NBER WORKING PAPER SERIES

THE INFORMED INSIDER: A LEADING MEASURE OF QUALITY

Wei Cai
Dennis Campbell
Yaxuan Chen
Yufei Chen
Andrea Prat

Working Paper 35348
http://www.nber.org/papers/w35348

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2026

We thank Phil Berger, Jan Bouwens, Eddy Cardinaels, John Core, Wouter Dessein, Maximilian Margolin, Mario Milone, Frank Moers, Joe Moran, Venky Nagar, Jim Omartian, Gerardo Pérez Cavazos, Lin Qiu, Daniel Schaupp, Mani Sethuraman, Jenny Tucker, Marshall Vance (discussant), Rodrigo Verdi, Joe Weber, Joanna Wu, Gwen Yu, and participants at 2025 CEPR Paris Symposium, Online Seminar in Economics and Data Science, 2026 Berkeley Culture Conference, 2026 AAA Management Accounting Section Meeting, 2026 Columbia Management Accounting Conference, Fordham University, UC San Diego, MIT, University of Michigan, University of Florida, Erasmus University Rotterdam, University of Amsterdam, WU Vienna, Tilburg University, and Maastricht University, for their helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Wei Cai, Dennis Campbell, Yaxuan Chen, Yufei Chen, and Andrea Prat. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Informed Insider: A Leading Measure of Quality
Wei Cai, Dennis Campbell, Yaxuan Chen, Yufei Chen, and Andrea Prat
NBER Working Paper No. 35348
June 2026
JEL No. D83, M14

ABSTRACT

Product and service quality is fundamental to firm value creation, yet it is well recognized as difficult to observe ex ante. Existing quality proxies are limited in coverage, lack cross-firm comparability, and primarily rely on lagging indicators that capture product or service failures only after they materialize. Exploiting employees' informational advantage as informed insiders and firsthand observers of firms' internal operations, we develop and validate a novel, forward-looking measure of firm-year-level product and service quality using over 4.3 million employee reviews on Glassdoor. Leveraging machine learning models trained on a subset of firms with third-party customer satisfaction data, we construct quality indices for S&P 1500 firms spanning 2008 to 2023. The resulting quality measures exhibit meaningful variation across firms and within firms over time. In out-of-sample tests, our quality indices demonstrate strong predictive power for future quality provision, emerging as the single most important predictor relative to firm fundamentals and Glassdoor ratings. We validate our measure by examining its association with alternative quality metrics. We show that our quality measures are useful in predicting important quality-related firm outcomes such as product recalls, brand value, and profitability. We also construct an alternative set of quality measures using a zero-shot prompt-based approach and a supervised fine-tuning approach with GPT models to assess the potential of LLMs and generative AI in capturing firm-level quality provision. Our paper shows the value of employee voices as a powerful, forward-looking, and scalable signal of firm quality provision. The paper offers implications for stakeholders seeking to identify quality-related risks and opportunities before they become externally visible.

Wei Cai
Columbia University
wc2419@columbia.edu

Dennis Campbell
Harvard University
dcampbell@hbs.edu

Yaxuan Chen
Cornell University
College of Business
yaxuan.chen@cornell.edu

Yufei Chen
Columbia University
Columbia Business School
Accounting Division
yc4340@gsb.columbia.edu

Andrea Prat
Columbia University
Columbia Business School
and NBER
andrea.prat@columbia.edu

1 Introduction

Product and service quality is fundamental to firm value creation and long-term performance, and has long been a central topic in a substantial body of research across diverse literatures. Beyond firm performance, quality has far-reaching consequences for a broad range of stakeholders: it shapes investor valuations and consumer welfare, and its failure can impose significant societal costs, ranging from financial losses and public health risks to, at the extreme, loss of life. Despite its first-order importance, quality is notoriously difficult to observe ex ante. The Boeing 737 MAX crashes serve as a stark reminder: Boeing's public trust index was twice as high as that of Airbus prior to the 737 MAX crisis, masking serious underlying quality issues that were not revealed until after the tragic loss of hundreds of lives.[1]

The difficulty in observing quality ex ante is well-recognized in economic theory. Much of this research treats quality as private information about the attributes of a firm's products or services held by sellers but unobservable to buyers (e.g., Akerlof, 1970; Milgrom and Roberts, 1986; Spence, 1978). This has motivated a rich literature on signaling, screening mechanisms, and incentive mechanisms such as reputation for dealing with quality uncertainty. Existing approaches in applied and empirical research for measuring quality include indicators such as voluntary quality disclosures, product durability, customer satisfaction and complaints, product recalls, and litigation. While these measures provide useful information for quantifying quality to some extent, they face several important limitations. First, they are lagging indicators—they capture quality issues only after a transaction has occurred or after quality issues become visible to the public. Second, such measures are not universally available for all firms. Many have limited coverage and employ different sampling and scoring methodologies, resulting in a lack of comparability across firms and time periods. Third, voluntary quality disclosures by firms are not always consistent with actual internal control practices

[1] See https://www.cnbc.com/2024/01/30/boeings-reputation-is-taking-a-hit-survey-shows.html

and processes, reducing their credibility as faithful quality signals (e.g., Chen, 2024; Graham et al., 2022; Guiso et al., 2015). These challenges underscore the need to develop alternative forward-looking and scalable measures of firm quality provision.

The consensus reached by prior literature across various disciplines is that quality provision is shaped by firms' internal operating and control environment (e.g., Bloom et al., 2013; Ichniowski et al., 1997; Ittner et al., 2001; Lazear, 2000; Guiso et al., 2015), and it is precisely the information asymmetry surrounding these internal environments that makes quality provision difficult to observe from the outside. Employees have direct exposure to these internal processes and environment as a natural result of their routine involvement in day-to-day operations, and therefore, they are well-positioned to offer insider insights into the quality of the products and services that their firms provide (Cai et al., 2025). Furthermore, the rise of crowdsourced platforms enables employees to talk and share their experiences about their firms. These platforms aggregate anonymous employee reviews from a large number of users, allowing for the potential identification of systematic patterns in firms' internal operational and control environments based on more candid and diverse perspectives (Cai et al., 2026; Campbell and Shang, 2022; Green et al., 2019; Hales et al., 2018; Huang et al., 2020). Accordingly, information extracted from crowdsourced employee reviews may offer valuable insights for predicting firm quality provision.

We construct firm-year-level measures of firm quality provision by extracting information from employee written reviews on Glassdoor using machine learning approaches that are well established in prior literature (e.g., Frankel et al., 2022; Siano, 2025). Specifically, we apply random-forest regression-tree (RF) and Gradient Boosting model (XGBoost) given their superior performance in handling complex non-linear relationships. We obtain over 4.3 million employee written reviews from both "pros" and "cons" sections of Glassdoor for S&P 1500 firms over the period of 2008 to 2023. We also obtain customer satisfaction data from American Customer Satisfaction Index (ACSI) as our primary quality indicator. We identify a subset of firms with both Glassdoor employee reviews and ACSI scores and use

it to train our machine learning models. We map each firm's employee reviews in year t to ACSI scores in year $t+1$. The models therefore are trained to use employee reviews from a given year to predict one-year-ahead ACSI scores. To prevent look-ahead bias, we split the sample chronologically, using the 2008-2021 subsample for model training and holding out the 2022-2023 subsample for testing purposes. For both the random forest and XGBoost models, we train and develop separate models based on reviews from the "pros" and "cons" sections, respectively. We construct text-based quality indices for S&P 1500 firms from 2008 to 2023 by computing weighted sums of unigrams and bigrams in employee reviews for each firm-year, where the weights correspond to the feature importance values derived from the trained models.

We begin by examining the distribution and variation of our text-based quality measures. We use Shapley values to examine the percentage of variation explained by different sets of fixed effects. Year fixed effects explain a small share of the variation, accounting for 2.6% to 5.5% across different quality indices and model specifications. Industry fixed effects explain 1.1% to 2.5% of the variation. The interaction of industry and year fixed effects accounts for a modest portion, explaining 7.9% to 11.0% of the variation. Collectively, these aggregate-level factors (year, industry, and industry-year effects) explain 12.3% to 18.4% of the total variation, implying that 81.6% to 87.7% of the variation occurs at the firm level. Adding firm fixed effects reveals that time-invariant differences across firms within industries account for 29.9% to 47.6% of the variation. The remaining 35.3% to 57.8% represents time-varying, firm-specific variation that is unexplained by the fixed effects.

We next evaluate the predictive power of our firm quality provision measures by regressing leading industry-adjusted ACSI scores on our text-based quality measures in the hold-out sample. We find that our quality indices are significantly associated with future ACSI scores and that such relationships are robust with the inclusion of firm observable characteristics, lagged ACSI scores, Glassdoor ratings, industry fixed effects, and year fixed effects. The adjusted R^2 increases substantially when our quality measures are included—ranging from 10.1

to 22.8 percentage points across specifications. Further out-of-sample performance analyses suggest that across all model specifications, our text-based quality measures consistently emerge as the most important predictor of future ACSI scores, contributing more to explanatory power than any other readily observable firm characteristics such as size, leverage, sales growth, operating performance, analyst following, institutional ownership, and Glassdoor ratings.

We perform a battery of validation tests to examine the association between our text-based measures of firm quality provision and a set of alternative quality proxies, including MSCI product safety and quality scores, and Fortune's rankings of best quality firms and worst quality firms. We find that, across most specifications and a wide range of proxies for future quality, our text-based quality measures effectively distinguish between higher- and lower-quality firms in addition to the inclusion of observable firm characteristics and fixed effects. These results lend strong support to our findings that information extracted from employee written reviews is useful for predicting firm quality provision.

We then explore the implications of firm quality provision on business outcomes. We show that firms with higher quality scores exhibit a lower likelihood of product recalls, greater brand value, and higher future profitability. Additional analyses demonstrate that our measures offer incremental explanatory power beyond management disclosure and corporate culture. In assessing measurement error, we find that the share of measurement error in our quality measures is at least comparable to, and in some cases lower than, those documented for other firm-level variables under similar assumptions in prior studies (e.g., Bloom et al., 2018; Collard-Wexler, 2011; Hassan et al., 2019). This suggests that our measures capture economically meaningful heterogeneity in firm-level quality provision.

In light of recent developments in large language models and generative AI, we construct an alternative set of measures using GPT models to explore the potential and limitations of generative AI for our measurement task. We apply a zero-shot prompt-based approach and a supervised fine-tuning approach to construct GPT-based quality measures, respec-

tively. Our findings show that GPT-based measures can perform reasonably well but are not uniformly effective across tasks and approaches. Specifically, we find that the zero-shot quality measures are directionally correlated with customer satisfaction and explain variation in firm-level quality provision after controlling for firm covariates. Nonetheless, they produce relatively less precise predictions and their incremental predictive power is largely subsumed by Glassdoor ratings. The results highlight that the zero-shot approach can serve as a practical and accessible option when more technically demanding tools are unavailable, and users should exercise caution in applications where high predictive accuracy is critical. In contrast, the fine-tuned quality measures yield generally consistent results with our main machine-learning-based measures. However, it is best suited for tasks that require ordinal classification at a coarser scale as opposed to continuous scores. Taken together, our exploration underscores the importance of considering precision, accessibility, prediction consistency, and measurement granularity when deploying generative AI for text-based measurement tasks.

We contribute to the literature in two primary ways. First, we contribute to the management accounting literature on the measurement of economically important yet difficult-to-observe dimensions of firm operations and performance. Due to information frictions, quality is difficult to observe and verify prior to or at the point of transaction (e.g., Akerlof, 1970; Milgrom and Roberts, 1986; Spence, 1978). Existing proxies capturing firms' product and service quality using observable attributes, customer reviews, and firm voluntary disclosures are often limited in coverage, difficult to compare across firms, and lagging in nature. We address these limitations and add to the extant measures by developing and validating a more broadly available, forward-looking measure of firms' product and service quality based on large-scale employee-written reviews. We demonstrate that our text-based quality measures have strong predictive power for future firm quality provision beyond conventional observables.

Second, we contribute to the accounting literature on alternative information channels by

showing that employee-generated text provides valuable information about firms' important attributes such as quality provision. Prior literature suggests that employee-generated information on crowdsourced platforms such as Glassdoor is informative about firm attributes and outcomes that are difficult for external stakeholders to observe (e.g., Cai et al., 2026; Campbell and Shang, 2022). Related work also shows that employee activity on such platforms reflects economically meaningful responses to firms' financial reporting (Dehaan et al., 2023). Our paper adds to the literature by showing that crowdsourced employee platforms are important information outlets and that employee reviews provide incremental insights into firms' underlying quality provision beyond formal managerial communication and standard summary measures.

Our measures may be useful to a wide range of audiences including researchers, consumers, regulators, and management. For example, researchers interested in the antecedents and consequences of firm quality provision, or who wish to control for firm quality provision in their analyses, may directly benefit from our measures. Consumers may benefit from more forward-looking and systematic measures about firms' quality provision when making purchase decisions, while regulators may leverage such measures to identify areas where greater attention or closer monitoring is warranted. Management may also benefit from our measures as an internal diagnostic tool that helps top management identify deteriorating quality performance and correct course before problems escalate into major crises.

2 Literature Review

2.1 Importance and Uncertainty of Quality Provision

Provision of high-quality goods and services is universally recognized as an important determinant of competitive advantage and long-term success for firms. Studies across various disciplines suggest that quality contributes to product differentiation, enables price premiums, enhances operational efficiency and cost control, creates higher entry barriers

that deter potential competitors, builds brand reputation, and fosters customer trust and loyalty (e.g., Chamberlin, 1953; Porter, 1985; Crosby, 1979; Deming, 1982). Through these channels, firms can achieve greater sales, profitability, and market share, ultimately securing long-term performance and survival. There is substantial empirical evidence supporting these theoretical predictions. For example, Bloom et al. (2013) provide causal evidence on the impact of adopting structured management practices, such as quality control systems, on firm productivity by conducting a randomized field experiment among textile firms in India. They find that treated firms experienced a 17% increase in productivity or over $300,000 in annual profitability within the first year, primarily driven by improved quality and operational efficiency, followed by sustained growth in firm size over the longer term. Hendricks and Singhal (1997) document strong evidence of performance gains driven by robust quality by comparing quality-award-winning firms to matched control counterparts. Nagar and Rajan (2001) find that product quality measures such as external failure costs and defect rates serve as significant leading indicators of future sales and profitability. Ittner and Larcker (1998) provide similar evidence that customer-perceived quality, as measured by customer satisfaction, translates into improved future accounting performance and financial outcomes.

Going beyond firm-level productivity and performance, quality has important implications for investors, consumers, and broader social outcomes. For investors, quality serves as an indicator for robust internal controls, profitability, management competence, and long-term value creation, thereby influencing investment decisions, firm valuation, and capital market outcomes (Hendricks and Singhal, 1996). For consumers, quality directly affects their welfare and utilities through its impacts on purchasing experience, satisfaction, and safety. At a broader level, quality failures such as product recalls, service breakdowns, and extreme downstream hazards can impose significant societal costs, including public health risks, mistrust towards market institutions, financial losses, and even loss of life. The Boeing 737 Max crashes serve as a prominent example of the devastating societal consequences that can result from quality deterioration.

2.2 Challenges in Measuring Quality Provision

Despite its desirability, in contrast to other product or service attributes such as price and design features that are more observable and verifiable, quality is difficult to observe and verify prior to or at the point of transaction. This challenge arises because quality is shaped by firms' internal operating and control environment, which remains largely inaccessible to external stakeholders. Within the economics literature, considerable thought has been given to quality uncertainty as asymmetric information. Arrow (1965) examines how quality uncertainty, particularly in the healthcare market, can lead to moral hazards and market inefficiencies when the patient is unable to assess the true value of medical services before having consumed the services. Akerlof (1970) further analyzes this problem in the context of the used car market. The seminal "market for lemons" model highlights that product quality, as asymmetric information, can lead to adverse selection, drive out high-quality providers, and ultimately cause market failure.

The literature approaches the information problem of quality uncertainty primarily in two ways. Through the lens of information economics, researchers investigate mechanisms aimed at reducing information asymmetry such as signaling, reputation, disclosures, and regulations (e.g., Nelson, 1970; Spence, 1978; Shapiro, 1983; Milgrom and Roberts, 1992; Tadelis, 1999). These efforts intend either to distinguish high-quality providers from their lower-quality counterparts through signals that are too costly for lower-quality firms to imitate (e.g., labels, licenses, certifications, or warranties) or to motivate firms to provide higher-quality goods to reap the benefits of price premiums. Firms can also voluntarily disclose information about their product and services quality through websites, conference calls, and other communication channels (e.g., Guiso et al., 2015; Li et al., 2021, 2024). Regulations mitigate information problems by setting quality standards and penalizing irregularities, lowering consumers' searching costs (Leland, 1979; Ronnen, 1991).

Applied researchers, on the other hand, have been exploring observable signals that are

available to external stakeholders to gauge firm quality provision. Examples range from positive indicators—such as product durability, consumer satisfaction, products rating and services ranking indices, and quality awards—to negative indicators such as consumer complaints, product recalls, penalties, and litigation. While these proxies offer an avenue for quantifying firm quality provision to some extent, they are subject to several important limitations.

First, these are retrospective measures that are lagging in nature as they capture quality only after outcomes are realized, which limits the ability of external stakeholders to make informed decisions. For example, consumers need to rely on prior buyers' experience to infer quality, which is a disadvantage to early users. In addition, internal operational and control practices may have deteriorated long before quality issues become visible to the public, thereby putting consumers who rely on historical quality indicators at risk. The consequences of untimely quality information can be particularly severe in industries of higher stakes such as healthcare, aviation, and financial service. For instance, before the outbreak of Wells Fargo's 2016 cross-selling scandal, the company outperformed its industry peers across multiple dimensions and was known for its sound management practices (Tayan, 2019). Similarly, Boeing's public trust index was twice as high as that of Airbus prior to the 737 MAX crisis that tragically resulted in the loss of hundreds of lives. Drug recalls typically only take place after millions of patients have consumed the problematic medications.

Second, these measures are not universally available for all firms and often do not share consistent evaluation criteria. For example, some commonly used customer satisfaction indices such as ACSI have limited scope and cover only selected industries and firms that meet certain criteria or opt into survey participation. Moreover, these indices often employ different sampling and scoring methodologies, making it difficult to compare quality performance across firms, industries, or sectors.

Third, firms' voluntary disclosures can be influenced by strategic considerations or reporting incentives and may not always be consistent with actual internal control practices

and norms, thereby limiting the reliability of such disclosures as faithful indicators of quality (Chen, 2024; Graham et al., 2022; Guiso et al., 2015).

2.3 Opportunities with Crowdsourced Employee Information

The preceding discussion boils down to a question that is at the heart of measuring firm quality provision: who has access to information about firms' internal operations and control systems that shape product and service quality, and whether such information can be used to predict firm quality provision? Drawing on the recent growing body of literature on crowdsourced employee information, we propose that crowdsourced employee reviews offer a novel and promising solution to overcome information barriers and gain insight into firms' internal quality-related operational and control practices.

First, employees have the best access to insider information about their firms as a natural result of their routine involvement in day-to-day operations and the development of crowdsourced platforms enables employees to talk and share their experiences about their firms. For instance, Cai et al. (2026) leverage Glassdoor employee reviews to develop and validate measures of organization capital, an economic construct that is intangible in nature. Most importantly, employee reviews are likely to contain forward-looking information about the underlying factors that drive firm quality provision, which may be useful for developing leading indicators of firm quality provision. For example, following the 737 Max crisis, subsequent investigations suggested that Boeing engineers were aware of deteriorating quality and internal control environments before the crisis unfolded (George, 2024; Sterman and Quinn, 2023). Such concerns are likely to be embedded in employee reviews, particularly when employees raise issues related to quality control processes, management practices, or organizational culture and norms.

Second, compared with firms' voluntary disclosures, employee reviews are less subject to distorted reporting incentives. Employee review websites typically allow employees to leave reviews anonymously, which significantly lowers the cost of sharing negative information

about their firms. In addition, employee reviews contributed by a broad and diverse base of users as a form of crowdsourced information may offer a comprehensive and reliable view of firm quality provision. Taking the reviews to the aggregate level, we can leverage the combined knowledge and perspectives of a large number of individuals to potentially uncover systematic patterns about their firms' internal operational and control environments.

Lastly, popular employee crowdsourced information-sharing platforms such as Glassdoor attract numerous users across firms, industries, and sectors. This broad coverage enables us to construct measures of quality provision that are available for a large sample of firms over time.

3 Data and Sample

3.1 Glassdoor Employee Reviews

We collect data on employee reviews from Glassdoor. Since its launch in 2008, Glassdoor has become one of the most developed and popular company review websites covering more than 2 million companies where employees can share anonymous reviews about their firms.[2] Glassdoor's review-sharing process is highly structured. When sharing their reviews, employees are asked to provide ratings on several different dimensions on a scale of one to five stars. Importantly, employees are also asked to leave written comments about the pros and cons of their firms in different sections and their advice to the management.[3] The feature of separating positive and negative reviews in different sections offers an advantage when constructing text inputs for our machine learning models. By separating positive and negative sentiments in a natural way, this format helps mitigate measurement errors commonly associated with bag-of-words assumptions, which tend to overlook contextual distinctions when opposing sentiments appear together within the same passage.

[2] https://www.glassdoor.com/about/

[3] It is optional to provide a commentary in the "advice to the management" section on Glassdoor.

Several other features of Glassdoor's review-sharing process add controls to ensure review reliability and reduce selection bias. First, Glassdoor requires either email verification from an active address or authentication through a valid social media account in order for a user to post a review. Second, the review can only be submitted anonymously, which reduces the high costs associated with sharing negative aspects of firms such as protecting employees from potential repercussions from their employers (Dyck et al., 2010). Third, Glassdoor implements a "give-to-get" policy in which users are prompted to submit an employer review before gaining full access to the website.[4] Lastly, Glassdoor employs a moderation process. In addition to the automatic system with technology filters and algorithms, a content moderation team manually reviews 50% of submitted reviews before a review is published on the website.[5]

3.2 American Customer Satisfaction Index (ACSI)

Prior literature has widely used customer satisfaction as a proxy for firm quality provision, drawing on sources such as internal records of customer feedback, self-administered customer surveys, and national survey systems. A common drawback of these measures is their limited coverage and lack of comparability across firms and sectors. These measures are not universally available, and often cover only selected industries or firms that either meet specific eligibility criteria or opt into survey participation. In addition, they often employ different sampling procedures and scoring methodologies, which undermines consistency and makes it difficult to benchmark quality performance across firms, industries, or sectors.

Among these measures, ACSI is one of the most widely used and validated proxies in prior studies (e.g., Anderson and Mansi, 2009; Otto et al., 2020; Wolter et al., 2019). ACSI is a system of customer satisfaction measurement available for U.S. firms with an approximate coverage of 400 companies across 10 major economic sectors.[6] Each year, ACSI surveys

[4] https://help.glassdoor.com/s/article/Give-to-get-policy
[5] https://help.glassdoor.com/s/article/Review-Integrity-Fraud-Detection
[6] https://theacsi.org/company/building-the-cross-industry-index/

around 200,000 U.S. consumers regarding their satisfaction with purchased goods and services. Participants are randomly selected via email and screened based on recent purchases within specified timeframes, which vary by the products or services purchased. Qualified respondents are then asked to identify the company or brand from which they purchased, and their responses are used to generate company-specific customer satisfaction scores. The ACSI score is calculated as a weighted average of responses to three survey questions—customer expectations (a measure of customer's anticipation of the quality of a company's products or services), perceived quality (a measure of customer's evaluation of the quality of a company's products or services), and perceived value (a measure of quality relative to price paid)—each capturing a distinct facet of satisfaction with a product or service.

We use ACSI as our primary quality outcome variable to train our machine learning model for two main reasons. First, ACSI covers both product- and service-oriented firms, which allows us to develop a measurement that is not confined to a particular industrial sector and is more generalizable across the broader universe of firms. Second, ACSI is a continuous index ranging from 0 to 100 that reflects the full spectrum from inferior to superior quality performance. Event-based outcome variables such as product failures or recalls, in contrast, are largely confined to manufacturing sectors and identify only severe downstream quality breakdowns without capturing the rich variation in quality levels. Thus, training our machine learning models on the ACSI allows us to construct a quality measure that (1) applies more generally across the universe of firms and beyond some specific industry sectors, and (2) reflects continuous variation in firm quality provision rather than focusing exclusively on extreme negative events.

3.3 Sample Description

One of the goals of this study is to develop a firm-year-level text-based quality measure that can be applied across the broader universe of firms, including non-consumer-facing firms and firms not covered by ACSI. To achieve this, we rely on two distinct samples. The first

sample is a training/testing sample that consists of firm-years with both sufficient Glassdoor reviews ($\geq$ 10) and ACSI scores, which we use to train our machine learning model and test the trained model's out-of-sample performance. The second is an application sample consisting of all S&P 1500 firm-years with sufficient Glassdoor reviews but *without* ACSI coverage, to which we apply the trained model to generate text-based quality measures and assess their broader applicability.

We begin by collecting all Glassdoor reviews for firms that are currently in the S&P 1500 or were included at any point between 2008 and 2023, resulting in approximately 4.4 million reviews across 2,575 unique firms. We aggregate individual reviews to the firm–year level, yielding 31,780 firm-years. We require each firm-year to have at least ten reviews, which leaves us with a final sample of 20,567 firm-years based on over 4.3 million reviews from 2,137 unique firms.

We obtain customer satisfaction data from the ACSI website. Among the firms in our Glassdoor sample, 181 companies (1,886 firm-years) are covered by ACSI. This subsample, which we label as the training/testing sample, includes both the input used to train our machine learning models and the hold-out set used to evaluate out-of-sample performance.

Finally, we obtain financial statement data from Standard & Poor's Compustat. Data on Analyst following is obtained from I/B/E/S. Data on institutional ownership is obtained from LSEG (formerly Thomson Reuters) Institutional Holdings database. We follow the definition in Bushee (2001) to identify transient investors. Data on conference call transcripts are obtained from Capital IQ's Transcripts Database.

4 Methodology

4.1 Machine Learning Based Measure: RF and XGBoost

We use RF and XGBoost to construct our quality index from employee reviews because of their superior performance in handling complex, non-linear relationships and their robustness

to overfitting compared to traditional regression methods. The input to our models consists of the "Pros" or "Cons" text from Glassdoor reviews. The outcome variable is the ACSI score adjusted for industry. Specifically, to enable cross-industry comparison, we calculate industry-adjusted ACSI score as the difference between a firm's customer satisfaction score and the mean score within its Fama-French 48 industry classification, scaled by the industry's standard deviation.[7] We map each firm's employee reviews in year t to ACSI scores in year $t+1$. The models therefore are trained to use employee reviews from a given year to predict one-year-ahead industry-adjusted ACSI scores.

To prevent look-ahead bias, we implement a chronological split of the ACSI sample: 2008-2021 for model training, and 2022-2023 as a hold-out sample for testing. After training the models on the 2008-2021 ACSI sample, we generate our text-based quality indices by calculating weighted sums of unigrams and bigrams present in reviews for each firm-year, where weights correspond to the feature importance values derived from our trained models. This approach enables us to generate quality indices for all firm-years with Glassdoor coverage, including both the hold-out test sample and the broader application sample of S&P 1500 firms without available ACSI scores.

