Board Diversity Appearance and Firm Performance: An Image-Based Deep Learning Approach

Lukas Greger* Hendrik Scholz† Anja Stiller‡ Nicolas Webersinke§

April 2025

Abstract

We introduce a new measure of board diversity appearance as perceived by stakeholders based on portrait pictures of board members. Using an image-based deep learning approach, we receive a score that quantifies the level of visually observable diversity on a board. We argue that stakeholders may alter their behavior towards a firm based on board diversity appearance, which ultimately affects firm performance. Building portfolios based on a sample of S&P 100 firms, factor model analyses show indication that firms with more diverse appearing boards and less public attention indeed tend to outperform.

Keywords: board diversity, firm performance, machine learning, stakeholder attention

JEL Classifications: G12, G32, G41

*Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Chair of Finance and Banking, Lange Gasse 20, 90403 Nürnberg, Germany, e-mail: lukas.greger@fau.de.

†Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Chair of Finance and Banking, Lange Gasse 20, 90403 Nürnberg, Germany, e-mail: hendrik.scholz@fau.de.

‡Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Chair of Finance and Banking, Lange Gasse 20, 90403 Nürnberg, Germany, e-mail: anja.stiller@fau.de.

§Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Chair of Finance and Banking, Lange Gasse 20, 90403 Nürnberg, Germany, e-mail: nicolas.webersinke@fau.de.

1 Introduction

Over the past decade, diversity has become an increasingly important topic in social, political, and corporate environments. Prompted by public pressure through social movements such as "Black Lives Matter" or "#MeToo", an increasing number of companies have adopted voluntary commitments to the promotion of diversity. Furthermore, some legislators have introduced mandatory policies to promote diversity, such as Norway or California, where female representation on corporate boards is regulated through gender quotas. However, it is important to recognize that both voluntary commitments and mandatory regulations can only address observable and thus measurable dimensions of diversity, such as the representation of women on boards. Although diversity is often defined by demographic factors such as differences in gender or ethnicity, these characteristics alone do not drive firm performance. Instead, benefits for a firm are more likely to result from unobservable characteristics, which includes differences in perspectives, skills, and knowledge (i.e., cognitive diversity) or a more inclusive corporate culture, potentially leading to greater innovation or better problem-solving skills. This highlights the need to differentiate between observable factors, e.g. demographic diversity, and unobservable factors, e.g. corporate culture or cognitive diversity. However, investors and other stakeholders face a challenge as these unobservable characteristics, while influential, remains difficult to observe.

In light of the inherent complexities of observing such factors like cognitive diversity and corporate culture, we suggest that stakeholders may use observable characteristics - namely the visual appearance of a board - as a proxy for unobservable characteristics. Drawing from concepts in social and behavioral psychology, stakeholders may assume - whether accurately or not - that a more visually diverse board reflects, for example, a higher level of cognitive diversity within the board. As a consequence, stakeholders could change their behavior towards a firm based on this observable characteristic of a diverse appearing board. While we suggest that stakeholders may use board diversity appearance as a proxy for cognitive diversity, we neither assume nor deny a causal relationship between visually observable diversity and cognitive diversity. Thus, we contribute to a strand of literature that examines the relationship between board diversity and firm performance by introducing a new measure for board diversity appearance. To achieve this, we use state-of-the-art deep learning methods to measure diversity as it is perceived by stakeholders when they visually observe a board.

