample period (and zero otherwise). For each company i , we include data from twelve years before the insolvency date t_{-12} up to t_0 , and drop all data beyond t_0 . This is the maximum period possible, given our sample construction.

We summarize the main results in Table 3, where column 1 reports coefficient estimates for model $\mathcal{M}^{\text{Trad}}$, while columns 2 to 5 show results for various specifications of $\mathcal{M}^{\text{Emp}}$. In columns 2 to 3, we include *Employee Rating* and *Employee Benefit Score* individually. In column 4, we use the variable *Latent Employee Rating* variable derived from the PCA analysis (see Section 2.3). Column 5 reports a combination of *EMP* variables *Employee Rating* and *Employee Benefit Score*.

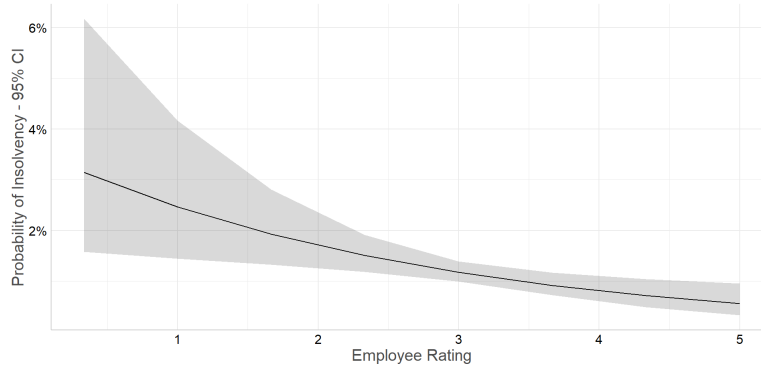
For various specifications of $\mathcal{M}^{\text{Emp}}$, the employee rating variables are statistically significant and have the expected signs. Including the three basic variables *EMP* individually shows that each is strongly significant, with a negative coefficients ranging from -0.16 to -0.71 , indicating that a higher rating is associated with a significantly lower odds of insolvency. Combining *Employee Rating* and *Employee Benefit Score*, as in specification 5, shows that both variables contribute significantly. Furthermore, when using the *Latent Employee Rating* derived from the PCA analysis (Section 2.3), we again find a highly significant negative coefficient. The *Employee Benefit Score* shows the largest coefficient magnitude (negative log-odds of 0.71), followed by *Employee Rating* (negative log-odds of 0.47). Note that each specification uses the full set of control variables, sector fixed effects and the *Industry Risk* variable.

Considering the model fit criteria, we find notable improvements. Nagelkerke R^2 increases from 0.19 for $\mathcal{M}^{\text{Trad}}$ to 0.23 and 0.27 for the employee model, $\mathcal{M}^{\text{Emp}}$ (McFadden R^2 increases from 0.12 to 0.15 and 0.18). Likewise, we find substantial declines in the Akaike information criterion (AIC) for

$\mathcal{M}^{\text{Emp}}$ compared to $\mathcal{M}^{\text{Trad}}$, with the strongest decline for the combination of *EMP* variables in the model. To assess the economic significance of the results, note that the coefficient estimates in Table 3 are interpreted as decreases in the log odds of insolvency. Thus, a one-unit increase in the *Employee Rating* variables decreases the log odds or logit of being insolvent by -0.47 . This equals a 38% lower probability of insolvency once the rating increases by one unit (assuming that all other variables remain fixed).

To elaborate on the impact of our estimation results, we show the marginal effects plot in Figure 2. As the *Employee Rating* variable decreases, the marginal effect on a company’s insolvency increases. On average, a unit change in the *Employee Rating* changes the predicted probability of the company being insolvent by approximately half a percentage point.³⁰

Figure 2
Marginal Effects Based on Predicted Probabilities of Companies Being Insolvent



This figure illustrates the marginal effects of a unit change in the *Employee Rating* variable on the probability of a company being insolvent. All other variables are held constant at their mean values. The purpose of this visualization is to show absolute probabilities of insolvency for third-party comparison (e.g., with rating agency predictions) and to highlight the overall size rather than relative probability changes (as defined by odds). The probability of insolvency is displayed on the vertical axis, including a 95 percent confidence interval (CI), while the *Employee Rating* variable is on the horizontal axis, ranging from 1 (very bad) to 5 (very good).

Concerning the control variables, for $\mathcal{M}^{\text{Trad}}$, we find that all financial variables are statistically significant, except *Working Capital to Assets* and *Debt to Equity*. Interestingly, the variable *Equity to Assets* shows the highest marginal effect of all variables. In addition, we observe significant controls Δ *Unemployment Rate* (*UR*). and *Log Age*. These findings are in line with existing research. Distressed companies usually show a lower level of equity and smaller size (see, for instance, Ohlson, 1980), and are younger (usually they have a higher rate of insolvency up to five years after foundation (see, for instance, Altman et al., 2016)). Moreover, we observe a higher probability of insolvency during economic downturns (see for instance Bonfim, 2009). We also acknowledge the appropriate explanatory power of the traditional models, which are standard in the credit industry. This is, for example, shown by Nagelkerke R^2 of 0.19 in the $\mathcal{M}^{\text{Trad}}$ model.

³⁰ We correct the sample imbalance and thus the probability following Saeuens et al. (2002).

Table 3
Discriminatory Power – Logistic Regression Results

Model	$Y_{i,t}$: Insolvency of Company				
	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$			
Specification	(1)	(2)	(3)	(4)	(5)
Intercept	−1.84*** (0.24)	−2.14*** (0.24)	−2.00*** (0.25)	−2.02*** (0.25)	−2.17*** (0.26)
Employee Benefit Score		−0.71*** (0.07)			−0.61*** (0.07)
Employee Rating			−0.47*** (0.06)		−0.19*** (0.07)
Latent Employee Rating				−0.16*** (0.02)	
Net Income to Assets	−0.46*** (0.13)	−0.43*** (0.13)	−0.43*** (0.12)	−0.43*** (0.12)	−0.43*** (0.13)
Working Capital to Assets	−0.01 (0.06)	−0.07 (0.06)	−0.01 (0.06)	−0.01 (0.06)	−0.06 (0.06)
Equity to Assets	−0.85*** (0.11)	−0.85*** (0.11)	−0.88*** (0.11)	−0.88*** (0.11)	−0.86*** (0.11)
Debt to Equity	−0.001 (0.04)	0.004 (0.04)	−0.01 (0.04)	−0.01 (0.04)	−0.002 (0.04)
Sales to Assets	−0.21*** (0.06)	−0.26*** (0.06)	−0.25*** (0.06)	−0.25*** (0.06)	−0.27*** (0.06)
Log Age	−0.39*** (0.08)	−0.31*** (0.08)	−0.37*** (0.08)	−0.37*** (0.08)	−0.32*** (0.08)
Unemployment Rate (UR)	0.20*** (0.06)	0.23*** (0.06)	0.16*** (0.06)	0.17*** (0.06)	0.21*** (0.07)
Cluster Fixed-Effects (Sector) included?	Yes	Yes	Yes	Yes	Yes
Industry Risk variable included?	Yes	Yes	Yes	Yes	Yes
Nagelkerke's R ²	0.19	0.26	0.23	0.23	0.27
McFadden's R ²	0.12	0.18	0.15	0.15	0.18
AIC	1992	1868	1926	1925	1862
Total Observations in Model	2063	2063	2063	2063	2063

This table presents the traditional model $\mathcal{M}^{\text{Trad}}$ (1) and the employee models $\mathcal{M}^{\text{Emp}}$ (2-5) incorporating various employee-related variables. The models assess whether employee variables (*EMP*) explain insolvency. Comparison based on different models of the information set (*EMP*, *FIN*, *CON* variables). Models based on complete sample data. Cluster fixed-effects and the *Industry Risk* variable are included but not reported. Standard errors (clustered for sector and year) are in parentheses. The sample period spans January 2007 - December 2018. Log-odds as coefficients.

Significance: *p<0.1, **p<0.05, ***p<0.01.

As the next step, we now turn to the question of whether the information in *EMP* has predictive power relative to traditional models $\mathcal{M}^{\text{Trad}}$. To evaluate the prediction capability, we use the same *EMP* models as shown in Table 3. Based on equation (1), we predict using all observations beginning twelve years before the insolvency date (t_0). This means we focus on a 12-year prediction horizon.

Table 4
Discriminatory Power – Prediction Metrics

Prediction Metrics	Model				
	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Emp}}$
Base Failure Rate	0.25	0.25	0.25	0.25	0.25
No Information Rate (NIR)	0.75	0.75	0.75	0.75	0.75
Accuracy	0.76	0.74	0.74	0.74	0.74
Balanced Accuracy	0.71	0.76	0.74	0.74	0.75
$\Delta \mathcal{M}^{\text{Trad}} \text{ vs } \mathcal{M}^{\text{Emp}}$		$+0.06$	$+0.03$	$+0.03$	$+0.04$
AUC	0.78	0.81	0.81	0.80	0.81
$\Delta \mathcal{M}^{\text{Trad}} \text{ vs } \mathcal{M}^{\text{Emp}}$		$+0.03$	$+0.03$	$+0.02$	$+0.03$
Sensitivity	0.61	0.78	0.73	0.73	0.78
$\Delta \mathcal{M}^{\text{Trad}} \text{ vs } \mathcal{M}^{\text{Emp}}$		$+0.17$	$+0.12$	$+0.12$	$+0.17$
Specificity	0.81	0.73	0.75	0.75	0.73
$\Delta \mathcal{M}^{\text{Trad}} \text{ vs } \mathcal{M}^{\text{Emp}}$		-0.08	-0.06	-0.06	-0.08
Total observations in model	2063	2063	2063	2063	2063

This table shows the prediction metrics for the traditional model $\mathcal{M}^{\text{Trad}}$ and the employee model $\mathcal{M}^{\text{Emp}}$ based on the full data sample. Comparison dependent on various time periods of the information set (*EMP*, *FIN*, *CON* variables). Models based on all relevant time periods, which comprise all values starting with the insolvency event (t_0) until 12 years before the insolvency. This table presents the traditional model $\mathcal{M}^{\text{Trad}}$ (1) and the employee models $\mathcal{M}^{\text{Emp}}$ (2-5), incorporating various employee-related variables based on the same specifications as Table 3. Rows labeled by Δ report the difference between the performance metrics of the $\mathcal{M}^{\text{Emp}}$ model and the traditional model $\mathcal{M}^{\text{Trad}}$. The number of observations equals the number of distinct companies per year with a (yearly) value in the sample. The sample period ranges from January 2007 to December 2018.

First, Balanced Accuracy in $\mathcal{M}^{\text{Emp}}$ is approximately equal to the No Information Rate (plus/minus one percentage point) for the $\mathcal{M}^{\text{Emp}}$ model, however is five percentage points lower for $\mathcal{M}^{\text{Trad}}$ model. Second, we observe that the employee model, i.e., $\mathcal{M}^{\text{Emp}}$ shows a higher Balanced Accuracy (BA) than the traditional model $\mathcal{M}^{\text{Trad}}$. The improvement shows an increase between three and five percentage points. This means that our models $\mathcal{M}^{\text{Emp}}$ show an incremental improvement in prediction. Regarding the Area Under the Curve (AUC) metric, the employee model $\mathcal{M}^{\text{Emp}}$ is consistently above 0.80 with a maximum of 0.81. This means that there is an 81% chance that the model can distinguish between the healthy class and the insolvent class (a random model is 50% and the No Information Rate here is 75%) up to twelve years before insolvency. More interestingly, the relative improvement in $\mathcal{M}^{\text{Emp}}$ is up to three percentage points, a considerable increase. To put our results into perspective: Classical studies, such as, e.g., Altman et al. (2010) find AUC values for default prediction of 0.71 to 0.75. They find that qualitative information, such as legal action by creditors, company filing histories, and audit report data, improves default prediction, increasing the AUC by approximately four percentage points. Similarly, a recent study by Berg et al. (2020) examines the value of a digital footprint for predicting consumer default. They find an AUC value of 0.68, which improves by five percentage points. Our results, in particular, the improvement in the AUC, are of comparable magnitude and significance. The

absolute magnitude is higher, though, with an AUC above 0.80, compared to a range of 0.68 to 0.75 reported in the other studies.

Furthermore, we evaluate Sensitivity and Specificity. Sensitivity is the ratio of companies that the model correctly predicts as insolvent. Both metrics are based on the optimal threshold of the ROC curve. Sensitivity is between 0.73 and 0.78, which is a good level overall. More interestingly, the relative improvement by the inclusion of *EMP* variables is between 12 and 17 percentage points. Specificity, which is the ratio of correctly predicted healthy companies, shows slightly different patterns. Despite the higher level compared to Sensitivity, Specificity decreases when *EMP* variables are included. Specificity shows an absolute value of 0.81 for the $\mathcal{M}^{\text{Trad}}$ model and values between 0.73 and 0.75 for the $\mathcal{M}^{\text{Emp}}$ model. The relative decrease of up to eight percentage points in $\mathcal{M}^{\text{Emp}}$ compared to $\mathcal{M}^{\text{Trad}}$.

In summary, the evidence, allows us to draw the following two conclusions, which are in line with our hypothesis that employee ratings improve business failure prediction models. First, *EMP* variables help to improve the prediction of insolvency. Second, these variables are especially helpful for identifying companies that are, in fact, insolvent. In the next section, we implement proper forecasting regressions and conduct extensive out-of-sample testing to corroborate these findings.