We develop separate models for the pros and cons review sections, yielding two distinct quality indices for each firm-year: *Quality Index - Pros* and *Quality Index - Cons*. We also construct an aggregate quality index, *Quality Index - Agg*, for each firm-year by taking a weighted average of *Quality Index - Pros* and *Quality Index - Cons*, where the weights are determined by the relative lengths of each review section's content for the firm in that year. Specifically, we calculate *Quality Index - Agg* as follows:

$$\textit{Quality Index - Agg}_{i,t} = \frac{\sum_s L^s_{i,t} \cdot (\textit{Quality Index - s})_{i,t}}{\sum_s L^s_{i,t}}$$

where i denotes the firm, t denotes the year, $s \in \{\text{Pros}, \text{Cons}\}$ denotes the review section

[7] We use the Fama-French 48 industry classifications for this adjustment, which offer a good balance between providing sufficient granularity for meaningful within-industry comparisons and maintaining a large enough sample size within each industry to ensure reliable standardization.

type, and $L^s_{i,t}$ denotes the length of section s content for firm i in year t. The lengths $L^s_{i,t}$ are determined after removing numbers, punctuation, stop words, and non-English words.

Figure 1 provides a visual representation of our sample construction and methodology.

We take several steps to clean and organize the words used in these comments. First, we convert text into lowercase characters. Second, we remove any numbers, punctuations, stop words (e.g., "a", "the", and "in"). Third, we lemmatize each word using the `WordNetLemmatizer` to reduce the influence of sparsity. We restrict our analysis to unigrams and bigrams that appear in the reviews of at least 5% of firms in each year, separately for each review section (i.e., pros and cons), to reduce overfitting and enhance predictive accuracy. This yields 4,190 words and phrases for the pros section and 5,742 for the cons section.[8]

We implement the random forest model using the `RandomForestRegressor` class from Python's scikit-learn library. Following prior literature (Frankel et al., 2022; Siano, 2025), we train the models using the default setting of `max_features = sqrt` and set the number of trees (`n_estimators`) to 3000. We train the Gradient Boosting model using the `XGBRegressor` class from the XGBoost Python library. When selecting hyperparameters for the XGBoost model, we balance prediction performance with computational efficiency by tuning only a limited set of parameters most likely to affect out-of-sample accuracy. We perform a grid search over the parameter space: the number of boosting rounds varies from 500 to 1500, learning rates from 0.01 to 0.05, and tree depths from 2 to 4. To preserve temporal structure and mitigate look-ahead bias during cross-validation, we adopt a five-fold `TimeSeriesSplit`.[9] The optimal configuration is selected based on the lowest average mean squared error (`MSE`) across validation folds. Additional details of the hyperparameter tuning

[8]Research shows that this approach produces a better regression fit in natural language processing tasks (Loughran and McDonald, 2016). For example, Siano (2025) select the top 1,500 unigrams and bigrams ranked by TF-IDF from disclosure text as input to a non-linear Gradient Boosting model to predict market reactions around earnings announcement dates.

[9]Given the panel structure of our data, randomly splitting observations as in standard k-fold cross-validation may introduce look-ahead bias by allowing future information to inform model training. We therefore employ time-series cross-validation, which preserves the temporal ordering of the data and ensures that each training fold is restricted to information available prior to the validation period.

process are provided in Online Appendix.

Tables A1 and A2 of Appendix A categorize and present the top 300 words that are most influential in constructing our quality measure for the pros and cons sections, respectively.[10] To provide intuition for the kinds of review content that are most predictive of quality provision, we apply our trained models to individual reviews to investigate which types of content are most predictive of high or low quality scores. Table A3 of Appendix A presents examples from the pros section with the highest predicted quality scores, as well as from the cons section with the lowest predicted scores.

4.2 Descriptive Statistics

4.2.1 Summary Statistics

Panel A of Table 1 presents the industry distribution of our sample across the 48 Fama-French industry classifications. Panel B reports yearly distribution of the training/testing sample (ACSI score available) and the application sample (ACSI score unavailable). Panel C reports descriptive statistics for all key variables used in our analysis. We note that *Industry-Adjusted ACSI* shows a mean and median above zero, indicating that across all firms covered by ACSI, S&P 1500 firms on average score higher than non-S&P 1500 firms within their respective industries.

Detailed variable definitions are provided in Appendix B. All continuous variables are winsorized at the 1% and 99% levels.

[Insert Table 1]

[10]For n-grams whose appropriate category is not readily apparent on its own, we manually examine 100 employee reviews in which the n-gram appears to assess its most common contextual meaning. Each n-gram is then assigned to the category that best reflects its predominant usage in the context of employee reviews.

4.2.2 Distribution of the Quality Score

We begin by examining the distribution of our predicted product/service quality score across firm-years. Figure 2 presents histograms of the quality scores by review section and machine learning model, which are approximately normally distributed.

[Insert Figure 2]

4.2.3 Variance Decomposition

We use Shapley values to decompose the variance of our quality measure across different fixed effects (Armstrong et al., 2025).[11] Table 2 reports the results.

Year fixed effects explain a small share of the variation, accounting for 2.6% to 5.5% across different quality indices and model specifications. Industry fixed effects explain 1.1% to 2.5% of the variation.[12] The interaction of industry and year fixed effects accounts for a modest portion, explaining 7.9% to 11.0% of the variation. Collectively, these aggregate-level factors (year, industry, and industry-year effects) explain 12.3% to 18.4% of the total variation, meaning that 81.6% to 87.7% of the variation occurs at the firm level. Adding firm fixed effects reveals that time-invariant differences across firms within industries account for 29.9% to 47.6% of the variation across models and indices. The remaining 35.3% to 57.8% represents time-varying, firm-specific variation that is unexplained by the fixed effects.

[Insert Table 2]

[11] Shapley values are derived by averaging a variable's marginal contribution to the model's explained variance across all possible combinations of explanatory variables. For each subset of explanatory variables, the method evaluates how much the inclusion of a given variable increases the variance explained by the model. By considering all possible variable combinations, Shapley values thus provide a reliable and order-invariant decomposition of explanatory power across model components.

[12] This is expected as we adjusted customer satisfaction scores within industry before training the models.

5 Validation

5.1 Out-of-Sample Performance

We take several steps to validate our measure. First, we validate our measure by examining its explanatory power for industry-adjusted ACSI scores in out-of-sample tests. Specifically, we estimate the following OLS regression:

$$\textit{Industry-Adjusted ACSI}_{i,t+1} = \beta_0 + \beta_1 \textit{Quality Index}_{i,t} + \gamma \mathbf{X}_{i,t} + \delta_{j,t} + \epsilon_{i,t} \tag{1}$$

where $\textit{Industry-Adjusted ACSI}_{i,t+1}$ is the industry-adjusted ACSI score of firm i in year $t+1$. $\textit{Quality Index}_{i,t}$ is our text-based quality measure derived from all employee reviews for firm i in year t. $\mathbf{X}_{i,t}$ is a vector of firm controls, including firm size (*Size*), leverage (*Leverage*), profitability (*ROA*), market-to-book ratio (*M/B Ratio*), and sales growth (*Sales Growth*). Prior literature suggests that analyst coverage and short-term pressure on management can negatively impact product quality (Li et al., 2024). To account for these factors, we include two additional controls: *Analyst Following*, defined as the number of analysts issuing EPS forecasts for the fiscal year end, and *% Transient Investor*, measured as the percentage of outstanding shares held by transient investors. In addition, we control for Glassdoor ratings (either the overall one-to-five rating or subratings across specific categories) to assess whether employee reviews contain incremental information about product/service quality beyond what is captured by numerical ratings alone. δ_{jt} denotes year fixed effects and Fama-French 48 industry fixed effects.

The results are presented in Table 3. Our results reveal a strong positive association between the text-based quality index and industry-adjusted ACSI scores in the hold-out sample, with economically meaningful explanatory power. Specifically, univariate regressions yield adjusted R^2 of 34.0% and 31.8% for pros-based quality indices (random forest and XGBoost, respectively), 27.9% (random forest) and 29.0% (XGBoost) for cons-based quality

indices, and 32.1% (random forest) and 33.3% (XGBoost) for aggregate quality indices.[13]

Table 4 further compares the adjusted R^2 from regressions with and without our text-based quality index. The results show substantial incremental explanatory power from our quality index. The pros-based quality index improves adjusted R^2 by 11.1% to 22.1% across specifications. The cons-based index contributes 10.1% to 19.5% in explanatory power. The aggregate index shows the largest improvement, increasing the adjusted R^2 by 12.4% to 22.8% across specifications. These incremental improvements translate to substantial percentage increases in explanatory power ($\%\Delta R^2$), with the quality indices generating 34.9% to 137.2% relative improvements in adjusted R^2 across different model specifications. Overall, the out-of-sample results suggest that our quality index captures meaningful information about product and service quality provision.

[Insert Table 3 and Table 4]

Figure 3 shows the relative contribution of each variable in the out-of-sample performance regressions. We calculate the contribution of each variable by subtracting the R^2 when that variable is excluded from the full model R^2. The figure shows that our text-based quality index consistently provides the largest contribution across all specifications, outperforming traditional firm-level controls such as *Size*, *Leverage*, and *ROA*.

[Insert Figure 3]

5.2 Correlation with MSCI Quality Indicators

We further validate our measure by examining its association with product quality related indicators from the MSCI ESG database and MSCI ESG KLD STATS database.[14] MSCI

[13]The magnitudes of our results are economically meaningful and consistent with prior literature on text-based predictive modeling. For example, Siano (2025) employs Gradient Boosting to capture the information content of quarterly earnings announcement press releases for predicting two-day cumulative abnormal stock return. Out-of-sample analysis shows that this Gradient Boosting-based CAR prediction explains approximately 12% of the variation in CAR[0,1].

[14]The last year of data availability is 2019.

ESG provides disaggregated scores for its environmental, social, and governance pillars, with product safety & quality identified as a key issue within the social pillar. Companies are evaluated on their exposure to possible recalls or product safety concerns, the robustness of their supply chain and sourcing systems, their quality management efforts in manufacturing, and their responsible marketing practices.

We obtain *Quality Management*, which assesses companies' proactive versus reactive approaches to product quality management. Companies that proactively manage product quality score higher. We also obtain *Quality Risk Exposure*, which captures a company's exposure to product safety and quality risks based on products/services, operational locations, and business nature. We derive *#Strengths - #Concerns* by subtracting the number of product characteristic concerns from the number of product characteristic strengths. We also create two binary indicators: *Has Product Concern*, which equals one if the firm has product characteristic concerns in a given year, and *Has Safety Concern*, which equals one if the firm has product safety concerns in a given year.

Table 5 presents the correlations of our text-based quality index and MSCI product quality-related scores. The results show negative and statistically significant correlations between our text-based quality measures and *Quality Risk Exposure*, *Has Product Concern*, and *Has Safety Concern*, indicating that companies with higher quality index tend to have lower exposure to product safety risks and fewer product-related concerns. We find positive and significant correlations with *#Strengths - #Concerns* and *Quality Management*, suggesting that companies with higher Glassdoor-derived quality indices also exhibit stronger product management practices and more favorable product characteristic assessments.

Overall, this comparison with an independent data source provides additional confirmation of our measure's ability to capture meaningful quality variations across firms and time.

[Insert Table 5]

5.3 Comparing Top- and Non-Top-Ranked Firms in Product/Service Quality

To further assess the external validity of our quality measure, we compare the quality index across firms with and without high reputational standing in product or service quality, based on the *Fortune's World's Most Admired Companies* list.

Fortune's World's Most Admired Companies is an annual ranking that provides a comprehensive evaluation of corporate reputations. The rankings are compiled through a survey conducted by Fortune in partnership with Korn Ferry, in which top executives, directors, and financial analysts rate companies within their own industries and across industries on nine key attributes: (1) Innovation, (2) People management, (3) Use of corporate assets, (4) Social responsibility, (5) Quality of management, (6) Financial soundness, (7) Long-term investment value, (8) Quality of products and services, and (9) Global competitiveness. From 2008 through 2012, Fortune published lists of the 10 most admired companies for each attribute.[15] This enables us to identify firms with the highest ratings in product and service quality for this five-year period. We obtain companies listed among the best in product/service quality and the full list of participating firms from *Fortune's World's Most Admired Companies* rankings on their website.[16]

From our application sample, we manually identify firms that appeared among the top ten in the "quality of products and services" attribute at any point between 2008 and 2012. We also identify firms that participated in the rankings during this period but were not listed among the top ten in product/service quality. This yields 14 unique firms in the top-quality group and 231 unique firms in the comparison group. We then compare the mean quality indices between the two groups and test for pairwise differences.

The results are reported in Table 6. The pros-based, cons-based, and aggregate quality indices all show that the top-quality group consistently has the higher mean values than the

[15] Beginning in 2013, Fortune discontinued listing companies by specific attributes on its website.
[16] https://money.cnn.com/magazines/fortune/mostadmired/

comparison group. This relationship is robust across quality indices constructed using both machine learning models. All pairwise comparisons are statistically significant, providing strong evidence that our text-based quality measures effectively distinguish between firms with different reputational standings in product/service quality.

[Insert Table 6]

6 Implications of Quality

In this section, we examine the economic applications of our quality measures in predicting several important quality-related business outcomes.

6.1 Quality and Future Product Recall Events

Product recalls represent one of the most visible and costly manifestations of quality deterioration, with adverse implications for consumer safety, firm reputation, and shareholder value. We examine whether our text-based quality measures are associated with future recall events. We expect firms that experience greater quality deterioration (measured as larger year-to-year declines in the quality index) to have more product recalls in subsequent periods.

We obtain data on recall events over 2000-2024 from the Consumer Product Safety Commission (CPSC) website. Following Chen et al. (2009), we exclude all retailer recalls (e.g., Walmart, Target) and focus exclusively on manufacturer recalls, as CPSC regulations on reporting product safety issues differ between manufacturers and retailers. We restrict our sample to firms in industries where the CPSC recorded at least one consumer product recall during the entire sample period.

We estimate the following regression using Poisson pseudo-maximum likelihood (PPML)[17]:

$$\begin{aligned} \#\ \textit{(Reactive) Recalls}_{i,t+1} = \beta_0 + \beta_1 \textit{Quality Change Measure}_{i,t} + \beta_2 \textit{Prior (Reactive) Recall}_{i,t} \\ + \gamma \mathbf{X}_{i,t} + \delta_{j,t} + \epsilon_{i,t} \end{aligned} \tag{2}$$

where the dependent variable is the number of (reactive) product recalls in the next year. Our primary independent variable is the change from year $t-1$ to year t in our text-based quality index, or an indicator equals one if the change in quality index from year $t-1$ to year t falls in the bottom quartile of the sample. We include a set of standard firm-level control variables as in prior analysis, along with *R&D Intensity*, motivated by prior research suggesting a potential link between innovation intensity and recall propensity (Cockrell et al., 2024; Zou et al., 2023). We also include recall history controls: the number of (reactive) product recalls in each of the three preceding years (i.e., at $t-1$, $t-2$, and $t-3$) to account for potential persistence in firms' recall behavior.

Table 7 presents our regression results. Panel A reports results using the quality index derived from the cons section. Columns (1)-(4) of Panel A show that the change in the text-based quality index is negatively associated with future product recall events, that is, firms experiencing increases (decreases) in quality index are less (more) likely to have product recalls in the subsequent year. Columns (5)–(8) indicate that this negative association is stronger in magnitude when we focus on reactive recalls, defined as recall events preceded by customer complaints, incident reports, or injuries prior to the official announcement. We posit that these cases reflect more severe quality failures, as issues were identified by consumers rather than detected proactively by the firm. In terms of economic magnitudes, a one-standard-deviation improvement in quality index corresponds to a 14.2%-19.4% reduction in the number of recalls in the next year.[18] The effect is more pronounced for reactive

[17] We employ Poisson pseudo-maximum likelihood (PPML) estimation because it yields consistent estimates under relatively weak distributional assumptions and is particularly appropriate for modeling non-negative count outcomes with substantial zero inflation, as is the case for our recall variable.

[18] Random Forest: $\exp(-1.549 \times 0.139) - 1 \approx -19.4\%$; XGBoost: $\exp(-0.496 \times 0.309) - 1 \approx -14.2\%$

recalls, where a one-standard-deviation improvement in the quality index corresponds to a 21.9%-22.9% reduction in the number of reactive recalls in the next year.[19]

While the coefficients are in the expected directions, the pros-based quality index (Panel B) shows weaker predictive power compared to the cons-based index. This finding is intuitive given that the pros section typically emphasizes favorable aspects of the workplace rather than highlighting specific concerns related to operational weaknesses, and may therefore be less informative about risks or deficiencies that could lead to severe quality failures such as product recalls.

The aggregate quality index (Panel C), which combines information from both pros and cons sections, also shows strong and significant predictive power for future recalls. The coefficients are statistically significant and economically meaningful, with magnitudes comparable to or even greater than those observed for the cons-based index. Overall, these findings suggest that declines in our text-based quality indices are predictive of subsequent adverse product-quality outcomes.

[Insert Table 7]

6.2 Quality and Future Brand Value

While quality deterioration can lead to costly operational failures such as product recalls, consistently high quality provision facilitates the creation of long-term value. Specifically, high-quality products and services help foster consumer trust, strengthen brand reputation, and create competitive advantages that can contribute to stronger brand value (e.g., Hendricks and Singhal, 1996). We therefore expect a positive association between quality and future brand value.

We follow Li et al. (2021) and examine this association using Brand Finance's global brand rankings, which identify firms with the highest brand valuations worldwide. We obtain brand ranking data from 2008 to 2025 through Branddirectory and manually identify

[19] Random Forest: $\exp(-1.869 \times 0.139) - 1 \approx -22.9\%$; XGBoost: $\exp(-0.800 \times 0.309) - 1 \approx -21.9\%$.

brands owned by S&P 1500 firms in our sample.[20] We then construct an indicator variable, *Top Brand*, which equals one if a firm appears in Brand Finance's annual Top 500 global brand rankings, and zero otherwise.

We examine the relationship between our quality index and the likelihood of being listed on the Top 500 global brand rankings. Specifically, we estimate the following regression:

$$(Ltm)\ Top\ Brand_{i,t+1} = \beta_0 + \beta_1 Quality\ Index_{i,t} + \gamma \mathbf{X}_{i,t} + \delta_{j,t} + \epsilon_{i,t} \tag{3}$$

where $Quality\ Index_{i,t}$ is our text-based quality measure derived from all employee reviews for firm i in year t. $\mathbf{X}_{i,t}$ is a vector of firm controls, including firm size (*Size*), leverage (*Leverage*), profitability (*ROA*), market-to-book ratio (*M/B Ratio*), and sales growth (*Sales Growth*). δ_{jt} denotes year fixed effects and Fama-French 48 industry fixed effects.

The results are reported in Table 8. We find that firms with higher quality indices derived from Glassdoor reviews are significantly more likely to appear in the Top 500 brand rankings. In terms of economic magnitude, a one–standard–deviation increase in our quality index increases the probability of appearing in the Top 500 by 2.1–2.6 percentage points in the subsequent year and by 1.1–2.8 percentage points over the next three years. These effects represent increases of 22.1–43.8% relative to the unconditional mean probability for the one-year horizon and 22.6–58.6% for the three-year horizon.[21] This suggests that quality as indicated by insider information has persistent associations with future brand value.

[Insert Table 8]

6.3 Quality and Future Profitability

Prior research provides evidence that quality management practices improve operating performance (Hendricks and Singhal, 1997) and that customer satisfaction serves as a leading

[20] https://brandirectory.com/

[21] The economic magnitudes are computed based on Probit regressions. Results based on OLS are quantitatively similar.

indicator of future profitability (Ittner and Larcker, 1998). We build on this literature by examining how our quality measures help predict future profitability.

We estimate the following equation:

$$\textit{(Ltm) Profitability}_{i,t+1} = \beta_0 + \beta_1 \textit{Quality Index}_{i,t} + \gamma \mathbf{X}_{i,t} + \delta_{j,t} + \epsilon_{i,t} \tag{4}$$

where *(Ltm) Profitability* refers to either (long-term) return on assets (*ROA*) or operating margin (*Operating Margin*). $X_{i,t}$ is a set of control variables including *Size*, *Leverage*, *M/B ratio*, and *Sales Growth*.

Table 9 presents the results. Panel A reports the association between quality and future return on assets (*ROA*), and Panel B reports the association between quality and future operating margin (*Operating Margin*). We find that quality indices are positively and significantly associated with future profitability in both the short and long term. In terms of economic magnitude, a one–standard–deviation increase in our quality index is associated with a 0.4–0.6 percentage-point improvement in ROA in the subsequent year and a 0.2–0.5 percentage-point improvement in average ROA over the next three years, corresponding to 5.0–9.6% increases in one-year ROA and 5.2–8.3% increases in three-year ROA relative to the unconditional means. A one–standard–deviation increase in our quality index is associated with a 0.5–0.7 percentage-point improvement in operating margin in the subsequent year and a 0.5–0.6 percentage-point improvement in average operating margin over the next three years. These effects correspond to 6.2–10.6% increases in one-year operating margin and 5.1–11.1% increases in three-year operating margin relative to the unconditional means. Overall, our results suggest that sustained quality provision plays an important role in driving future financial performance.

[Insert Table 9]

7 Additional Analyses

7.1 Quality in Product versus Service Industries

We examine the predictive power of employee reviews for quality provision separately by industry type. This test is motivated by the idea that the determinants of quality may vary depending on the nature of the industry. In service industries, where customer experience is heavily influenced by employee behavior and frontline interactions, employee perceptions may be particularly informative about quality outcomes. To explore this heterogeneity, we classify industries as either product-oriented or service-oriented based on the names and descriptions of sub-industry groupings within the Fama-French 48 industry classification.[22] We use ChatGPT to generate initial classifications (see prompt in Online Appendix), and then manually review and refine the assignments.

We train separate machine learning models on subsamples split by industry orientation and evaluate their out-of-sample performance. Results for service-oriented industries are presented in Table 10, and product-oriented industries in Table 11. The models exhibit stronger predictive performance in service-oriented industries, suggesting that employee reviews may provide more informative signals about quality in service contexts.

[Insert Table 10 and Table 11]

7.2 Assessing Informativeness Beyond Management Disclosure

To assess whether our quality index captures bottom-up insights from employees beyond what is already reflected in the messaging or discourse of management, we re-estimate our out-of-sample performance test with additional controls for conference call-based cultural values. The cultural values include five dimensions: innovation, integrity, quality, respect,

[22] Available at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_48_ind_port.html

and teamwork, measured using word embedding models following Li et al. (2021). Specifically, we construct firm-year-level scores by first training a Word2Vec model on earnings call transcripts to expand a set of seed words for each cultural dimension, following the predefined lists in Li et al. (2021). We then apply a TF-IDF weighting scheme to compute the relative emphasis placed on each cultural dimension in the language used by top management. Each score quantifies the extent to which the respective cultural dimension is conveyed in the tone and language of top management during conference calls.

The results are reported in Table 12. We find that controlling for cultural values conveyed by top management in conference calls does not significantly affect the predictive power of our quality index in out-of-sample tests. In particular, even the culture score on *quality* derived from conference calls does not subsume the predictive power of our quality index. This suggests that our quality index captures information beyond what is reflected in top management's explicit messaging about quality, highlighting the unique value of employee perspectives in assessing actual product or service quality.

[Insert Table 12]

7.3 Assessing Informativeness Beyond Glassdoor-Implied Cultural Values

As a robustness test, we re-estimate our out-of-sample performance test with additional controls for cultural values derived from Glassdoor reviews. These values are constructed following the same procedure as described in the previous section, but based on the Glassdoor review corpus. Each score reflects the extent to which a firm's employee reviews reflect the respective cultural dimension.

The results are presented in Table 13. We find that including controls for cultural values derived from Glassdoor reviews has no significant impact on the predictive power of our quality index in out-of-sample tests. This suggests that our index captures information

beyond any individual cultural value explicitly discussed in employee reviews—including *quality* culture—or any linear combination of several cultural values.

[Insert Table 13]

7.4 Assessing Measurement Error

A potential concern is that part of the documented variation in our quality indices—in particular, the sizable firm-level variation—may simply reflect idiosyncratic measurement error. We perform a battery of tests to address this concern. First, we observe that our extensive validation tests reveal strong and significant associations between the quality index and relevant outcomes, indicating that the variation reflected in our quality index is not simply noise but captures real economic content.

Second, we follow Hassan et al. (2019) and Sautner et al. (2023) and assess the extent of measurement error in firm-level variation by assuming that a firm's underlying product or service quality follows a first-order autoregressive (AR) process. We further assume that our constructed quality indices are noisy signals of this latent quality, with the noise taking the form of classical i.i.d. measurement error. We exploit the fact that measurement error attenuates the estimated persistence coefficient in an AR(1) regression. Specifically, when we regress our quality index on its lagged value using OLS, the presence of measurement error biases the autocorrelation coefficient toward zero. However, if a valid instrument for the lagged quality index ($Quality\ Index_{i,t-1}$) can be identified, IV estimation can correct for this attenuation bias and allows us to recover the true persistence parameter. By comparing the OLS and IV estimates of the autocorrelation parameter, we can quantify the share of variation due to measurement error as $1 - (\hat{\beta}_{OLS}/\hat{\beta}_{IV})$. Intuitively, a larger divergence between the OLS and IV coefficients indicates that a greater proportion of the observed variation in our quality measures reflects noise rather than true underlying quality differences.

Table 14 provides three specifications implementing this approach. Columns (1) and (5) present the OLS estimates of the autocorrelation in the *Quality Index*. Columns (2) and (6)

instrument the quality index using its second lag. Columns (3)–(4) and (7)–(8) instrument the quality index using the first and second lags constructed from the alternative Glassdoor review section (i.e., for the pros-based index, we use the cons-based index as the instrument, and vice versa).[23] The estimates of the implied share of measurement error in our quality indices suggest that between 13.5% and 21.1% of the overall variation in the random forest–constructed indices and between 41.7% and 50.6% of the overall variation in the XGBoost–constructed indices can be attributed to measurement error. For firm-level variations, between 16.2% and 23.3% of the firm-level variation in the random forest–constructed indices and between 49.9% and 56.5% of the firm-level variation in the XGBoost–constructed indices can be attributed to measurement error. We note that the estimated measurement error shares of the quality indices constructed using random forests compare favorably to those documented for other firm-level variables under similar assumptions, such as the measures of total factor productivity in Bloom et al. (2018) and Collard-Wexler (2011), and the measures of firm-level political risk constructed by Hassan et al. (2019). The XGBoost-constructed indices yield measurement error shares that are broadly comparable to those reported in these studies. For example, Hassan et al. (2019) report measurement error shares ranging from 40.6% to 49.6% for overall variation and from 44.5% to 54.1% for firm-level variation in their measure of firm-level political risk. Taken together, the evidence provides reassurance that the variation in our quality indices indeed reflects economically meaningful differences in firm quality provision and is not merely noise.

[Insert Table 14]

[23] The underlying assumption is that, while each index is a noisy proxy for the same latent firm-level quality, the measurement errors in the pros- and cons-based indices are plausibly independent, since they are derived from fundamentally different textual inputs and are generated using separate machine learning models.

8 Measuring Quality with Generative AI

In light of recent developments in large language models and generative AI, we construct alternative measures of firm quality provision using GPT models from OpenAI. To explore the potential and limitations of generative AI for our measurement task, we use two approaches to construct quality measures: (1) a zero-shot, prompt-based approach and (2) a supervised fine-tuning approach.