Unobservable characteristics, such as the corporate culture or the level of cognitive diversity on a firm's board, may attract stakeholder interest, as fostering diversity can potentially lead to the inclusion of individuals with different perspectives, experiences, and skills. This in turn can enhance innovation and problem-solving, ultimately contributing to improved firm performance. Indeed, "diversity wins" is a common phrase and the name of a well-known report by the consulting firm McKinsey (Hunt et al., 2020), which stresses that the business case for diversity is "stronger than ever". While several other practitioners and initiatives share this opinion, the academic literature is rather divided. Indeed, Green and Hand (2021) conclude that caution is warranted in relying on the findings of Hunt et al. (2020). Further academic literature on boardroom diversity has predominantly used only one-dimensional measures with a strong focus on gender. However, consensus among the studies regarding the impact of gender diversity is low. Some studies find a positive (Campbell & Mínguez-Vera, 2008), whereas others report a negative (R. B. Adams & Ferreira, 2009) or non-existent (Carter et al., 2010) relationship between gender diversity and performance. As the specific role of gender within the broader concept of diversity remains unclear, some researchers have attempted to gain a deeper understanding by using alternative one-dimensional measures as proxies for boardroom diversity, such as ethnicity (Dodd et al., 2022), age (Xu et al., 2022) or financial expertise (Minton et al., 2014). In order to achieve a more comprehensive understanding of diversity, some studies have incorporated multiple measures, for instance ethnicity alongside gender (Carter et al., 2003; Erhardt et al., 2003). However, when these measures are examined separately, they fail to consider potential interactions between different dimensions (Herring, 2009). To overcome this issue, some researchers have developed "diversity indices" to incorporate multiple measures to approximate diversity. For instance, both Anderson et al. (2011) and Bernile et al. (2018) use demographic measures, including gender, ethnicity, and age, as well as cognitive attributes, including education, experience, expertise and profession. Overall, both studies find a positive relationship between their diversity measure and firm performance, as measured by Tobin's Q. However, Bernile et al. (2018) even admit (with commendable honesty) that the specific dimensions of their diversity index are driven by data availability. In contrast, Edmans et al. (2023) take a more holistic approach to measuring diversity. Instead of relying on demographic data, they construct a measure for diversity, equity, and inclusion (DEI) based on survey responses. This measure shows low correlation with traditional measures such as gender and ethnic diversity, suggesting that it captures additional dimensions. They find that their diversity measure is associated with higher future accounting performance and higher valuation ratios, as measured by ROA and Tobin's Q.

In summary, the board diversity puzzle remains unsolved. The existing literature on the relationship between diversity and financial performance does not provide a general consensus on whether diversity enhances, reduces, or has no impact on firm performance. The major caveat of both practitioner and academic studies is that they tend to overlook important aspects of diversity due to a predominant focus on readily available, one-dimensional measures such as gender or ethnicity. However, even a focus on these more readily available measures often results in loss of observations due to missing data.

As a remedy, we introduce an innovative computer vision approach to measure diversity as it is perceived by stakeholders when they visually observe a board. For this, we leverage state-of-the-art deep learning methods to assess how diverse a board appears by analyzing portrait pictures of board members. By providing an appearance-based perspective on board diversity, we argue that our approach offers a more informative and stakeholder-relevant measure, particularly when traditional diversity measures are sparse or unavailable. The use of machine learning to measure aspects that were previously unobservable or difficult to observe is becoming increasingly popular in the field of finance, as researchers are adopting these methods more frequently (Obaid & Pukthuanthong, 2022), including the analysis of facial images (Lu & Teo, 2022). We use this unstructured data to calculate our measure of perceived board diversity, the

Board Diversity Appearance (BDA) score. We suggest that the visual appearance of a board may serve as a valuable proxy or signal to stakeholders for unobservable characteristics, e.g. corporate culture or the level of cognitive diversity within a board. On the one hand, stakeholders may recognize that these signals do not directly represent those unobservable characteristics, but they may still use them as indicators in the absence of better alternatives, which is in line with the signaling theory proposed by Spence (1978). On the other hand, stakeholders may even subconsciously assume a relationship between these observable signals and unobservable information, which is in line with a cognitive bias known as the halo effect (Nisbett & Wilson, 1977; Thorndike, 1920). These channels are explained in more detail in section 2.

When stakeholders use their perceptions of a more diverse appearing board as a proxy for unobservable characteristics and thus firm performance, they may be more inclined to work with the firm on better terms or at all. As a result, firms with more diverse appearing boards could be able to gain benefits that may be not available to firms with less diverse appearing boards. We argue that the BDA score is particularly valuable to stakeholders when the firm does not attract much public attention, as the access to more detailed and informative proxies for unobservable characteristics (e.g., information about the corporate culture of the firm or cognitive diversity within the board) may be difficult or expensive in such cases. To measure public attention, we use an attention measure (ATT) derived from the Google Search Volume (GSV) index, which is based on Google search data. The GSV index has been used in previous studies to capture (retail) investor attention (e.g., Da et al., 2011) or public attention to a topic (e.g., Giannetti and Wang, 2023).