3.2 In-sample forecasting

The approach in the previous section was based on fitting a classification model and assessing the predictive power conditional on belonging to either the insolvent or the healthy company group in the information set. In contrast, this section implements the arguably even stricter test of formulating a dedicated forecasting regression model. More precisely, we now define the default indicator $Y_{i,t}$ as one for the year in which the default event occurs, i.e., at time $t = t_0$, and zero otherwise, i.e. for all years prior to default, $t < t_0$. We regress $Y_{i,t}$ on the lagged values of the explanatory variables, sequentially considering horizons $h = 1, \dots, 3$. Due to data limitations, we consider lags of up to three years. As in the previous section, we continue to maintain the realistic assumption that financial variables are only available with an additional one-year time lag.³¹ Thus, we adapt the workhorse empirical model of (1) as:

$$\text{logit } Y_{i,t+h} = \beta_0 + \beta_1 EMP_{i,t} + \beta_2 FIN_{i,t-1} + \beta_3 CON_{i,t} + \delta_{i,j} + \gamma_{i,k} + \varepsilon_{i,t}, \quad (2)$$

where, *EMP*, *FIN*, and *CON* are vectors summarizing our employee rating variables, financial accounting data, and additional control variables, respectively. Again, we include sector fixed effects $\delta_{i,j}$ for sectors j and an industry risk variable $\gamma_{i,k}$. For forecasting, we will focus exclusively on the primary variable of interest, *Employee Rating*, which has a larger effect size than *Latent Employee Rating* but not as large as *Employee Benefit Score*. However, we favor *Employee Rating* over *Employee Benefit Score* because of the methodological distinction between different employee ratings.

³¹ We investigate the stability of our results with respect to the lagged financial variables as additional robustness check in Section 4.2 By removing the lag from the financial variables, one adopts a more theoretical perspective rather than that of a decision-maker in the real world. Our results remain qualitatively unchanged from either perspective.

This specification of the regression model tests whether the information available at time t has predictive value for the default event that occurs exactly h years from t . Instead of asking for a prediction of default to occur exactly *in* year $t+h$, we may also test the forecasting power to predict default to occur sometime *within* year t and $t+h$. To do this, we perform a separate estimate, where we define the default indicator $Y_{i,t+h}$ to be one when the default event t_0 occurs between t and $t+h$, i.e. $t_0 \in [t, t+h]$, and zero otherwise. Put it differently, for example, if $h = 3$, we test whether employee and financial information available at time t can forecast that the company i will default at any time within the next three years. We label the corresponding specifications by $h \in \{1, 2\}$ and $h \in \{1, 2, 3\}$, respectively. For $h \in \{1, 2\}$ we lag the variables by two years; for $h \in \{1, 2, 3\}$, three years. We split the data into 80% for estimation (in-sample, IS) and 20% for evaluating model performance (out-of-sample, OOS). The in-sample training period includes 80% of randomly selected companies.³² We subsequently validated the model within an out-of-sample period with the remaining 20% of the companies. To obtain robust findings, we run a Monte Carlo simulation across sample splits ($n = 1,000$) and present the average results in our tables. We present the results from the out-of-sample period in Section 3.3. We also measure the time relative to the insolvency event, that is, the time before and after it. In-sample and out-of-sample data comprise data from twelve years before the insolvency ($t_0 - 12$) to the insolvency (t_0). We summarize the results in Table 5.

We find that our *EMP* variables have the expected negative sign and are (strongly) significant across all lag specifications. We find that the coefficient estimates are of comparable magnitude even for longer horizons (i.e. coefficients on *Employee Rating* are -0.445 for $h = 3$, -0.408 for $h \in \{1, 2\}$ and -0.504 for $h \in \{1, 2, 3\}$). The finding that the improvement of the forecast power of $\mathcal{M}^{\text{Emp}}$ relative to the traditional model $\mathcal{M}^{\text{Trad}}$ tends to increase as we attempt to forecast default at longer horizons is also supported by considering the metrics for model fit. Although for the short horizon $h = 1$, we observe only an increase from 0.05 to 0.06 in R^2 (Nagelkerke), this improvement in the model fit increases to 0.09 in $h = 2$ and 0.07 in $h = 3$. A similar conclusion can be drawn from the specifications where we consider default anytime up to horizons of two and three years. Here, we observe an increase to 0.11 (from 0.08) and even to 0.09 (from 0.05) for $h \in \{1, 2\}$ and $h \in \{1, 2, 3\}$, in the employee model $\mathcal{M}^{\text{Emp}}$.³³ *FIN* variables *Net Income to Assets* and *Working Capital to Assets* lose their significance and *Equity to Assets* loses part of their effect size. The effect size of *Sales to Assets* on the opposite stays comparatively constant around 0.200 and 0.300, indicating that size is a stable predictor (regardless of time). The fact that model fit, as measured by Nagelkerke's R^2 , is generally lower in the forecasting setup than in the previous results from Table 3 should not come as a surprise as the forecasting tests are much stronger.

Taken together, our (in-sample) forecasting regression results strengthen the evidence from the previous sections, and lend support to the additional finding that employee rating information provides relevant information for forecasting default, particularly over longer horizons, compared to the benchmark

³² For a discussion of time- versus random splits, see, e.g. Morita et al. (2022).

³³ As further observation, we can compare the results from column (1) in Table 3 and column (3) in Table 5

Table 5
Forecasting – Logistic Regression Results

$Y_{i,t+h}$: Insolvency of Company										
Forecast Horizon	$h = 1$		$h = 2$		$h = 3$		$h \in \{1, 2\}$		$h \in \{1, 2, 3\}$	
Model	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$
Intercept	-3.266*** (0.538)	-3.305*** (0.505)	-3.157*** (0.541)	-3.204*** (0.491)	-3.392*** (0.569)	-3.488*** (0.507)	-3.024*** (0.494)	-3.099*** (0.447)	-2.793*** (0.408)	-2.937*** (0.382)
Employee Rating		-0.250*** (0.088)		-0.351*** (0.096)		-0.445*** (0.172)		-0.408*** (0.116)		-0.504*** (0.150)
Net Income to Assets	-0.199 (0.138)	-0.192 (0.137)	-0.043 (0.095)	-0.031 (0.088)	0.008 (0.050)	0.020 (0.048)	-0.039 (0.108)	-0.022 (0.102)	0.040 (0.054)	0.059 (0.054)
Working Capital to Assets	-0.114 (0.145)	-0.116 (0.143)	-0.087 (0.133)	-0.100 (0.130)	-0.016 (0.108)	-0.036 (0.107)	-0.080 (0.123)	-0.096 (0.122)	0.047 (0.089)	0.034 (0.089)
Equity to Assets	-0.355*** (0.086)	-0.351*** (0.092)	-0.462*** (0.111)	-0.467*** (0.117)	-0.429*** (0.123)	-0.438*** (0.128)	-0.523*** (0.124)	-0.531*** (0.131)	-0.472*** (0.160)	-0.493*** (0.175)
Debt to Equity	-0.106* (0.063)	-0.108* (0.057)	0.067 (0.076)	0.061 (0.064)	0.108*** (0.042)	0.097*** (0.035)	0.060 (0.060)	0.053 (0.049)	0.090** (0.040)	0.079** (0.034)
Sales to Assets	-0.189* (0.101)	-0.209** (0.099)	-0.319*** (0.102)	-0.337*** (0.094)	-0.290** (0.113)	-0.312*** (0.106)	-0.311*** (0.096)	-0.333*** (0.091)	-0.164 (0.100)	-0.193* (0.100)
Log Age	-0.299* (0.163)	-0.290* (0.157)	-0.333** (0.156)	-0.311** (0.148)	-0.354** (0.144)	-0.337** (0.140)	-0.367*** (0.132)	-0.344*** (0.125)	-0.387*** (0.109)	-0.371*** (0.109)
Δ Unemployment Rate (UR)	-0.647 (0.447)	-0.726 (0.508)	-0.248 (0.397)	-0.311 (0.433)	-0.113 (0.371)	-0.171 (0.396)	-0.236 (0.359)	-0.304 (0.388)	0.079 (0.372)	0.039 (0.388)
Sector Cluster Fixed-Effects included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Risk Variable included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nagelkerke's R^2	0.05	0.06	0.06	0.09	0.04	0.07	0.08	0.11	0.05	0.09
McFadden's R^2	0.04	0.05	0.05	0.07	0.03	0.06	0.06	0.09	0.03	0.07
Brier Score	0.057	0.056	0.059	0.058	0.055	0.054	0.080	0.078	0.094	0.091
AIC	742	733	755	737	724	705	957	928	1097	1060
Total Observations in Model	1650	1650	1650	1650	1646	1646	1650	1650	1646	1646

This table presents the traditional model $\mathcal{M}^{\text{Trad}}$ and the employee models $\mathcal{M}^{\text{Emp}}$ with different forecasting horizons, i.e. we evaluate the forecasting of default events in t given the information sets of $t - h$ respectively. Comparison based on different time periods of the information set (EMP , FIN , CON variables). Models based on data in-sample (IS - 80% of companies). Sample split is done by a Monte Carlo simulation with $n = 1,000$ splits. Lag $h = 1$ refers to a one-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 2$ refers to a two-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 3$ refers to a three-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h \in \{1, 2\}$ refers to a two-year forecast horizon. This model forecast whether the company i defaults anytime within the next two years, latest at the default event t_0 . Lag $h \in \{1, 2, 3\}$ refers to a three-year forecast horizon. This model forecasts whether the company i defaults anytime within the next three years, latest at the default event t_0 . Cluster fixed-effects and the *Industry Risk* variable are included but not reported. Standard errors (clustered for sector and year) are in parentheses. The sample period spans January 2007 - December 2018. Log-odds as coefficients.
Significance: *p<0.1, **p<0.05, ***p<0.01.

model based on traditional financial information. Next, we evaluate the forecasting model introduced in this section using the classification metrics from the previous section and present the results in Table 6. In addition to the results for $\mathcal{M}^{\text{Emp}}$ and $\mathcal{M}^{\text{Trad}}$ for the different lag specifications, we add the random model $\mathcal{M}^{\text{Random}}$ for comparison in column 2. With respect to the forecasting horizons of one and two years ($h = 1$, $h = 2$), we observe that the Balanced Accuracy of $\mathcal{M}^{\text{Emp}}$ is only one percentage point better than $\mathcal{M}^{\text{Trad}}$. The periods $h = 3$, $h \in \{1, 2\}$ and $h \in \{1, 2, 3\}$ show an increase in BA between 0.02 and 0.03. This improvement is primarily driven by Sensitivity, whereas Specificity shows even a slight decrease of 0.01 in $h \in \{1, 2, 3\}$. Except $h = 1$, Sensitivity improves in $\mathcal{M}^{\text{Emp}}$ model between three and six percentage points, which we consider to be sizable. Although showing a lower effect size, these results are in line what we have observed for prediction. *EMP* variables especially help to better identify and forecast companies, which are insolvent. Together, this suggests that *EMP* variables substantially improve the forecasting power beyond $\mathcal{M}^{\text{Trad}}$, especially for forecasting true positives.

Table 6
Forecasting – In-Sample Prediction Metrics

Forecast Horizon	$Y_{i,t+h}$: Insolvency of Company										
	$h = 1$			$h = 2$		$h = 3$		$h \in \{1, 2\}$		$h \in \{1, 2, 3\}$	
	IS			IS		IS		IS		IS	
Sample											
Model	$\mathcal{M}^{\text{Random}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$
Base Failure Rate (Prevalence)	0.02	0.07	0.07	0.07	0.07	0.06	0.06	0.10	0.10	0.12	0.12
No Information Rate (NIR)	0.98	0.93	0.93	0.93	0.93	0.94	0.94	0.90	0.90	0.88	0.88
Accuracy	0.96	0.76	0.78	0.75	0.75	0.72	0.72	0.74	0.75	0.69	0.69
Balanced Accuracy (BA)	0.50	0.72	0.73	0.72	0.73	0.69	0.72	0.71	0.73	0.67	0.70
Δ BA $\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$			0.01		0.01		0.03		0.02		0.03
AUC	0.49	0.77	0.77	0.76	0.77	0.73	0.76	0.76	0.78	0.71	0.74
Δ AUC $\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$			0.00		0.01		0.03		0.02		0.03
Sensitivity (SEN)	0.02	0.68	0.68	0.68	0.71	0.66	0.72	0.68	0.71	0.64	0.70
Δ SEN $\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$			0.00		0.03		0.06		0.03		0.06
Specificity (SPEC)	0.98	0.76	0.78	0.76	0.76	0.72	0.72	0.74	0.76	0.70	0.69
Δ SPEC $\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$			0.02		0.00		0.00		0.02		-0.01
Total Observations in Model	2079	1644	1644	1641	1641	1640	1640	1641	1641	1639	1639