8.1 Zero-Shot GPT-based Quality Measures

We prompt ChatGPT to categorize firm quality provision for each employee review.[24] We carefully review prior studies determinants of quality provision and compile a detailed explanation of the rubric for classifying firm quality provision. We then provide GPT with the task statement, rubric, and employee reviews. We instruct GPT to respond the following: a label of firm quality provision, explanation, relevance, and confidence level. For each employee review, based on the rubric provided, GPT selects one label from three categories: (1) Q0 – low quality, (2) Q1 – moderate quality, and (3) Q2 – high quality. In terms of review relevance labeling, GPT determines whether a review explicitly mentions quality issues, implicitly suggests them, or is irrelevant to firm quality provision. GPT also outputs a confidence level, classified as low, moderate, or high. The prompt submitted to the OpenAI GPT API is presented in Online Appendix.

We find that zero-shot, GPT-based quality measures are directionally correlated with industry-adjusted ACSI scores, with correlations ranging from 0.046 to 0.118 (all significant at $p < 0.10$). We further evaluate the effectiveness of zero-shot, GPT-based quality measures by examining their predictive power for industry-adjusted ACSI scores. This test corresponds to the out-of-sample performance test of machine learning-based quality mea-

[24] We use GPT-4.1, the flagship GPT model at the time of analyzing the data, as the language model to process texts. Due to computational resource constraints, we randomly sample 20% of reviews for each firm-year for analysis.

sures. The results are presented in Table 15, which matches Table 3. We find that zero-shot, GPT-based quality measures demonstrate modest performance in explaining variation in firm-level quality provision after controlling for firm characteristics and Glassdoor overall ratings. However, these associations tend to be mostly subsumed with the inclusion of sub-dimensional Glassdoor ratings. We also conduct the same set of validation and implication tests as in Section 5 and Section 6 to more comprehensively assess the effectiveness of the zero-shot, GPT-based quality measures. In analyses that examine firm-level quality variation as measured by MSCI quality scores, distinguish firms by their reputational standing in quality, and predict future product recalls, the zero-shot, GPT-based quality measures reproduce patterns broadly consistent with those obtained with the machine-learning-based measures. In other settings, however, their performance is noticeably attenuated and exhibits lower precision. We present the results of additional validation and implication tests of zero-shot, GPT-based measures in the Online Appendix. Taken together, these results suggest that, while less precise than our supervised machine-learning measures in predicting firm-level quality provision, the zero-shot, GPT-based quality measures deliver moderate predictive accuracy.

[Insert Table 15]

8.2 Fine-Tuned GPT-based Quality Measures

To construct GPT-based quality measures using supervised fine-tuning GPT models, we first transform industry-adjusted ACSI scores into five ordinal quality categories to formulate a classification task suitable for fine-tuning: (1) Low, (2) Medium-Low, (3) Medium, (4) Medium-High, and (5) High. These categories are defined based on quintiles of the industry-adjusted ACSI distribution within our sample. We randomly sample reviews for firm-years where the total token count exceeds 15,000 due to token limitations and cost considerations. This constraint filters out approximately 3% of firm-years for pros reviews and 5% for cons reviews. To ensure the model infers quality provision based on generalizable quality patterns

rather than firm-specific identifiers, we employ a transformer-based Named Entity Recognition (NER) model to de-identify all references to organizations, persons, geopolitical entities, and products in the employee reviews.

We fine-tune two separate GPT-4o-mini models, one based on the pros section of employee reviews and the other based on the cons section of employee reviews. For each model, we structure the fine-tuning task as follows. The training and testing sample split mirrors that used for the machine-learning-based measures. At the firm-year level, we provide the model with employee reviews as input, accompanied by a system prompt instructing the model to assess firm quality provision based on the information reflected in the reviews. The corresponding industry-adjusted ACSI quintile serves as the target label. Each training example thus consists of firm-year-aggregated employee reviews as the user input and the observed quality quintile as the model's target response. The prompts are provided in the Online Appendix.

After fine-tuning, we apply the models to both the hold-out test sample and the broader application sample of S&P 1500 firms without ACSI coverage. For each firm-year, the model outputs a predicted quality category. We convert these predicted categories into an ordinal numeric score ranging from 1 to 5, which serves as our fine-tuned GPT quality measure. As with the machine-learning-based measures, we construct separate pros-based and cons-based indices and an aggregate index that is calculated as the token-count-weighted average of the two.

We start by examining the predictive power of fine-tuned, GPT-based quality measures for industry-adjusted ACSI scores in out-of-sample tests. The results are presented in Table 16. We find that the fine-tuned GPT quality measures show strong out-of-sample predictive performance across all specifications, indicating that supervised fine-tuning enables the GPT model to better map employee language to quality outcomes. However, their predictive performance, as measured by adjusted R-squared, remains relatively weaker than that of our main machine-learning-based measures.

We also conduct the same validation and implication tests performed for our main machine learning-based measures, with results reported in the Online Appendix. The findings are largely consistent with those from the machine learning-based measures. Specifically, the fine-tuned GPT quality measures show consistent correlations with MSCI quality-related indicators, distinguish between firms ranked as best quality versus others, and are positively associated with future brand value and profitability as measured by ROA. We find no significant results for the relationship between quality deterioration (measured by the change in fine-tuned GPT measures) and future product recalls. This may be explained by the coarseness of the five-category quality classification, which provides insufficient granularity to capture quality deterioration. Overall, these results provide complementary evidence on the link between internal employee perspectives and a firm's external quality provision, and demonstrate the potential of fine-tuned generative AI models to extract meaningful quality signals from employee reviews.

[Insert Table 16]

8.3 Discussion of Measuring Quality with Generative AI

Our findings show that GPT-based measures can perform reasonably well but are not uniformly effective across tasks and approaches. Importantly, while the zero-shot, prompt-based approach serves as a practical and accessible solution, it generally produces relatively less precise predictions. On the other hand, the supervised fine-tuning approach exhibits performance that is generally consistent with our main machine-learning-based approach, but it is limited to tasks that require ordinal classification at a coarser scale rather than continuous scores.

The relatively lower precision of the zero-shot GPT-based approach in our context can be explained by several underlying factors. First, while GPT models excel at extracting surface-level semantic patterns, firm quality provision is a latent construct, and employee reviews often do not explicitly mention quality outcomes. In our data, GPT classifies only

3.80% of the reviews in the pros section and 8.65% of the reviews in the cons section as explicitly mentioning quality outcomes. Supervised machine learning models, by contrast, are trained on pre-labeled data and can learn non-obvious predictive features. Second, zero-shot GPT models (not providing the model with any pre-labeled examples), even with carefully designed instructions, may default to general sentiment analysis or cultural tone classification instead of focusing on operational quality indicators. This is consistent with our two observations: (1) zero-shot GPT-based quality measures are strongly correlated with zero-shot GPT-based cultural measures, with correlations ranging from 0.707 to 0.743 (all significant at $p < 0.01$); (2) the predictive ability of zero-shot GPT-based quality measures is largely captured by Glassdoor ratings. These findings highlight that zero-shot GPT-based approaches offer a practical and accessible alternative when more technically demanding tools such as machine-learning models are unavailable, and users should exercise caution in applications where high predictive accuracy is critical.

Turning to the supervised fine-tuning approach, Figure OA6.1 in the Online Appendix illustrates the relationship between true and predicted quality quintiles using fine-tuned GPT models in the testing sample. We find that the predicted quality quintile generally increases with the true quality quintile, with higher true quintiles more likely to be classified into higher predicted quintiles. In addition, we observe two notable patterns. First, predicted classifications are concentrated in the middle categories, particularly the *Medium-Low* quintile. Second, the models tend to systematically under-predict. For example, the number of observations classified into the *High* quintile (n = 18 for the pros-based index and n = 13 for the cons-based index) is substantially smaller than the number of firms in the true high-quality group (n = 65). These observations suggest that while the fine-tuned GPT models capture meaningful quality variation, they show a conservative classification tendency that generally favors intermediate or lower categories.

An additional point to note is that we formulate the prediction problem as an ordinal classification task rather than a continuous score prediction when using fine-tuned GPT

models. This design choice aligns with the architectural nature of Large Language Models. LLMs are fundamentally designed as autoregressive classifiers that predict discrete tokens based on cross-entropy loss, rather than regression models that minimize continuous numerical error (McCoy et al., 2024). Prior work shows that language models can be poorly calibrated in numeric contexts (Lovering et al., 2025). To address these limitations, related research discretizes continuous targets into bins to form valid predictive distributions compatible with token generation (Requeima et al., 2024). Consistent with this literature, our fine-tuned GPT quality index uses a category-based scale to ensure robustness and stable inference. One limitation of doing so is that the resulting measure is relatively coarse and may obscure fine-grained variation in quality provision.

Our findings add to the emerging body of research leveraging generative large language models (GLLMs) such as ChatGPT as alternative tools to perform various textual analysis tasks. For example, Cai et al. (2026) use a zero-shot approach with a ChatGPT model to generate synthetic employee reviews and construct an alternative measure of organizational capital based on ChatGPT-generated reviews using embedding models. Berger et al. (2024) use a zero-shot approach with a ChatGPT model to construct measures of exposure to Generative AI from job postings. Armstrong et al. (2025) use a zero-shot approach with a ChatGPT model to generate alternative measures of firm exposure to government agencies from 10-Ks. Lu et al. (2025) fine-tune a GPT model to measure a firm's engagement in climate solutions from 10-K filings. Bernard et al. (2024) build a GPT model fine-tuned on narrative disclosures and inline XBRL tags to construct and validate a business-complexity measure. Our analysis adds further nuance to the literature by underscoring the importance of aligning GLLM applications with the required level of precision and with the desired trade-off between accessibility, consistency, and the granularity of measurement.

9 Conclusion

Product and service quality is critical to firm valuation creation, yet it is difficult for external stakeholders to observe. We examine whether qualitative information extracted from employee reviews is useful in predicting the quality of the products and services that their firms provide. Using machine learning approaches, we develop and validate a forward-looking, scalable, and time-varying firm-specific measurement of quality provision by utilizing employee-written reviews on Glassdoor as a credible and timely source of insider information on firm quality provision.

We construct text-based quality indices from over 4.3 million employee-written reviews for S&P 1500 firms spanning 2008 to 2023. These indices are trained to predict future product and service quality, as measured by ACSI scores. Our findings show that the text-based quality measures significantly enhance the explanatory power of regression models, outperforming all other readily observable firm characteristics such as size, leverage, and analyst coverage.

We validate our quality measures through out-of-sample performance tests and a series of robustness checks. The results consistently demonstrate that our text-based indices are the most important predictors of future ACSI scores across all model specifications. Furthermore, we find strong associations between our measures and alternative quality proxies, including MSCI product quality related scores and third-party product/service quality rankings.

We also show that our quality measures are associated with meaningful firm outcomes, including future product recalls, brand value, and profitability. Firms with higher quality scores experience fewer product recalls, greater brand strength, and superior future profitability, suggesting that our measures capture forward-looking information about firms' operational reliability and market performance. These results highlight the economic relevance of our quality indices and confirm that employee reviews offer insights into firm quality provision not captured by conventional metrics.

Finally, our exploration of Generative AI models provides additional perspective on the evolving toolkit for text-based measurement. We assess the potential and limitations of generative AI for our measurement task by constructing an alternative set of quality measures using a zero-shot prompt-based approach and a supervised fine-tuning approach with GPT models, respectively. Our results suggest that GPT-based measures perform reasonably well, but their effectiveness varies by task and implementation. Our exploration highlights the importance of considering the task's required precision, accessibility, prediction consistency, and measurement granularity when applying GLLMs.

We acknowledge that our measurement of firm quality provision is limited to firms that are covered by Glassdoor. Thus, the generalizability of our approach may be constrained by the availability and representativeness of employee reviews, particularly for smaller or less visible firms. Despite these limitations, we believe that our measurement offers a promising new lens through which academic researchers, investors, and other external stakeholders can directly assess firm quality provision in a forward-looking manner. Moreover, our study contributes to the growing field of crowdsourced employee information by highlighting the untapped potential of employee reviews in understanding organizational aspects that are less observable to external stakeholders in nature.

References

Akerlof, George A (1970) 'The market for 'lemons': quality uncertainty and the market mechanism.' *Market Failure or Success*

Anderson, Eugene W, and Sattar A Mansi (2009) 'Does customer satisfaction matter to investors? findings from the bond market.' *Journal of Marketing Research* 46(5), 703–714

Armstrong, Daphne M, Stephen Glaeser, and Jeffrey L Hoopes (2025) 'Measuring firm exposure to government agencies.' *Journal of Accounting and Economics* 79(1), 101703

Arrow, Kenneth J. (1965) 'Uncertainty and the welfare economics of medical care: Reply (the implications of transaction costs and adjustment lags).' *The American Economic Review* 55(1/2), 154–158

Berger, Philip G, Wei Cai, Lin Qiu, and Cindy Xinyi Shen (2024) 'Employer and employee responses to generative ai: Early evidence.' *Available at SSRN 4874061*

Bernard, Darren, Elizabeth Blankespoor, Ties de Kok, and Sara Toynbee (2024) 'Using gpt models to measure the complexity of business transactions.' *Available at SSRN 4480309*

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts (2013) 'Does management matter? evidence from india.' *The Quarterly Journal of Economics* 128(1), 1–51

Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J Terry (2018) 'Really uncertain business cycles.' *Econometrica* 86(3), 1031–1065

Bushee, Brian J (2001) 'Do institutional investors prefer near-term earnings over long-run value?' *Contemporary Accounting Research* 18(2), 207–246

Cai, Wei, Andrea Prat, and Jiehang Yu (2026) 'Measuring organizational capital.' *Journal of Accounting Research, forthcoming*

Cai, Wei, Dennis Campbell, and Jiehang Yu (2025) 'The signaling value of internal employee coordination.' *Journal of Accounting Research* 63(5), 1953–1993

Campbell, Dennis W, and Ruidi Shang (2022) 'Tone at the bottom: Measuring corporate misconduct risk from the text of employee reviews.' *Management Science* 68(9), 7034–7053

Chamberlin, Edward H. (1953) 'The product as an economic variable.' *The Quarterly Journal of Economics* 67(1), 1–29

Chen, Yaxuan (2024) 'Inconsistency in organizational culture.' *Working paper*

Chen, Yubo, Shankar Ganesan, and Yong Liu (2009) 'Does a firm's product-recall strategy affect its financial value? an examination of strategic alternatives during product-harm crises.' *Journal of Marketing* 73(6), 214–226

Cockrell, Seth, Wesley Friske, Clay M Voorhees, and Roger J Calantone (2024) 'The effects of innovation on product recall likelihood.' *Journal of Business Research* 173, 114452

Collard-Wexler, Allan (2011) 'Productivity dispersion and plant selection in the ready-mix concrete industry.' *US Census Bureau Center for Economic Studies Paper No. CES-WP-11-25*

Crosby, Philip B (1979) 'Quality is free-if you understand it.' *Winter Park Public Library History and Archive Collection*

Dehaan, Ed, Nan Li, and Frank S Zhou (2023) 'Financial reporting and employee job search.' *Journal of Accounting Research* 61(2), 571–617

Deming, W E (1982) *Quality, productivity and competitive position* (Cambridge: MIT Press)

Dyck, Alexander, Adair Morse, and Luigi Zingales (2010) 'Who blows the whistle on corporate fraud?' *The Journal of Finance* 65(6), 2213–2253

Frankel, Richard, Jared Jennings, and Joshua Lee (2022) 'Disclosure sentiment: Machine learning vs. dictionary methods.' *Management Science* 68(7), 5514–5532

George, Bill (2024) Available at: https://www.library.hbs.edu/working-knowledge/why-boeings-problems-with-737-max-began-more-than-25-years-ago

Graham, John R, Jillian Grennan, Campbell R Harvey, and Shivaram Rajgopal (2022) 'Corporate culture: Evidence from the field.' *Journal of Financial Economics* 146(2), 552–593

Green, T Clifton, Ruoyan Huang, Quan Wen, and Dexin Zhou (2019) 'Crowdsourced employer reviews and stock returns.' *Journal of Financial Economics* 134(1), 236–251

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2015) 'The value of corporate culture.' *Journal of Financial Economics* 117(1), 60–76

Hales, Jeffrey, James R Moon Jr, and Laura A Swenson (2018) 'A new era of voluntary disclosure? empirical evidence on how employee postings on social media relate to future corporate disclosures.' *Accounting, Organizations and Society* 68, 88–108

Hassan, Tarek A, Stephan Hollander, Laurence Van Lent, and Ahmed Tahoun (2019) 'Firm-level political risk: Measurement and effects.' *The Quarterly Journal of Economics* 134(4), 2135–2202

Hendricks, Kevin B, and Vinod R Singhal (1996) 'Quality awards and the market value of the firm: An empirical investigation.' *Management Science* 42(3), 415–436

_ (1997) 'Does implementing an effective tqm program actually improve operating performance? empirical evidence from firms that have won quality awards.' *Management Science* 43(9), 1258–1274

Huang, Kelly, Meng Li, and Stanimir Markov (2020) 'What do employees know? evidence from a social media platform.' *The Accounting Review* 95(2), 199–226

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi (1997) 'The effects of human resource management practices on productivity: A study of steel finishing lines.' *The American Economic Review* 87(3), 291–313

Ittner, Christopher D, and David F Larcker (1998) 'Are nonfinancial measures leading indicators of financial performance? an analysis of customer satisfaction.' *Journal of Accounting Research* 36, 1–35

Ittner, Christopher D, Venky Nagar, and Madhav V Rajan (2001) 'An empirical examination of dynamic quality-based learning models.' *Management Science* 47(4), 563–578

Lazear, Edward P (2000) 'Performance pay and productivity.' *American Economic Review* 90(5), 1346–1361

Leland, Hayne E (1979) 'Quacks, lemons, and licensing: A theory of minimum quality standards.' *Journal of Political Economy* 87(6), 1328–1346

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan (2021) 'Measuring corporate culture using machine learning.' *The Review of Financial Studies* 34(7), 3265–3315

Li, Yinghua, Yupeng Lin, Xiaoqiao Wang, and Shijie Yang (2024) 'Wall street and product quality: The duality of analysts.' *The Accounting Review* 99(5), 387–420

Loughran, Tim, and Bill McDonald (2016) 'Textual analysis in accounting and finance: A survey.' *Journal of Accounting Research* 54(4), 1187–1230

Lovering, Charles, Michael Krumdick, Viet Dac Lai, Varshini Reddy, Seth Ebner, Nilesh

Kumar, Rik Koncel-Kedziorski, and Chris Tanner (2025) 'Language model probabilities are not calibrated in numeric contexts.' In 'Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)' pp. 29218–29257

Lu, Shirley, George Serafeim, Simon Xu, and MarcAntonio Awada (2025) 'Tracking business opportunities for climate solutions using ai in regulated accounting reports.' *Nature Communications* 16(1), 9769

McCoy, R Thomas, Shunyu Yao, Dan Friedman, Mathew D Hardy, and Thomas L Griffiths (2024) 'Embers of autoregression show how large language models are shaped by the problem they are trained to solve.' *Proceedings of the National Academy of Sciences* 121(41), e2322420121

Milgrom, Paul, and John Roberts (1986) 'Relying on the information of interested parties.' *The RAND Journal of Economics* pp. 18–32

_ (1992) *Economics, Organization and Management,* vol. 7 (Prentice-hall Englewood Cliffs, NJ)

Nagar, Venky, and Madhav V Rajan (2001) 'The revenue implications of financial and operational measures of product quality.' *The Accounting Review* 76(4), 495–513

Nelson, Phillip (1970) 'Information and consumer behavior.' *Journal of Political Economy* 78(2), 311–329

Otto, Ashley S, David M Szymanski, and Rajan Varadarajan (2020) 'Customer satisfaction and firm performance: insights from over a quarter century of empirical research.' *Journal of the Academy of Marketing Science* 48(3), 543–564

Porter, M E (1985) *The competitive advantage: Creating and sustaining superior performance* (NY: Free Press)

Requeima, James, John Bronskill, Dami Choi, Richard Turner, and David K Duvenaud (2024) 'Llm processes: Numerical predictive distributions conditioned on natural language.' *Advances in Neural Information Processing Systems* 37, 109609–109671

Ronnen, Uri (1991) 'Minimum quality standards, fixed costs, and competition.' *The RAND Journal of Economics* pp. 490–504

Sautner, Zacharias, Laurence Van Lent, Grigory Vilkov, and Ruishen Zhang (2023) 'Firm-level climate change exposure.' *The Journal of Finance* 78(3), 1449–1498

Shapiro, Carl (1983) 'Premiums for high quality products as returns to reputations.' *The Quarterly Journal of Economics* 98(4), 659–679

Siano, Federico (2025) 'The news in earnings announcement disclosures: Capturing word context using llm methods.' *Management Science*

Spence, Michael (1978) 'Job market signaling.' In 'Uncertainty in Economics' (Elsevier) pp. 281–306

Sterman, John, and James Quinn (2023) 'Boeing's 737 max 8 disasters'

Tadelis, Steven (1999) 'What's in a name? reputation as a tradeable asset.' *American Economic Review* 89(3), 548–563

Tayan, Brian (2019) 'The wells fargo cross-selling scandal.' Available at: https://corpgov.law.harvard.edu/2019/02/06/the-wells-fargo-cross-selling-scandal-2/

Wolter, Jeremy S, Dora Bock, Jeremy Mackey, Pei Xu, and Jeffery S Smith (2019) 'Employee satisfaction trajectories and their effect on customer satisfaction and repatronage intentions.' *Journal of the Academy of Marketing Science* 47, 815–836

Zou, Fan, Yan Dong, Sining Song, and Manus Rungtusanatham (2023) 'Product recalls and

supply base innovation.' *Manufacturing & Service Operations Management* 25(5), 1931–1946

Figures

Figure 1. Sample Construction and Machine Learning Methodology

Notes: This figure illustrates the sample construction and machine learning procedure used to generate firm-year level quality indices.

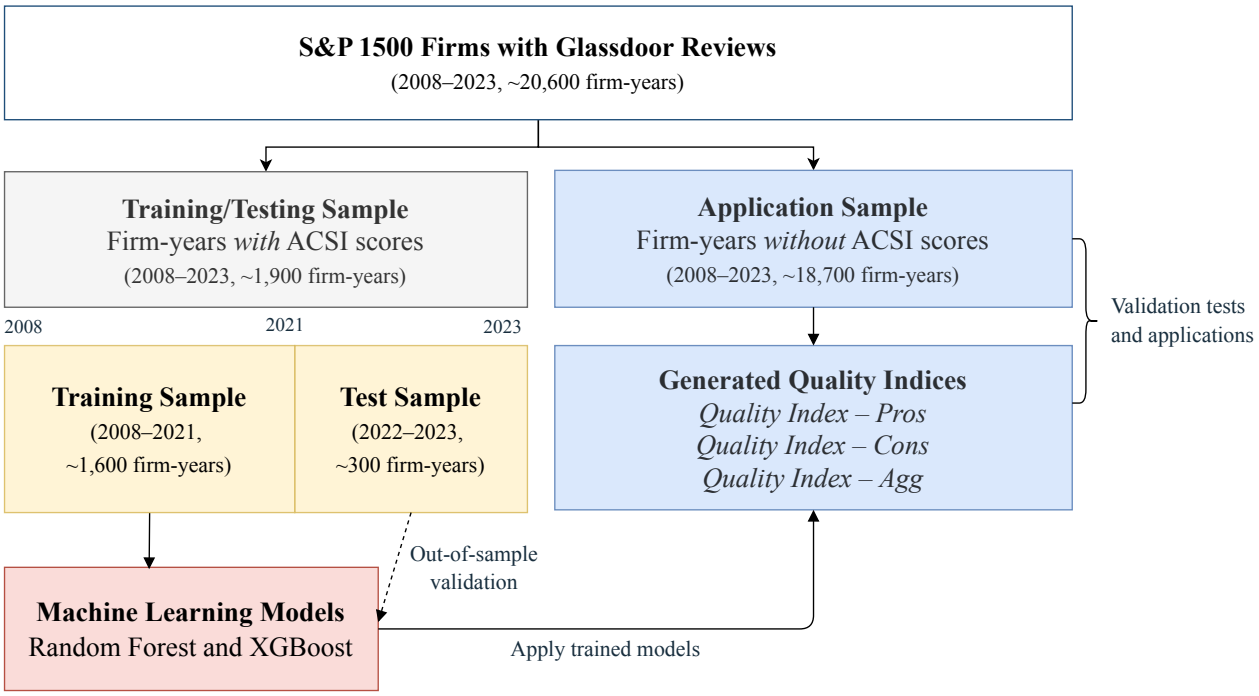
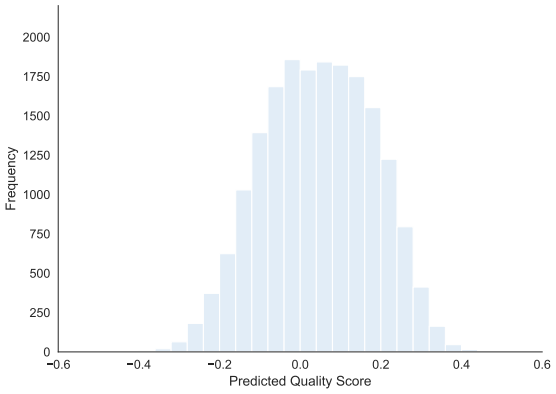
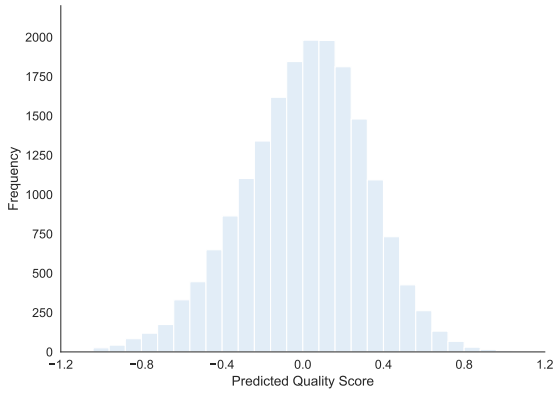


Figure 2. Distribution of Predicted Quality Scores

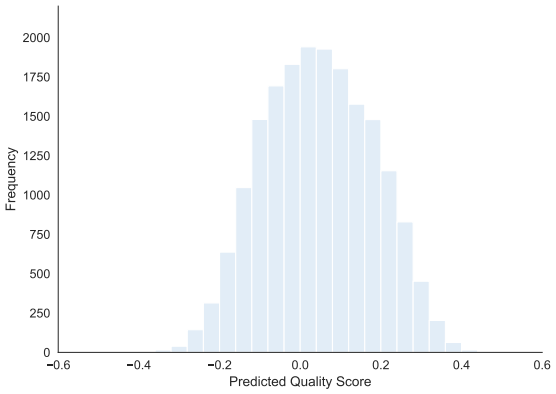
Notes: This figure shows the distribution of predicted quality scores in the application sample for both the pros and cons sections using random forest and XGBoost models. The sample period is 2008 to 2023. Panel (a) plots the distribution of the RF predictions for the pros section, and Panel (b) plots the distribution of XGBoost predictions for the pros section. Panel (c) plots the distribution of RF predictions for the cons section, and Panel (d) plots the distribution of XGBoost predictions for the cons section. Panel (e) plots the distribution of RF predictions for the aggregate quality index (weighted average of pros and cons sections by section length), and Panel (f) plots the distribution of XGBoost predictions for the aggregate quality index.