We independently double-sort the firms in our sample consisting of S&P 100 firms using our BDA scores and ATT indices as sorting dimensions. Using the six-factor model of Fama and French (2018), we find a statistically significant and positive abnormal return for the portfolio consisting of firms that are perceived to have more diverse appearing boards (high-BDA), while receiving relatively less overall stakeholder attention (low-ATT). The effect persists over the short-term and near-mid-term. Furthermore, we show that these findings can be attributed to information in our BDA score beyond pure board gender diversity. To explain this finding, we examine measures for firm valuation and accounting performance and find some indications that the outperformance of high-BDA and low-ATT companies is reflected in those measures as well.

The remainder of the paper is organized as follows. Section 2 discusses the business case for diversity, along with the relevant impact channels. Section 3 introduces our new BDA measure and explains how we measure board diversity appearance. In section 4, we explain our empirical approach using BDA scores and ATT for building double-sorted portfolios. Section 5 provides a description of our sample, followed by the presentation and discussion of the results in section 6. Finally, section 7 concludes.

2 Is there a business case for board diversity?

The primary functions of boards are to monitor and control the firm while providing essential resources, such as experience and expertise (Hillman & Dalziel, 2003). Through these functions, boards directly influence the strategic actions of firms (Rindova, 1999), which may subsequently affect their financial performance. The actions of boards are in turn shaped by the skills, abilities, and knowledge of its members, often referred to as human capital (Becker, 1964). In general, the unique combination of human capital that each individual, such as a board member, brings to the table can influence the dynamics within a team, such as a board. However, academic literature does not provide a clear indication of whether more heterogeneity in teams affects outcomes positively or negatively, as there are multiple drivers involved in this setting. On the one hand, diversity within the board may promote divergent thinking by incorporating individuals with different sets of human capital and therefore different skills, experiences, and viewpoints into decision-making processes (Westphal & Milton, 2000). This inclusion of diverse perspectives can challenge conventional wisdom within the group, prompting others to question assumptions that have implicitly shaped their thinking. Consequently, minority team members may encourage their majority counterparts to consider a broader range of potential solutions (Nemeth, 1986), which may lead to increased creativity and innovation within the team. For instance, Bernile et al. (2018) show a positive correlation between board diversity and impactful innovation output, as measured by innovation investment and patent activity. Furthermore, diverse teams may be better equipped to solve problems. In this context, Hoffman (1959) suggests that heterogeneous groups, in terms of personality, tend to produce higher quality solutions to problems compared to homogeneous groups. Even in scenarios designed to evoke emotional conflict, heterogeneous groups demonstrate more effective problem-solving skills, highlighting the benefits of diverse perspectives in addressing complex challenges (Hoffman & Maier, 1961). However, while a diverse group may encourage more diverse opinions and critical thinking, this diversity may also lead to friction within the team, resulting in a more time-consuming and less efficient decision-making process. Moreover, diversity may escalate conflict within the group, weaken group cohesion, and increase employee turnover, thus potentially hindering consensus-building in the group's decision-making process (Becker, 1957). This can be particularly challenging for firms in situations where quick reactions are required, such as during economic downturns (Hambrick et al., 1996).

Overall, there is no general consensus on whether the impact of unobservable characteristics of diversity, such as cognitive diversity of board members, on the decision-making process within a board and thus firm performance is positive or negative. However, stakeholders may expect those unobservable characteristics to have somehow an impact on firm performance and seek to evaluate them, but they typically lack insight into internal dynamics. The signaling theory proposed by Spence (1978) provides an explanatory framework for behavior under conditions of information asymmetry. Though the theory was initially formulated to understand how potential employees signal their abilities to employers, it can also be applied to explore how firms signal certain characteristics to stakeholders. Thus, stakeholders have to rely on observable signals and use them as proxies for unobservable information to assess a firm's capabilities (Fombrun & Shanley, 1990). We argue that stakeholders may use the visual appearance of a board as such an observable signal to approximate unobservable information, e.g. cognitive diversity within the board or corporate culture, and their impact on future financial performance of the firm. Consistent with this, Albinger and Freeman (2000) show that a firm's commitment to diversity, as measured by representation of women and minorities, serves as an informational signal for job applicants when comparing firms. On the one hand, stakeholders may recognize that these signals do not directly represent unobservable characteristics, but they may still use them as indicators in the absence of better alternatives. On the other hand, stakeholders may even subconsciously assume a relationship between these observable signals on demographic diversity or board appearance and the unobservable information on cognitive diversity or corporate culture, e.g. This is consistent with a cognitive bias known as the halo effect (Nisbett & Wilson, 1977; Thorndike, 1920), where one aspect, such as the board diversity appearance, influences perceptions of other (unobservable) dimensions, and ultimately perceptions of the future financial performance of the firm.