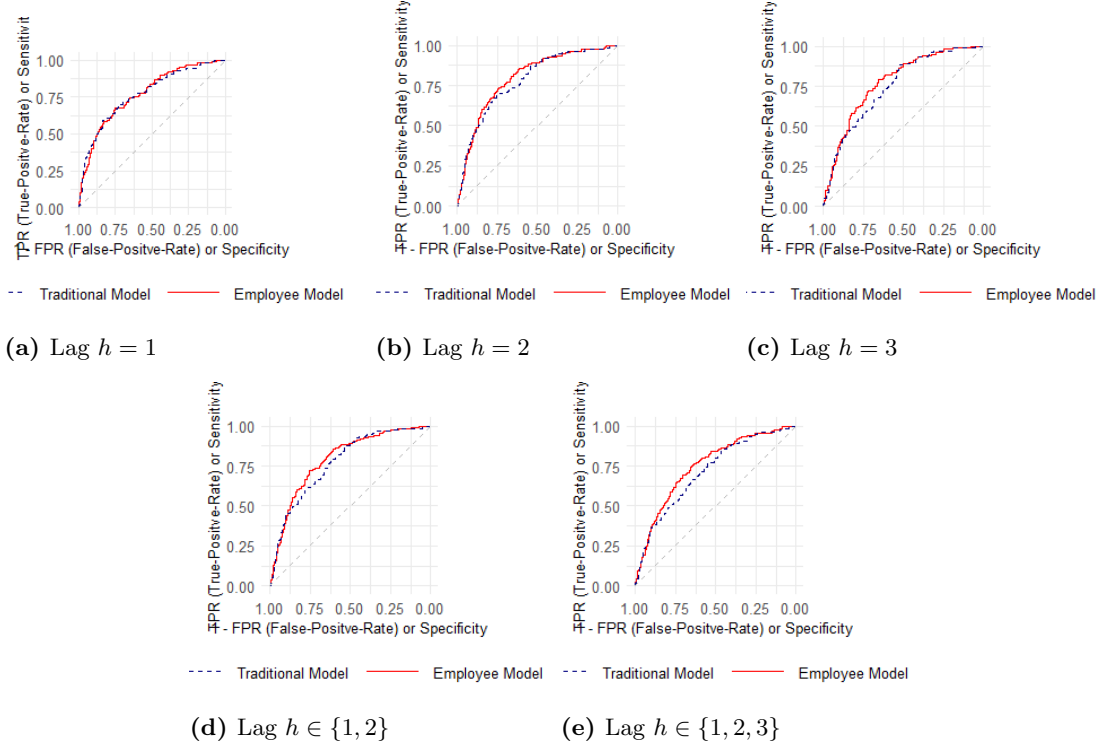
This table shows the forecast metrics for the random classifier $\mathcal{M}^{\text{Random}}$ the traditional model $\mathcal{M}^{\text{Trad}}$ and the employee model $\mathcal{M}^{\text{Emp}}$. Comparison based on different time periods of the information set (*EMP*, *FIN*, *CON* variables). Models based on data in-sample (IS - 80% of companies). Sample split is done by a Monte Carlo simulation with $n = 1,000$ splits. Lag $h = 1$ refers to a one-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 2$ refers to a two-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 3$ refers to a three-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h \in \{1, 2\}$ refers to a two-year forecast horizon. This model forecast whether the company i defaults anytime within the next two years, latest at the default event t_0 . Lag $h \in \{1, 2, 3\}$ refers to a three-year forecast horizon. This model forecasts whether the company i defaults anytime within the next three years, latest at the default event t_0 . The number of observations corresponds to the number of distinct companies per year that have a (yearly) value in the model. The sample period ranges from January 2007 to December 2018.

The AUC of the employee model improves by up to three percentage points and shows a pattern similar to Balanced Accuracy. Comparable results and values are obtained for the cumulative time periods. In fact, in line with the previous literature, we consider the AUC to be the most relevant metric for assessing the forecasting power. As discussed in Section 3.1 we aim to have an AUC value greater than 0.70. The minimum AUC in our $\mathcal{M}^{\text{Emp}}$ specifications is 0.74, with a maximum AUC of 0.78 for $h \in \{1, 2\}$. We provide a graphical representation of these results by plotting the ROC curve for the different lag specifications in Figure 3.

Despite the overall positive assessment, we still note that the Balanced Accuracy of $\mathcal{M}^{\text{Emp}}$ does not outperform the No Information Rate. For the three-year horizon $h = 3$, we observe 0.72 for Balanced Accuracy compared to 0.94 for the No Information Rate, which – when taken at face value – would suggest that the predictive model is performing worse than a simple baseline guess that always predicts non-default. However, because the sample is highly imbalanced (as shown by the prevalence), this conclusion is misguided. Still, we follow, e.g., Agarwal and Taffler (2007), and report the conditional probability perspective here. As a representative example, we report the confusion matrix for the three-year forecast horizon $h = 3$, which is included in Table A.5 of the Appendix. The conditional probability that a company is insolvent given the forecasted probability of being insolvent (which is equal to Sensitivity) of

Figure 3

Receiver operating characteristic (ROC) curve in-sample forecasting



These figures show the the different forecast horizons for the in-sample data. ^(a) Lag $h = 1$: This period refers to a one-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. ^(b) Lag $h = 2$: This period refers to a two-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. ^(c) Lag $h = 3$: This period refers to a three -year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. ^(d) Lag $h \in \{1, 2\}$: This period refers to a two-year forecast horizon. This model forecast whether the company i defaults anytime within the next two years, latest at the default event t_0 . ^(e) Lag $h \in \{1, 2, 3\}$: This period refers to a three-year forecast horizon. This model forecast whether the company i defaults anytime within the next three years, latest at the default event t_0 .

$\mathcal{M}^{\text{Emp}}$ is 0.72. This is significantly better than the Base Failure Rate with 0.06. The values for Balanced Accuracy (which are in the range of 0.70–0.80) suggest the model captures the minority class well enough while not overly penalizing the majority class performance.

In summary, the employee model $\mathcal{M}^{\text{Emp}}$ demonstrates a superior performance to $\mathcal{M}^{\text{Trad}}$ on all horizons evaluated. It delivers higher Balanced Accuracy (indicating a better balance of Sensitivity and Specificity). The higher AUC value of $\mathcal{M}^{\text{Emp}}$ further underscores its overall stronger classification ability across different horizons (except $h = 1$). These performance differences suggest that $\mathcal{M}^{\text{Emp}}$ provides a more reliable and effective tool for insolvency risk forecasting, especially for longer-range predictions, where its improvements in detecting companies at risk can be critical. We find that most improvements come from identifying troubled companies sooner, without significantly sacrificing the correct identification of healthy companies. As a final step in our analysis, the next section complements our findings with out-of-sample evidence.

3.3 Out-of-sample forecasting

Similar to the previous section, we split our sample into an in-sample (IS) training dataset that incorporates 80% of the data and an out-of-sample (OOS) test dataset that includes 20% of the data. We evaluate OOS performance using the classification metrics introduced above and summarize the results in Table 7 across different forecast horizons. For ease of comparison, we repeat the in-sample (IS) results from Table 6 along with the OOS results for each lag specification.

First, by inspecting Balanced Accuracy, we find that the employee model $\mathcal{M}^{\text{Emp}}$ achieves values in the OOS test between 0.64 and 0.69, which is substantially above the random model. More importantly, we find that BA lies between two and four percentage points above the traditional model $\mathcal{M}^{\text{Trad}}$ (except for $h = 1$). Compared with the IS results, these improvements are slightly better. Interestingly, the relative improvement in $\mathcal{M}^{\text{Emp}}$ compared to $\mathcal{M}^{\text{Trad}}$ increases with the length of the forecast horizon. Next, we consider AUC as the most important metric for our purposes. Again, the crucial quantity is the difference between $\mathcal{M}^{\text{Emp}}$ and $\mathcal{M}^{\text{Trad}}$, and here we find similar results compared to BA. In line with the pattern for BA, we find an increasing performance with the length of the forecasting horizon. ΔAUC increases from +0.00 for $h = 1$, to +0.02 for $h = 2$, and to +0.04 for $h = 3$. Consequently, the improvement in the two- and three-year windows is three and six percentage points, respectively. We infer that *EMP* helps improve forecasting of insolvency, especially when the default indicator $Y_{i,t+h}$ spans a longer time (here: three years). With respect to Sensitivity (correct identification of insolvent companies), we find OOS results that are positive across all horizons (except $h = 1$) and again sizable with improvements between three and six percentage points. This represents only a slightly weaker outcome for $\mathcal{M}^{\text{Emp}}$, as in the IS test, where Sensitivity improvements are on average one percentage point better. It offers the interpretation that the employee model shows the highest relative improvement in detecting insolvent companies. Specificity, on the other hand, shows no or only a moderate improvement, ranging from one to two percentage points. Only $h \in \{1, 2, 3\}$ is with four percentage points considerably higher. Summing up, we find that the performance metrics from the OOS tests show somewhat weaker absolute values in BA and AUC, but which is to be expected given the smaller number of observations in the model as well as the general fact that OOS forecasting is arguably the most challenging test setting. Against this background, we consider it as even more reassuring, that the improvement of the employee model against $\mathcal{M}^{\text{Trad}}$ remains of similar magnitude as in the IS results. Note that we conducted this analysis under the assumption that financial variables are in reality only available with a significant delay, and therefore we estimated all specifications with financial variables lagged by one additional year. In the robustness checks section, we briefly discuss the results when we drop these assumptions.

Table 7
Forecasting – Out-of-Sample Prediction Metrics

		$Y_{i,t+h}$: Insolvency of Company											
Forecast Horizon		$h = 1$				$h = 2$				$h = 3$			
Sample	Model	IS		OOS		IS		OOS		IS		OOS	
		$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$
Base Failure Rate (Prevalence)		0.02	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.06	0.07	0.10	0.12
		0.98	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.94	0.93	0.88	0.88
No Information Rate (NIR)													
Accuracy		0.96	0.76	0.78	0.73	0.73	0.75	0.71	0.72	0.72	0.65	0.69	0.62
Balanced Accuracy (BA)		0.50	0.72	0.73	0.68	0.72	0.73	0.67	0.69	0.69	0.64	0.67	0.60
Δ BA ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)			0.01			0.01	0.02		0.02	0.03	0.04	0.03	0.04
AUC		0.49	0.77	0.77	0.69	0.76	0.77	0.66	0.68	0.73	0.67	0.76	0.59
Δ AUC ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)			0.00	0.00		0.01	0.02		0.02	0.03	0.04	0.03	0.06
Sensitivity (SEN)		0.02	0.68	0.68	0.63	0.62	0.68	0.71	0.65	0.66	0.68	0.71	0.57
Δ SEN ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)			0.00	0.00		-0.01	0.03		0.03	0.06	0.05	0.03	0.06
Specificity (SPEC)		0.98	0.76	0.78	0.74	0.75	0.76	0.76	0.73	0.72	0.65	0.74	0.62
Δ SPEC ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)			0.02			0.01	0.01		0.02	0.00	0.02	0.02	-0.01
Total Observations in Model		2079	1644	1644	1644	1644	1641	1641	1641	1640	1640	1641	1639

This table shows the forecast metrics for the traditional model $\mathcal{M}^{\text{Trad}}$ and the employee model $\mathcal{M}^{\text{Emp}}$. Comparison based on different time periods of the information set (EMP , FIN , CON variables). Models based on data in-sample (IS - 80% of companies) and out-of-sample (OOS - 20% of companies). Sample split is done by a Monte Carlo simulation with $n = 1,000$ splits. Lag $h = 1$ refers to a one-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 2$ refers to a two-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 3$ refers to a three-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h \in \{1, 2\}$ refers to a two-year forecast horizon. This model forecasts whether the company i defaults anytime within the next two years, latest at the default event t_0 . Lag $h \in \{1, 2, 3\}$ refers to a three-year forecast horizon. This model forecasts whether the company i defaults anytime within the next three years, latest at the default event t_0 . The number of observations corresponds to the number of distinct companies per year that have a (yearly) value in the model. The sample period ranges from January 2007 to December 2018.

4 Additional robustness tests

We subject our findings to a couple of additional robustness tests. First, to assess the impact of the imputation procedure, we perform our entire analysis on the original dataset without imputed values. Second, we examine the rather theoretical perspective that financial data are available in the year they refer to (rather than one year later). Third, we analyze the impact of the latent *EMP* variable. We conclude our robustness tests with a model focused on the largest sector in our sample. All regression tables are shown in Table 8. Forecasting results are displayed in Table 9.

4.1 Results without imputation

As noted above, we use imputation to reduce the number of missing values. We thereby use imputation via the k -nearest neighbors algorithm to predict the outcome of a new observation as the average of the five most similar observations. Thus, it is vital to check whether the imputation procedure is potentially responsible for overstating our results. We rerun our analysis without the imputed data points. Note that without imputation, the number of valid data points in the model decreases on average by more than 50%; for example, with $h = 1$, it drops from 1.650 to 646. First, we compare descriptive statistics before and after imputation and observe, as expected, a decrease in the standard deviation and effect size. We find that our qualitative results remain almost unchanged. We report results for the most stringent test the proper forecasting regression setup as in Section 3.2 in Panel A of Table 8, and we document the corresponding OOS results in Panel A of Table 9. The significance level for *Employee Rating* is comparable, and nearly all coefficients show the same or even a higher magnitude (e.g., for $h = 1$, the coefficient on *Employee Rating* is -0.400 before and -0.450 after imputation. Only results for $h \in \{1, 2, 3\}$ are insignificant and show a lower effect size. Most importantly, the goodness-of-fit measures show comparable increases relative to $\mathcal{M}^{\text{Trad}}$, or even stronger improvements, as in the case of Pseudo- R^2 . Here we see, for instance, a value of 0.11 ($h = 2$) or 0.17 ($h = 3$) before imputation compared to 0.09 in $h = 2$ or 0.07 in $h = 3$ after imputation. Similar observations can be made regarding AUC results. The AUC is 0.74 ($h = 3$) before imputation and 0.67 after imputation. The relative improvement between $\mathcal{M}^{\text{Emp}}$ and $\mathcal{M}^{\text{Trad}}$ is six percentage points before and four after imputation. The cost of having a lower goodness of fit (imputation reduces the average variability as can be seen in Table A.4 in the Appendix) is more than compensated for the benefits of having more data points in the regression. Note that the model without imputation has between $N = 646$ and $N = 714$ cases, which is below the suggested threshold (1.571 cases)³⁴, and could therefore lead to biased coefficients. Fortunately, our robustness checks can discard these concerns. Therefore, we conclude that the imputation process does not change any material results and is necessary to reach a certain minimum data sample size. We take these results as justification for replacing missing values through imputation for our main findings.

³⁴ Peduzzi et al. (1996) suggests for logistic regression $N = 10k/p$ as the minimum number of cases N . Here, p is the smallest proportion of negative or positive cases in the population and k is the number of independent variables. In our case, we have p on average at 0.07 (mean of insolvent ratings and companies) and $k = 11$ resulting in a minimum number of 1.571 cases.