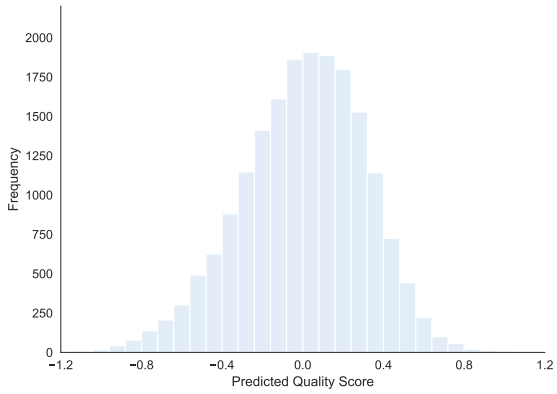
(a) *Pros* Section, RF



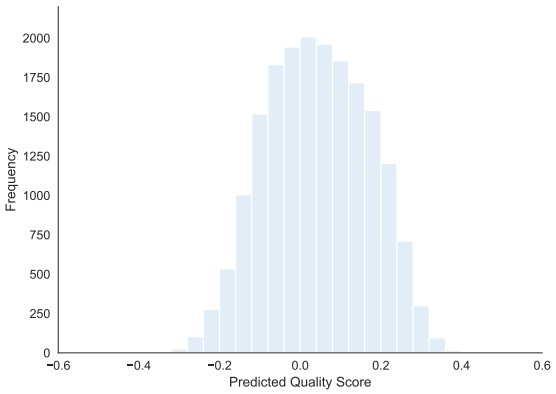
(b) *Pros* Section, XGBoost



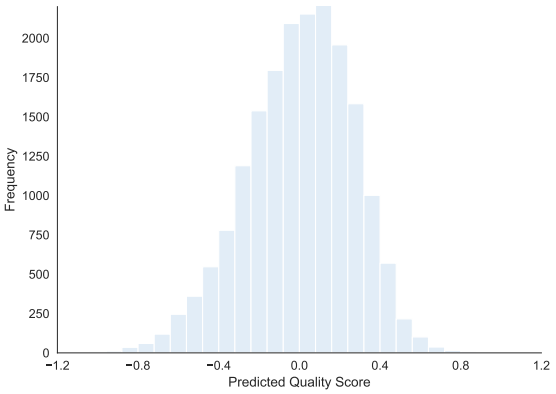
(c) *Cons* Section, RF



(d) *Cons* Section, XGBoost



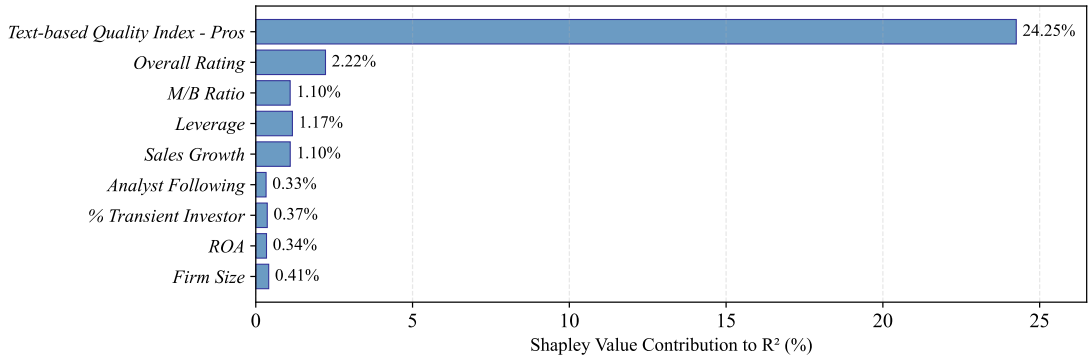
(e) Aggregate, RF



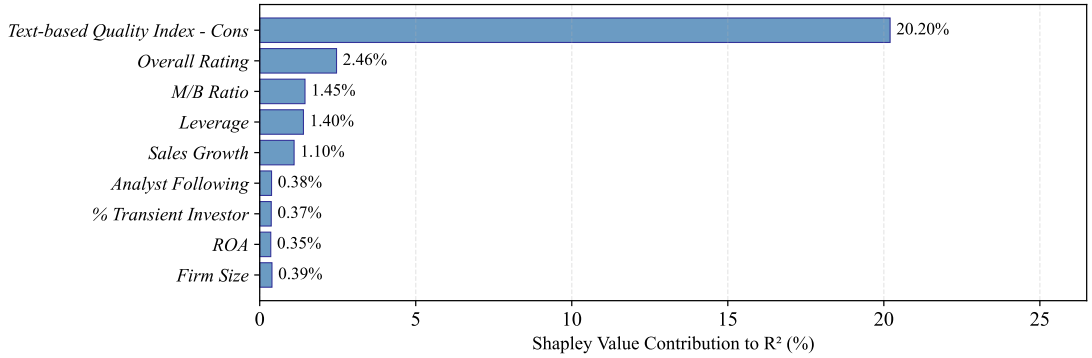
(f) Aggregate, XGBoost

Figure 3. Variable Contributions to R^2 in Out-of-Sample Tests

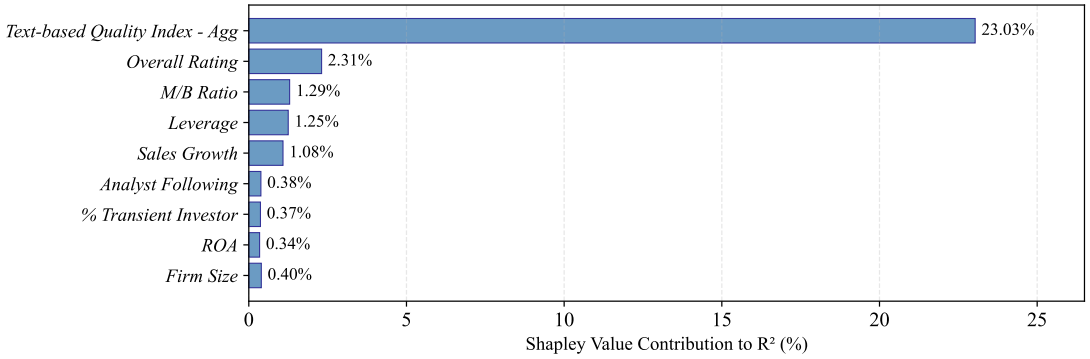
Notes: This figure presents the marginal contribution of each variable to the overall R^2 in out-of-sample prediction tests. The dependent variable is *Industry-Adjusted ACSI*. Industry and year fixed effects are included. Standard errors are clustered at the firm level. All continuous variables are winsorized at the 1st and 99th percentiles. The marginal contribution is calculated as the difference between the full model R^2 and the R^2 obtained when excluding each variable. Panel (a) shows results of quality index based on pros section constructed using random forest model. Panel (b) shows results of quality index based on cons section constructed using random forest model. Panel (c) shows results for the aggregate quality index constructed using random forest model. Panel (d) shows results of quality index based on pros section constructed using XGBoost model. Panel (e) shows results of quality index based on cons section constructed using XGBoost model. Panel (f) shows results for the aggregate quality index constructed using XGBoost model.



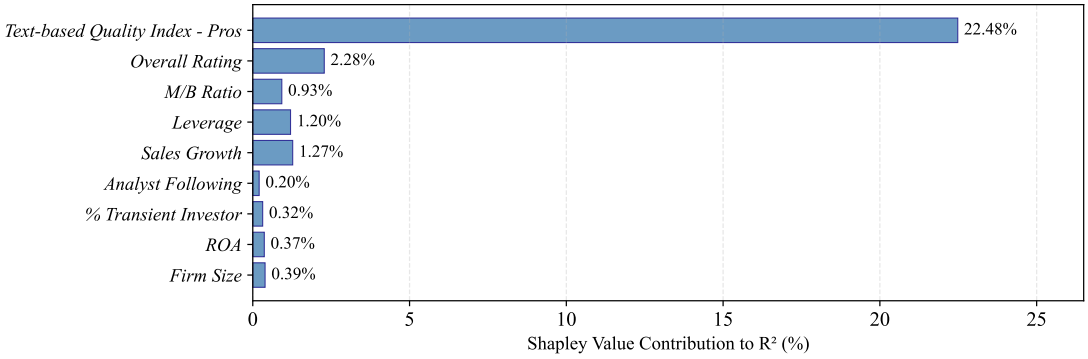
(a) *Pros* Section, RF



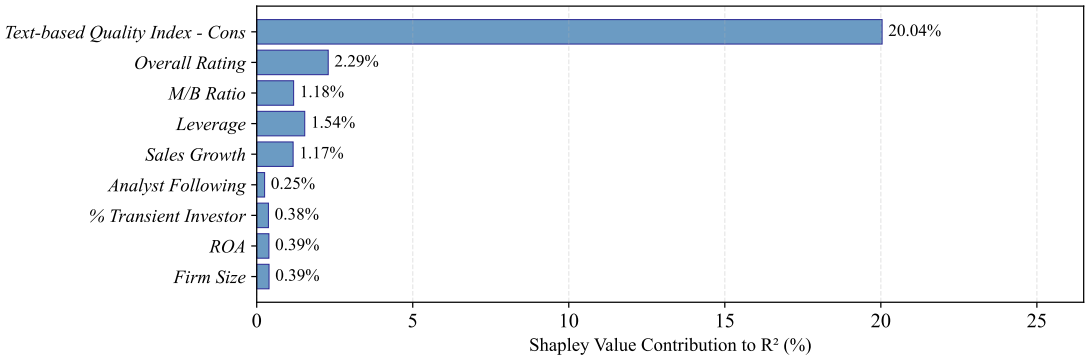
(b) *Cons* Section, RF



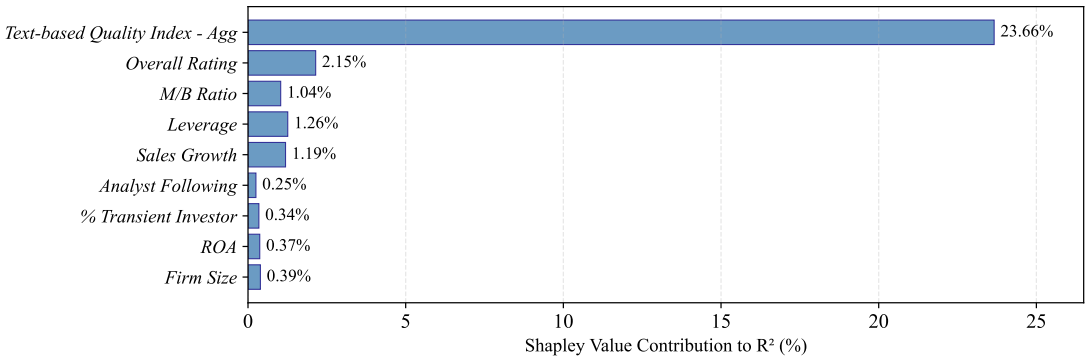
(c) Aggregate, RF



(d) *Pros* Section, XGBoost



(e) *Cons* Section, XGBoost



(f) Aggregate, XGBoost

Tables

Table 1. Summary Statistics

Notes: This table presents summary statistics for our sample. Panel A shows the industry distribution based on Fama-French 48 industry classification. Panel B reports yearly distribution of the training/testing sample (ACSI score available) and the application sample (ACSI score unavailable). Panel C reports descriptive statistics for key variables used in the analysis. All variables are defined in Appendix B. All continuous variables are winsorized at the 1% and 99% levels.

Panel A: Industry Distribution

No.	Fama-French 48 Industry	Count	Percent
1	Agriculture	48	0.23%
2	Food Products	350	1.70%
3	Candy & Soda	69	0.34%
4	Beer & Liquor	77	0.37%
5	Tobacco Products	42	0.20%
6	Recreation	106	0.52%
7	Entertainment	272	1.32%
8	Printing and Publishing	171	0.83%
9	Consumer Goods	361	1.76%
10	Apparel	339	1.65%
11	Healthcare	483	2.35%
12	Medical Equipment	552	2.68%
13	Pharmaceutical Products	590	2.87%
14	Chemicals	408	1.98%
15	Rubber and Plastic Products	81	0.39%
16	Textiles	25	0.12%
17	Construction Materials	313	1.52%
18	Construction	260	1.26%
19	Steel Works Etc	163	0.79%
20	Fabricated Products	17	0.08%
21	Machinery	656	3.19%
22	Electrical Equipment	177	0.86%
23	Automobiles and Trucks	320	1.56%
24	Aircraft	190	0.92%
25	Shipbuilding, Railroad Equipment	36	0.18%
26	Defense	44	0.21%
27	Precious Metals	26	0.13%
28	Non-Metallic and Industrial Metal Mining	53	0.26%
29	Coal	22	0.11%
30	Petroleum and Natural Gas	586	2.85%
31	Utilities	500	2.43%
32	Communication	441	2.14%
33	Personal Services	247	1.20%
34	Business Services	2,937	14.28%
35	Computers	668	3.25%
36	Electronic Equipment	1,023	4.97%
37	Measuring and Control Equipment	363	1.76%
38	Business Supplies	156	0.76%
39	Shipping Containers	92	0.45%
40	Transportation	599	2.91%
41	Wholesale	665	3.23%
42	Retail	1,827	8.88%
43	Restaurants, Hotels, Motels	619	3.01%
44	Banking	1,337	6.50%
45	Insurance	856	4.16%
46	Real Estate	133	0.65%
47	Trading	1,069	5.20%
48	Other	198	0.96%
	Total	20,567	100.00%

Table 1. Summary Statistics (Continued)

Panel B: Yearly Distribution by Sample Type

Year	Training/Testing Sample	Application Sample	Total
2008	67	243	310
2009	72	315	387
2010	86	557	643
2011	88	683	771
2012	97	833	930
2013	101	982	1,083
2014	105	1,246	1,351
2015	140	1,416	1,556
2016	140	1,481	1,621
2017	112	1,509	1,621
2018	144	1,442	1,586
2019	142	1,415	1,557
2020	150	1,478	1,628
2021	156	1,711	1,867
2022	151	1,710	1,861
2023	135	1,660	1,795
Total	1,886	18,681	20,567

Table 1. Summary Statistics (Continued)

Panel C: Descriptive Statistics

Variable	N	Mean	SD	P25	Median	P75	Sample
Compustat Firm Characteristics							
Size	17,194	8.677	1.722	7.441	8.570	9.833	Full
Leverage	16,938	0.293	0.220	0.115	0.268	0.421	Full
ROA	17,191	0.041	0.083	0.010	0.041	0.081	Full
M/B Ratio	14,103	2.277	1.619	1.282	1.733	2.620	Full
Sales Growth	17,038	0.080	0.199	-0.015	0.056	0.143	Full
R&D Intensity	9,482	0.043	0.059	0.001	0.020	0.064	Full
Operating Margin	14,530	0.111	0.124	0.051	0.106	0.175	Full
Ltm ROA	16,818	0.042	0.072	0.010	0.040	0.078	Full
Ltm Operating Margin	14,206	0.113	0.115	0.051	0.105	0.175	Full
Other Firm Characteristics							
% Transient Investor	20,567	0.118	0.117	0.000	0.115	0.203	Full
Analyst Following	20,567	1.715	1.222	0.000	2.079	2.773	Full
Glassdoor Ratings							
Overall Rating	20,564	3.328	0.496	3.000	3.347	3.670	Full
Balance Subrating	20,564	3.301	0.490	2.980	3.320	3.634	Full
Culture Subrating	18,453	3.283	0.559	2.905	3.303	3.667	Full
Opportunities Subrating	20,564	3.115	0.482	2.788	3.109	3.439	Full
Compensation Subrating	20,564	3.350	0.505	3.015	3.372	3.700	Full
Management Subrating	20,564	2.944	0.519	2.596	2.931	3.277	Full
Customer Satisfaction							
Industry-Adjusted ACSI	1,886	0.098	0.897	-0.415	0.187	0.717	Train/Test
Quality Indices							
Quality Index - Pros (RF)	18,681	0.045	0.140	-0.056	0.046	0.149	App
Quality Index - Cons (RF)	18,681	0.046	0.139	-0.056	0.044	0.147	App
Quality Index - Agg (RF)	18,681	0.044	0.131	-0.053	0.042	0.141	App
Quality Index - Pros (XGB)	18,681	0.008	0.310	-0.190	0.027	0.221	App
Quality Index - Cons (XGB)	18,681	0.003	0.309	-0.197	0.023	0.221	App
Quality Index - Agg (XGB)	18,681	0.003	0.267	-0.170	0.020	0.191	App
Validation Variables							
Quality Risk Exposure	7,878	4.171	2.225	2.900	4.200	5.600	App
Has Product Concern	1,695	0.256	0.437	0.000	0.000	1.000	App
Has Safety Concern	6,294	0.068	0.252	0.000	0.000	0.000	App
#Strengths - #Concerns	1,695	-0.135	0.749	0.000	0.000	0.000	App
Quality Management	2,662	4.924	2.442	3.200	5.200	6.700	App
Top Brand	18,681	0.059	0.236	0.000	0.000	0.000	App
Ltm Top Brand	17,021	0.048	0.213	0.000	0.000	0.000	App
# Recall	3,259	0.084	0.601	0.000	0.000	0.000	App
# Passive Recall	3,259	0.060	0.498	0.000	0.000	0.000	App

Table 1. Summary Statistics (Continued)

Variable	N	Mean	SD	P25	Median	P75	Sample
Glassdoor-Based Cultural Value Scores							
Innovation Score GD	283	0.004	0.011	-0.001	0.004	0.008	Test
Integrity Score GD	283	0.006	0.013	-0.001	0.003	0.011	Test
Quality Score GD	283	-0.008	0.009	-0.013	-0.008	-0.003	Test
Respect Score GD	283	0.000	0.010	-0.005	-0.001	0.006	Test
Teamwork Score GD	283	0.026	0.012	0.018	0.024	0.031	Test
Conference Call-Based Cultural Value Scores							
Innovation Score CC	283	0.011	0.007	0.005	0.009	0.014	Test
Integrity Score CC	283	0.004	0.003	0.002	0.003	0.005	Test
Quality Score CC	283	0.007	0.005	0.003	0.006	0.009	Test
Respect Score CC	283	0.004	0.002	0.003	0.004	0.005	Test
Teamwork Score CC	283	0.003	0.002	0.002	0.002	0.004	Test

Table 2. Variance Decomposition of Predicted Quality Scores

Notes: This table reports the Shapley value decomposition of the variance in predicted quality scores from random forest and XGBoost models. The decomposition includes year fixed effects, industry fixed effects (Fama-French 48), industry-year fixed effects, and firm fixed effects. The "firm-level" variation represents the combined contribution of permanent firm differences and unexplained within-firm variation over time.

	Quality Index - Pros		Quality Index - Cons		Quality Index - Agg		Average
	RF	XGBoost	RF	XGBoost	RF	XGBoost	
Year FE	4.9%	5.4%	2.6%	3.3%	3.9%	5.5%	4.3%
Industry FE	2.5%	1.8%	2.0%	1.1%	2.5%	1.7%	2.0%
Industry × Year FE	11.0%	10.6%	8.1%	7.9%	9.7%	10.5%	9.6%
"Firm-level"	**81.6%**	**82.2%**	**87.3%**	**87.7%**	**83.9%**	**82.3%**	**84.1%**
Permanent differences across firms (Firm FE)	46.3%	33.6%	45.3%	29.9%	47.6%	36.1%	39.8%
Within firms over time (Residual)	35.3%	48.6%	42.0%	57.8%	36.3%	46.2%	44.3%

Table 3. Out-of-Sample Validation of the Text-Based Quality Index

Notes: This table reports regression results of industry-adjusted ACSI scores on the text-based quality index in the hold-out sample. The dependent variable is the industry-adjusted ACSI score in year $t+1$. Panel A (B) presents results using the quality index derived from the pros (cons) section of employee reviews. Panels A.1 (B.1) and A.2 (B.2) show results of quality indices constructed using the random forest model and the XGBoost model, respectively. Panel C reports results using the aggregate quality index. Panels C.1 and C.2 present results from the random forest and XGBoost models, respectively. All continuous variables are winsorized at the 1st and 99th percentiles. T-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Quality Index Based on Pros Section

Panel A.1: Random Forest					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Pros	2.344***	2.254***	2.007***	1.940***	0.802***
	(10.79)	(7.64)	(6.75)	(6.48)	(3.07)
Size			0.002	0.048	0.031
			(0.03)	(0.59)	(0.61)
Leverage			-0.365	-0.356	-0.237
			(-1.47)	(-1.35)	(-1.38)
ROA			-0.268	-0.027	0.253
			(-0.21)	(-0.02)	(0.30)
M/B Ratio			0.106*	0.088	0.023
			(1.84)	(1.51)	(0.63)
Sales Growth			0.740**	0.742**	0.408
			(2.28)	(2.25)	(1.38)
% Transient Investor			0.768	0.471	0.340
			(0.92)	(0.59)	(0.65)
Analyst Following			-0.071	-0.054	-0.031
			(-1.16)	(-0.88)	(-0.77)
Overall Rating			0.369		
			(1.56)		
Balance Subrating				-0.222	-0.059
				(-0.73)	(-0.23)
Culture Subrating				1.047**	0.572
				(2.35)	(1.58)
Opportunities Subrating				-0.353	0.009
				(-0.77)	(0.03)
Compensation Subrating				-0.230	-0.119
				(-0.78)	(-0.58)
Management Subrating				-0.173	-0.273
				(-0.39)	(-0.87)
Industry-Adjusted $ACSI_{t-1}$					0.524***
					(6.75)
Constant	0.250***	0.151***	-1.376	-0.823	-0.598
	(5.15)	(2.84)	(-1.24)	(-0.83)	(-0.85)
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	286	286	250	250	250
Adjusted R^2	0.340	0.387	0.426	0.439	0.594

Table 3. Out-of-Sample Validation of the Text-Based Quality Index (Continued)

Panel A.2: XGBoost					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Pros	1.191***	1.115***	0.957***	0.911***	0.401***
	(10.51)	(8.11)	(6.32)	(6.09)	(2.78)
Size			-0.028	0.013	0.020
			(-0.39)	(0.18)	(0.40)
Leverage			-0.391	-0.361	-0.225
			(-1.54)	(-1.34)	(-1.30)
ROA			0.073	0.272	0.384
			(0.06)	(0.24)	(0.45)
M/B Ratio			0.085	0.068	0.013
			(1.48)	(1.13)	(0.35)
Sales Growth			0.844**	0.847**	0.447
			(2.51)	(2.46)	(1.48)
% Transient Investor			0.360	0.120	0.192
			(0.43)	(0.15)	(0.36)
Analyst Following			-0.022	-0.010	-0.014
			(-0.40)	(-0.18)	(-0.38)
Overall Rating			0.403*		
			(1.79)		
Balance Subrating				-0.268	-0.067
				(-0.91)	(-0.27)
Culture Subrating				1.179***	0.600*
				(2.74)	(1.71)
Opportunities Subrating				-0.364	0.008
				(-0.79)	(0.02)
Compensation Subrating				-0.136	-0.083
				(-0.51)	(-0.43)
Management Subrating				-0.335	-0.334
				(-0.74)	(-1.06)
Industry-Adjusted $ACSI_{t-1}$					0.531***
					(6.89)
Constant	0.370***	0.260***	-1.140	-0.507	-0.456
	(7.66)	(4.72)	(-1.06)	(-0.53)	(-0.65)
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	286	286	250	250	250
Adjusted R^2	0.318	0.376	0.415	0.428	0.595

Table 3. Out-of-Sample Validation of the Text-Based Quality Index (Continued)

Panel B: Quality Index Based on Cons Section

Panel B.1: Random Forest					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Cons	2.224***	2.013***	1.785***	1.784***	0.655***
	(7.35)	(6.30)	(5.84)	(6.00)	(2.87)
Size			-0.017	0.035	0.021
			(-0.22)	(0.43)	(0.42)
Leverage			-0.519**	-0.469*	-0.290*
			(-2.18)	(-1.79)	(-1.70)
ROA			-0.471	-0.201	0.198
			(-0.36)	(-0.17)	(0.23)
M/B Ratio			0.135**	0.113**	0.032
			(2.59)	(2.12)	(0.92)
Sales Growth			0.677**	0.702**	0.393
			(1.99)	(2.05)	(1.32)
% Transient Investor			0.734	0.428	0.315
			(0.82)	(0.53)	(0.58)
Analyst Following			-0.077	-0.064	-0.031
			(-1.23)	(-1.07)	(-0.79)
Overall Rating			0.434*		
			(1.71)		
Balance Subrating				-0.193	-0.057
				(-0.62)	(-0.22)
Culture Subrating				1.377***	0.715*
				(2.98)	(1.93)
Opportunities Subrating				-0.437	-0.015
				(-0.99)	(-0.05)
Compensation Subrating				-0.141	-0.078
				(-0.48)	(-0.39)
Management Subrating				-0.535	-0.422
				(-1.22)	(-1.33)
Industry-Adjusted $ACSI_{t-1}$					0.530***
					(6.75)
Constant	0.214***	0.122**	-1.405	-0.803	-0.588
	(4.12)	(2.21)	(-1.21)	(-0.79)	(-0.82)
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	286	286	250	250	250
Adjusted R^2	0.279	0.356	0.409	0.435	0.589

Table 3. Out-of-Sample Validation of the Text-Based Quality Index (Continued)

Panel B.2: XGBoost					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Cons	1.120***	1.056***	0.889***	0.846***	0.248**
	(9.83)	(8.16)	(5.79)	(5.80)	(2.21)
Size			-0.013	0.018	0.010
			(-0.18)	(0.23)	(0.20)
Leverage			-0.528**	-0.465	-0.295
			(-2.00)	(-1.65)	(-1.65)
ROA			0.011	0.183	0.341
			(0.01)	(0.15)	(0.40)
M/B Ratio			0.109*	0.090	0.022
			(1.96)	(1.60)	(0.62)
Sales Growth			0.750**	0.762**	0.406
			(2.28)	(2.29)	(1.33)
% Transient Investor			0.759	0.538	0.334
			(0.86)	(0.67)	(0.63)
Analyst Following			-0.051	-0.043	-0.020
			(-0.89)	(-0.77)	(-0.53)
Overall Rating			0.388		
			(1.51)		
Balance Subrating				-0.319	-0.105
				(-1.04)	(-0.42)
Culture Subrating				1.303***	0.690*
				(2.79)	(1.89)
Opportunities Subrating				-0.345	0.033
				(-0.80)	(0.10)
Compensation Subrating				-0.076	-0.052
				(-0.26)	(-0.26)
Management Subrating				-0.485	-0.402
				(-1.12)	(-1.25)
Industry-Adjusted $ACSI_{t-1}$					0.549***
					(7.52)
Constant	0.296***	0.207***	-1.241	-0.644	-0.535
	(5.97)	(3.87)	(-1.08)	(-0.65)	(-0.76)
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	286	286	250	250	250
Adjusted R^2	0.290	0.362	0.401	0.418	0.584

Table 3. Out-of-Sample Validation of the Text-Based Quality Index (Continued)

Panel C: Aggregate Quality Index

Panel C.1: Random Forest					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Agg	2.403***	2.217***	1.987***	1.942***	0.755***
	(8.57)	(6.53)	(6.09)	(6.15)	(3.00)
Size			-0.003	0.046	0.027
			(-0.04)	(0.57)	(0.54)
Leverage			-0.431*	-0.405	-0.264
			(-1.80)	(-1.56)	(-1.55)
ROA			-0.428	-0.163	0.204
			(-0.33)	(-0.14)	(0.24)
M/B Ratio			0.123**	0.102*	0.029
			(2.26)	(1.86)	(0.82)
Sales Growth			0.699**	0.717**	0.401
			(2.12)	(2.15)	(1.36)
% Transient Investor			0.775	0.466	0.333
			(0.89)	(0.58)	(0.62)
Analyst Following			-0.080	-0.064	-0.033
			(-1.27)	(-1.04)	(-0.81)
Overall Rating			0.388		
			(1.57)		
Balance Subrating				-0.193	-0.055
				(-0.62)	(-0.21)
Culture Subrating				1.219***	0.657*
				(2.67)	(1.78)
Opportunities Subrating				-0.408	-0.011
				(-0.92)	(-0.03)
Compensation Subrating				-0.181	-0.096
				(-0.62)	(-0.47)
Management Subrating				-0.378	-0.362
				(-0.86)	(-1.15)
Industry-Adjusted $ACSI_{t-1}$					0.523***
					(6.64)
Constant	0.228***	0.133**	-1.384	-0.829	-0.601
	(4.57)	(2.45)	(-1.21)	(-0.82)	(-0.84)
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	286	286	250	250	250
Adjusted R^2	0.321	0.378	0.422	0.442	0.592

Table 3. Out-of-Sample Validation of the Text-Based Quality Index (Continued)

Panel C.2: XGBoost					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Agg	1.272***	1.231***	1.084***	1.031***	0.393***
	(10.40)	(8.48)	(6.62)	(6.45)	(2.87)
Size			-0.001	0.034	0.022
			(-0.01)	(0.44)	(0.43)
Leverage			-0.401	-0.368	-0.252
			(-1.60)	(-1.38)	(-1.45)
ROA			0.021	0.201	0.345
			(0.02)	(0.17)	(0.39)
M/B Ratio			0.094*	0.077	0.020
			(1.66)	(1.34)	(0.53)
Sales Growth			0.789**	0.799**	0.432
			(2.44)	(2.41)	(1.43)
% Transient Investor			0.629	0.395	0.305
			(0.73)	(0.49)	(0.57)
Analyst Following			-0.051	-0.039	-0.023
			(-0.88)	(-0.69)	(-0.60)
Overall Rating			0.320		
			(1.32)		
Balance Subrating				-0.263	-0.083
				(-0.87)	(-0.33)
Culture Subrating				1.155**	0.635*
				(2.54)	(1.74)
Opportunities Subrating				-0.383	-0.000
				(-0.89)	(-0.00)
Compensation Subrating				-0.098	-0.063
				(-0.35)	(-0.32)
Management Subrating				-0.380	-0.364
				(-0.89)	(-1.14)
Industry-Adjusted $ACSI_{t-1}$					0.524***
					(6.74)
Constant	0.337***	0.241***	-1.090	-0.566	-0.500
	(7.05)	(4.55)	(-0.96)	(-0.57)	(-0.70)
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	286	286	250	250	250
Adjusted R^2	0.333	0.395	0.428	0.442	0.591

Table 4. Adjusted R^2 Comparison: With vs. Without Predicted Quality Score

Notes: This table compares adjusted R^2 values from out-of-sample regressions across different model specifications. For each specification, we report results from models with and without the inclusion of the predicted quality index. Panel A presents results using the quality index based on the pros section of employee reviews. Panel B presents results using the quality index based on the cons section of employee reviews. Panel C presents results using the aggregate quality index, constructed by weighting the pros- and cons-based quality indices by their respective lengths.