Thus, a firm could benefit to some extent from observable diversity even in the absence of unobservable characteristics that ultimately influence firm performance. For example, a firm's board may exhibit a high level of observable diversity, such gender diversity or a diverse appearance, but the members may have similar skills and abilities, resulting in a low level of cognitive diversity within the board. However, stakeholders would reasonably assume a high level of cognitive diversity based on the observable signal of a high level of diversity appearance and consequently may change their behavior towards the firm.

Obviously, there are more accurate indicators of some unobservable characteristics within a board than demographic diversity or the appearance of a board. For instance, stakeholders could assess the previous work experience and the educational background of board members or the corporate culture within the firm. In fact, these characteristics may actually provide a better picture of a board's capabilities compared to the visibly observable attributes of its members. However, accessing this detailed information can be difficult or expensive, especially for smaller companies or those that attract less public attention. Thus, we argue that the described channel - stakeholders using board appearance as an observable signal to approximate unobservable characteristics - is more pronounced for firms that receive less public attention. Following this, stakeholders could be willing to engage with such firms at more favorable conditions than with other firms. In contrast, stakeholders are more likely to have access to more detailed and informative signals of firms that receive more public attention, reducing the need to use board appearance as a proxy for unobservable characteristics such as cognitive diversity or corporate culture. Consequently, we argue that board appearance is less likely to be a crucial factor when evaluating, e.g., the abilities and decision-making processes of boards in firms with higher public attention.

Thus, this study neither assumes nor denies that important drivers for a firm's financial performance, such as a board's capability or corporate culture, are directly influenced by the demographic diversity or the appearance of its members. Instead, we suggest that stakeholders may alter their behavior toward a firm based on their perception of those drivers, which may be influenced by the board's visual appearance in lack of better signals. For instance, job applicants might be more inclined to work for a firm, or potential customers and suppliers might be more willing to engage in business, leading to lower recruitment costs and a stronger position in negotiations, which, in turn, would impact the firm's financial performance.

3 Measuring firm-level board diversity appearance

We aim to develop a board diversity appearance measure at the firm level that (i) measures the diversity appearance of a board directly (i.e., through the channel of stakeholder perception), (ii) is pure, i.e. captures many dimensions of diversity and interactions between them, and (iii) can be reduced to a firm-year observation with only little missing values for further hypotheses testing.[1] These properties are difficult to achieve with traditional demographic diversity measures, such as gender, age, ethnicity, or educational background. Therefore, we turn to unstructured data, namely

[1]Previous literature focuses primarily on demographic diversity measures based on a single dimension, such as gender or age (e.g., Zhang, 2020). While these approaches are valuable for analyzing the attribution of these single dimensions, we argue that diversity is inherently multidimensional and should therefore be measured multidimensionally. Moreover, a single dimension seems inappropriate for measuring the appearance of board diversity, as a company's stakeholders are most likely to consider multiple dimensions simultaneously in their perceptions.

photos of corporate board members. Photos as input for a board diversity appearance measure have several advantages over traditional diversity measures. First, they measure board diversity appearance directly, as stakeholders most likely use them for their perception, too. Second, they not only approximate traditional diversity dimensions such as gender, age, or ethnicity, but also expand to additional dimensions of diversity. Therefore, a diversity measure derived from photos is arguably purer. Finally, due to advances in computer vision over the last decade and the wide availability of corporate board member photos, we can derive diversity appearance scores in an intuitive way at the firm-level with only little missing values.