Table 8
Robustness Checks – Forecasting Regression Results

Panel A: <i>EMP</i> Variables Without Imputation										
Variable Name	$h = 1$		$h = 2$		$h = 3$		$h \in \{1, 2\}$		$h \in \{1, 2, 3\}$	
	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$
Intercept	-3.091*** (0.576)	-2.159*** (0.584)	-3.282*** (0.639)	-2.667*** (0.702)	-3.550*** (0.611)	-3.849*** (0.559)	-3.163*** (0.601)	-2.640*** (0.705)	-3.018*** (0.482)	-3.527*** (0.549)
Employee Rating		-0.400*** (0.106)		-0.454*** (0.149)		-0.448* (0.242)		-0.380** (0.163)		-0.237 (0.197)
Controls	Included									
Nagelkerke's R^2	0.09	0.07	0.10	0.11	0.05	0.17	0.12	0.20	0.07	0.09
Δ NK R^2 ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)		-0.02		+0.01		+0.12		+0.08		+0.02
McFadden's R^2	0.07	0.06	0.08	0.09	0.04	0.14	0.10	0.15	0.05	0.07
Brier Score	0.056	0.066	0.061	0.066	0.059	0.055	0.081	0.081	0.099	0.083
AIC	702	347	723	349	704	316	905	406	1068	452
Total Observations in Model	1574	646	1549	679	1538	714	1549	679	1538	714
Panel B: <i>FIN</i> Variables Without Lags of One Year										
Variable Name	$h = 1$		$h = 2$		$h = 3$		$h \in \{1, 2\}$		$h \in \{1, 2, 3\}$	
	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$
Intercept	-3.649*** (0.575)	-3.656*** (0.554)	-3.048*** (0.561)	-3.053*** (0.544)	-3.370*** (0.646)	-3.504*** (0.535)	-2.651*** (0.440)	-2.675*** (0.446)	-2.678*** (0.422)	-2.790*** (0.402)
Employee Rating		-0.279*** (0.102)		-0.364*** (0.093)		-0.514*** (0.138)		-0.427*** (0.109)		-0.574*** (0.140)
Controls	Included									
Nagelkerke's R^2	0.13	0.15	0.05	0.07	0.06	0.10	0.07	0.11	0.07	0.13
Δ NK R^2 ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)		+0.02		+0.02		+0.04		+0.04		+0.06
McFadden's R^2	0.12	0.13	0.04	0.06	0.05	0.08	0.06	0.08	0.05	0.10
Brier Score	0.040	0.040	0.057	0.056	0.059	0.058	0.081	0.079	0.099	0.094
AIC	557	550	744	728	756	732	972	943	1136	1084
Total Observations in Model	1662	1662	1650	1650	1650	1650	1650	1650	1650	1650
Panel C: Latent Employee Rating as <i>EMP</i>										
Variable Name	$h = 1$		$h = 2$		$h = 3$		$h \in \{1, 2\}$		$h \in \{1, 2, 3\}$	
	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$
Intercept	-3.265*** (0.538)	-3.309*** (0.506)	-3.157*** (0.541)	-3.210*** (0.493)	-3.394*** (0.569)	-3.501*** (0.504)	-3.022*** (0.494)	-3.106*** (0.446)	-2.794*** (0.408)	-3.052*** (0.386)
Latent Employee Rating		-0.086*** (0.028)		-0.114*** (0.028)		-0.145*** (0.052)		-0.135*** (0.036)		-0.076* (0.044)
Controls	Included									
Nagelkerke's R^2	0.05	0.07	0.06	0.09	0.04	0.07	0.08	0.12	0.05	0.10
Δ NK R^2 ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)		+0.02		+0.03		+0.03		+0.04		+0.05
McFadden's R^2	0.04	0.05	0.05	0.07	0.03	0.06	0.06	0.09	0.03	0.08
Brier Score	0.057	0.056	0.059	0.058	0.055	0.054	0.080	0.078	0.094	0.090
AIC	742	732	755	737	724	704	957	927	1097	1049
Total Observations in Model	1650	1650	1650	1650	1646	1646	1650	1650	1646	1646

This table presents three sets of logistic regression models forecasting insolvency: (A) Without Imputation, (B) Without Lags of Financial Variables, and (C) With *Latent Employee Rating* as *EMP* variable. The models compare traditional models ($\mathcal{M}^{\text{Trad}}$) and employee models ($\mathcal{M}^{\text{Emp}}$). Comparison based on different time periods of the information set (*EMP*, *FIN*, *CON* variables). Models based on data in-sample (IS - 80% of companies) and out-of-sample (OOS - 20% of companies). Sample split is done by a Monte Carlo simulation with $n = 1,000$ splits. Lag $h = 1$ refers to a one year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 2$ refers to a two-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h \in \{1, 2\}$ refers to a two-year forecast horizon. This model forecasts whether the company i defaults anytime within the next two years, latest at the default event t_0 . Lag $h \in \{1, 2, 3\}$ refers to a three-year forecast horizon. This model forecasts whether the company i defaults anytime within the next three years, latest at the default event t_0 . All specifications include the set of financial and macroeconomic control variables. Cluster fixed-effects and the *Industry Risk* variable are included but not reported. Standard errors (clustered for sector and year) are in parentheses. The sample period spans January 2007 - December 2018. Log-odds as coefficients.

Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.2 *FIN* variables without lag

We assess the robustness of the results to the timing assumption for the financial variables. While our main results are derived from the perspective of a real-world decision-maker who will have access to financial information from annual reports only with a significant time delay, one may contrast these results with the (theoretical) perspective of the econometrician who takes the values as being given in the period to which they refer. As explained in Section 2.5, other studies have analyzed the topic and recommend including a one-year delay for the availability of financial accounts. Although we share the view that the time delay is a more realistic assumption, we drop the time lag of the financial variables in this robustness check and re-specify the model equation accordingly:

$$\text{logit } Y_{i,t} = \beta_0 + \beta_1 EMP_{i,t} + \beta_2 FIN_{i,t} + \beta_3 CON_{i,t} + \delta_{j,i} + \gamma_{k,i} + \varepsilon_{i,t}, \quad (3)$$

where $Y_{i,t}$ is the default indicator variable for company i in sector j at time t . EMP , FIN , and CON are vectors summarizing our employee rating variables, the financial accounting data, and additional control variables, respectively. We include sector fixed-effects $\delta_{j,i}$ for sector j and an industry risk variable $\gamma_{k,i}$ that captures the market riskiness k of the industry to which company i belongs.

We report results in Panel B of Table 8, and the corresponding OOS results in Panel B of Table 9.

We find qualitatively unchanged results, where all EMP variables remain significant and comparable in effect size once we exclude the time-lag for FIN variables. As expected, Nagelkerke's R^2 in the $\mathcal{M}^{\text{Emp}}$ without time lag increases in all models (except $h = 2$). Not surprisingly, the largest increase with 0.09 is in $h = 1$, but also for $h = 2$, we find a R^2 of 0.07 in the $\mathcal{M}^{\text{Emp}}$ model with time-lag as compared to 0.10 without time-lag. Consequently, the AIC decreases from 733 to 557 ($h = 1$). We also calculated the forecasting model and obtained the following results. The relative improvement of BA and AUC ($\mathcal{M}^{\text{Emp}}$ vs. $\mathcal{M}^{\text{Trad}}$) of the model without lag is with one to four percentage points, comparable to the $\mathcal{M}^{\text{Emp}}$ model with time lag. However, taking into account the absolute magnitude of BA and AUC, the performance is slightly better once we take out the one lag for FIN variables (on average, a two percentage points increase). For example, $h = 1$ shows an OOS BA value of 0.75 without a one-year delay compared to the OOS BA value of 0.68. This is no surprise to us, as the practitioner's view (= later data availability) is the more conservative approach. We infer that some of the information from earlier available theoretical financial data is also represented in EMP variables. One key advantage is the substantially better (more than three percentage points) Sensitivity of $\mathcal{M}^{\text{Emp}}$ models starting from $h = 2$. When taking out the one-year delay for the FIN variables, this positive impact shifts one year and starts the first time in $h = 3$. Put differently, within the two-year forecasting horizon, FIN variables show good performance in identifying insolvent companies, and EMP variables help improve the models from two years onward. In sum, although, as suggested by intuition, we find that the FIN variables gain in explanatory importance without the time lag, this does not invalidate the qualitative results for the EMP variables. Therefore, since the availability of financial information in real time is practically unrealistic, we consider our main results the relevant ones and conclude that it is meaningful to maintain the assumption of a time lag for the FIN variables.

Table 9
Robustness Checks – Out-of-Sample Prediction Metrics

Forecast Horizon		Y_{t+h} : Insolvency of Company																			
		$h = 1$				$h = 2$				$h = 3$				$h \in \{1, 2\}$				$h \in \{1, 2, 3\}$			
Sample		IS		OOS		IS		OOS		IS		OOS		IS		OOS		IS		OOS	
Model		$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$
Without imputation of EMP variables																					
Base Failure Rate (Prevalence)		0.07	0.09	0.07	0.09	0.08	0.08	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.04	0.06	0.04	0.04
Accuracy		0.77	0.73	0.74	0.67	0.73	0.73	0.69	0.69	0.70	0.67	0.68	0.71	0.56	0.68	0.71	0.56	0.64	0.77	0.53	0.49
Balanced Accuracy (BA)		0.73	0.76	0.70	0.69	0.73	0.77	0.70	0.72	0.71	0.68	0.66	0.69	0.55	0.66	0.69	0.55	0.63	0.64	0.71	0.54
$\Delta \text{BA } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		0.03	0.03	-0.01	-0.01	0.04	0.02	0.02	0.07	0.06	0.06	0.03	0.03	0.08	0.08	0.07	0.08	0.07	0.07	0.01	0.01
AUC		0.79	0.81	0.72	0.68	0.79	0.82	0.71	0.77	0.85	0.68	0.69	0.73	0.56	0.69	0.71	0.56	0.63	0.67	0.71	0.60
$\Delta \text{AUC } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		0.02	0.02	-0.04	-0.04	0.03	0.00	0.08	0.08	0.06	0.06	0.04	0.04	0.07	0.63	0.67	0.54	0.61	0.65	0.55	0.61
Sensitivity (SEN)		0.68	0.80	0.66	0.72	0.74	0.82	0.68	0.77	0.73	0.82	0.63	0.67	0.54	0.63	0.67	0.54	0.61	0.65	0.65	0.61
$\Delta \text{SEN } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		0.12	0.12	0.06	0.06	0.08	0.09	0.09	0.09	0.09	0.09	0.04	0.04	0.07	0.06	0.07	0.06	0.00	0.00	0.06	0.06
Specificity (SPEC)		0.78	0.72	0.75	0.67	0.73	0.72	0.71	0.67	0.69	0.75	0.67	0.70	0.56	0.69	0.71	0.56	0.66	0.63	0.77	0.53
$\Delta \text{SPEC } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		-0.06	-0.06	-0.08	-0.08	-0.01	-0.04	0.06	0.06	0.06	0.03	0.03	0.02	0.10	0.02	0.02	0.10	0.14	0.07	0.14	-0.04
Total Observations in Model		1563	621	1563	621	1539	654	1532	692	1532	692	1539	1539	1539	1539	1539	1539	1566	775	1566	775
Without one year lag of F/N variables																					
Base Failure Rate (Prevalence)		0.05	0.05	0.05	0.05	0.07	0.07	0.07	0.07	0.07	0.07	0.10	0.10	0.11	0.10	0.10	0.11	0.13	0.13	0.13	0.13
Accuracy		0.81	0.81	0.78	0.79	0.76	0.76	0.75	0.75	0.75	0.70	0.76	0.78	0.73	0.76	0.78	0.73	0.75	0.73	0.75	0.68
Balanced Accuracy (BA)		0.77	0.78	0.74	0.75	0.72	0.73	0.68	0.69	0.72	0.74	0.67	0.73	0.68	0.70	0.74	0.68	0.70	0.70	0.72	0.64
$\Delta \text{BA } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.04
AUC		0.83	0.83	0.76	0.77	0.77	0.78	0.68	0.70	0.76	0.66	0.70	0.78	0.68	0.70	0.79	0.68	0.71	0.75	0.77	0.64
$\Delta \text{AUC } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.04	0.04	0.01	0.03	0.03	0.01	0.03	0.02	0.02	0.05	0.05
Sensitivity (SEN)		0.74	0.76	0.70	0.71	0.68	0.69	0.62	0.62	0.68	0.73	0.63	0.69	0.62	0.66	0.68	0.62	0.62	0.65	0.69	0.64
$\Delta \text{SEN } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.05	0.03	0.03	-0.01	0.00	0.03	-0.01	0.00	0.00	0.04	0.04	0.05
Specificity (SPEC)		0.81	0.81	0.79	0.79	0.77	0.77	0.74	0.76	0.75	0.71	0.73	0.77	0.74	0.73	0.79	0.74	0.77	0.74	0.76	0.72
$\Delta \text{SPEC } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.02	0.02	0.02	0.03	0.02	0.02	0.03	0.02	0.02	0.02	0.02
Total Observations in Model		1648	1648	1648	1648	1644	1644	1644	1641	1641	1641	1644	1644	1644	1641	1644	1644	1641	1641	1641	1641
With latent EMP variable																					
Base Failure Rate (Prevalence)		0.07	0.07	0.07	0.07	0.07	0.07	0.06	0.06	0.07	0.07	0.10	0.10	0.10	0.10	0.10	0.10	0.12	0.12	0.12	0.12
Accuracy		0.76	0.78	0.73	0.74	0.75	0.75	0.72	0.72	0.65	0.66	0.74	0.75	0.69	0.72	0.75	0.69	0.72	0.69	0.70	0.62
Balanced Accuracy (BA)		0.72	0.73	0.68	0.68	0.72	0.73	0.69	0.69	0.64	0.68	0.71	0.73	0.66	0.69	0.73	0.66	0.69	0.67	0.70	0.60
$\Delta \text{BA } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		0.01	0.01	0.00	0.00	0.01	0.01	0.02	0.03	0.04	0.04	0.02	0.02	0.03	0.03	0.02	0.03	0.03	0.03	0.03	0.06
AUC		0.77	0.78	0.69	0.69	0.76	0.77	0.66	0.68	0.73	0.76	0.68	0.76	0.66	0.68	0.78	0.66	0.69	0.71	0.76	0.59
$\Delta \text{AUC } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		0.01	0.01	0.00	0.00	0.01	0.01	0.02	0.03	0.03	0.05	0.02	0.02	0.03	0.03	0.02	0.03	0.03	0.05	0.05	0.08
Sensitivity (SEN)		0.69	0.68	0.63	0.61	0.68	0.71	0.62	0.65	0.66	0.73	0.68	0.71	0.63	0.69	0.71	0.63	0.65	0.64	0.70	0.57
$\Delta \text{SEN } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		-0.01	-0.01	-0.02	-0.02	0.03	0.03	0.03	0.03	0.07	0.06	0.06	0.03	0.02	0.06	0.03	0.02	0.02	0.06	0.06	0.09
Specificity (SPEC)		0.76	0.78	0.74	0.75	0.76	0.75	0.71	0.73	0.72	0.71	0.66	0.74	0.70	0.73	0.76	0.70	0.72	0.70	0.70	0.62
$\Delta \text{SPEC } (\mathcal{M}^{\text{Trad}} \text{ vs. } \mathcal{M}^{\text{Emp}})$		0.02	0.02	0.01	0.01	-0.01	-0.01	0.02	0.02	-0.01	0.01	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.00	0.00	0.03
Total Observations in Model		1644	1644	1644	1644	1641	1641	1641	1639	1639	1639	1641	1641	1641	1639	1641	1641	1639	1639	1639	1639