Panel A: Quality Index Based on Pros Section

	Full Model		w/o Index	Incremental Adj. R^2		% ΔAdj. R^2	
	RF	*XGB*		*RF*	*XGB*	*RF*	*XGB*
Industry FE + Year FE	38.7%	37.6%	16.7%	22.1%	20.9%	132.6%	125.7%
Industry FE + Year FE + Firm Controls + Glassdoor Overall Rating	42.6%	41.5%	29.2%	13.4%	12.4%	46.0%	42.4%
Industry FE + Year FE + Firm Controls + Glassdoor Subratings	43.9%	42.8%	31.7%	12.2%	11.1%	38.4%	34.9%

Panel B: Quality Index Based on Cons Section

	Full Model		w/o Index	Incremental Adj. R^2		% ΔAdj. R^2	
	RF	*XGB*		*RF*	*XGB*	*RF*	*XGB*
Industry FE + Year FE	35.6%	36.2%	16.7%	18.9%	19.5%	113.7%	117.4%
Industry FE + Year FE + Firm Controls + Glassdoor Overall Rating	40.9%	40.1%	29.2%	11.7%	10.9%	40.1%	37.3%
Industry FE + Year FE + Firm Controls + Glassdoor Subratings	43.5%	41.8%	31.7%	11.8%	10.1%	37.2%	31.9%

Panel C: Aggregate Quality Index

	Full Model		w/o Index	Incremental Adj. R^2		% ΔAdj. R^2	
	RF	*XGB*		*RF*	*XGB*	*RF*	*XGB*
Industry FE + Year FE	37.8%	39.5%	16.7%	21.1%	22.8%	126.8%	137.2%
Industry FE + Year FE + Firm Controls + Glassdoor Overall Rating	42.2%	42.8%	29.2%	13.1%	13.6%	44.8%	46.7%
Industry FE + Year FE + Firm Controls + Glassdoor Subratings	44.2%	44.2%	31.7%	12.5%	12.4%	39.4%	39.2%

Table 5. MSCI Validation Tests

Notes: This table reports correlations of quality indices and MSCI product quality related scores. The last column (% of Sample Covered) indicates the percentage of our sample for which the corresponding MSCI metric is available. p-values are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Pr. Sign	Pros		Cons		Aggregate		% of Sample Covered
		RF	XGB	RF	XGB	RF	XGB	
Quality Risk Exposure	-	-0.044***	-0.018*	-0.102***	-0.046***	-0.081***	-0.041***	42.2%
		(0.000)	(0.087)	(0.000)	(0.000)	(0.000)	(0.000)	
Has Product Concern	-	-0.168***	-0.105***	-0.110***	-0.028	-0.146***	-0.076**	9.1%
		(0.000)	(0.001)	(0.000)	(0.373)	(0.000)	(0.015)	
Has Safety Concern	-	-0.064***	-0.042***	-0.087***	-0.046***	-0.082***	-0.054***	33.7%
		(0.000)	(0.002)	(0.000)	(0.001)	(0.000)	(0.000)	
#Strengths - #Concerns	+	0.106***	0.074**	0.070**	0.013	0.094***	0.052*	9.1%
		(0.001)	(0.018)	(0.024)	(0.669)	(0.002)	(0.093)	
Quality Management	+	0.057***	0.034*	0.048**	0.022	0.051***	0.030	14.3%
		(0.003)	(0.078)	(0.012)	(0.242)	(0.007)	(0.121)	

Table 6. Quality Index Comparison Across Groups

Notes: This table reports the comparisons of quality indices between firms listed among the best in product/service quality and other firms based on *Fortune's World's Most Admired Companies* rankings. The "Best Quality" group includes firms listed among the best in product/service quality from 2008 to 2012 in the Fortune rankings, while the "Other" group includes all remaining firms that participated in the rankings during this period but were not listed among the best in product/service quality. Panel A shows the results of the pros-based quality index, Panel B shows the results of the cons-based quality index, and Panel C shows the results of the aggregate quality index. p-values are reported in parentheses under the differences. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Group Mean		Diff.
	(1)	(2)	(1)-(2)
	Best Quality	Other	Best vs Other
Panel A: Pros-based Quality Index			
Random Forest			
Mean	0.017	-0.004	0.021**
			(0.014)
XGBoost			
Mean	0.032	-0.044	0.076***
			(0.001)
Panel B: Cons-based Quality Index			
Random Forest			
Mean	0.023	0.007	0.015**
			(0.048)
XGBoost			
Mean	0.047	-0.021	0.067***
			(0.001)
Panel C: Aggregate Quality Index			
Random Forest			
Mean	0.020	0.002	0.018**
			(0.019)
XGBoost			
Mean	0.039	-0.031	0.070***
			(0.000)

Table 7. Quality Index and Product Recalls

Notes: This table reports tests on the relationship between quality indices and CPSC product recalls. Panels A and B report regression results on the relationship between changes in cons-based (Panel A) and pros-based (Panel B) quality indices and product recalls in the subsequent year. Panel C reports regression results on the relationship between changes in aggregate quality indices and product recalls in the subsequent year. The dependent variable in columns (1)–(4) is the total number of recalls in the subsequent year, and the dependent variable in columns (5)–(8) is the total number of reactive recalls over the same period. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Changes in Cons-Based Quality Indices and Future Recall Events

	# Recalls				# Reactive Recalls			
	RF	XGB	RF	XGB	RF	XGB	RF	XGB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ*Quality Index - Cons*	-1.549**	-0.496**			-1.869*	-0.800***		
	(-1.99)	(-2.49)			(-1.96)	(-3.57)		
Large Quality Drop			0.311	0.426***			0.313	0.547***
			(1.54)	(2.65)			(1.29)	(3.21)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Recall History Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,259	3,259	3,259	3,259	2,937	2,937	2,937	2,937
Pseudo R^2	0.468	0.469	0.468	0.470	0.450	0.453	0.449	0.453

Panel B: Changes in Pros-Based Quality Indices and Future Recall Events

	# Recalls				# Reactive Recalls			
	RF	XGB	RF	XGB	RF	XGB	RF	XGB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ*Quality Index - Pros*	-1.265	-0.493**			-0.814	-0.407		
	(-1.57)	(-2.08)			(-0.82)	(-1.32)		
Large Quality Drop			0.233	0.406***			0.140	0.469**
			(1.36)	(2.82)			(0.69)	(2.47)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Recall History Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,259	3,259	3,259	3,259	2,937	2,937	2,937	2,937
Pseudo R^2	0.467	0.468	0.467	0.469	0.447	0.448	0.447	0.451

Panel C: Changes in Aggregate Quality Indices and Future Recall Events

	# Recalls				# Reactive Recalls			
	RF	XGB	RF	XGB	RF	XGB	RF	XGB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ*Quality Index - Agg*	-2.119**	-0.866***			-2.193*	-1.161***		
	(-2.25)	(-3.03)			(-1.88)	(-3.18)		
Large Quality Drop			0.303**	0.456***			0.303	0.452**
			(2.11)	(2.81)			(1.50)	(2.46)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Recall History Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,259	3,259	3,259	3,259	2,937	2,937	2,937	2,937
Pseudo R^2	0.469	0.470	0.468	0.470	0.450	0.453	0.448	0.451

Table 8. Quality Index and Brand Value

Notes: This table presents tests on the relationship between quality indices and the likelihood of appearing in Brand Finance's Global 500 brand rankings. Panel A (B) shows regression results of a dummy indicating whether the firm is listed in Brand Finance's Global 500 brand rankings on the pros-based (cons-based) quality index. Panel C shows regression results of a dummy indicating whether the firm is listed in Brand Finance's Global 500 brand rankings on the aggregate quality index. *Top Brand* equals 1 if the firm appears in Brand Finance's Global 500 rankings in the following year, and 0 otherwise. *Ltm Top Brand* equals 1 if the firm appears in the Global 500 rankings in the next three years, and 0 otherwise. Panel A.1, B.1, and C.1 report results using quality indices constructed from the random forest model. Panel A.2, B.2, C.2 report results using quality indices constructed from the XGBoost model. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Pros-Based Quality Index and Top 500 Brand

Panel A.1: RF				
	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Pros	0.178***	1.127**	0.178***	1.664***
	(4.28)	(2.42)	(4.31)	(3.07)
Size	0.092***	1.181***	0.080***	1.321***
	(11.94)	(10.30)	(10.24)	(9.81)
Leverage	-0.097***	-0.208	-0.091***	-0.183
	(-3.69)	(-0.50)	(-3.55)	(-0.39)
ROA	-0.050	2.181***	-0.046	2.643***
	(-1.29)	(3.05)	(-1.27)	(2.91)
M/B Ratio	0.020***	0.259***	0.018***	0.262***
	(6.23)	(6.30)	(5.79)	(5.67)
Sales Growth	-0.048***	-0.430***	-0.040***	-0.383**
	(-3.77)	(-2.87)	(-3.30)	(-2.04)
Constant	-0.651***	-12.435***	-0.613***	-14.327***
	(-10.16)	(-9.66)	(-9.88)	(-9.48)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,475	10,869	11,456	9,629
Adjusted/Pseudo R^2	0.256	0.540	0.238	0.579

Table 8. Quality Index and Brand Value (Continued)

Panel A.2: XGBoost

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Pros	0.061***	0.449***	0.062***	0.682***
	(3.97)	(2.70)	(4.12)	(3.32)
Size	0.089***	1.173***	0.077***	1.308***
	(12.03)	(9.88)	(10.26)	(9.28)
Leverage	-0.097***	-0.216	-0.090***	-0.190
	(-3.67)	(-0.52)	(-3.54)	(-0.40)
ROA	-0.048	2.167***	-0.045	2.612***
	(-1.23)	(3.08)	(-1.22)	(2.93)
M/B Ratio	0.020***	0.256***	0.017***	0.258***
	(6.09)	(6.22)	(5.63)	(5.61)
Sales Growth	-0.046***	-0.422***	-0.038***	-0.377**
	(-3.58)	(-2.81)	(-3.14)	(-2.04)
Constant	-0.618***	-12.369***	-0.580***	-14.239***
	(-10.20)	(-9.33)	(-9.96)	(-9.04)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,475	10,869	11,456	9,629
Adjusted/Pseudo R^2	0.254	0.540	0.235	0.579

Panel B: Cons-Based Quality Index and Top 500 Brand

Panel B.1: RF

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Cons	0.197***	1.282***	0.177***	1.336**
	(5.02)	(2.74)	(4.66)	(2.35)
Size	0.093***	1.177***	0.080***	1.299***
	(12.07)	(10.15)	(10.25)	(9.48)
Leverage	-0.093***	-0.220	-0.088***	-0.220
	(-3.54)	(-0.53)	(-3.42)	(-0.47)
ROA	-0.062	2.003***	-0.056	2.451***
	(-1.62)	(2.81)	(-1.54)	(2.76)
M/B Ratio	0.020***	0.260***	0.018***	0.264***
	(6.24)	(6.34)	(5.75)	(5.74)
Sales Growth	-0.047***	-0.422***	-0.039***	-0.386**
	(-3.71)	(-2.80)	(-3.23)	(-2.03)
Constant	-0.657***	-12.378***	-0.611***	-14.084***
	(-10.31)	(-9.51)	(-9.90)	(-9.18)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,475	10,869	11,456	9,629
Adjusted/Pseudo R^2	0.258	0.541	0.238	0.577

Table 8. Quality Index and Brand Value (Continued)

Panel B.2: XGBoost

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Cons	0.065***	0.332**	0.055***	0.324*
	(4.61)	(2.14)	(4.11)	(1.69)
Size	0.089***	1.163***	0.077***	1.286***
	(12.07)	(9.71)	(10.21)	(9.08)
Leverage	-0.095***	-0.221	-0.089***	-0.218
	(-3.61)	(-0.53)	(-3.48)	(-0.46)
ROA	-0.049	2.150***	-0.044	2.641***
	(-1.28)	(3.04)	(-1.20)	(2.99)
M/B Ratio	0.020***	0.257***	0.017***	0.262***
	(6.05)	(6.27)	(5.58)	(5.69)
Sales Growth	-0.042***	-0.412***	-0.035***	-0.398**
	(-3.36)	(-2.72)	(-2.92)	(-2.11)
Constant	-0.616***	-12.216***	-0.573***	-13.958***
	(-10.27)	(-9.14)	(-9.91)	(-8.84)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,475	10,869	11,456	9,629
Adjusted/Pseudo R^2	0.255	0.538	0.234	0.574

Panel C: Aggregate Quality Index and Top 500 Brand

Panel C.1: Random Forest

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Agg	0.219***	1.449***	0.206***	1.755***
	(4.89)	(2.76)	(4.70)	(2.81)
Size	0.093***	1.182***	0.081***	1.310***
	(12.02)	(10.29)	(10.26)	(9.68)
Leverage	-0.094***	-0.212	-0.088***	-0.202
	(-3.59)	(-0.51)	(-3.46)	(-0.43)
ROA	-0.061	2.045***	-0.056	2.480***
	(-1.58)	(2.86)	(-1.54)	(2.77)
M/B Ratio	0.020***	0.259***	0.018***	0.262***
	(6.25)	(6.31)	(5.78)	(5.70)
Sales Growth	-0.048***	-0.422***	-0.040***	-0.376**
	(-3.78)	(-2.82)	(-3.30)	(-2.00)
Constant	-0.662***	-12.425***	-0.619***	-14.191***
	(-10.26)	(-9.64)	(-9.89)	(-9.35)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,475	10,869	11,456	9,629
Adjusted/Pseudo R^2	0.258	0.541	0.239	0.578

Table 8. Quality Index and Brand Value (Continued)

Panel C.2: XGBoost				
	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Agg	0.091***	0.539**	0.084***	0.668**
	(4.65)	(2.53)	(4.44)	(2.55)
Size	0.090***	1.168***	0.078***	1.294***
	(12.07)	(9.82)	(10.26)	(9.20)
Leverage	-0.095***	-0.215	-0.089***	-0.199
	(-3.62)	(-0.52)	(-3.48)	(-0.42)
ROA	-0.056	2.097***	-0.051	2.572***
	(-1.44)	(2.97)	(-1.39)	(2.92)
M/B Ratio	0.019***	0.255***	0.017***	0.258***
	(6.03)	(6.22)	(5.57)	(5.62)
Sales Growth	-0.044***	-0.405***	-0.036***	-0.376**
	(-3.48)	(-2.69)	(-3.04)	(-2.01)
Constant	-0.629***	-12.278***	-0.587***	-14.042***
	(-10.27)	(-9.25)	(-9.93)	(-8.95)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,475	10,869	11,456	9,629
Adjusted/Pseudo R^2	0.256	0.540	0.237	0.577

Table 9. Quality Index and Future Profitability

Notes: This table presents tests on the relationship between quality indices and firm profitability. Panel A shows regression results of return on assets (*ROA*) on quality indices. Panel B shows regression results of operating margin on quality indices. *ROA* is the firm's return on assets in the following year, and *Ltm ROA* is the average return on assets over the subsequent three years. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Quality Index and ROA

Panel A.1: Pros-Based Quality Index

	ROA		*Ltm ROA*	
	RF	XGB	RF	XGB
	(1)	(2)	(3)	(4)
Quality Index - Pros	0.028***	0.015***	0.017**	0.010***
	(3.56)	(4.81)	(2.12)	(3.28)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,068	12,068	12,071	12,071
Adjusted R^2	0.229	0.230	0.257	0.257

Panel A.2: Cons-Based Quality Index

	ROA		*Ltm ROA*	
	RF	XGB	RF	XGB
	(1)	(2)	(3)	(4)
Quality Index - Cons	0.043***	0.013***	0.034***	0.012***
	(5.39)	(4.62)	(4.35)	(4.42)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,068	12,068	12,071	12,071
Adjusted R^2	0.231	0.230	0.259	0.258

Panel A.3: Aggregate Quality Index

	ROA		*Ltm ROA*	
	RF	XGB	RF	XGB
	(1)	(2)	(3)	(4)
Quality Index - Agg	0.044***	0.021***	0.033***	0.017***
	(5.20)	(5.73)	(3.84)	(4.70)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,068	12,068	12,071	12,071
Adjusted R^2	0.231	0.231	0.258	0.259

Table 9. Quality Index and Future Profitability (Continued)

Panel B: Quality Index and Operating Margin

Panel B.1: Pros-Based Quality Index

	Operating Margin		*Ltm Operating Margin*	
	RF	XGB	RF	XGB
	(1)	(2)	(3)	(4)
Quality Index - Pros	0.049***	0.018***	0.041***	0.016***
	(3.62)	(3.58)	(3.01)	(3.12)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,040	12,040	12,043	12,043
Adjusted R^2	0.253	0.252	0.277	0.276

Panel B.2: Cons-Based Quality Index

	Operating Margin		*Ltm Operating Margin*	
	RF	XGB	RF	XGB
	(1)	(2)	(3)	(4)
Quality Index - Cons	0.046***	0.017***	0.039***	0.016***
	(3.52)	(3.67)	(3.02)	(3.48)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,040	12,040	12,043	12,043
Adjusted R^2	0.253	0.252	0.277	0.277

Panel B.3: Aggregate Quality Index

	Operating Margin		*Ltm Operating Margin*	
	RF	XGB	RF	XGB
	(1)	(2)	(3)	(4)
Quality Index - Agg	0.055***	0.026***	0.047***	0.024***
	(3.84)	(4.24)	(3.21)	(3.84)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,040	12,040	12,043	12,043
Adjusted R^2	0.253	0.253	0.277	0.277

Table 10. Out-of-Sample Validation of the Text-Based Quality Index in Service-Oriented Industries

Notes: This table presents regressions of industry-adjusted ACSI scores on the text-based quality index for the hold-out sample of service-oriented firms. The dependent variable is the industry-adjusted ACSI score. The text-based quality indices are derived from models trained on service-oriented firms. Panel A (B) presents results using the quality index derived from the pros (cons) section of employee reviews. Panel C presents results using the aggregate quality index. Panel A.1, B.1, and C.1 report results using quality indices constructed from the random forest model. Panel A.2, B.2, C.2 report results using quality indices constructed from the XGBoost model. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Quality Index Based on Pros Section

Panel A.1: Random Forest					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Pros	2.286***	2.268***	2.087***	2.039***	0.877***
	(10.39)	(7.81)	(6.19)	(6.15)	(2.67)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	173	173	137	137	137
Adjusted R^2	0.417	0.476	0.488	0.518	0.683

Panel A.2: XGBoost					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Pros	2.178***	1.987***	1.754***	1.683***	0.531*
	(8.93)	(6.86)	(5.85)	(5.34)	(1.89)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	173	173	137	137	137
Adjusted R^2	0.270	0.376	0.391	0.417	0.664

Table 10. Out-of-Sample Validation of the Text-Based Quality Index in Service-Oriented Industries (Continued)

Panel B: Quality Index Based on Cons Section

Panel B.1: Random Forest					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Cons	2.040***	1.960***	1.844***	1.848***	0.748***
	(7.37)	(5.94)	(5.46)	(5.82)	(2.76)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	173	173	137	137	137
Adjusted R^2	0.369	0.457	0.486	0.534	0.679

Panel B.2: XGBoost					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Cons	1.788***	1.879***	1.842***	1.779***	0.729***
	(6.28)	(8.01)	(6.28)	(6.06)	(3.27)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	173	173	137	137	137
Adjusted R^2	0.262	0.447	0.488	0.516	0.679

Table 10. Out-of-Sample Validation of the Text-Based Quality Index in Service-Oriented Industries (Continued)

Panel C: Aggregate Quality Index

Panel C.1: Random Forest

	(1)	(2)	(3)	(4)	(5)
Quality Index - Agg	2.229***	2.172***	2.022***	1.997***	0.845***
	(8.57)	(6.42)	(5.65)	(5.93)	(2.73)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	173	173	137	137	137
Adjusted R^2	0.403	0.476	0.496	0.535	0.682

Panel C.2: XGBoost

	(1)	(2)	(3)	(4)	(5)
Quality Index - Agg	2.378***	2.360***	2.304***	2.234***	0.920***
	(7.78)	(8.12)	(6.09)	(6.03)	(2.97)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	173	173	137	137	137
Adjusted R^2	0.324	0.473	0.499	0.521	0.680

Table 11. Out-of-Sample Validation of the Text-Based Quality Index in Product-Oriented Industries

Notes: This table presents regressions of industry-adjusted ACSI scores on the text-based quality index for the hold-out sample of product-oriented firms. The dependent variable is the industry-adjusted ACSI score. The text-based quality indices are derived from models trained on product-oriented firms. Panel A (B) presents results using the quality index derived from the pros (cons) section of employee reviews. Panel C presents results using the aggregate quality index. Panel A.1, B.1, and C.1 report results using quality indices constructed from the random forest model. Panel A.2, B.2, C.2 report results using quality indices constructed from the XGBoost model. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Quality Index Based on Pros Section

Panel A.1: Random Forest					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Pros	1.565***	1.915***	1.605**	2.371**	1.049
	(3.70)	(2.97)	(2.02)	(2.61)	(1.33)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	113	113	113	113	113
Adjusted R^2	0.151	0.263	0.314	0.354	0.450

Panel A.2: XGBoost					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Pros	0.856***	0.983***	0.681**	0.990***	0.442
	(4.35)	(3.70)	(2.45)	(3.26)	(1.37)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	113	113	113	113	113
Adjusted R^2	0.163	0.271	0.311	0.351	0.449

Table 11. Out-of-Sample Validation of the Text-Based Quality Index in Product-Oriented Industries (Continued)

Panel B: Quality Index Based on Cons Section

Panel B.1: Random Forest					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Cons	1.779***	1.847**	1.543	1.861*	0.122
	(3.54)	(2.29)	(1.58)	(1.72)	(0.13)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	113	113	113	113	113
Adjusted R^2	0.144	0.236	0.298	0.317	0.438

Panel B.2: XGBoost					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Cons	0.899***	0.856***	0.583*	0.621*	0.086
	(3.95)	(2.78)	(1.72)	(1.79)	(0.26)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	113	113	113	113	113
Adjusted R^2	0.156	0.256	0.297	0.311	0.438

Table 11. Out-of-Sample Validation of the Text-Based Quality Index in Product-Oriented Industries (Continued)

Panel C: Aggregate Quality Index

Panel C.1: Random Forest					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Agg	1.719***	2.021**	1.810*	2.455**	0.750
	(3.61)	(2.62)	(1.89)	(2.32)	(0.81)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	113	113	113	113	113
Adjusted R^2	0.152	0.254	0.310	0.342	0.442

Panel C.2: XGBoost					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Agg	0.982***	1.113***	0.867**	1.021***	0.345
	(4.29)	(3.44)	(2.48)	(2.83)	(0.89)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	113	113	113	113	113
Adjusted R^2	0.181	0.287	0.317	0.342	0.444

Table 12. Out-of-Sample Validation of the Text-Based Quality Index (Controlling for Conference Call Cultural Value Scores)

Notes: This table reports regression results of industry-adjusted ACSI scores on the text-based quality index in the hold-out sample. The analysis replicates Table 3 with additional controls for conference call-based cultural value scores. Panel A (Panel B) reports results using the quality index derived from the pros (cons) section and controlling for conference call-based cultural value scores. Panel C reports results using the aggregate quality index and controlling for conference call-based cultural value scores. Panel A.1, B.1, and C.1 report results using quality indices constructed from the random forest model. Panel A.2, B.2, C.2 report results using quality indices constructed from the XGBoost model. The dependent variable is the industry-adjusted ACSI score in year $t+1$. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Quality Index Based on Pros Section

Panel A.1: RF						
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Pros	2.064***	2.018***	1.980***	1.903***	0.844***	0.862***
	(6.94)	(6.99)	(6.50)	(6.60)	(3.23)	(3.24)
Quality Score CC	-5.604	1.600	-7.178	2.248	-5.346	-0.094
	(-0.52)	(0.13)	(-0.66)	(0.19)	(-0.73)	(-0.01)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes	Yes
Other CC Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	247	247	247	247	247	247
Adjusted R^2	0.442	0.466	0.448	0.479	0.603	0.613

Panel A.2: XGBoost						
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Pros	0.967***	0.925***	0.917***	0.857***	0.413***	0.400***
	(6.37)	(6.10)	(6.13)	(5.96)	(2.86)	(2.74)
Quality Score CC	-6.958	-0.693	-9.053	-0.311	-5.876	-1.234
	(-0.61)	(-0.05)	(-0.82)	(-0.03)	(-0.80)	(-0.15)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes	Yes
Other CC Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	247	247	247	247	247	247
Adjusted R^2	0.427	0.445	0.434	0.460	0.604	0.612