3.1 An image-based deep learning approach

The field of computer vision has seen significant progress in recent years, driven by advances in deep learning. These advances have enabled machines to process visual information with remarkable accuracy, paving the way for numerous applications. The goal of computer vision is to enable machines to understand, interpret, and analyze visual data from the real world. Deep-learning-based computer vision makes it possible to develop sophisticated models and algorithms that can perform complex tasks such as image classification, face detection and verification, and semantic segmentation with unprecedented accuracy. As a result, researchers in accounting and finance have also turned to computer vision. While some researchers still use traditional computer vision approaches with pre-defined features, e.g. to assess executives' facial trustworthiness (Hsieh et al., 2020) or to analyze masculine behavioral traits of fund managers (Lu & Teo, 2022) and CEOs (Jia et al., 2014; Kamiya et al., 2019), deep-learning-based computer vision without pre-defined features is becoming increasingly common. For example, Obaid and Pukthuanthong (2022) gauge investor sentiment using photos in newspapers and Ahmed et al. (2023) examine CEOs' facial attractiveness by using deep convolutional neural networks (CNN).

We use portrait photos of the board members to approximate diversity appearance.[2] Since board members are typically public figures, portrait photos of them are widely available, for instance via Google's image search. Within these portrait photos, we focus exclusively on faces for three reasons. First, this helps the algorithm focus on what is important and not pick up other spurious patterns (e.g., background colors). Second, photos of faces are harder to manipulate to give the impression of diversity. While it is easy for companies to change the background in photos of the board members or to change their clothes and even their hair, it is much more difficult to perform this "diversity dressing" with faces. Finally, with faces, we can make use of off-the-shelf pretrained deep learning models trained for common tasks such as face detection and verification, which increases not only objectivity but also accuracy.

[Figure 1 about here.]

Figure 1 illustrates our approach. In general, we use methods from face verification or recognition, as we argue that measuring board diversity appearance can be seen as the opposite of verifying that people are the same. For each collection of board member photos, we first perform face detection. To detect faces, we use "face_recognition", a Python wrapper for the C++ library dlib, which employs a deep CNN (King, 2009).[3] Then, we take only the detected faces and encode them by computing 128-dimensional face embeddings of the sub-images.[4] For this task, we use dlib's "ResNet-29" CNN implementation, which is pretrained on more than 3 million faces (King, 2017).[5] It achieves competitive performance on the face verification task.[6] This results in a face embedding matrix X of shape $n \times 128$, where n is the number of board members. We then calculate the pairwise euclidean distance between the n row vectors and take the

[2] Portrait photos of board members are ideal for computer vision because of their clean and tidy nature. We can expect to suffer much less from problems arising from noisy data, such as blurred images (Levi & Hassner, 2015).

[3] See github.com/ageitgey/face_recognition for "face_recognition".

[4] We use 128-dimensional embeddings because Schroff et al. (2015) find no statistically significant difference in performance on the face verification task for more dimensions.

[5] It must be noted that the accuracy of face recognition algorithms may be influenced by biases present in the training data, which may impact recognition outcomes across different demographic groups (Kolla & Savadamuthu, 2023).

[6] See paperswithcode.com/sota/face-verification-on-labeled-faces-in-the.

average of these to obtain a scalar value for each company board.[7]

One obvious question is what the 128 feature dimensions learned by the CNN represent. Neural networks are remarkably good at automatically finding relevant features. Even when not directly supervised to learn a particular feature, neural networks pick up that feature if it is deemed relevant, such as the sentiment of the text (Radford et al., 2017). While we cannot know for sure what these 128 dimensions represent, there are common dimensions of demographic diversity. Gardenswartz and Rowe (2003) provide a four-layer model to help understand the differences and similarities of people within an organization. It provides an overview of relevant diversity dimensions. These include age, ethnic background and nationality, gender and gender identity, physical and mental abilities, religion and worldview, sexual orientation, and social background. The literature has shown that many of these dimensions can be captured in facial images using computer vision. For instance, Levi and Hassner (2015) and Mazières et al. (2021) demonstrate that age and gender classification work well. Wang and Kosinski (2018) show that faces also contain information about sexual orientation that can be extracted by deep neural networks. Some researchers find even more dimensions in facial images (e.g., Kachur et al., 2020; Tkachenko & Jedidi, 2020), although care must be taken to avoid potential multicollinearity.