This table shows the robustness models forecast metrics for $\mathcal{M}^{\text{Random}}$ the traditional model $\mathcal{M}^{\text{Trad}}$ and the employee model $\mathcal{M}^{\text{Emp}}$. Comparison based on different time periods of the information set (*EMP*, *F/N*, *CON* variables). Models based on data in-sample (IS - 80% of companies) and out-of-sample (OOS - 20% of companies). Sample split is done by a Monte Carlo Simulation with $n = 1,000$ splits. Lag $h = 1$ refers to a one-year forecast horizon. The default indicator $Y_{t,h}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 2$ refers to a two-year forecast horizon. The default indicator $Y_{t,h}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 3$ refers to a three-year forecast horizon. The default indicator $Y_{t,h}$ is the insolvency of the company for the year at which the default events occurs. Lag $h \in \{1, 2\}$ refers to a two-year forecast horizon. This model forecasts whether the company i defaults anytime within the next two years, latest at the default event t_0 . Lag $h \in \{1, 2, 3\}$ refers to a three-year forecast horizon. This model forecasts whether the company i defaults anytime within the next three years, latest at the default event t_0 . The number of observations corresponds to the number of distinct companies per year that have a (yearly) value in the model. The sample period ranges from January 2007 to December 2018.

4.3 Latent *EMP* variable

In Section 3.1, we introduced the *Latent Employee Rating* variable as the first principal component of the generic employee information. We demonstrated the strong discriminatory power of *Latent Employee Rating* in the baseline logistic regression, but restricted ourselves to generic employee variables in the subsequent analysis. In this robustness section, we also briefly summarize the latent variable’s forecasting performance. We report the corresponding results in Panels C of Tables 8 and 9.

We obtain, both in-sample and out-of-sample, qualitatively similar, and even stronger, results than with the generic information. Forecasting metrics are comparable to our original model and even stronger. As shown in Table 9, the model with a latent *EMP* variable in period $h \in \{1, 2, 3\}$ achieves the highest relative improvement compared to all other robustness models. AUC increases by eight percentage points from $\mathcal{M}^{\text{Trad}}$ to $\mathcal{M}^{\text{Emp}}$. Overall, we conclude that the extraction of a latent variable tends to deliver even stronger support for our main results. Since the *Latent Employee Rating* variable is not readily observable, we nevertheless hesitated to base our most important findings on it, but we consider it a reassuring confirmation of our results.

In summary, our results are robust across respect to different checks. With an improvement of the $\mathcal{M}^{\text{Emp}}$ versus the $\mathcal{M}^{\text{Trad}}$ model of up to eight percentage points in BA, we find in all robustness tests very positive forecasting capabilities. This means that, especially from the long-term view, *EMP* variables help improve forecasting, which concludes that employees possess valuable intangible information and may observe a distress situation in companies much sooner than it is reflected in financial key performance indicators.

4.4 Further robustness tests

We have performed a couple of additional robustness checks, from which we report a sector effect here. As pointed out in the introductory section 2.1, we consider it to be crucially important to control for potential peer-group effects. We have already addressed this concern through our unique sample construction and by controlling for cluster effects with sector fixed-effect dummies. To strengthen this line of reasoning, we rerun our analysis on a restricted sub-sample of companies from a particular sector. As our sample is concentrated in the manufacturing sector (287 companies out of 380 are in the manufacturing sector, see Table S.4 in the Supplementary Material for further details), we choose this sector to achieve a sufficient sample size. The models show that we can replicate the same findings also within a particular sector. We consider this as relevant information and report the corresponding results in Table A.6 and Table A.7 in the Appendix. Coefficients for *EMP* variables show the same magnitude, although with somewhat lower significance levels (which, again, is likely due to the smaller sample size). We therefore regard these results as strong confirmation of our earlier findings.

5 Conclusions

This study investigates whether crowd-sourced employee ratings can enhance the prediction of company default risk, complementing traditional financial and macroeconomic indicators. While prior research has linked employee sentiment to company performance, stock returns, and governance outcomes, its role in predicting business failure has remained unexplored. Our analysis addresses this gap by systematically assessing the predictive value of employee ratings for corporate insolvency.

Methodologically, we construct a peer-group-matched dataset by leveraging employee ratings from KUNUNU and mandatory financial distress reports for companies operating in Germany, Austria, and Switzerland. We employ logistic regression models that integrate traditional financial risk factors with employee sentiment metrics, and evaluate both in-sample classification accuracy and out-of-sample forecasting performance. To mitigate potential endogeneity concerns, we use peer-group matching strategy, ensuring a robust comparison between companies that eventually default and their non-distressed counterparts.

Our results demonstrate that employee ratings provide significant incremental predictive power. Companies with lower employee ratings have a higher probability of default, with statistically significant differences observable up to twelve years before insolvency. We observe a Balanced Accuracy increase of at least three percentage points and improve insolvency prediction with an AUC of up to 0.81. The forecast results show an out-of-sample AUC of 0.69 (in robustness models of 0.77) and a relative improvement in the $\mathcal{M}^{\text{Emp}}$ model for a three-year horizon of three percentage points or more. These findings have several important implications. Three observations are particularly noteworthy in this context: First, *EMP* variables improve long-term forecasting and thus complement classical *FIN* variables, which perform well on shorter horizons. Second, once we extend $Y_{i,t+h}$ from one up to three years the results improve considerably, suggesting that *EMP* variables capture more intangible knowledge from employees and translate it into relevant forecasting information far ahead of an actual insolvency event. Third, *EMP* variables help to improve the forecasting of true positives the most: these are companies which go insolvent and thus possess the highest economic risk.

From a risk management perspective, our results suggest that credit rating agencies, investors, and financial institutions could integrate employee sentiment data into their credit risk models to improve early detection of financial distress. From a corporate governance standpoint, boards and executives should recognize that employee perceptions reflect organizational health and can influence external assessments of company stability.

Despite these contributions, our study has limitations. First, the dataset is regionally focused on German-speaking countries, and its applicability to other markets remains to be tested. Second, while we control for endogeneity using peer-group matching, unobserved company-specific factors may still influence both employee ratings and financial distress. Third, employee ratings may be subject to selection bias, as dissatisfied employees might be more likely to leave reviews. Future research could explore alternative data sources, such as text-based sentiment analysis of employee comments or additional labor market indicators, to refine and expand upon our findings.

Overall, our study provides compelling evidence that employee sentiment serves as a forward-looking predictor of corporate distress, reinforcing the growing relevance of alternative data in financial risk assessment and corporate decision-making.

References

- Abdulsalam, K., Christensen, D. M., Graffin, S. D., and Li, J. (2025). Do Boards Reward and Punish CEOs Based on Employee Satisfaction Ratings? *Organization Science*, 36(2):881–902.
- Agarwal, V. and Taffler, R. J. (2007). Twenty-five years of the Taffler z-score model: Does it really have predictive ability? *Accounting and Business Research*, 37(4):285–300.
- Altman, E. I., Iwanicz-Drozdzowska, M., Laitinen, E. K., and Suvas, A. (2016). Financial and non-financial variables as long-horizon predictors of bankruptcy. *Journal of Credit Risk*, 12(4):1–30.
- Altman, E. I., Sabato, G., and Wilson, N. (2010). The value of non-financial information in SME risk management. *Journal of Credit Risk*, 6(2):1–33.
- Beal, D. J. and Dawson, J. F. (2007). On the use of Likert-type scales in multilevel data: Influence on aggregate variables. *Organizational Research Methods*, 10(4):657–672.
- Berg, T., Burg, V., Gombović, A., and Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. *Review of Financial Studies*, 33(7):2845–2897.
- Bonfim, D. (2009). Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics. *Journal of Banking & Finance*, 33(2):281–299.
- Campbell, D. W. and Shang, R. (2022). Tone at the Bottom: Measuring Corporate Misconduct Risk from the Text of Employee Reviews. *Management Science*, 68(9):7034–7053.
- Chemmanur, T. J., Rajaiya, H., and Sheng, J. (2019). How does soft information affect external firm financing? Evidence from online employee ratings. *Available at SSRN 3507695*.
- Chi, W. and Chen, Y. (2021). Employee satisfaction and the cost of corporate borrowing. *Finance Research Letters*, 40:101666.
- Dube, S. and Zhu, C. (2021). The disciplinary effect of social media: Evidence from firms’ responses to Glassdoor reviews. *Journal of Accounting Research*, 59(5):1783–1825.
- Farhadi, R. and Nanda, V. (2021). What do employees know? quality perception and ‘over-satisfaction’ in firms going public. *Journal of Corporate Finance*, 66:101779.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.
- Gambacorta, L., Huang, Y., Li, Z., Qiu, H., and Chen, S. (2023). Data versus collateral. *Review of Finance*, 27(2):369–398.
- Gornall, W., Gredil, O. R., Howell, S. T., Liu, X., and Sockin, J. (2025). Do employees cheer for private equity? the heterogeneous effects of buyouts on job quality. *Management Science*, 71(5):4287–4317.
- Green, T. C., Huang, R., Wen, Q., and Zhou, D. (2019). Crowdsourced employer reviews and stock returns. *Journal of Financial Economics*, 134(1):236–251.
- Grunert, J., Norden, L., and Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance*, 29(2):509–531.
- Guiso, L., Sapienza, P., and Zingales, L. (2015). The value of corporate culture. *Journal of Financial Economics*, 117(1):60–76.
- Hillegeist, S. A., Keating, E. K., Cram, D. P., and Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1):5–34.
- Huang, K., Li, M., and Markov, S. (2020). What do employees know? Evidence from a social media platform. *The Accounting Review*, 95(2):199–226.
- Huang, M., Li, P., Meschke, F., and Guthrie, J. P. (2015). Family firms, employee satisfaction, and corporate performance. *Journal of Corporate Finance*, 34:108–127.

- Jagtiani, J. and Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management*, 48(4):1009–1029.
- Jing, C., Keasey, K., Lim, I., and Xu, B. (2019). Financial constraints and employee satisfaction. *Economics Letters*, 183:108599.
- Kallberg, J. G. and Udell, G. F. (2003). The value of private sector business credit information sharing: The US case. *Journal of Banking & Finance*, 27(3):449–469.
- Keasey, K. and Watson, R. (1987). Non-financial symptoms and the prediction of small company failure: A test of Argenti’s hypotheses. *Journal of Business Finance & Accounting*, 14(3):335–354.
- Keasey, K. and Watson, R. (1991). Financial distress prediction models: A review of their usefulness. *British Journal of Management*, 2(2):89–102.
- Lawrence, E. C. (1983). Reporting delays for failed firms. *Journal of Accounting Research*, 21(2):606–610.
- Lee, Y., Ng, S., Shevlin, T., and Venkat, A. (2021). The effects of tax avoidance news on employee perceptions of managers and firms: Evidence from Glassdoor.com ratings. *The Accounting Review*, 96(3):343–372.
- Li, H., Yu, B. X., Li, G., and Gao, H. (2023). Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews. *Tourism Management*, 96:104707.
- Li, J. (2022). The effect of employee satisfaction on effective corporate tax planning: Evidence from Glassdoor. *Advances in Accounting*, 57:100597.
- Li, M. and Pinto, J. (2025). How Does Going Public Affect Employee Satisfaction? Evidence from Glassdoor Reviews. *Management Science*, in print.
- Lussier, R. N. (1995). A nonfinancial business success versus failure prediction model for young firms. *Journal of Small Business Management*, 1(33):8–24.
- Morita, K., Mizuno, T., and Kusuhara, H. (2022). Investigation of a data split strategy involving the time axis in adverse event prediction using machine learning. *Journal of Chemical Information and Modeling*, 62(17):3982–3992.
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692.
- Naumzik, C., Feuerriegel, S., and Weinmann, M. (2022). I will survive: Predicting business failures from customer ratings. *Marketing Science*, 41(1):188–207.
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1):109–131.
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., and Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12):1373–1379.
- Platt, H. D. and Platt, M. B. (1990). Development of a class of stable predictive variables: The case of bankruptcy prediction. *Journal of Business Finance & Accounting*, 17(1):31–51.
- Saerens, M., Latinne, P., and Decaestecker, C. (2002). Adjusting the outputs of a classifier to new a priori probabilities: a simple procedure. *Neural computation*, 14(1):21–41.
- Sheng, J. (2025). Asset pricing in the information age: Employee expectations and stock returns. *The Review of Asset Pricing Studies*, 15(1):74–101.
- Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1):101–124.
- Symitsi, E., Stamolampros, P., and Daskalakis, G. (2018). Employees’ online reviews and equity prices. *Economics Letters*, 162:53–55.