Table 12. Out-of-Sample Validation of the Text-Based Quality Index (Controlling for Conference Call Cultural Value Scores) (Continued)

Panel B: Quality Index Based on Cons Section

Panel B.1: RF						
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Cons	1.914***	1.894***	1.867***	1.807***	0.745***	0.758***
	(6.17)	(6.26)	(6.04)	(6.25)	(3.32)	(3.25)
Quality Score CC	-8.670	-3.778	-10.198	-2.446	-6.765	-2.199
	(-0.80)	(-0.30)	(-1.00)	(-0.21)	(-0.96)	(-0.27)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes	Yes
Other CC Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	247	247	247	247	247	247
Adjusted R^2	0.434	0.457	0.450	0.480	0.600	0.609

Panel B.2: XGBoost						
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Cons	0.962***	0.930***	0.908***	0.851***	0.317***	0.301***
	(6.62)	(6.37)	(6.52)	(6.22)	(3.11)	(2.92)
Quality Score CC	-8.637	-3.077	-11.206	-3.044	-7.189	-2.376
	(-0.83)	(-0.25)	(-1.10)	(-0.25)	(-1.01)	(-0.29)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes	Yes
Other CC Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	247	247	247	247	247	247
Adjusted R^2	0.426	0.445	0.437	0.461	0.595	0.603

Table 12. Out-of-Sample Validation of the Text-Based Quality Index (Controlling for Conference Call Cultural Value Scores) (Continued)

Panel C: Aggregate Quality Index

Panel C.1: RF						
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Agg	2.096***	2.061***	2.016***	1.947***	0.835***	0.854***
	(6.31)	(6.42)	(6.17)	(6.35)	(3.33)	(3.30)
Quality Score CC	-6.721	-1.090	-8.187	0.044	-5.919	-1.118
	(-0.64)	(-0.09)	(-0.80)	(0.00)	(-0.83)	(-0.14)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes	Yes
Other CC Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	247	247	247	247	247	247
Adjusted R^2	0.445	0.468	0.456	0.485	0.602	0.612

Panel C.2: XGBoost						
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Agg	1.142***	1.102***	1.080***	1.016***	0.451***	0.436***
	(7.14)	(6.95)	(6.85)	(6.71)	(3.40)	(3.23)
Quality Score CC	-7.109	-2.096	-9.136	-1.685	-6.306	-1.879
	(-0.68)	(-0.17)	(-0.89)	(-0.14)	(-0.87)	(-0.23)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes	Yes
Other CC Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	247	247	247	247	247	247
Adjusted R^2	0.450	0.467	0.457	0.479	0.603	0.610

Table 13. Out-of-Sample Validation of the Text-Based Quality Index (Controlling for Glassdoor Cultural Value Scores)

Notes: This table reports regression results of industry-adjusted ACSI scores on the text-based quality index in the hold-out sample. The analysis replicates Table 3 with additional controls for Glassdoor review-based cultural value scores. Panel A (Panel B) reports results using the quality index derived from the pros (cons) section and controlling for Glassdoor review-based cultural value scores. Panel C reports results using the aggregate quality index and controlling for Glassdoor review-based cultural value scores. Panel A.1, B.1, and C.1 report results using quality indices constructed from the random forest model. Panel A.2, B.2, C.2 report results using quality indices constructed from the XGBoost model. The dependent variable is the industry-adjusted ACSI score in year $t+1$. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Quality Index Based on Pros Section

Panel A.1: RF

	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Pros	2.030***	2.008***	1.956***	1.934***	0.814***	0.825***
	(6.68)	(7.01)	(6.49)	(6.53)	(3.13)	(3.13)
Quality Score GD	3.537	2.223	3.220	2.192	2.054	2.382
	(0.68)	(0.38)	(0.67)	(0.41)	(0.50)	(0.56)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes	Yes
Other GD Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	250	250	250	250	250	250
Adjusted R^2	0.425	0.424	0.437	0.440	0.592	0.592

Panel A.2: XGBoost

	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Pros	0.973***	0.955***	0.926***	0.910***	0.412***	0.420***
	(6.23)	(6.59)	(6.12)	(6.35)	(2.85)	(2.97)
Quality Score GD	4.230	3.219	4.266	3.591	2.613	3.125
	(0.76)	(0.52)	(0.84)	(0.65)	(0.64)	(0.73)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes	Yes
Other GD Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	250	250	250	250	250	250
Adjusted R^2	0.415	0.415	0.427	0.432	0.594	0.595

Table 13. Out-of-Sample Validation of the Text-Based Quality Index (Controlling for Glassdoor Cultural Value Scores) (Continued)

Panel B: Quality Index Based on Cons Section

Panel B.1: RF						
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Cons	1.805***	1.748***	1.804***	1.782***	0.669***	0.690***
	(5.80)	(5.85)	(5.98)	(5.77)	(2.99)	(3.03)
Quality Score GD	3.274	2.199	3.584	2.958	2.057	2.600
	(0.59)	(0.35)	(0.73)	(0.53)	(0.50)	(0.60)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes	Yes
Other GD Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	250	250	250	250	250	250
Adjusted R^2	0.407	0.405	0.434	0.437	0.588	0.589

Panel B.2: XGBoost						
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Cons	0.897***	0.878***	0.855***	0.842***	0.254**	0.259**
	(5.82)	(5.98)	(5.89)	(5.96)	(2.31)	(2.41)
Quality Score GD	2.848	1.579	3.270	2.408	1.735	2.237
	(0.54)	(0.27)	(0.68)	(0.44)	(0.42)	(0.52)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes	Yes
Other GD Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	250	250	250	250	250	250
Adjusted R^2	0.399	0.398	0.417	0.420	0.582	0.582

Table 13. Out-of-Sample Validation of the Text-Based Quality Index (Controlling for Glassdoor Cultural Value Scores) (Continued)

Panel C: Aggregate Quality Index

Panel C.1: Random Forest						
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Agg	2.012***	1.972***	1.962***	1.947***	0.770***	0.791***
	(6.03)	(6.14)	(6.14)	(6.01)	(3.10)	(3.15)
Quality Score GD	3.657	2.478	3.641	2.838	2.146	2.592
	(0.68)	(0.41)	(0.75)	(0.52)	(0.52)	(0.61)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes	Yes
Other GD Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	250	250	250	250	250	250
Adjusted R^2	0.421	0.419	0.441	0.443	0.590	0.591

Panel C.2: XGBoost						
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index - Agg	1.102***	1.088***	1.047***	1.043***	0.404***	0.419***
	(6.64)	(7.05)	(6.56)	(6.80)	(3.02)	(3.27)
Quality Score GD	4.275	3.101	4.419	3.606	2.424	2.884
	(0.81)	(0.52)	(0.92)	(0.67)	(0.60)	(0.68)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Glassdoor Controls	Overall	Overall	Subratings	Subratings	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes	Yes
Other GD Value Scores	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	250	250	250	250	250	250
Adjusted R^2	0.427	0.427	0.441	0.445	0.590	0.590

Table 14. Quantifying Measurement Error in Quality Index

Notes: This table shows AR(1) regressions of quality indices. Regressions are estimated at the firm-year level. *Implied Share Measurement Error* is calculated as $1-(\hat{\beta}_{OLS}/\hat{\beta}_{IV})$, where $\hat{\beta}_{OLS}$ is the estimated coefficient in $\textit{Quality Index}_{i,t} = \alpha + \beta \textit{Quality Index}_{i,t-1} + \varepsilon$ and $\hat{\beta}_{IV}$ is the coefficient on the instrumented $\textit{Quality Index}_{i,t}$ in the same specification. To obtain bootstrapped standard errors for *Implied Share Measurement Error*, we repeat the following procedure 500 times: draw a random sample of the same sample size (with replacement and clustered by firm) from our regression sample, run the two regressions, and obtain the implied share of measurement error. All regressions in Panel A are estimated without industry and year fixed effects. All regressions in Panel B include industry and year fixed effects. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Overall Variation

Panel A.1 Pros

	Quality Index - Pros$_t$							
	RF				XGBoost			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Quality Index - Pros$_{t-1}$	0.720***	0.892***	0.833***	0.896***	0.524***	0.939***	0.899***	0.956***
	(64.23)	(87.38)	(47.53)	(78.84)	(57.04)	(68.65)	(54.80)	(50.58)
Observations	16,050	13,964	16,050	13,964	16,050	13,964	16,050	13,964
Specification	OLS	IV	IV	IV	OLS	IV	IV	IV
Instrument		Pros, t-2	Cons, t-1	Cons, t-2		Pros, t-2	Cons, t-1	Cons, t-2
Implied Share M.E.		0.193	0.135	0.196		0.442	0.417	0.452
		(0.018)	(0.008)	(0.019)		(0.014)	(0.012)	(0.014)

Panel A.2 Cons

	Quality Index - Cons$_t$							
	RF				XGBoost			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Quality Index - Cons$_{t-1}$	0.686***	0.870***	0.801***	0.862***	0.434***	0.870***	0.872***	0.879***
	(63.45)	(81.78)	(49.16)	(68.84)	(44.06)	(49.99)	(52.99)	(45.24)
Observations	16,050	13,964	16,050	13,964	16,050	13,964	16,050	13,964
Specification	OLS	IV	IV	IV	OLS	IV	IV	IV
Instrument		Cons, t-2	Pros, t-1	Pros, t-2		Cons, t-2	Pros, t-1	Pros, t-2
Implied Share M.E.		0.211	0.144	0.204		0.501	0.502	0.506
		(0.018)	(0.008)	(0.020)		(0.016)	(0.014)	(0.015)

Panel B: Firm-Level Variation

Panel B.1 Pros

	Quality Index - Pros$_t$							
	RF				XGBoost			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Quality Index - Pros$_{t-1}$	0.685***	0.889***	0.817***	0.885***	0.437***	0.933***	0.871***	0.943***
	(54.25)	(73.70)	(39.20)	(68.09)	(42.14)	(50.85)	(40.39)	(36.23)
Observations	16,050	13,964	16,050	13,964	16,050	13,964	16,050	13,964
Specification	OLS	IV	IV	IV	OLS	IV	IV	IV
Instrument		Pros, t-2	Cons, t-1	Cons, t-2		Pros, t-2	Cons, t-1	Cons, t-2
Implied Share M.E.		0.230	0.162	0.226		0.532	0.499	0.537
		(0.021)	(0.010)	(0.022)		(0.016)	(0.015)	(0.017)

Panel B.2 Cons

	Quality Index - Cons$_t$							
	RF				XGBoost			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Quality Index - Cons$_{t-1}$	0.655***	0.856***	0.784***	0.854***	0.370***	0.838***	0.838***	0.851***
	(56.02)	(71.84)	(40.98)	(58.25)	(38.80)	(38.43)	(39.83)	(32.48)
Observations	16,050	13,964	16,050	13,964	16,050	13,964	16,050	13,964
Specification	OLS	IV	IV	IV	OLS	IV	IV	IV
Instrument		Cons, t-2	Pros, t-1	Pros, t-2		Cons, t-2	Pros, t-1	Pros, t-2
Implied Share M.E.		0.235	0.165	0.233		0.558	0.558	0.565
		(0.021)	(0.010)	(0.023)		(0.017)	(0.015)	(0.017)

Table 15. Validation of the Zero-Shot GPT Quality Index

Notes: This table reports regression results of industry-adjusted ACSI scores on the zero-shot GPT quality index. Panel A (B) presents results using the quality index derived from the pros (cons) section of employee reviews. Panel C reports results using the aggregate quality index. Panel A.1, B.1, and C.1 report results using zero-shot GPT quality indices constructed using all sampled reviews. Panel A.2, B.2, C.2 report results using zero-shot GPT quality indices constructed using relevant reviews (i.e., those identified by GPT as explicitly mentioning quality issues or implicitly suggesting them). The dependent variable is the industry-adjusted ACSI score in year $t+1$. All continuous variables are winsorized at the 1st and 99th percentiles. T-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Quality Index Based on Pros Section

Panel A.1: Quality Index - Overall					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Pros	0.353***	0.447***	0.276**	0.197	0.093
	(2.70)	(3.50)	(2.43)	(1.35)	(1.14)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	1,608	1,608	1,608	1,309	1,309
Adjusted R^2	0.008	0.131	0.218	0.274	0.592

Panel A.2: Quality Index - Relevant Only					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Pros	0.495***	0.527***	0.232*	0.196	0.238**
	(3.60)	(3.82)	(1.80)	(1.28)	(2.44)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	1,566	1,566	1,566	1,290	1,290
Adjusted R^2	0.009	0.131	0.225	0.284	0.595

Table 15. Validation of the Zero-Shot GPT Quality Index (Continued)

Panel B: Quality Index Based on Cons Section

Panel B.1: Quality Index - Overall					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Cons	0.163	0.457***	0.044	-0.133	-0.010
	(1.19)	(3.61)	(0.35)	(-0.92)	(-0.10)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	1,608	1,608	1,608	1,309	1,309
Adjusted R^2	0.001	0.133	0.214	0.274	0.592

Panel B.2: Quality Index - Relevant Only					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Cons	0.348**	0.519***	0.136	0.130	0.246**
	(2.31)	(3.34)	(1.00)	(0.74)	(2.24)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	1,588	1,588	1,588	1,300	1,300
Adjusted R^2	0.004	0.129	0.215	0.275	0.591

Table 15. Validation of the Zero-Shot GPT Quality Index (Continued)

Panel C: Aggregate Quality Index

Panel C.1: Quality Index - Overall					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Agg	0.216**	0.387***	0.148*	0.024	0.037
	(2.37)	(4.25)	(1.91)	(0.27)	(0.59)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	1,608	1,608	1,608	1,309	1,309
Adjusted R^2	0.007	0.141	0.217	0.273	0.592

Panel C.2: Quality Index - Relevant Only					
	(1)	(2)	(3)	(4)	(5)
Quality Index - Agg	0.376***	0.499***	0.206**	0.175	0.226***
	(3.79)	(4.80)	(2.30)	(1.49)	(2.84)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. $ACSI_{t-1}$	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	1,539	1,539	1,539	1,275	1,275
Adjusted R^2	0.013	0.144	0.226	0.287	0.594

Table 16. Out-of-Sample Validation of the Fine-Tuned GPT Quality Index

Notes: This table reports out-of-sample regression results of industry-adjusted ACSI scores on the quality index constructed through fine-tuned GPT-4o-mini models. The dependent variable is the industry-adjusted ACSI score in year $t+1$. Panel A (B) presents results using the quality index derived from the pros (cons) section of employee reviews. Panel C reports results using the aggregate quality index. All continuous variables are winsorized at the 1st and 99th percentiles. T-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Quality Index Based on Pros Section

	(1)	(2)	(3)	(4)	(5)
Quality Index - Pros	0.284***	0.282***	0.255***	0.253***	0.127***
	(7.17)	(6.08)	(4.80)	(4.69)	(3.23)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	286	286	250	250	250
Adjusted R^2	0.196	0.319	0.379	0.398	0.596

Panel B: Quality Index Based on Cons Section

	(1)	(2)	(3)	(4)	(5)
Quality Index - Cons	0.302***	0.295***	0.266***	0.261***	0.104**
	(7.53)	(5.75)	(4.87)	(4.94)	(2.23)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	286	286	250	250	250
Adjusted R^2	0.219	0.321	0.385	0.407	0.590

Panel C: Aggregate Quality Index

	(1)	(2)	(3)	(4)	(5)
Quality Index - Agg	0.354***	0.361***	0.348***	0.344***	0.161***
	(8.31)	(6.48)	(5.55)	(5.59)	(3.00)
Firm Controls	No	No	Yes	Yes	Yes
Glassdoor Controls	No	No	Overall	Subratings	Subratings
Ind. Adj. ACSI_{t-1}	No	No	No	No	Yes
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Observations	286	286	250	250	250
Adjusted R^2	0.252	0.359	0.413	0.433	0.599

Appendix A. Most Predictive N-Grams and Examples of Reviews

Table A1. Top 300 N-grams Predictive of Product/Service Quality from the "Pros" Section

Category	Random Forest	XGBoost
Human Capital and Skill Development	first, skill, train, first job, great train, smart, learn, experience, training, improve, opportunity, teach, position, knowledge, career, improvement, starter, start, growth, choice, advancement, skill good, place learn, development, learn work, option, career growth, know, exposure, learn lot	easy hire, job training, college student, enable, lot career, jump, first, opportunity well, know job, company promote, learn opportunity, lot young, career good, development good, prepare, skill, plenty room, extensive training, develop skill, foundation, opportunity many, first start, subject, teach, first class, insight, provide opportunity, starter, development program, program work, great train, another job, great advancement, skill good, deep, give opportunity, train program, chance advancement, best training, want see, promote within, career development, learn new, choice, skill set
Operational Processes and Disciplines	work, job, challenge, safety, steady, process, good job, project, engineer, reason work, stable, work work, server, steady work, service, customer, quality, business, program, product, hard, use, problem, solid, drive, provide, system, technology, equipment, ability, customer service, line business, innovation, plant	server, cell, ready, great customer, mobile, different function, store, technology work, challenge, lot customer, material, work still, challenge work, job, work, technique, technology, system, objective, wear, reason work, safety, analyst, dress, assign, new technology, need do, capable, data, work work, warehouse, developer, guest, contractor, good job, line business, array, good system, functional, immediate, capability
Incentives and Pay	pay, good pay, good salary, salary, free food, great benefit, great pay, tip, make good, money, food, decent, pay good, healthcare, pay benefit, free, benefit, offer, perk, wage, benefit great, travel, paycheck, salary work, good benefit, health, compensation, decent pay, competitive, pay work, base, raise, pay great, stock, salary benefit, reward, bonus, benefit good, pay well, lunch, package, vacation	leverage, make decent, free food, flight benefit, management pay, tip, medical dental, wage, food, provide good, ok benefit, decent health, pay holiday, make good, eligible, good salary, healthcare, opportunity travel, offer great, membership, benefit people, great benefit, breakfast, profit, money benefit, travel, salary work, benefit tuition, competitive benefit, financial, food good, include medical, perk good, paycheck, good pay, buck, pension, great pay, employee pay, cheap, pay well, amount money
Workplace Culture and Employee Satisfaction	work good, good place, job good, atmosphere, pro, good work, good good, fun, company work, culture, casual, environment, meaningful, place work, community, place, work environment, work great, open, nice, overall great, value, proud, office, great place, amaze, willing, diversity, exciting, awesome, great work, love, political, atmosphere great	work good, work excellent, want work, philosophy, overall good, et great, environment company, discipline, best best, look forward, many good, interesting work, good amount, project good, ethical, meaningful, much good, encouragement, fun atmosphere, job best, job good, incredible, job bad, lucky enough, encouraging, job want, always great, available good, right thing, political, overall great, lot interesting

Table A1. Top 300 N-grams Predictive of Product/Service Quality from the "Pros" Section (Continued)

Category	Random Forest	XGBoost
Staffing and Resource Allocation	overtime, hour, work life, work home, home, life balance, year, flexible work, life, balance, part, flexible, flexible good, time, fast, schedule, flexibility, flexibility work, time employee, day, part time, long, available, allow, close home, full	overtime, move fast, put time, longer, hour day, work school, work around, time take, flex time, home good, need time, work personal, go fast, time employee, remote, time need, pretty flexible, location work
Teamwork and Internal Coordination	team, people, good people, friend, employee, everyone, people work, great team, communication, inclusive, family, colleague, friendly, great people, people great, coworkers, group, worker, team great, team work, co worker, nice people, co, employee customer, smart people, engage, help, people good, support, collaborative	proactive, woman, collaboration, man, team, people employee, people want, everyone work, diverse workforce, people diverse, people, associate work, meet lot, inclusive, nice helpful, right people, employee customer, group good, team work, people love, engage
Organizational Structure and Change	merger, organization, new, local, scale, change, area, move, acquisition, level, location, corporate, department, union	merger, within organization, work large, lot change, large scale, size, union, bureaucracy, region, trend, acquisition, company try, turnover rate, department work, scale
Leadership and Managerial Oversight	supervisor, leadership, transparent, management, encourage, care, manager, treat, care employee, decision, top, professional, bos, plan	transparent, micro manage, board, corporate office, management seem, open door, meritocracy, really take, listen employee, management generally, expectation, make decision, management team, management level
Company Brand and Reputation	reputation, great company, impact, big, company, big company, large, great reputation, good company, large company, strong, name, company great, brand, great corporate, work company	impact, strength, global brand, global reach, reputation, great corporate, large organization, name recognition, great reputation, large company, big company, community involvement, global company, worldwide, great company, strong work, image, world great, decade
Industry or Firm Specific	america, world, commerce, united, industry, seattle, fast food	seattle, united, commerce, america, starbucks, downtown, pharmacy
General Sentiment	good, great, ok, best, well, really, okay, interesting, excellent, high, wonderful, bad, easy, pretty, low, really good, right	generally easy, ok, terrible, really want, great
Uncategorized	none, nothing, overall, lot, make, com, many, reason, thing, one, around, think, everyday, include, never, need, within, non, take, back, look, get, still, meet, different, give, want, much, see, anywhere, go, spend, way, find, do, keep	concept, none, true, nothing, prefer, news, table, initial, avoid, per, last year, pick, even part, occasion, com, unit, throw, aside, onside, relative, specific, one, thru, several year, sound, use good, round, seat, reason, half, wonder, lose, quite, company still, nearly, bus, mix, hit, wait, proper, meet

Table A2. Top 300 N-grams Predictive of Product/Service Quality from the "Cons" Section

Category	Random Forest	XGBoost
Human Capital and Skill Development	promotion, seniority, promote, position, failure, advancement, opportunity, start, know, first month, become, career, future, grow, growth, development, advance, experience, training, job security	promotion process, promote base, grow fast, first month, people move, network, seniority, annual review, new employee, english, starter, external, tenure, move within, first year, long term, job security, limited opportunity, need improve, teach company, climb ladder, month advance, opportunity move, know job, fast track, failure
Operational Processes and Disciplines	equipment, plant, smell, depend, safety, customer, dirty, challenge, server, temp, rude customer, loss, customer rude, work, consumer, chaotic, process, still work, complete, system, call, action, work extremely, job, use, supply, hard, business, technology, difficult, force, problem, project, service, practice, guest, working, product, repetitive, tool, facility, function, disconnect, complaint, information, access, maintenance	smell, plant, ship, big problem, equipment, selling, thing customer, error, dirty, safety, temp, paper work, document, customer rude, credit card, challenge work, system process, server, divide, consumer, sign, user, call, deal customer, production, work difficult, loss, maintenance, working, technician, tool, factory, project management, chaotic, repetitive, difficult make, challenging, specialized, random, data, still work, functional, automatically, pile, get work, disconnect
Incentives and Pay	health, tip, benefit, base, deal, low pay, money, bonus, pay benefit, pay, medical, food, raise, fairness, bad pay, health insurance, salary, gas, compensation, vacation	compensation plan, tip, illness, high salary, benefit work, health, health care, yearly, money make, million dollar, pay market, pension, pay long, base, loan, compensation low, pay competitive, paycheck, lunch, prescription, financial, medical, overhead, compensation benefit, sub par, fee
Workplace Culture and Employee Satisfaction	toxic, rude, environment, workplace, culture, privacy, boring, bore, place work, place, office, room, open, angry	entrepreneurial, dissatisfaction, toxic, formal, insane, privacy, workplace, look forward, politics, section, mold, pit, principle, lot bureaucracy, autonomy, company politics, dysfunctional, mindset, easy lose, crowd, hell
Staffing and Resource Allocation	day week, fast, weekend, night, lack, night weekend, broken, fast pace, overtime, balance, life, life balance, work life, enough, pace, shift, hour, short, year, long, move fast, schedule, load, day, week, excessive, long time, long hour, limit, lot work, work hard, late, time, weekend holiday, hard work, weekend night, full, month	move fast, give day, fast pace, night weekend, fast, time really, press, work death, day week, fast paced, store hour, heavily, broken, much work, require work, work holiday, retail hour, sunday, skew, night, short hand, overtime, employee long, much job, pace, monthly, pressure sale, pattern, one year, week day, schedule, put hour, irregular, tough, least month
Teamwork and Internal Coordination	disrespect, people, employee, worker, respect, communication, team, feel, people work, help, everyone, group, peer, personal, many people, opinion, others, support, co, staff	people need, work people, way treat, people see, disrespect, gender, criticism, depend group, work employee, colleague, respect employee, folk, peer, cheat, people make, subordinate, staffed, condescend, relation, people work

Table A2. Top 300 N-grams Predictive of Product/Service Quality from the "Cons" Section (Continued)

Category	Random Forest	XGBoost
Organizational Structure and Change	merger, organization, location, implementation, change, issue, unstable, slow, acquisition, merge, bureaucracy, integrate, integration, move, reorganization, relocate, acquire, uncertainty, corporate, department, layoffs, massive, organizational structure, new, strategy, level, hire	job cut, merger, spin, commercial, layoffs, branch, location, establish, offs, national, turnover high, integrate, challenge time, unstable, implementation, relocate, startup, turnover rate, compare competitor, within department, reorganization, orgy, company thing, lot turnover, crisis, slow change, employee cut, moving, international, change direction, government, union, replace, overseas, strategy, policy procedure
Leadership and Managerial Oversight	leadership, poor management, speak, chief, upper, care, management poor, upper management, decision, management, many chief, ce, bad management, management employee, manager, plan, treat, accountability, senior, vision, supervisor, direction, management bad, decision making, expect, senior management, communication management, expectation, management style	area manager, many chief, make decision, superior, yes men, downhill, regional manager, management respect, pay attention, lead lot, chief, management look, manager manager, manager tend, decision make, open door, respond, sr management, incompetence, hour management, mid management, speak, effectively, manager level, leadership lack, officer, ce, slow react, management poor
Company Brand and Reputation	reputation, company, big, impact, mission, large, company lot	reputation, low compare, company grow, company go, image, company, company year, huge company
Industry or Firm Specific	fast food, amazon, america, chicago, industry, united	fast food, corporate america, amazon, retail job, bank, chicago, america, united
General Sentiment	poor, bad, con, terrible, horrible, good, low, great, high, well, really, extremely, best	poor, terribly, excellent, hard good, quite bit, interesting, shady, stupid, nonexistent, truly, exactly, increasingly, finally, good thing, substantial
Uncategorized	everything, none, hot, seem, average, continue, make, frequent, sometimes, want, may, small, many, table, one, lot, constant, must, give, much, still, per, old, quite, multiple, making, go, little, do, within, thing, think, play, way, take, find, due, eat, keep, try, nothing, tend, might, look, back, run, come, unless, need, already, least, outside, get, another, even, bit, band, constantly, news	sometimes seem, let, making, differ, fat, side, refer, want, lean, news, stay away, thought, doubt, somewhat, hard make, recommend, black, original, hot, initial, com, company even, band, wait, list, way thing, none, everything, two, primary

Table A3. Examples of Reviews Most Predictive of High and Low Quality Scores

Notes: This table presents examples of employee reviews with the highest and lowest predicted quality scores from our trained models. Panel A shows ten examples of reviews from the pros section that receive the highest quality scores. Panel B shows ten examples of reviews from the cons section that receive the lowest quality scores. Company names are anonymized as "[Company Name]". Lengthy reviews are presented in excerpted form due to space constraints; omitted content is indicated by "[. . .]".