3.2 Validation

Since our approach is completely unsupervised, a direct systematic validation is not available. While the underlying CNNs have been evaluated for their original tasks and found to perform very well, this performance cannot be directly transferred to the task of diversity measurement. We therefore demonstrate the validity of our approach indirectly by using media reports to identify companies whose board is perceived as particularly diverse or as particularly non-diverse. Then, we visualize our approach

[7]We compute the mean pairwise Euclidean distance because this is the metric that the authors of the dlib library found to work best for the similarity task. Therefore, we assume that this metric will also work well for our dissimilarity or diversity task. Moreover, Schakel and Wilson (2015) find that not only the direction, but also the length of an embedding vector may carry important information.

for them by using Principal Component Analysis (PCA).

A company that got particular attention for its diversity activities in general and for its diverse board is “Merck & Co., Inc.” (S. Adams, 2013; Wingard, 2021). On the other hand, “Berkshire Hathaway Inc.” is considered one of the companies with the least diverse boards (S. Adams, 2013). Therefore, we use the 2022 boards (including both officers and directors) of these two companies to visualize our approach for validation. We first calculate 128-dimensional face embeddings for all board members. For dimensionality reduction, we perform a PCA and keep the three principal components that explain most of the variance in the data. Figure 2 shows the 3D scatter plots derived from these three principal components (PC). The board of “Merck & Co., Inc.” is on the left side and the board of “Berkshire Hathaway Inc.” is on the right side. All directors are indicated by orange dots, officers are indicated by teal dots, and members of the board who are both officers and directors are indicated by blue dots.

[Figure 2 about here.]

The dispersion of board members differs significantly between the two companies. The figure shows that Merck’s board members are more dispersed than Berkshire Hathaway’s. Combined, the three PCs explain 23.7 % of the total variance, suggesting that these components already explain a substantial proportion of the variance, but also highlighting the need for additional dimensions to capture a more adequate proportion of the total variance. After looking at the portrait pictures of the board members[8], the dispersion of board members across the two companies suggests that PC1 likely represents characteristics associated with gender, PC2 likely represents characteristics associated with age, and PC3 possibly represents characteristics associated with ethnicity.

To verify the hypothesis that PC1 is associated with gender and PC3 is associated with ethnicity, we conduct a panel regression analysis using “classical” diversity measures to determine a potential relationship between the PC’s and the respective

[8] Unfortunately, we cannot display the pictures of the board members in the figure due to copyright issues.

dimension of diversity.[9] To get a board value for each PC, we calculate the values of PC1 to PC3 for each individual board member per board per year and take the simple average of the values of the PC per member of each board in a respective year. The results are shown in Table 1.

[Table 1 about here.]

The results for PC1 in the first column show that board gender diversity indeed has a positive and significant coefficient, indicating that PC1 can partially be explained using board gender diversity. However, since our BDA measure and therefore PC1 is much more nuanced than "just" percentage of female board members, it is not surprising that a large part of the variation in PC1 remains unexplained, as reflected in the R^2 of approximately 37%. Additionally, ethnic diversity seems to contribute to explaining PC1. For PC3 in the third column, our assumption seems to be validated as well, as ethnic diversity is the only variable with a statistically significant and positive coefficient. Again, as our BDA score and therefore PC3 is much more nuanced than just a percentage number, the relatively low R^2 of about 12% is not surprising. In addition, skills diversity has a positive and significant coefficient. Turning to PC2 in the second column, we suspect that this PC is associated with age. However, as age is not a readily available data point, we use the other available dimensions to explain it. Once again, ethnic diversity has a statistically significant and positive coefficient, as it has with all three PCs.