- Symitsi, E., Stamolampros, P., Daskalakis, G., and Korfiatis, N. (2021). The informational value of employee online reviews. *European Journal of Operational Research*, 288(2):605–619.
- Tao, J. and Zhou, L. (2023). Can online consumer reviews signal restaurant closure: A deep learning-based time-series analysis. *IEEE Transactions on Engineering Management*, 70(3):834–848.
- Thölke, P., Mantilla-Ramos, Y.-J., Abdelhedi, H., Maschke, C., Dehgan, A., Harel, Y., Kemtur, A., Mekki Berrada, L., Sahraoui, M., Young, T., Bellemare Pépin, A., El Khantour, C., Landry, M., Pascarella, A., Hadid, V., Combrisson, E., O’Byrne, J., and Jerbi, K. (2023). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage*, 277:120253.
- Wang, D., Zhu, Q., Avolio, B. J., Shen, W., and Waldman, D. (2023). Do employees’ views matter in corporate governance? The relationship between employee approval and CEO dismissal. *Strategic Management Journal*, 44(5):1328–1354.
- Whittred, G. and Zimmer, I. (1984). Timeliness of financial reporting and financial distress. *The Accounting Review*, 59(2):287–295.
- Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22:59–82.

Appendix

Table A.1

Description of Variables Used in the Study

Variable	Description	Transformation / Standardization	Source
Employee Rating Variables			
Employee Rating	Average employee rating based on 13 subcategories (1 = very bad, 5 = very good)	Scaled and k(5)-nearest neighbors imputation	New variable
Employee Benefit Score	Number of employee benefits (e.g., free parking), ranging from 0 to 16	Scaled and k(5)-nearest neighbors imputation	New variable
Financial Variables			
Net Income to Assets	One-year lagged net income divided by lagged assets (percentage)	Scaled	Grunert et al. (2005)
Working Capital to Assets	One-year lagged working capital (receivables + stock - payables) divided by lagged assets (percentage)	Scaled	Altman et al. (2010)
Equity to Assets	One-year lagged equity (book value) divided by lagged assets (percentage)	Scaled	Grunert et al. (2005)
Debt to Equity	One-year lagged debt divided by lagged equity (book value) (percentage)	Scaled	Altman et al. (2010)
Sales to Assets	One-year lagged sales divided by lagged assets (percentage)	Scaled	Altman et al. (2010)
Macroeconomic Variables			
Δ Unemployment Rate (UR)	Yearly change in the German labor market unemployment rate (percentage)	HP-Prescott Filter and scaled	Bonfim (2009)
Control Variables			
Age	Log of the company age as a proxy for maturity risk	Logarithm and scaled	Altman et al. (2010)
Sector	Industry classification: 'Manufacturing', 'Energy', 'Logistics', 'Trade', 'Service', and 'Construction'	Categorical	Altman et al. (2016)
Industry Risk	Five-scale risk classification: 'average risk', 'increased industry risk', 'temporal industry decline', 'shrinking industry', and 'industry disruption'	Categorical	—

Note: This table provides an overview of key variables, including employee rating variables, financial indicators, macroeconomic factors, and control variables. These variables are utilized in models $\mathcal{M}^{\text{Trad}}$ and $\mathcal{M}^{\text{Emp}}$ to examine their explanatory power.

Table A.2
Explanation of KUNUNU Rating Criteria

Rating Criteria	Explanation: Questions on KUNUNU
Management Behavior	How does management behave in conflict situations? Does leadership set realistic expectations, communicate clear goals, and involve employees in decision-making?
Internal Communication	How effective is internal communication? Are you informed about company results, successes, and challenges through regular updates?
Work Atmosphere	What is the overall work atmosphere and corporate culture like? Does management give positive feedback? Does the company foster a fair and trustful work environment?
Company Image	How is the company's image? Do you speak positively about your employer? Does the company's external image align with reality?
Overall Compensation	Do you feel fairly compensated for your work? Does your salary reflect your responsibilities? Does the company offer social benefits, and does it pay salaries on time?
Career Development	What are your career growth and professional development prospects? Are promotion criteria transparent? Are employees supported with training?
Older Colleagues	Does the company hire and support older workers? Are senior colleagues given equal opportunities and appreciation?
Challenging Work	How challenging is your work? Is the workload fairly distributed? Do you have autonomy in organizing your responsibilities?
Environmental Friendliness	To what extent does the company demonstrate environmental awareness? Does it support sustainable practices, such as fair trade?
Gender Equality	Are women treated equally and provided the same career opportunities? Do women re-turning from maternity leave have equal chances for advancement?
Office/Work Environment	Is the work environment comfortable and suitable for your tasks? Are ventilation, lighting, temperature control, and technology adequate?
Teamwork	How well do co-workers collaborate? Is there honest and direct communication?
Work-Life Balance	How does the company support work-life balance? Are family needs considered? Is there pressure to work long hours? Can employees take a vacation freely?
Employee Benefits	
Benefits	What benefits does the company offer? Examples range from free parking to remote work options.

Note: This table provides a detailed explanation of the KUNUNU rating criteria. The German version of the criteria differs slightly from the U.S. version in selected categories.

Table A.3Definition of the *Industry Risk* Variable

Industry Type	Explanation	Relevance of Porter's Five Forces (I)	Development of Market Metrics (II)	Example Industries
Average Industry Risk (1)	No significant risk factors present in the industry.	Average competition.	Average growth and profitability.	Food, regional energy suppliers.
Increased Industry Risk (2)	Industry exhibits heightened risk.	Higher competition intensity and/or rivalry with new players and new distribution channels.	Saturated market (slow or no growth), lower profitability due to price or volume shifts.	Energy suppliers, hospitals, chemicals.
Temporal Industry Decline (3)	Industry experiences temporary decline due to economic recessions or industry cycles.	Increased industry rivalry, high bargaining power of suppliers and buyers.	Temporary negative growth, decreasing profitability.	Automotive, wood, plant construction.
Shrinking Industry (4)	Industry faces overcapacity and/or constant price and volume decline.	High industry rivalry.	Negative market growth and decreasing profitability.	Shipping, solar manufacturing, shoe retail.
Industry Disruption (5)	Industry is undergoing significant structural changes.	Substitutes, changing customer needs, and innovations.	Negative market growth and profitability.	CD/DVD manufacturing.

Note: This table provides an overview of the *Industry Risk* variable's levels, including explanations and industry examples.

The methodology for assigning companies to an industry risk level follows a three-step approach: (I) Porter's Five Forces framework assesses industry attractiveness by analyzing five competitive factors: bargaining power of suppliers, bargaining power of buyers, threats of new entrants, threat of substitutes, and overall industry rivalry. (II) Market size, growth, and profitability metrics are evaluated based on restructuring reports. (III) Findings from management interviews in restructuring reports are used to validate the industry risk classification.

Table A.4
Descriptive Statistics Before and After Imputation

Variable	Healthy Firms			Insolvent Firms		
	Mean	SD	SE	Mean	SD	SE
Employee Rating Before Imputation						
Employee Rating ^(a)	3.10	0.873	0.034	2.69	0.963	0.069
E. Benefit Score ^(b)	5.52	3.329	0.129	3.83	2.648	0.190
Older Colleagues	3.42	1.042	0.041	3.13	1.040	0.076
Management Behavior	2.88	1.077	0.042	2.42	1.191	0.085
Career Development	2.79	1.081	0.042	2.36	1.124	0.081
Challenging Work	3.36	0.969	0.038	3.02	1.080	0.077
Company Image	3.13	1.077	0.042	2.58	1.168	0.084
Environmental Friendliness	3.18	1.034	0.040	2.73	1.145	0.082
Gender Equality	3.24	1.022	0.040	2.95	1.132	0.082
Internal Communication	2.76	1.034	0.040	2.36	1.132	0.081
Work Atmosphere	3.01	1.085	0.042	2.59	1.179	0.084
Office/Work Environment	3.23	1.012	0.039	2.87	1.077	0.077
Overall Compensation	3.06	1.011	0.039	2.56	1.057	0.076
Teamwork	3.58	0.884	0.034	3.46	0.951	0.068
Work-Life Balance	2.99	1.043	0.041	2.55	1.120	0.081
Employee Rating After Imputation						
Employee Rating ^(a)	3.06	0.681	0.017	2.73	0.719	0.032
E. Benefit Score ^(b)	5.35	2.624	0.064	4.16	2.097	0.094
Older Colleagues	3.36	0.793	0.019	3.17	0.780	0.035
Management Behavior	2.85	0.832	0.020	2.43	0.891	0.040
Career Development	2.73	0.823	0.020	2.39	0.831	0.037
Challenging Work	3.34	0.749	0.018	3.03	0.818	0.037
Company Image	3.10	0.849	0.021	2.67	0.882	0.040
Environmental Friendliness	3.12	0.810	0.020	2.75	0.868	0.039
Gender Equality	3.20	0.796	0.019	2.98	0.837	0.038
Internal Communication	2.73	0.784	0.019	2.42	0.836	0.037
Work Atmosphere	2.96	0.848	0.021	2.62	0.880	0.039
Office/Work Environment	3.20	0.784	0.019	2.89	0.821	0.037
Overall Compensation	3.04	0.780	0.019	2.64	0.809	0.036
Teamwork	3.55	0.680	0.017	3.43	0.715	0.032
Work-Life Balance	2.96	0.817	0.020	2.61	0.829	0.037

Note: This table presents the mean, standard deviation (SD), and standard error (SE) of sample variables before and after imputation. Employee ratings range from 1 (very bad) to 5 (very good), except for the *Employee Benefit Score*, which represents the number of benefits. The sample period spans from January 2007 to December 2018. The table includes data only up to six years prior to insolvency to avoid dilution effects.

The imputation process aims to preserve key metrics. By comparing mean, SD, and SE values, no significant differences are observed, confirming that the imputation does not introduce bias.

^(a) *Employee Rating* is the average of 13 rating variables from KUNUNU.

^(b) *Employee Benefit Score* represents the number of benefits offered to employees, such as free parking.

Note: The table illustrates the effect of the 4-step matching process by comparing the standard deviation (=SD) of selected variables before (=Between) and after (= Within) the matching process. Statistics are the mean and SD of sample variables for the period 2007-2018, based on yearly averages (to balance companies with many and/or few ratings) and to match the financial (yearly) variables. All variables are ratio values - only age and average assets are shown in years and million euros, respectively. The sample variables include healthy and insolvent companies. The Wilcoxon test statistic determines whether the matching process leads to a significant reduction in SD, as an indicator for a successful matching process.

^(a) 'Between' refers to the sampled group by Sector. This represents the (highest) level of aggregation before the companies are matched.

^(b) 'Within' represents the KPIs after the 4-step matching process.

Table A.5Confusion Matrix for Three-Year Forecast Horizon (In-Sample) – $\mathcal{M}^{\text{Emp}}$ Model

	Observed Data		Total
Prediction	Healthy	Insolvent	
$\mathcal{M}^{\text{Emp}}$ Model Predictions			
Healthy	1106	28	1134
Insolvent	433	73	506
Total	1539	101	1640
No Information Rate ^(a)	94%	–	–
Base Failure Rate ^(b)	–	6%	–
Conditional Probability (Sensitivity) ^(c)	–	72%	–
Conditional Probability (Specificity) ^(d)	–	72%	–

Note: This table presents the confusion matrix for the $\mathcal{M}^{\text{Emp}}$ model's insolvency prediction over a three-year forecast horizon. The sample period spans from January 2007 to December 2018.

^(a) No Information Rate represents the proportion of the most frequent class: $\frac{1539}{1640} = 94\%$.

^(b) Base Failure Rate represents the proportion of insolvent companies: $\frac{101}{1640} = 6\%$.

^(c) Conditional Probability (Sensitivity) is the probability of correctly identifying insolvency: $\frac{73}{101} = 72\%$.

^(d) Conditional Probability (Specificity) is the probability of correctly identifying healthy firms: $\frac{1106}{1539} = 72\%$.