Panel A: Pros Reviews with High Predicted Quality Scores

No.	Review
1	Great Benefits, Great training, Pay is decent, Great name in the industry and overall great products and services to customers.
2	Overall, great company! They have a great reputation and provide high quality products. They offer both a pension and 401k plans which is very hard to find in the market today. They take great care of their employees.
3	Great benefits including educational assistance. Good core values. The company recently acquired [Company Name] which was a great company too. Hopefully it will be a good thing in the long run. Both companies were known for quality products, and the integration should bring an even higher quality of customer service.
4	Great benefits (even for part-timers), you can get benefits for your whole family (even your pets), 401 K, stocks, bonds, flexible work schedule, great employees, room for advancement if you are dedicated to making it into a career, and wonderful training. Most employees are hard working and will help you at the drop of a dime! Everything is centered around providing great customer service.
5	Fair Compensation, Great benefits, great people overall. The products were excellent and it felt good to say that I worked for [Company Name]. Over all leadership was very strong and forward looking. Very supportive of their employees.
6	During my career at [Company Name], I have discovered so many opportunities to develop and advance my career. The training, tools, and support provided have helped me to be successful through the journey. [Company Name] has a strategy that guides us to be more innovative and future focused as we continue to provide reliable service to our customers. Our core values are focused on our employees, our customers and the communities we serve, allowing us to care for others and do the right things while improving the way we communicate and interact with our customers. I'm proud to be a [Company Name] employee and for the work we do.
7	The core values really set [Company Name] apart from other companies. It can be an amazing company to work for if you are under the right leadership. There is training lots of support from the corporate office and the benefits are great. It is a growing yet stable company. Upper management does care about their employees in a sense of trying to get to know everyone.
8	Very stable company; nation-wide branches; multiple subsidiary companies offer many opportunities; focus on customer service; fair location-adjusted wages, good benefits; high community involvement; relaxed yet focused company culture; supports ongoing education of employees

Table A3. Examples of Reviews Most Predictive of High and Low Quality Scores (Continued)

No.	Review
9	The company has good morals and intentions with safety of the workforce and quality of the product. There is no doubt in my mind that this company will continue to be successful in the future. The work atmosphere is good, and the people in all the departments I have been in are very helpful and generally happy to help. The company also places a lot of value in training their employees.
10	[Company Name] has great training programs in place so you feel prepared to work when you step onto the floor. The teachings continue throughout your employment so you are always learning how to better yourself. The open door policy is great making it very easy for people at a lower ranking to make suggestions to people with a higher job title. When making suggestions I felt as though they actually took my thoughts and suggestions seriously. The promote from within policy is a great incentive to work hard. The culture of [Company Name] is great and makes you proud to work for a company with such a great history.

Table A3. Examples of Reviews Most Predictive of High and Low Quality Scores (Continued)

Panel B: Cons Reviews with Low Predicted Quality Scores

No.	Review
1	Compensation and health benefits are not competitive. Lack of respect and appreciation for employees from upper management [...]. Unmanageable workload with no end in sight has resulted in low morale and quality of work has been compromised across the board. Support and communication from HR or the CEO are virtually non-existent.
2	Post merger all downhill. Unfortunately, [Company Name] upper management was given control and these people only care about what they can add to their resume. Absolutely no managerial training, no support for the middle managers, no customer values and the worst customer service I have ever seen!! [...]
3	[...] Quality of work by off-shore teams is poor. You get what you pay for. Quality of outsourcing is terrible, on both the North America and India sides. This is an issue if you're on a major project and you're doing your job well, but that team has no idea what they're doing and is completely botching the client relationship [...]. Limited training opportunities [...].
4	Job is a customer service yet they are doing everything possible to make that impossible for you. To start the system they are now moving everyone onto still doesn't work properly, causes lots of issues, goes down or doesn't work making your life hell as they tell you not to let the customers know and to lie to them. Different departments won't take responsibility for their own mistakes and then just comes down to you to deal with it as best as you can [...]. Since the merger with [Company Name] lots of things have changed, lots of belt tightening and morale is unsure as people are wary of jobs being lost [...].
5	Low pay. No benefits. Horrible health insurance. Unrealistic work expectations, constant change and then they get mad when it's not done the way they wanted it but they just changed it again. Upper management rarely comes in but is quick to judge when they do [...]. Always short staffed [...].
6	Prior to the [Company Name] merger [Company Name] was a revered and respected brand but the [Company Name], in its present form, has no future for any current or prospective employee [...]. Executive management has over promised to the street and when the results of their failure become evident management's reaction, in one trick pony fashion, is to enact massive irrational and incoherent layoffs across the enterprise.
7	Awful management and leadership practices, very ineffective [...]. Poor engagement scores among employees despite advertising otherwise [...]. Highly toxic work environment [...]. Management decisions regularly compromise advertised corporate values. Highly political environment, very difficult to get anything done. There's a large dictatorship faction in the company. Management functions with a strong preference for command & control, and doesn't heed advice from front-line employees [...].
8	Lower pay than what I've experienced in the industry. The health insurance benefits are the worst I've ever had. Workload for new employees sets you up for failure. I'm very competent and good at what I do, however I find myself the most stressed I've ever been at a job [...].

Table A3. Examples of Reviews Most Predictive of High and Low Quality Scores (Continued)

No.	Review
9	[...] [Company Name] overworks its employees [...]. [Company Name] loves to lay off people [...]. [Company Name] executives are short sighted and distracted by the stock price and earnings per share at the expense of their employees and customers.
10	[...] Business processes are old fashioned and people resource heavy. Very little spend has been made in the past 10 years to provide systems and technology to streamline and make the customer experience better [...]. Customers again the losers in lower quality of service [...]. Software product releases are inadequately tested, fail to meet customer expectations and often damage the customer's business or interrupt work flows [...]. Lack of vision, standards and leadership on the development side are likely culprits.

Appendix B. Variable Definition

Table B1. Variable Definition

Panel A: Compustat Firm Characteristics

Variable	Definition	Source
Size	The natural logarithm of total assets	Compustat
Leverage	Total debt divided by total assets	Compustat
ROA	Return on assets, calculated as income before extraordinary items divided by total assets	Compustat
M/B Ratio	Market value of equity divided by book value of equity	Compustat
Sales Growth	Year-over-year percentage change in sales	Compustat
Operating Margin	Operating income after depreciation divided by sales revenue	Compustat
R&D Intensity	Research and development expenses divided by total assets	Compustat
Ltm ROA	Average return on assets (calculated as income before extraordinary items divided by total assets) over the subsequent three years	Compustat
Ltm Operating Margin	Average operating margin (calculated as operating income after depreciation divided by sales revenue) over the subsequent three years	Compustat

Panel B: Other Firm Characteristics

Variable	Definition	Source
% Transient Investor	The percentage of outstanding shares held by transient investors. We follow the definition in Bushee (2001) to identify transient investors	F13
Analyst Following	The natural logarithm of the number of analysts following the firm; missing values are replaced with zero	I/B/E/S

Panel C: Customer Satisfaction

Variable	Definition	Source
Industry-Adjusted ACSI	Industry-adjusted American Customer Satisfaction Index score, standardized by Fama-French 48 industries	ACSI

Panel D: Quality Index

Variable	Definition	Source
Quality Index	Text-based firm-year quality measure derived from employee written reviews on Glassdoor. Higher scores indicate better quality provision	
Large Quality Drop	Indicator equals one if the change in quality index from year $t-1$ to year t falls in the bottom quartile of the sample	

Table B1. Variable Definition (Continued)

Panel E: Glassdoor Ratings

Variable	Definition	Source
Overall Rating	Glassdoor overall rating averaged by firm-year	Glassdoor
Balance Subrating	Glassdoor work-life balance rating averaged by firm-year	Glassdoor
Culture Subrating	Glassdoor culture and values rating averaged by firm-year	Glassdoor
Opportunities Subrating	Glassdoor career opportunities rating averaged by firm-year	Glassdoor
Compensation Subrating	Glassdoor compensation and benefits rating averaged by firm-year	Glassdoor
Management Subrating	Glassdoor senior management rating averaged by firm-year	Glassdoor

Panel F: Product Recalls

Variable	Definition	Source
# Recalls	The number of product recalls during the year	CPSC
# Reactive Recalls	The number of reactive product recalls during the year	CPSC

Panel G: MSCI

Variable	Definition	Source
Quality Management	This score evaluates the extent to which companies are at risk of facing major product recalls or losing customer trust through major product quality concerns. Companies that proactively manage product quality by achieving certification to widely acceptable standards, undertaking extensive product testing and building processes to track raw materials or components score higher. Companies that take a reactive approach to managing recalls and product quality concerns score lower. (Score: 0-10)	MSCI
Quality Risk Exposure	Measures a company's vulnerability to ESG risks related to product safety and quality based on products/services, operational locations, and business nature. Higher scores indicate greater risk exposure. (Score: 0-10)	MSCI
#Strengths - #Concerns	Calculate by subtracting the number of product characteristic concerns from the number of product characteristic strengths	MSCI
Has Product Concern	Indicator variable equal to one if the firm has product characteristic concerns in a given year, and zero otherwise	MSCI
Has Safety Concern	Indicator variable equal to one if the firm has product safety concerns in a given year, and zero otherwise	MSCI

Panel H: Brand Rankings

Variable	Definition	Source
Top Brand	Indicator variable equal to one if the firm is listed in the Brand Finance Global 500 ranking in the given year, and zero otherwise	Brand Finance
Ltm Top Brand	Indicator variable equal to one if the firm is listed in the Brand Finance Global 500 ranking in the next three years, and zero otherwise	Brand Finance

Table B1. Variable Definition (Continued)

Panel I: Glassdoor-Based Cultural Value Scores

Variable	Definition	Source
Innovation Score GD	Difference in weighted frequency of innovation-related words between the Pros and Cons sections of Glassdoor reviews	Glassdoor
Integrity Score GD	Difference in weighted frequency of integrity-related words between the Pros and Cons sections of Glassdoor reviews	Glassdoor
Quality Score GD	Difference in weighted frequency of quality-related words between the Pros and Cons sections of Glassdoor reviews	Glassdoor
Respect Score GD	Difference in weighted frequency of respect-related words between the Pros and Cons sections of Glassdoor reviews	Glassdoor
Teamwork Score GD	Difference in weighted frequency of teamwork-related words between the Pros and Cons sections of Glassdoor reviews	Glassdoor

Panel J: Conference Call-Based Cultural Value Scores

Variable	Definition	Source
Innovation Score CC	Weighted-frequency count of innovation-related words in conference call transcripts	Capital IQ
Integrity Score CC	Weighted-frequency count of integrity-related words in conference call transcripts	Capital IQ
Quality Score CC	Weighted-frequency count of quality-related words in conference call transcripts	Capital IQ
Respect Score CC	Weighted-frequency count of respect-related words in conference call transcripts	Capital IQ
Teamwork Score CC	Weighted-frequency count of teamwork-related words in conference call transcripts	Capital IQ

Online Appendix:

The Informed Insider: A Leading Measure of Quality

This online appendix contains additional robustness checks and supplementary analyses for the main paper.

Contents

OA1: Machine Learning Implementation Details **2**

OA2: Chat-GPT Prompts for Industry Classification **3**

OA3: Zero-Shot GPT Prompts for Quality Scoring **4**

OA4: Zero-Shot GPT Quality Index: Other Validation and Implication Tests **9**

Table OA4.1: Zero-Shot GPT Quality Index: MSCI Validation Tests 9

Table OA4.2: Zero-Shot GPT Quality Index Comparison Across Groups 10

Table OA4.3: Zero-Shot GPT Quality Index and Product Recalls 11

Table OA4.4: Zero-Shot GPT Quality Index and Brand Value 13

Table OA4.5: Zero-Shot GPT Quality Index and Future Profitability 17

OA5: Prompts for Constructing Fine-Tuned GPT Measures **19**

OA6: Fine-Tuned GPT Quality Index: Other Validation and Implication Tests ... **20**

Figure OA6.1: True vs. Predicted Quality Quintiles in the Testing Sample 20

Table OA6.1: Fine-Tuned GPT Quality Index: MSCI Validation Tests 21

Table OA6.2: Fine-Tuned GPT Quality Index Comparison Across Groups 22

Table OA6.3: Fine-Tuned GPT Quality Index and Product Recalls 23

Table OA6.4: Fine-Tuned GPT Quality Index and Brand Value 25

Table OA6.5: Fine-Tuned GPT Quality Index and Future Profitability 27

OA1: Machine Learning Implementation Details

We implement the random forest model using the `RandomForestRegressor` class from Python's scikit-learn library.[25] We set `n_estimators = 3000` and `max_features = sqrt`.

We implement the XGBoost model using the `XGBRegressor` class from the XGBoost Python library.[26] We set `subsample` at 0.8 and `colsample_bytree` at 0.8, which introduces controlled randomness through bootstrap sampling of observations and features during each boosting iteration. We concentrate our hyperparameter search on critical parameters to balance predictive performance with computational efficiency. Hyperparameter tuning is conducted using `GridSearchCV` from `scikit-learn`, combined with a time series split cross-validation strategy to preserve the temporal ordering of observations. We use negative mean squared error as the scoring metric, and evaluate the following parameter grid:

`n_estimators`:	[500, 800, 1000, 1500]
`learning_rate`:	[0.01, 0.03, 0.05]
`max_depth`:	[2, 3, 4]

The grid search evaluates all parameter combinations across the training folds defined by the time series split and selects the configuration that minimizes validation error. The final model is retrained on the full training set using the best-performing hyperparameters.

[25] Available at http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html.

[26] Available at https://xgboost.readthedocs.io/en/stable/.

OA2: Chat-GPT Prompts for Industry Classification

We employed OpenAI's GPT-4o model for automated industry classification. The following prompt was used to classify industries into product-oriented or service-oriented categories:

> Classify the following industry as either 'product-oriented' or 'service-oriented'.
> Definitions:
> - A product-oriented industry primarily involves the production, processing, or distribution of physical goods.
> - A service-oriented industry primarily involves delivering intangible services to consumers or businesses.
> You are given the industry group name and a specific industry description. Respond with exactly one of the following terms: 'product-oriented' or 'service-oriented'.

OA3: Zero-Shot GPT Prompts for Quality Scoring

The following prompt was used to classify reviews from the pros section:

System Prompt:
Assume you are a helpful research assistant whose task to label the quality of products and services that a firm provides based on information extracted from employee-written reviews about their firms. You need to label the given reviews according to the rubric below. You must base your label only on the content of the provided review.
Quality rubric:
Q2 – High quality: label reviews Q2 if they contain clear, explicit, and specific praise for the firm's internal operating processes, workplace systems, support, or management practices that positively contribute to firm's ability to provide high-quality, promised, consistent, reliable, or safe products or services. Assess whether the review reflects excellent quality conditions by:
- looking for explicit mentions of the product or service performance. This includes signals such as praises of excellent product features or reliable service, positive customer feedback and interactions, structured and sound internal control practices (e.g., thorough testing process; machine maintenance, establishing standard procedures for operations) OR
- looking for mentions of aspects that are positively linked to quality-related consequences in an implicit way. This could include signals such as:
- firm's financial position: healthy financial positions and cash flow, good inventory management and no shortfalls, proactive investment in workplace safety so that the consequences for product/service quality are described or strongly implied;
- working conditions and work pressure: respect for personal boundaries, manageable workload, flexibility, and support for well-being so that the consequences for product/service quality are described or strongly implied;
- pay and compensation: competitive pay, aligned incentive pay, meaningful bonuses and benefits so that the consequences for product/service quality are described or strongly implied;
- other organizational aspects: high employee satisfaction which positively affects service standards, quality-driven/integrity-driven culture, operational clarity and stability so that the consequences for product/service quality are described or strongly implied
Q1 – Mediocre quality: label reviews Q1 if they indicate average, inconsistent, or mixed signals regarding the firm's quality provision. Apply Q1 if the review either:
- Describes quality outcomes or internal operations that are adequate but not exemplary, or show occasional success without sustained reliability OR
- Suggest that some internal quality control systems or practices are in place, but they are either applied unevenly, underdeveloped, or not enforced consistently OR
- Lacks strong evidence of proactive quality investment, OR
- Shows mixed, moderate or ambiguous indicators of the signals provided in evaluating Q2.
Q0 – Low quality: label reviews Q0 if they contain no substantive praises related to quality provision or only raise generic or non-operational issues. The absence of substantive praises is informative and can signal low quality. Apply Q0 if the review either:
- Include filler, minimal, or placeholder text (e.g., "N/A", "None", "Can't think of anything") or explicitly states things like there are "no real pros" or "nothing significant," OR
- Contain generic praise (e.g., "It's fine", "No issues", "Pretty good overall") with no elaboration and unrelated to firm operations, product and service delivery, OR

- Mentions only trivial, personal preferences, or context-specific factors (e.g., perks or amenities) unrelated to firm operations, product and service delivery, OR
- Shows no strong positive indicators of the signals provided in evaluating Q2.

User Prompt:
Read the provided reviews. Then do sixteen things:
1. Reason step by step to decide which of the labels (Q0/Q1/Q2) from the quality rubric you were given applies to firm quality provision as suggested by the reviews. Report. Give an explanation for which label applies and report the label that you think fits best. Do not say zero or N/A.
2. Report only the label that you determined for the review, which should match the label in step 1. Do not reply N/A.
3. Given the reasoning in step 1, describe the extent to which the reviews reflect firm quality provision-the reviews explicitly mention the existence of quality issues, the reviews implicitly suggest the existence of quality issues, or the reviews are irrelevant to firm quality provision. Do not reply N/A.
4. Given the amount of speculation required in step 1, describe your certainty about the estimate-either high, moderate, or low. Do not reply N/A.
5. Given the reasoning in step 1, describe whether the reviews reflect firm quality provision by describing the firm's financial position. Do not reply N/A.
6. Given the reasoning in step 1 and your response to step 5, consider only the firm's financial position aspect described in the review. Report the label that you determined for the review, which may not necessarily match the label in step 1. Do not reply N/A.
7. Given the amount of speculation required in step 5, describe your certainty about the estimate-either high, moderate, or low. Do not reply N/A.
8. Given the reasoning in step 1, describe whether the reviews reflect firm quality provision by describing the firm's working conditions and work pressure. Do not reply N/A.
9. Given the reasoning in step 1 and your response to step 8, consider only the working conditions and work pressure aspect described in the review. Report the label that you determined for the review, which may not necessarily match the label in step 1. Do not reply N/A.
10. Given the amount of speculation required in step 8, describe your certainty about the estimate-either high, moderate, or low. Do not reply N/A.
11. Given the reasoning in step 1, describe whether the reviews reflect firm quality provision by describing the firm's pay and compensation. Do not reply N/A.
12. Given the reasoning in step 1 and your response to step 11, consider only the pay and compensation aspect described in the review. Report the label that you determined for the review, which may not necessarily match the label in step 1. Do not reply N/A.
13. Given the amount of speculation required in step 11, describe your certainty about the estimate-either high, moderate, or low. Do not reply N/A.
14. Given the reasoning in step 1, describe whether the reviews reflect firm quality provision by describing the firm's other organizational aspects. Do not reply N/A.
15. Given the reasoning in step 1 and your response to step 14, consider only the other organizational aspects aspect described in the review. Report the label that you determined for the review, which may not necessarily match the label in step 1. Do not reply N/A.

16. Given the amount of speculation required in step 14, describe your certainty about the estimate-either high, moderate, or low. Do not reply N/A.

The following prompt was used to classify reviews from the cons section:

System Prompt:
Assume you are a helpful research assistant whose task to label the quality of products and services that a firm provides based on information extracted from employee-written reviews about their firms. You need to label the given reviews according to the rubric below. You must base your label only on the content of the provided review.
Quality rubric:
Q0 – Low quality: label reviews Q0 if they indicate systematic quality problems or frequent operational failures that reflect poorly on the firm's ability to provide promised, consistent, reliable, or safe products/services. Assess whether the review reflects serious quality conditions by:
- looking for explicit mentions of the product or service performance. This includes mentions of product defects, safety issues, unreliable service, declined product and service standards, complaints regarding quality ignored, negative customer feedback/treatment, or flawed/deteriorated quality control practices OR
- looking for mentions of aspects that negatively linked to quality-related consequences in an implicit way. This could include signals such as:
- firm's financial position: financial constraints, high financial leverage, inventory shortfalls so that the consequences for product/service quality are described or strongly implied;
- working conditions and work pressure: understaffing, excessive workloads, high worker turnover, physical/emotional strain so that the consequences for product/service quality are described or strongly implied;
- pay and compensation: lack of incentive pays, lack of aligned incentives, unrealistic performance targets, lack of employment security and training, pay justices so that the consequences for product/service quality are described or strongly implied;
- other organizational aspects: low employee satisfaction which negatively affects service standards, quality-poor/integrity-poor culture (e.g., saying one thing but doing another), frequent restructuring, disrupting operating consistency so that the consequences for product/service quality are described or strongly implied;
Q1 – Mediocre quality: label reviews Q1 if they indicate average, inconsistent, or mixed signals regarding the firm's quality provision. Apply Q1 if the review either:
- Does not describe severe breakdowns or dysfunction of internal operating processes OR
- Suggest that some internal quality control systems or practices are in place, but they are either applied unevenly, underdeveloped, or not enforced consistently OR
- Lacks strong evidence of proactive quality investment, OR
- Shows mixed, moderate or ambiguous indicators of the signals provided in evaluating Q0.
Q2 – High quality: label reviews Q2 if they reveal no substantive concerns related to quality provision or only raise generic or non-operational issues. The absence of serious criticism is informative and can signal high quality. Apply Q2 if the review either:

- Include filler, minimal, or placeholder text (e.g., "N/A", "None", "Can't think of anything") or explicitly states things like there are "no real cons" or "nothing significant," OR
- Contain generic praise (e.g., "It's bad", "Pretty bad overall") with no elaboration and unrelated to firm operations, product and service delivery, OR
- Mentions only trivial, personal preferences, or context-specific factors (e.g., perks or amenities) unrelated to firm operations, product and service delivery, OR
- Shows no strong negative indicators of the signals provided in evaluating Q0.

User Prompt:
Read the provided reviews. Then do sixteen things:
1. Reason step by step to decide which of the labels (Q0/Q1/Q2) from the quality rubric you were given applies to firm quality provision as suggested by the reviews. Report. Give an explanation for which label applies and report the label that you think fits best. Do not say zero or N/A.
2. Report only the label that you determined for the review, which should match the label in step 1. Do not reply N/A.
3. Given the reasoning in step 1, describe the extent to which the reviews reflect firm quality provision-the reviews explicitly mention the existence of quality issues, the reviews implicitly suggest the existence of quality issues, or the reviews are irrelevant to firm quality provision. Do not reply N/A.
4. Given the amount of speculation required in step 1, describe your certainty about the estimate-either high, moderate, or low. Do not reply N/A.
5. Given the reasoning in step 1, describe whether the reviews reflect firm quality provision by describing the firm's financial position. Do not reply N/A.
6. Given the reasoning in step 1 and your response to step 5, consider only the firm's financial position aspect described in the review. Report the label that you determined for the review, which may not necessarily match the label in step 1. Do not reply N/A.
7. Given the amount of speculation required in step 5, describe your certainty about the estimate-either high, moderate, or low. Do not reply N/A.
8. Given the reasoning in step 1, describe whether the reviews reflect firm quality provision by describing the firm's working conditions and work pressure. Do not reply N/A.
9. Given the reasoning in step 1 and your response to step 8, consider only the working conditions and work pressure aspect described in the review. Report the label that you determined for the review, which may not necessarily match the label in step 1. Do not reply N/A.
10. Given the amount of speculation required in step 8, describe your certainty about the estimate-either high, moderate, or low. Do not reply N/A.
11. Given the reasoning in step 1, describe whether the reviews reflect firm quality provision by describing the firm's pay and compensation. Do not reply N/A.
12. Given the reasoning in step 1 and your response to step 11, consider only the pay and compensation aspect described in the review. Report the label that you determined for the review, which may not necessarily match the label in step 1. Do not reply N/A.
13. Given the amount of speculation required in step 11, describe your certainty about the estimate-either high, moderate, or low. Do not reply N/A.
14. Given the reasoning in step 1, describe whether the reviews reflect firm quality provision by describing the firm's other organizational aspects. Do not reply N/A.

15. Given the reasoning in step 1 and your response to step 14, consider only the other organizational aspects aspect described in the review. Report the label that you determined for the review, which may not necessarily match the label in step 1. Do not reply N/A.
16. Given the amount of speculation required in step 14, describe your certainty about the estimate-either high, moderate, or low. Do not reply N/A.

OA4: Zero-Shot GPT Quality Index: Other Validation and Implication Tests

Table OA4.1. Zero-Shot GPT Quality Index: MSCI Validation Tests

Notes: This table reports correlations of zero-shot GPT quality indices and MSCI product quality related scores. p-values are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Pr. Sign	Pros		Cons		Aggregate	
		Overall	Relevant Only	Overall	Relevant Only	Overall	Relevant Only
Quality Risk Exposure	-	0.014	0.015	-0.086***	-0.027***	-0.052***	-0.009
		(0.141)	(0.143)	(0.000)	(0.006)	(0.000)	(0.368)
Has Product Concern	-	0.013	0.009	0.003	0.039	0.011	0.038
		(0.597)	(0.714)	(0.898)	(0.120)	(0.643)	(0.145)
Has Safety Concern	-	0.022*	0.016	-0.016	0.033***	0.003	0.032**
		(0.071)	(0.196)	(0.170)	(0.006)	(0.831)	(0.011)
#Strengths - #Concerns	+	-0.010	0.000	0.037	0.004	0.019	0.016
		(0.694)	(0.990)	(0.126)	(0.861)	(0.432)	(0.542)
Quality Management	+	0.075***	0.068***	-0.010	-0.017	0.040**	0.017
		(0.000)	(0.000)	(0.530)	(0.320)	(0.015)	(0.333)

Table OA4.2. Zero-Shot GPT Quality Index Comparison Across Groups

Notes: This table reports the comparisons of zero-shot GPT quality indices between firms listed among the best in product/service quality and other firms based on *Fortune's World's Most Admired Companies* rankings. The "Best Quality" group includes firms listed among the best in product/service quality from 2008 to 2012 in the Fortune rankings, while the "Other" group includes all remaining firms that participated in the rankings during this period but were not listed among the best in product/service quality. Panel A shows the results of the zero-shot GPT quality index generated from "pros" section, Panel B shows the results of zero-shot GPT quality index generated from "cons" section, and Panel C shows the results of the aggregate zero-shot GPT quality index. p-values are reported in parentheses under the differences. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Group Mean		Diff.
	(1)	(2)	(1)-(2)
	Best Quality	Other	Best vs Other
Panel A: Pros, Zero-Shot GPT Quality Index			
Overall			
Mean	-0.169	-0.199	0.030**
			(0.026)
Relevant Only			
Mean	0.201	0.166	0.035***
			(0.003)
Panel B: Cons, Zero-Shot GPT Quality Index			
Overall			
Mean	0.004	-0.060	0.064***
			(0.000)
Relevant Only			
Mean	-0.285	-0.325	0.040***
			(0.000)
Panel C: Aggregate Zero-Shot GPT Quality Index			
Overall			
Mean	-0.165	-0.260	0.094***
			(0.000)
Relevant Only			
Mean	-0.082	-0.160	0.078***
			(0.000)

Table OA4.3. Zero-Shot GPT Quality Index and Product Recalls

Notes: This table reports tests on the relationship between zero-shot GPT quality indices and CPSC product recalls. Panels A and B report regression results on the relationship between changes in cons-based (Panel A) and pros-based (Panel B) zero-shot GPT quality indices and product recalls in the subsequent year. Panel C reports regression results on the relationship between changes in aggregate zero-shot GPT quality indices and product recalls in the subsequent year. The dependent variable in columns (1)–(4) is the total number of recalls in the subsequent year, and the dependent variable in columns (5)–(8) is the total number of reactive recalls over the same period. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Changes in Cons, Zero-Shot GPT Quality Index and Future Recall Events

	# Recalls				# Reactive Recalls			
	Overall	Relevant Only	Overall	Relevant Only	Overall	Relevant Only	Overall	Relevant Only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ*Quality Index - Cons*	-0.097	-0.159			0.065	0.069		
	(-0.52)	(-0.47)			(0.32)	(0.20)		
Large Quality Drop			0.129	0.005			-0.152	-0.389
			(0.69)	(0.02)			(-0.65)	(-0.98)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,245	4,775	5,240	4,672	5,240	4,256	5,245	4,775
Pseudo R^2	0.442	0.441	0.442	0.443	0.436	0.417	0.436	0.425

Panel B: Changes in Pros, Zero-Shot GPT Quality Index and Future Recall Events

	# Recalls				# Reactive Recalls			
	Overall	Relevant Only	Overall	Relevant Only	Overall	Relevant Only	Overall	Relevant Only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ*Quality Index - Pros*	-0.082	-0.224			0.184	0.473**		
	(-0.45)	(-0.92)			(0.98)	(2.19)		
Large Quality Drop			0.339*	0.420*			0.162	-0.238
			(1.88)	(1.78)			(0.89)	(-0.93)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,245	4,775	5,245	4,775	5,245	4,775	5,245	4,775
Pseudo R^2	0.451	0.451	0.452	0.452	0.437	0.425	0.437	0.424

Panel C: Changes in Aggregate Zero-Shot GPT Quality Index and Future Recall Events

	# Recalls				# Reactive Recalls			
	Overall	Relevant Only	Overall	Relevant Only	Overall	Relevant Only	Overall	Relevant Only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ*Quality Index - Agg*	0.029	-0.362			0.126	0.317**		
	(0.14)	(-1.32)			(0.93)	(2.16)		
Large Quality Drop			0.202	0.296**			0.031	-0.395
			(1.52)	(2.39)			(0.14)	(-1.10)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,245	4,775	5,240	4,469	5,240	4,074	5,245	4,775
Pseudo R^2	0.442	0.442	0.443	0.449	0.437	0.429	0.436	0.425

Table OA4.4. Zero-Shot GPT Quality Index and Brand Value

Notes: This table presents tests on the relationship between zero-shot GPT quality indices and the likelihood of appearing in Brand Finance's Global 500 brand rankings. Panel A (B) shows regression results of a dummy indicating whether the firm is listed in Brand Finance's Global 500 brand rankings on the pros (cons) zero-shot GPT quality index. Panel C shows regression results of a dummy indicating whether the firm is listed in Brand Finance's Global 500 brand rankings on the aggregate zero-shot GPT quality index. *Top Brand* equals 1 if the firm appears in Brand Finance's Global 500 rankings in the following year, and 0 otherwise. *Ltm Top Brand* equals 1 if the firm appears in the Global 500 rankings in the next three years, and 0 otherwise. Panel A.1, B.1, and C.1 report results using zero-shot GPT quality indices constructed using all sampled reviews. Panel A.2, B.2, C.2 report results using zero-shot GPT quality indices constructed using relevant reviews (i.e., those identified by GPT as explicitly mentioning quality issues or implicitly suggesting them). All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Pros, Zero-Shot GPT Quality Index and Top 500 Brand

Panel A.1: Overall

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Pros	-0.003	-0.000	-0.005	-0.079
	(-0.46)	(-0.00)	(-1.01)	(-0.38)
Size	0.102***	1.070***	0.089***	1.125***
	(16.37)	(11.71)	(13.67)	(10.65)
Leverage	-0.113***	-0.093	-0.105***	-0.007
	(-4.18)	(-0.29)	(-3.93)	(-0.02)
ROA	-0.093**	1.742***	-0.073**	2.368***
	(-2.52)	(3.09)	(-2.04)	(3.56)
M/B Ratio	0.026***	0.246***	0.024***	0.254***
	(7.47)	(7.14)	(6.75)	(6.33)
Sales Growth	-0.036***	-0.301**	-0.030***	-0.276
	(-3.08)	(-2.13)	(-2.66)	(-1.61)
Constant	-0.675***	-12.159***	-0.646***	-11.212***
	(-13.85)	(-11.17)	(-13.07)	(-10.38)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	15,996	14,253	14,719	12,263
Adjusted/Pseudo R^2	0.309	0.559	0.284	0.578

Table OA4.4. Zero-Shot GPT Quality Index and Brand Value (Continued)

Panel A.2: Relevant Only

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Pros	-0.006	-0.093	-0.009*	-0.162
	(-1.02)	(-0.79)	(-1.66)	(-1.20)
Size	0.106***	1.075***	0.094***	1.138***
	(16.61)	(11.51)	(13.92)	(10.59)
Leverage	-0.113***	-0.045	-0.106***	0.065
	(-4.01)	(-0.14)	(-3.77)	(0.18)
ROA	-0.096**	1.724***	-0.074*	2.373***
	(-2.42)	(3.04)	(-1.91)	(3.53)
M/B Ratio	0.027***	0.246***	0.025***	0.259***
	(7.43)	(7.10)	(6.81)	(6.42)
Sales Growth	-0.034***	-0.276*	-0.029**	-0.262
	(-2.72)	(-1.93)	(-2.39)	(-1.51)
Constant	-0.704***	-12.233***	-0.676***	-11.330***
	(-14.00)	(-11.09)	(-13.19)	(-10.68)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	14,831	13,146	13,619	11,430
Adjusted/Pseudo R^2	0.318	0.557	0.294	0.578

Panel B: Cons, Zero-Shot GPT Quality Index and Top 500 Brand

Panel B.1: Overall

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Cons	0.001	-0.094	0.000	-0.087
	(0.15)	(-0.71)	(0.00)	(-0.51)
Size	0.102***	1.072***	0.089***	1.127***
	(16.39)	(12.10)	(13.69)	(11.13)
Leverage	-0.112***	-0.101	-0.104***	-0.013
	(-4.18)	(-0.32)	(-3.92)	(-0.04)
ROA	-0.093**	1.744***	-0.072**	2.373***
	(-2.52)	(3.11)	(-2.03)	(3.58)
M/B Ratio	0.026***	0.247***	0.024***	0.255***
	(7.47)	(7.45)	(6.73)	(6.68)
Sales Growth	-0.035***	-0.302**	-0.029***	-0.276
	(-3.06)	(-2.14)	(-2.65)	(-1.62)
Constant	-0.675***	-12.211***	-0.646***	-11.241***
	(-13.91)	(-11.55)	(-13.10)	(-11.20)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	15,997	14,252	14,720	12,262
Adjusted/Pseudo R^2	0.309	0.559	0.284	0.578

Table OA4.4. Zero-Shot GPT Quality Index and Brand Value (Continued)

Panel B.2: Relevant Only

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Cons	0.004	0.039	0.004	0.059
	(0.65)	(0.29)	(0.69)	(0.33)
Size	0.104***	1.067***	0.091***	1.121***
	(16.52)	(11.84)	(13.78)	(10.91)
Leverage	-0.112***	-0.088	-0.104***	0.008
	(-4.09)	(-0.28)	(-3.83)	(0.02)
ROA	-0.093**	1.729***	-0.073*	2.364***
	(-2.45)	(3.08)	(-1.96)	(3.57)
M/B Ratio	0.026***	0.245***	0.024***	0.253***
	(7.44)	(7.27)	(6.72)	(6.50)
Sales Growth	-0.036***	-0.298**	-0.030**	-0.271
	(-2.99)	(-2.11)	(-2.58)	(-1.58)
Constant	-0.690***	-12.128***	-0.657***	-11.148***
	(-14.06)	(-11.38)	(-13.16)	(-11.07)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	15,584	13,891	14,344	11,993
Adjusted/Pseudo R^2	0.312	0.556	0.288	0.576

Panel C: Aggregate Zero-Shot GPT Quality Index and Top 500 Brand

Panel C.1: Overall

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Agg	-0.001	-0.046	-0.002	-0.077
	(-0.19)	(-0.58)	(-0.67)	(-0.74)
Size	0.102***	1.071***	0.089***	1.128***
	(16.38)	(11.81)	(13.68)	(10.79)
Leverage	-0.113***	-0.099	-0.105***	-0.015
	(-4.18)	(-0.31)	(-3.93)	(-0.04)
ROA	-0.093**	1.745***	-0.073**	2.374***
	(-2.52)	(3.09)	(-2.03)	(3.56)
M/B Ratio	0.026***	0.247***	0.024***	0.256***
	(7.47)	(7.29)	(6.75)	(6.51)
Sales Growth	-0.035***	-0.302**	-0.029***	-0.276
	(-3.07)	(-2.14)	(-2.66)	(-1.62)
Constant	-0.675***	-12.183***	-0.646***	-11.253***
	(-13.88)	(-11.27)	(-13.10)	(-10.73)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	15,991	14,248	14,714	12,258
Adjusted/Pseudo R^2	0.309	0.559	0.284	0.578

Table OA4.4. Zero-Shot GPT Quality Index and Brand Value (Continued)

Panel C.2: Relevant Only				
	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Agg	-0.001	0.028	-0.002	0.014
	(-0.16)	(0.39)	(-0.59)	(0.17)
Size	0.109***	1.073***	0.096***	1.136***
	(16.83)	(11.43)	(14.09)	(10.55)
Leverage	-0.114***	-0.034	-0.106***	0.078
	(-3.89)	(-0.11)	(-3.66)	(0.22)
ROA	-0.092**	1.702***	-0.071*	2.360***
	(-2.24)	(2.99)	(-1.78)	(3.51)
M/B Ratio	0.027***	0.244***	0.025***	0.256***
	(7.32)	(7.09)	(6.74)	(6.42)
Sales Growth	-0.036***	-0.268*	-0.031**	-0.265
	(-2.76)	(-1.89)	(-2.44)	(-1.53)
Constant	-0.735***	-12.261***	-0.704***	-11.361***
	(-14.31)	(-11.02)	(-13.43)	(-10.63)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	14,207	12,621	13,050	11,004
Adjusted/Pseudo R^2	0.323	0.553	0.299	0.575

Table OA4.5. Zero-Shot GPT Quality Index and Future Profitability

Notes: This table presents tests on the relationship between zero-shot GPT quality indices and firm profitability. Panel A (B) shows regression results of return on assets *ROA* (operating margin) on zero-shot GPT quality indices. Panel B shows regression results of operating margin on zero-shot GPT quality indices. *ROA* is the firm's return on assets in the following year, and *Ltm ROA* is the average return on assets over the subsequent three years. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Zero-Shot GPT Quality Index and ROA

Panel A.1: Pros, Zero-Shot GPT Quality Index

	ROA		*Ltm ROA*	
	Overall	Relevant Only	Overall	Relevant Only
	(1)	(2)	(3)	(4)
Quality Index - Pros	-0.003	-0.006**	-0.002	-0.004
	(-1.41)	(-1.98)	(-1.00)	(-1.52)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	15,476	14,345	15,481	14,350
Adjusted R^2	0.218	0.222	0.243	0.249

Panel A.2: Cons, Zero-Shot GPT Quality Index

	ROA		*Ltm ROA*	
	Overall	Relevant Only	Overall	Relevant Only
	(1)	(2)	(3)	(4)
Quality Index - Cons	0.001	0.007**	0.002	0.007**
	(0.46)	(2.46)	(0.82)	(2.47)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	15,477	15,078	15,482	15,083
Adjusted R^2	0.218	0.222	0.243	0.247

Panel A.3: Aggregate Zero-Shot GPT Quality Index

	ROA		*Ltm ROA*	
	Overall	Relevant Only	Overall	Relevant Only
	(1)	(2)	(3)	(4)
Quality Index - Agg	-0.001	0.001	-0.000	0.001
	(-0.54)	(0.72)	(-0.07)	(0.89)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	15,471	13,739	15,476	13,744
Adjusted R^2	0.218	0.225	0.243	0.253

Table OA4.5. Zero-Shot GPT Quality Index and Future Profitability (Continued)

Panel B: Zero-Shot GPT Quality Index and Operating Margin

Panel B.1: Pros, Zero-Shot GPT Quality Index

	Operating Margin		*Ltm Operating Margin*	
	Overall	Relevant Only	Overall	Relevant Only
	(1)	(2)	(3)	(4)
Quality Index - Pros	-0.000	-0.003	0.002	0.000
	(-0.01)	(-0.55)	(0.45)	(0.04)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	15,439	14,314	15,445	14,320
Adjusted R^2	0.192	0.202	0.217	0.227

Panel B.2: Cons, Zero-Shot GPT Quality Index

	Operating Margin		*Ltm Operating Margin*	
	Overall	Relevant Only	Overall	Relevant Only
	(1)	(2)	(3)	(4)
Quality Index - Cons	-0.002	0.001	-0.003	-0.001
	(-0.43)	(0.27)	(-0.87)	(-0.19)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	15,440	15,046	15,446	15,051
Adjusted R^2	0.192	0.194	0.217	0.219

Panel B.3: Aggregate Zero-Shot GPT Quality Index

	Operating Margin		*Ltm Operating Margin*	
	Overall	Relevant Only	Overall	Relevant Only
	(1)	(2)	(3)	(4)
Quality Index - Agg	-0.001	0.000	-0.001	0.000
	(-0.33)	(0.09)	(-0.33)	(0.10)
Firm Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	15,434	13,712	15,440	13,717
Adjusted R^2	0.192	0.206	0.217	0.231

OA5: Prompts for Constructing Fine-Tuned GPT Measures

The following prompt was used to label quality provision based on reviews from the pros section:

> You are a helpful research assistant. Your task is to infer a firm's product and service quality provision from the pros section of employee-written reviews about the firm.
> Carefully analyze how the information reflected in the reviews signals the firm's ability to deliver quality products and services to customers.
> Label the firm's most likely quality level as one of the following: Low, Medium-Low, Medium, Medium-High, or High.
> You must base your label only on the content of the provided review. Please respond with only the category label.

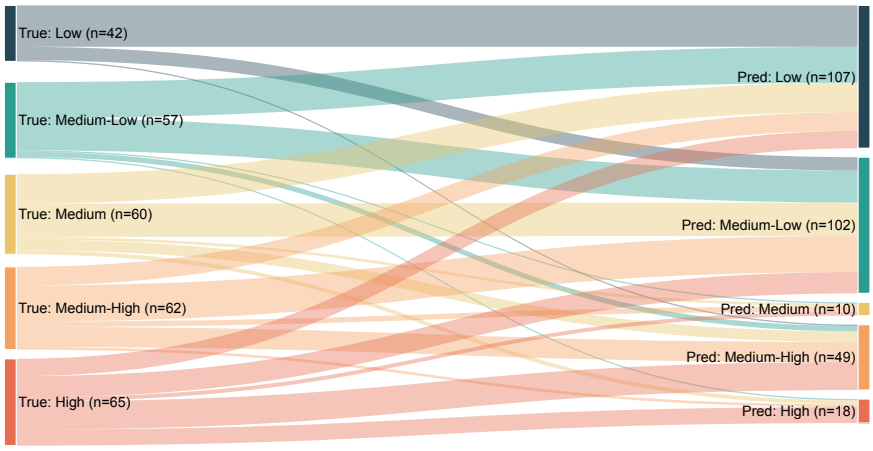
The following prompt was used to label quality provision based on reviews from the cons section:

> You are a helpful research assistant. Your task is to infer a firm's product and service quality provision from the cons section of employee-written reviews about the firm.
> Carefully analyze how the information reflected in the reviews signals the firm's ability to deliver quality products and services to customers.
> Label the firm's most likely quality level as one of the following: Low, Medium-Low, Medium, Medium-High, or High.
> You must base your label only on the content of the provided review. Please respond with only the category label.

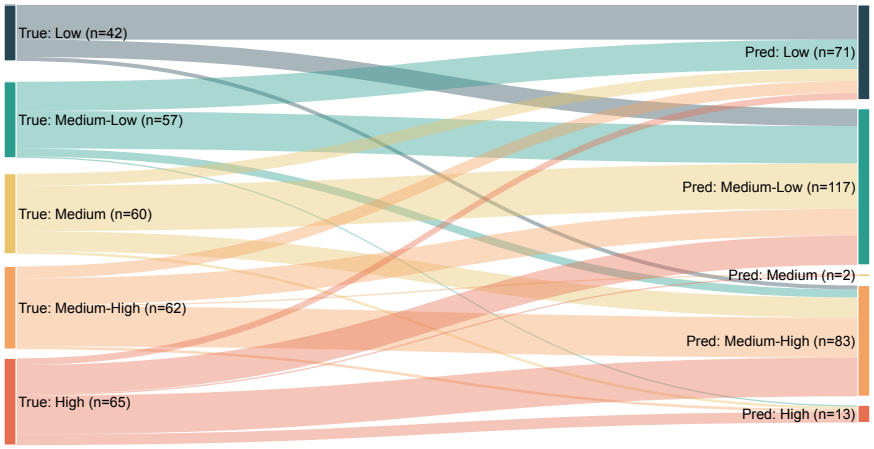
OA6: Fine-Tuned GPT Quality Index: Other Validation and Implication Tests

Figure OA6.1. True vs. Predicted Quality Quintiles in the Testing Sample

Notes: This figure illustrates the relationship between true and predicted firm-year quality quintiles in the testing sample. True quality quintiles are constructed using industry-adjusted ACSI scores. Predicted quality quintiles are generated by fine-tuned GPT models based on employee review text.



(a) Pros-Based Fine-Tuned GPT Quality Index



(b) Cons-Based Fine-Tuned GPT Quality Index

Table OA6.1. Fine-Tuned GPT Quality Index: MSCI Validation Tests

Notes: This table reports correlations of fine-tuned GPT quality indices and MSCI product quality related scores. p-values are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Pred. Sign	Pros	Cons	Aggregate
Quality Risk Exposure	-	-0.083***	-0.002	-0.040***
		(0.000)	(0.859)	(0.000)
Has Product Concern	-	-0.071**	-0.010	-0.052*
		(0.023)	(0.757)	(0.095)
Has Safety Concern	-	-0.076***	-0.012	-0.046***
		(0.000)	(0.373)	(0.001)
#Strengths Minus #Concerns	+	0.089***	0.013	0.064**
		(0.004)	(0.665)	(0.038)
Quality Management	+	0.088***	0.025	0.059***
		(0.000)	(0.180)	(0.002)

Table OA6.2. Fine-Tuned GPT Quality Index Comparison Across Groups

Notes: This table reports the comparisons of fine-tuned GPT quality indices between firms listed among the best in product/service quality and other firms based on *Fortune's World's Most Admired Companies* rankings. The "Best Quality" group includes firms listed among the best in product/service quality from 2008 to 2012 in the Fortune rankings, while the "Other" group includes all remaining firms that participated in the rankings during this period but were not listed among the best in product/service quality. Panel A shows the results of the fine-tuned GPT quality index generated from "pros" section, Panel B shows the results of fine-tuned GPT quality index generated from "cons" section, and Panel C shows the results of the aggregate fine-tuned GPT quality index. p-values are reported in parentheses under the differences. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Group Mean		Diff.
	(1)	(2)	(1)-(2)
	Best Quality	Other	Best vs Other
Panel A: Pros-Based Quality Index			
Mean	2.850	2.385	0.464***
			(0.000)
Panel B: Cons-Based Quality Index			
Mean	3.268	2.815	0.454***
			(0.000)
Panel C: Aggregate Quality Index			
Mean	3.092	2.650	0.442***
			(0.000)

Table OA6.3. Fine-Tuned GPT Quality Index and Product Recalls

Notes: This table reports tests on the relationship between fine-tuned GPT quality indices and CPSC product recalls. Panels A and B report regression results on the relationship between changes in cons-based (Panel A) and pros-based (Panel B) fine-tuned GPT quality indices and product recalls in the subsequent year. Panel C reports regression results on the relationship between changes in aggregate fine-tuned GPT quality indices and product recalls in the subsequent year. The dependent variable in columns (1)–(2) is the total number of recalls in the subsequent year, and the dependent variable in columns (3)–(4) is the total number of reactive recalls over the same period. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Changes in Cons, Fine-Tuned GPT Quality Index and Future Recall Events

	# Recalls		# Reactive Recalls	
	(1)	(2)	(3)	(4)
Δ*Quality Index - Cons*	-0.059		-0.055	
	(-0.65)		(-0.57)	
Large Quality Drop		-0.034		-0.089
		(-0.19)		(-0.41)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	3,259	3,259	2,937	2,937
Pseudo R^2	0.466	0.466	0.447	0.447

Panel B: Changes in Pros, Fine-Tuned GPT Quality Index and Future Recall Events

	# Recalls		# Reactive Recalls	
	(1)	(2)	(3)	(4)
Δ*Quality Index - Pros*	0.018		0.041	
	(0.24)		(0.50)	
Large Quality Drop		-0.011		-0.168
		(-0.04)		(-0.72)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	3,259	3,259	2,937	2,937
Pseudo R^2	0.466	0.466	0.447	0.447

Panel C: Changes in Aggregate Fine-Tuned GPT Quality Index and Future Recall Events

	# Recalls		# Reactive Recalls	
	(1)	(2)	(3)	(4)
Δ*Quality Index - Agg*	-0.057		-0.056	
	(-0.48)		(-0.43)	
Large Quality Drop		0.044		0.013
		(0.23)		(0.05)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	3,259	3,259	2,937	2,937
Pseudo R^2	0.466	0.466	0.447	0.447

Table OA6.4. Fine-Tuned GPT Quality Index and Brand Value

Notes: This table presents tests on the relationship between fine-tuned GPT quality indices and the likelihood of appearing in Brand Finance's Global 500 brand rankings. Panel A (B) shows regression results of a dummy indicating whether the firm is listed in Brand Finance's Global 500 brand rankings on the pros (cons) fine-tuned GPT quality index. Panel C shows regression results of a dummy indicating whether the firm is listed in Brand Finance's Global 500 brand rankings on the aggregate fine-tuned GPT quality index. *Top Brand* equals 1 if the firm appears in Brand Finance's Global 500 rankings in the following year, and 0 otherwise. *Ltm Top Brand* equals 1 if the firm appears in the Global 500 rankings in the next three years, and 0 otherwise. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Pros, Fine-Tuned GPT Quality Index and Top 500 Brand

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Pros	0.006	0.104**	0.008*	0.195***
	(1.20)	(2.02)	(1.86)	(3.63)
Size	0.087***	1.180***	0.075***	1.337***
	(11.93)	(9.75)	(10.16)	(9.26)
Leverage	-0.096***	-0.196	-0.089***	-0.171
	(-3.63)	(-0.47)	(-3.46)	(-0.36)
ROA	-0.035	2.236***	-0.032	2.743***
	(-0.91)	(3.11)	(-0.88)	(2.93)
M/B Ratio	0.019***	0.252***	0.016***	0.246***
	(6.07)	(6.07)	(5.52)	(5.29)
Sales Growth	-0.045***	-0.449***	-0.038***	-0.413**
	(-3.51)	(-2.94)	(-3.10)	(-2.13)
Constant	-0.611***	-12.705***	-0.584***	-15.082***
	(-9.54)	(-9.23)	(-9.51)	(-9.19)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,475	10,869	11,456	9,629
Adjusted/Pseudo R^2	0.250	0.539	0.231	0.581

Table OA6.4. Fine-Tuned GPT Quality Index and Brand Value (Continued)

Panel B: Cons, Fine-Tuned GPT Quality Index and Top 500 Brand

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Cons	0.013***	0.108**	0.012***	0.135**
	(3.07)	(2.38)	(3.12)	(2.41)
Size	0.087***	1.162***	0.075***	1.288***
	(12.06)	(9.72)	(10.22)	(9.12)
Leverage	-0.092***	-0.210	-0.086***	-0.210
	(-3.47)	(-0.51)	(-3.35)	(-0.44)
ROA	-0.045	2.186***	-0.041	2.666***
	(-1.15)	(3.06)	(-1.11)	(3.03)
M/B Ratio	0.019***	0.250***	0.016***	0.249***
	(5.92)	(6.05)	(5.47)	(5.47)
Sales Growth	-0.047***	-0.451***	-0.039***	-0.429**
	(-3.66)	(-3.01)	(-3.21)	(-2.28)
Constant	-0.644***	-12.608***	-0.603***	-14.479***
	(-9.93)	(-9.45)	(-9.60)	(-9.24)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,475	10,869	11,456	9,629
Adjusted/Pseudo R^2	0.252	0.539	0.233	0.576

Panel C: Aggregate Fine-Tuned GPT Quality Index and Top 500 Brand

	Top Brand		*Ltm Top Brand*	
	OLS	Probit	OLS	Probit
	(1)	(2)	(3)	(4)
Quality Index - Agg	0.013**	0.141**	0.014***	0.219***
	(2.24)	(2.32)	(2.62)	(3.13)
Size	0.088***	1.173***	0.075***	1.312***
	(11.99)	(9.80)	(10.19)	(9.29)
Leverage	-0.093***	-0.198	-0.086***	-0.187
	(-3.50)	(-0.48)	(-3.35)	(-0.39)
ROA	-0.041	2.181***	-0.039	2.670***
	(-1.06)	(3.02)	(-1.05)	(2.95)
M/B Ratio	0.019***	0.248***	0.016***	0.242***
	(5.93)	(5.96)	(5.43)	(5.27)
Sales Growth	-0.047***	-0.454***	-0.039***	-0.425**
	(-3.67)	(-3.01)	(-3.25)	(-2.23)
Constant	-0.641***	-12.792***	-0.607***	-14.972***
	(-9.55)	(-9.44)	(-9.41)	(-9.39)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,475	10,869	11,456	9,629
Adjusted/Pseudo R^2	0.251	0.540	0.232	0.580

Table OA6.5. Fine-Tuned GPT Quality Index and Future Profitability

Notes: This table presents tests on the relationship between fine-tuned GPT quality indices and firm profitability. Panel A shows regression results of return on assets (*ROA*) on quality indices. Panel B shows regression results of operating margin on quality indices. *ROA* is the firm's return on assets in the following year, and *Ltm ROA* is the average return on assets over the subsequent three years. All continuous variables are winsorized at the 1st and 99th percentiles. t-statistics are in parentheses. Standard errors are clustered at the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Quality Index and ROA

	ROA			*Ltm ROA*		
	Pros	Cons	Aggregate	Pros	Cons	Aggregate
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index	0.002*	0.004***	0.004***	0.002*	0.004***	0.004***
	(1.78)	(3.92)	(3.25)	(1.78)	(4.11)	(3.41)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,068	12,068	12,068	12,071	12,071	12,071
Adjusted R^2	0.228	0.230	0.229	0.257	0.259	0.258

Panel B: Quality Index and Operating Margin

	Operating Margin			*Ltm Operating Margin*		
	Pros	Cons	Aggregate	Pros	Cons	Aggregate
	(1)	(2)	(3)	(4)	(5)	(6)
Quality Index	-0.002	0.001	-0.000	-0.002	0.001	-0.000
	(-0.84)	(0.65)	(-0.09)	(-0.79)	(0.70)	(-0.03)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,040	12,040	12,040	12,043	12,043	12,043
Adjusted R^2	0.251	0.251	0.251	0.275	0.275	0.275