Table A.6**Robustness Logistic Regression Results for Forecasting (Sector Manufacturing)**

	Dependent Variable: Insolvency of Company									
	$h = 1$		$h = 2$		$h = 3$		$h = 4$		$h = 5$	
	Finance	Combined	Finance	Combined	Finance	Combined	Finance	Combined	Finance	Combined
Intercept	-2.787*** (0.215)	-2.859*** (0.242)	-2.639*** (0.207)	-2.726*** (0.246)	-2.699*** (0.242)	-2.841*** (0.295)	-2.317*** (0.175)	-2.433*** (0.225)	-2.017*** (0.240)	-2.163*** (0.262)
Employee Rating		-0.278*** (0.105)		-0.338*** (0.106)		-0.449** (0.217)		-0.407*** (0.139)		-0.479*** (0.178)
Net Income to Assets	-0.036 (0.114)	-0.025 (0.109)	0.063 (0.082)	0.073 (0.076)	0.030 (0.054)	0.051 (0.060)	0.051 (0.062)	0.065 (0.059)	0.061 (0.055)	0.088 (0.059)
Working Capital to Assets	-0.280** (0.136)	-0.274** (0.133)	-0.248 (0.153)	-0.242 (0.153)	-0.169 (0.147)	-0.160 (0.147)	-0.249* (0.151)	-0.244 (0.154)	-0.119 (0.130)	-0.108 (0.131)
Equity to Assets	-0.444*** (0.133)	-0.463*** (0.140)	-0.498*** (0.184)	-0.521*** (0.191)	-0.356** (0.146)	-0.389** (0.172)	-0.516*** (0.189)	-0.546*** (0.200)	-0.378** (0.179)	-0.418** (0.209)
Debt to Equity	-0.033 (0.072)	-0.025 (0.066)	-0.049 (0.055)	-0.041 (0.050)	-0.015 (0.041)	-0.021 (0.042)	-0.054 (0.035)	-0.049 (0.032)	0.006 (0.037)	0.007 (0.039)
Sales to Assets	-0.177* (0.095)	-0.185** (0.086)	-0.266** (0.118)	-0.275*** (0.106)	-0.159 (0.134)	-0.176 (0.121)	-0.215* (0.120)	-0.231** (0.110)	-0.034 (0.097)	-0.054 (0.096)
Δ Unemployment Rate (UR)	-0.399 (0.404)	-0.438 (0.437)	-0.063 (0.365)	-0.090 (0.387)	-0.017 (0.350)	-0.051 (0.371)	-0.103 (0.339)	-0.137 (0.359)	0.162 (0.363)	0.142 (0.373)
Log Age	-0.259* (0.152)	-0.253* (0.147)	-0.253* (0.150)	-0.243* (0.146)	-0.250* (0.144)	-0.252* (0.148)	-0.284** (0.134)	-0.276** (0.132)	-0.302*** (0.107)	-0.307*** (0.114)
Nagelkerke's R^2	0.05	0.06	0.07	0.09	0.02	0.05	0.08	0.11	0.04	0.08
McFadden's R^2	0.04	0.05	0.06	0.07	0.02	0.04	0.06	0.08	0.03	0.06
Brier Score	0.064	0.063	0.066	0.065	0.061	0.060	0.089	0.086	0.102	0.099
AIC	509	505	518	510	493	482	650	634	733	711
Total Observations in Model	1041	1041	1038	1038	1029	1029	1038	1038	1029	1029

This table presents the traditional model $\mathcal{M}^{\text{Trad}}$ and the employee models $\mathcal{M}^{\text{Emp}}$ with different forecasting horizons, i.e. we evaluate the forecasting of default events in t given the information sets of $t - h$ respectively. Comparison based on different time periods of the information set (EMP , FIN , CON variables). Models based on data in-sample (IS - 80% of companies). Sample split is done by a Monte Carlo Simulation with $n = 1,000$ splits. Lag $h = 1$ refers to a one-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 2$ refers to a two-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 3$ refers to a three-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 1 \rightarrow 2$ refers to a two-year forecast horizon. This model forecasts whether the company i defaults anytime within the next two years, latest at the default event t_0 . Lag $h = 1 \rightarrow 3$ refers to a three-year forecast horizon. This model forecasts whether the company i defaults anytime within the next three years, latest at the default event t_0 . Cluster fixed-effects and the *Industry Risk* variable are included but not reported. Standard errors (clustered for sector and year) are in parentheses. The sample period spans January 2007 - December 2018. Log-odds as coefficients.

Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7
Forecasting Results in the Training and Test Sample (Manufacturing Sector)

		$Y_{i,t+h}$: Insolvency of Company											
Forecast Horizon		$h = 1$				$h = 2$				$h = 3$			
Sample		IS		OOS		IS		OOS		IS		OOS	
Model		$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$	$\mathcal{M}^{\text{Trad}}$	$\mathcal{M}^{\text{Emp}}$
Base Failure Rate (Prevalence)		0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.07	0.07	0.11	0.11
No Information Rate (NIR)		0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.93	0.93	0.89	0.89
Accuracy		0.76	0.78	0.73	0.75	0.73	0.75	0.69	0.72	0.67	0.69	0.68	0.71
Balanced Accuracy (BA)		0.74	0.74	0.71	0.71	0.72	0.74	0.69	0.71	0.69	0.70	0.69	0.73
Δ BA ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)		0.00	0.00	0.00	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.01
AUC		0.77	0.77	0.70	0.70	0.76	0.77	0.68	0.70	0.72	0.73	0.68	0.70
Δ AUC ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)		0.00	0.00	0.00	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.03
Sensitivity (SEN)		0.71	0.70	0.69	0.67	0.71	0.72	0.69	0.70	0.71	0.70	0.70	0.67
Δ SEN ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)		-0.01	-0.02	-0.02	0.01	0.01	0.01	-0.01	0.00	-0.01	-0.01	-0.03	0.06
Specificity (SPEC)		0.76	0.78	0.74	0.75	0.73	0.76	0.69	0.72	0.66	0.69	0.72	0.70
Δ SPEC ($\mathcal{M}^{\text{Trad}}$ vs. $\mathcal{M}^{\text{Emp}}$)		0.02	0.01	0.01	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.05	-0.04
Total Observations in Model		1063	1063	1063	1063	1059	1059	1059	1059	1055	1055	1059	1055

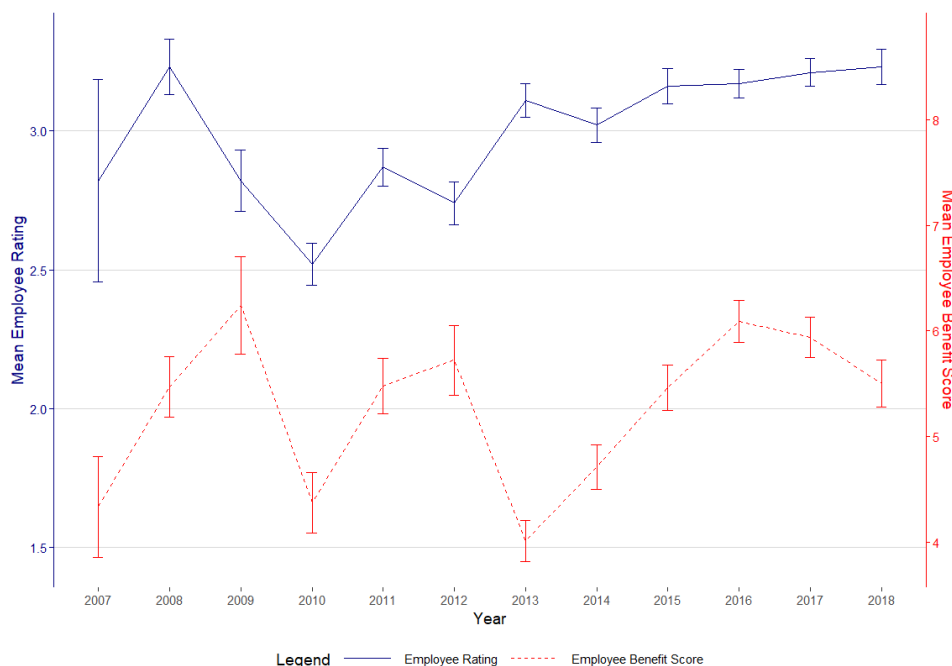
This table shows the forecast metrics for the random classifier $\mathcal{M}^{\text{random}}$, the traditional model $\mathcal{M}^{\text{Trad}}$ and the employee model $\mathcal{M}^{\text{Emp}}$. The model does only include companies from the manufacturing sector. Comparison based on different time periods of the information set (EMP , FIN , CON variables). Models based on data in-sample (IS - 80% of companies) and out-of-sample (OOS - 20% of companies). Sample split is done by a Monte Carlo Simulation with $n = 1,000$ splits. Lag $h = 1$ refers to a one year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 2$ refers to a two-year forecast horizon. The default indicator $Y_{i,t}$ is the insolvency of the company for the year at which the default events occurs. Lag $h = 1 \rightarrow 2$ refers to a two-year forecast horizon. This model forecast whether the company i defaults anytime within the next two years, latest at the default event t_0 . Lag $h = 1 \rightarrow 3$ refers to a three-year forecast horizon. This model forecasts whether the company i defaults anytime within the next three years, latest at the default event t_0 . The number of observations corresponds to the number of distinct companies per year that have a (yearly) value in the model. The sample period ranges from January 2007 to December 2018.

Supplementary Material to:

“Do employees foresee the looming iceberg ahead? Empirical evidence on the predictability of corporate default events.”

Figure S.1

Mean *Employee Rating* and *Employee Benefit Score* 2007-2018



Note: This figure shows the yearly averages of the *Employee Rating* (blue) and *Employee Benefit Score* variables before imputing missing values. Scale of *Employee Rating* (blue) ranges from 1 ('very bad') to 5 ('very good'). *Employee Benefit Score* (red) represents the number of benefits, whereas a benefit, for instance, is free parking for employees. The bars represent ± 1 standard error ('SE'). The horizontal axis values are years. The sample period ranges from January 2007 to December 2018. Data includes all values before and after the insolvency date ($t = 0$) as well as insolvent and healthy companies. Variable definition is explained in Table A.1 in the Appendix.

Table S.1
Number of *EMP* Ratings per Year and Industry

	Number of Companies												NN	Sum
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018		
Administration	2	27	19	32	29	33	43	40	36	44	47	28	0	380
Automotive	4	59	41	64	74	77	147	151	182	183	253	170	1	1406
Construction	0	6	4	4	4	3	20	13	21	26	30	23	0	154
Consumer Goods	1	20	28	36	61	48	115	109	123	167	227	150	0	1085
E-Commerce	0	3	7	9	18	23	32	33	35	28	38	24	0	250
Electrical Equipment	2	6	5	14	11	12	17	14	16	23	34	16	0	170
Energy	1	2	0	4	3	4	8	8	10	9	16	3	0	68
Energy Supply	0	4	3	9	7	3	23	31	24	38	44	21	0	207
Entertainment	0	2	2	10	8	2	8	5	8	15	21	18	0	99
Food	0	8	6	15	15	4	43	33	25	34	50	25	0	258
Glass	0	4	3	7	7	7	27	18	21	19	24	10	0	147
Healthcare	0	1	0	1	1	0	2	6	9	8	18	3	0	49
Industrial Goods	0	2	0	2	5	6	4	2	9	14	12	13	0	69
Information	2	17	13	17	25	12	42	37	37	43	44	32	0	321
Investigation	0	13	11	12	20	15	26	25	31	39	35	14	0	241
Machinery	0	17	8	20	33	19	71	59	68	80	107	38	0	520
Metals	1	14	10	19	22	25	63	54	58	84	86	37	0	473
Paper	0	0	0	0	0	0	1	2	2	4	5	2	0	16
Shipping	0	3	2	7	4	4	13	8	11	8	7	11	0	78
Solar	0	11	15	51	78	60	124	71	72	81	93	32	1	689
Tourism	0	1	0	0	1	0	0	3	1	3	4	3	0	16
Transportation	0	3	3	0	3	2	7	6	10	12	14	5	1	66
Wood	0	0	2	2	3	3	3	3	6	3	9	4	0	38
Sum	13	223	182	335	432	362	839	731	815	965	1218	682	3	6800

Note: The number of ratings represents the monthly mean to balance (big) companies with many ratings against (smaller) companies with few ratings. It can clearly be inferred that online employee ratings are strongly increasing. We expect this trend to be even stronger in the future. 2018 is only half a year. Original sample size in terms of number of ratings was 24,679; thereof 4,429 are excluded. The sample period ranges from January 2007 to December 2018. Variable definition is explained in Table A.1 in the Appendix.

Table S.2Yearly mean *EMP* Ratings per Industry per Year

	Yearly mean rating per industry											
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Administration	1.3	2.8	2.5	2.3	2.4	2.0	2.8	2.7	2.9	3.2	3.1	3.1
Automotive	3.2	3.1	3.1	2.5	2.9	2.9	3.2	3.2	3.2	3.3	3.3	3.1
Construction	NA	3.9	2.9	2.5	2.7	2.4	3.4	3.4	4.0	3.7	3.8	3.9
Consumer Goods	4.1	2.9	2.6	2.4	2.7	2.7	3.1	2.9	2.9	3.1	3.1	3.1
E-Commerce	NA	2.4	2.2	2.2	2.6	2.8	3.1	2.9	2.7	3.2	3.1	3.3
Electrical Equipment	1.3	2.9	2.4	2.4	2.8	2.0	2.9	2.8	3.5	2.5	3.1	3.4
Energy	4.5	3.5	NA	2.9	2.8	2.5	3.8	3.7	3.4	3.5	3.6	3.6
Energy Supply	NA	3.0	3.3	3.8	3.2	4.0	3.7	3.1	3.8	3.8	3.7	3.6
Entertainment	NA	1.5	1.8	2.6	2.0	2.3	2.6	3.5	2.7	3.0	3.3	3.2
Food	NA	3.0	2.3	2.6	2.3	1.7	2.5	2.5	2.6	2.8	3.1	3.4
Glass	NA	3.1	2.9	2.4	2.6	2.9	3.2	3.0	3.2	3.5	3.3	2.9
Healthcare	NA	4.2	NA	2.4	2.4	NA	2.8	3.6	3.5	3.2	2.7	3.9
Industrial Goods	NA	2.1	NA	1.9	2.3	2.3	3.3	2.5	2.6	2.9	3.6	3.4
Information	2.2	3.1	3.5	2.8	3.0	3.4	3.2	3.2	3.6	3.2	3.6	3.6
Investigation	NA	2.2	2.0	1.8	2.4	1.9	3.1	2.9	2.5	2.6	3.1	3.1
Machinery	NA	3.2	2.9	2.4	2.8	3.0	3.1	3.0	3.1	3.2	3.3	3.3
Metals	3.5	3.8	3.0	2.7	3.5	3.1	3.4	3.2	3.4	3.2	3.6	3.3
Paper	NA	NA	NA	NA	NA	NA	1.8	2.8	3.0	1.8	3.0	4.4
Shipping	NA	3.9	2.5	2.2	3.1	2.6	2.6	3.1	3.5	3.8	2.7	3.0
Solar	NA	2.8	3.3	2.5	3.0	2.9	3.1	3.2	3.3	3.4	3.6	3.5
Tourism	NA	3.2	NA	NA	3.0	NA	NA	3.6	4.8	3.4	2.8	4.4
Transportation	NA	2.9	2.7	NA	3.7	3.7	4.0	4.0	4.0	3.8	3.1	3.6
Wood	NA	NA	3.1	2.1	3.5	1.8	2.6	1.4	2.3	2.5	2.9	1.8
Total	2.7	3.0	2.8	2.5	2.8	2.7	3.1	3.0	3.1	3.2	3.3	3.3

Note: The number of ratings represents the monthly mean to balance (big) companies with many ratings against (smaller) companies with few ratings. The purpose of this table is to show the development of the average number of ratings. Especially the years 2007, 2009, and 2013 show lower averages compared to the other years in the sample period, which we attribute to the aftermath of the financial crisis. The original sample size, in terms of the number of ratings, was 24,679. The sample period ranges from January 2007 to December 2018.

Table S.3Yearly mean *EMP* Benefit Score per Industry per Year

	Yearly mean rating per industry											
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Administration	2.0	3.9	4.4	4.4	4.7	2.6	3.5	4.0	5.3	6.2	5.1	4.9
Automotive	5.8	6.0	8.2	5.0	6.1	7.0	4.8	5.9	5.7	6.6	6.7	5.4
Construction	NA	4.7	6.8	3.2	3.6	5.3	4.3	6.5	5.8	7.1	8.0	7.3
Consumer Goods	6.0	3.6	4.4	3.2	4.6	3.7	4.2	3.9	4.7	5.5	5.6	5.1
E-Commerce	NA	3.4	5.8	3.6	4.6	5.2	3.9	5.1	4.8	8.2	6.8	6.6
Electrical Equipment	2.0	6.1	5.4	6.0	6.3	3.0	3.5	3.9	7.5	5.4	5.6	7.1
Energy	3.0	6.5	NA	4.8	4.8	3.5	6.3	6.5	5.0	7.1	7.1	7.9
Energy Supply	NA	6.5	7.0	10.0	6.3	13.0	6.5	7.6	8.8	9.7	10.0	8.7
Entertainment	NA	1.0	3.5	3.4	1.9	3.5	1.5	2.5	4.4	5.5	5.5	5.0
Food	NA	4.1	4.9	4.7	3.1	2.9	2.9	3.5	4.0	4.9	3.6	3.9
Glass	NA	7.9	8.0	6.6	6.3	8.4	6.6	6.4	6.6	9.8	7.7	7.0
Healthcare	NA	8.0	NA	0.0	3.0	NA	6.5	4.2	6.2	5.6	4.4	6.7
Industrial Goods	NA	4.0	NA	1.0	4.0	4.2	3.0	4.0	4.6	5.4	7.8	5.7
Information	5.0	7.1	10.5	6.5	8.6	8.3	5.8	6.5	8.5	8.3	9.3	8.4
Investigation	NA	1.3	1.0	1.5	1.6	1.0	2.9	1.6	2.1	2.8	3.3	3.4
Machinery	NA	5.7	6.8	3.7	6.4	6.2	5.0	5.3	5.1	6.4	6.1	6.2
Metals	4.0	8.4	6.5	4.3	8.2	6.6	5.1	5.8	6.4	6.7	6.6	6.0
Paper	NA	NA	NA	NA	NA	NA	0.0	5.5	5.0	1.9	5.9	10.5
Shipping	NA	1.7	4.5	2.1	5.3	2.6	3.9	2.9	6.9	5.6	5.6	5.4
Solar	NA	4.7	9.2	4.4	7.2	7.3	4.6	6.6	6.3	7.9	8.3	6.6
Tourism	NA	8.5	NA	NA	3.0	NA	NA	2.0	7.0	7.7	5.2	8.3
Transportation	NA	7.3	8.7	NA	8.3	13.0	8.0	9.2	7.2	11.3	7.6	8.4
Wood	NA	NA	5.0	2.0	4.7	2.0	0.0	1.3	0.9	2.8	2.0	1.8
Total	4.2	5.2	6.4	4.5	5.7	5.6	4.6	5.2	5.6	6.5	6.4	5.8

Note: Yearly mean *Benefit Score* represents yearly mean of *Benefit Score* to balance (big) companies with a large amount of ratings with (smaller) companies with a small amount of ratings. The purpose of this table is to show the development of the average number of benefits. Especially the years 2007 and 2010 show lower averages than the other years in the sample period, which we attribute to the aftermath of the financial crisis. The original sample size, in terms of the number of ratings, was 24,679. The sample period ranges from January 2007 to December 2018.

Table S.4

Sample Size: Number of Companies per Industry, Sector, and Type

Industry/Sector	Number of Companies			
	Healthy	Insolvency	Total	
Administration	4	1	5	
Automotive	60	17	77	
Construction	12	4	16	
Consumer Goods	42	12	54	
E-Commerce	4	1	5	
Electrical Equipment	8	2	10	
Energy	4	2	6	
Energy Supply	8	2	10	
Entertainment	8	2	10	
Food	16	5	21	
Glass	4	1	5	
Healthcare	8	2	10	
Industrial Goods	4	1	5	
Information	4	1	5	
Investigation	4	1	5	
Machinery	27	7	34	
Metals	22	6	28	
Paper	4	1	5	
Shipping	6	2	8	
Solar	30	13	43	
Tourism	2	1	3	
Transportation	8	2	10	
Wood	4	1	5	
Total	Industry^(a)	293	87	380
Manufacturing	201	61	262	
Service	38	10	48	
Trade	32	9	41	
Construction	8	3	11	
Energy	8	2	10	
Logistics	6	2	8	
Total	Sector^(b)	293	87	380

Note: The table shows the number of firms, divided into healthy and insolvent firms, by industry. The sector is the 1st hierarchy level, and industry is the 2nd. The purpose of these figures is to make the sample composition transparent. 25.7% of the companies (87) in the data set are insolvent. The sample period ranges from January 2007 to December 2018. Variable definition is explained in Table A.1 in the Appendix.

^(a) *Industry distribution:* We observe an imbalance through the industries Automotive, Consumer Goods, and Solar. The higher insolvency rate in these industries drives this.

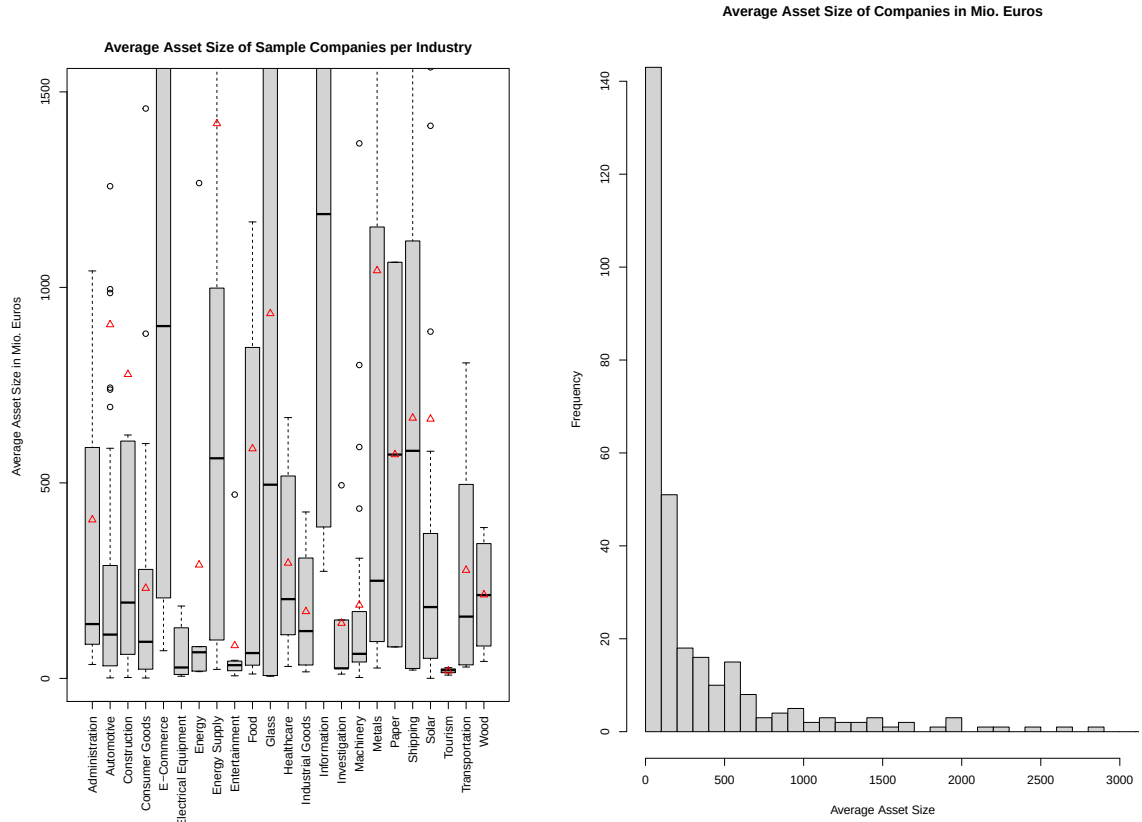
^(b) *Sector distribution:* As discussed, the 262 companies out of 380 companies originate from the sector Manufacturing (69 percent). Likewise, the majority of ratings also come from the Manufacturing Sector. We see here a comparable split with approximately 3.600 ratings out of 5.320.

Figure S.2

Average Asset Size of Sample Companies

(a) Average asset size per industry

(b) Asset size of sample companies

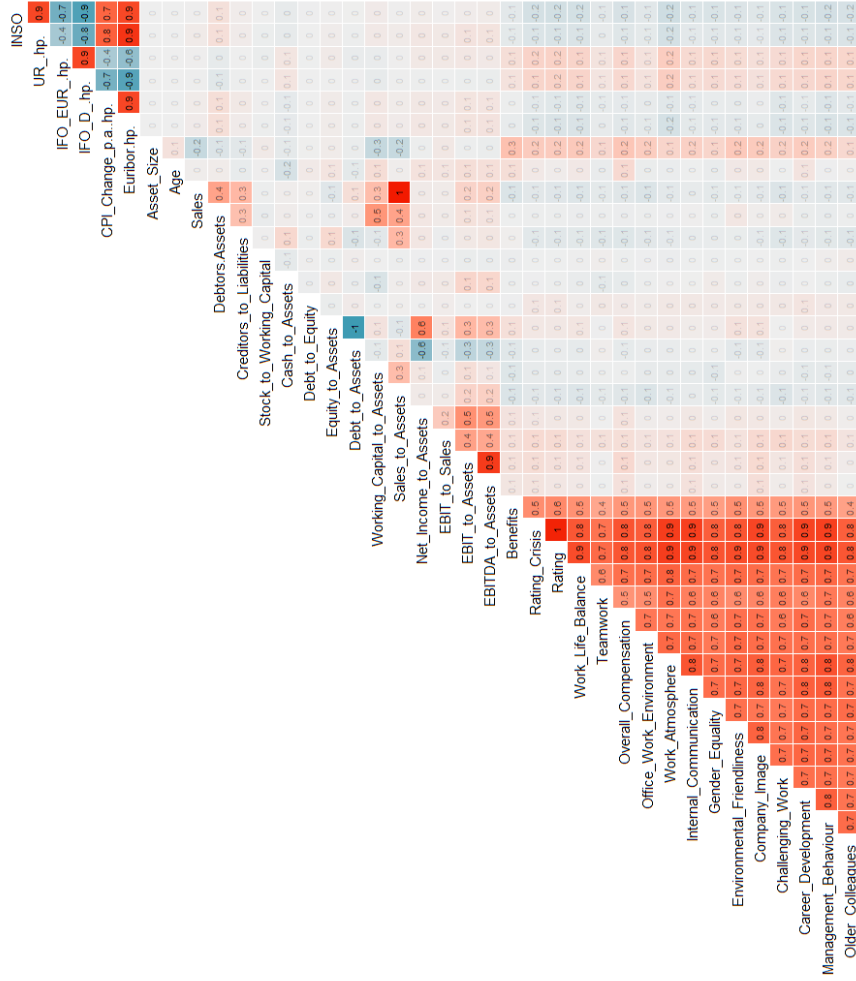


Note: These figures show the distribution of companies in the sample by sample size and industry. The purpose of this figures is to provide additional information about the sample composition, more specifically, the sample size of companies included in the sample. We consider these potential drivers of employee ratings. The sample period ranges from January 2007 to December 2018.

^(a) *Sample size per industry:* The figure shows the distribution of sample companies in terms of asset size per industry (mean equals the triangle)

^(a) *Sample size distribution:* The figure shows the distribution of the companies in the sample in terms of asset size in the year 2017, whereas 69 company values are missing out of 380.

Figure S.3
Correlations for Numerical Variables



Note: This figure shows the correlation matrix of the relevant numerical variables in the data set. The purpose of this figure is to provide visual support for the identification of linear relationships. Interestingly, our employee rating variables show very high correlations above 0.5. In addition, the two-dimensional view indicates a strong correlation between the macro and employee ratings variables within the group. We identified no strong correlation between these groups, nor with financial variables. This view supports our hypotheses of a strong additional value of employee rating variables. The correlation is measured by the Spearman correlation coefficient, rho. The sample period ranges from January 2007 to December 2018. Variable definition is explained in Table A.1 in the Appendix.