So overall we are confident that our score actually measures what it should, namely how diverse a board appears through the eye of a stakeholder. The validation suggests that our approach is indeed able to capture relevant dimensions of diversity in board member images, while being much more nuanced than pure percentage numbers. Thus, we continue with explaining our methodology using the BDA score to test our

[9] We use gender and ethnic diversity as explanatory variables, which are defined hereafter as the percentage of female board members and the percentage of board members classified under minority groups, respectively. Additionally, we use skills diversity, defined as percentage of board members who have either an industry specific background or a strong financial background. We use those three variables as they are readily available from LSEG Workspace as our data source. Unfortunately, age is not among the variables provided, so we are not able to test our observations regarding PC2.

hypothesis that the information of a diverse appearing board is a valuable signal to stakeholders when a company does not receive much stakeholder attention.

4 Methodology

We use a portfolio approach to test our hypothesis that stakeholders use the signal of a more or less diverse appearing board for firms that receive relatively less overall stakeholder attention. We build equal-weighted portfolios based on a double-sorting of the firms. The first dimension is our BDA score, which measures whether a board appears more or less diverse as shown in section 3. For stakeholder attention as the second dimension we use an attention measure (ATT) derived from the Google Search Volume (GSV) index as a proxy. The index is based on Google Trends, which reflects the amount of search queries on Google internet search. It was first proposed by Da et al. (2011) and has been used in prior literature to measure investor attention (e.g., Bijl et al. (2016), Preis et al. (2013), Yoshinaga and Rocco (2020), Ekinci and Bulut (2021), Chai et al. (2021), Mandasari et al. (2023)). Overall, prior literature suggests that ATT or the GSV index, respectively, most likely measures the attention of retail investors and that there are no effects in any particular direction on the stock prices of companies with a high (retail) investor attention based on this measure. Thus, there does not seem to be a significant difference in the performance of high attention and low attention companies, making it suitable for our double-sorting approach. Additionally, Vozlyublennaia (2014) finds that a higher investor attention approximated based on GSV diminishes predictability of stock returns. Most studies use the ticker symbol as search query when approximating investor attention, but since we want to measure overall stakeholder attention we use the company's name as search query.[10] Also, Google Trends allows to select categories. We select "Business & Industrial" to ensure that our results are not biased by "double-themed" names, e.g. for "Apple" the value

[10] Using Google Trends data to capture "public attention" instead of investor attention has been applied in prior research, including studies on board diversity. For example, Giannetti and Wang (2023) use the GSV index to measure public attention to gender equality.

would otherwise include both the search queries for the company and for the fruit. Additionally, Google Trends allows to select topics for specific queries. We select the search query with the most specific topic, e.g. for "Apple" we selected the theme "Technology company".

We collect monthly values for each company over our sample period using U.S. search queries. Since values for only five queries can be downloaded at a time and they are all relative to each other, i.e. from 0 to 100, we need to transform them to an attention measure (ATT), which is basically a standardized GSV relative to a "benchmark company". We used Berkshire Hathaway as a benchmark company, so Berkshire Hathaway is always one of the five queries. We then transformed the values for all firms using the following equation:

$$ATT_{i,t} = \frac{GSV_{i,t}}{GSV_{Benchmark,t}} \tag{1}$$

Here, i stands for a firm and t for a respective month.[11] We used those monthly ATT values to derive a yearly value by calculating the simple average of the monthly values. The yearly ATT value was then used for the portfolio sorting. A high ATT indicates high stakeholder attention and vice versa.

We build equal-weighted portfolios at the beginning of each year using the values for BDA and ATT of the prior year. We use independent double-sorting using terciles as cut-off points for both dimensions, which gives us a total of nine portfolios. We use the values of the previous year for re-balancing as (i) these are the information that the stakeholders have at the time of the re-balancing and (ii) it is likely that the effects of board appearance take some time until the benefits of the stakeholder's decision can be seen in a firm's performance. Because of the latter effect, we repeat some analyses with portfolios constructed using BDA and ATT values with a greater time lag.

The performance of every portfolio is measured as the abnormal return α, which we

[11] Thus, Berkshire Hathaway has for all observations an ATT of 1 since it is the benchmark company. Some companies had a relatively low (normal) GSV of "< 1" (but not 0), so we replaced those values with 0.5 to calculate ATT.

calculate using the six-factor model (Fama and French, 2018) based on the five-factor model from Fama and French (2015) plus the momentum factor from Carhart (1997):

$$\begin{aligned} ER_{p,t} = \alpha_p + \beta_p^{RMRF} \cdot RMRF_t + \beta_p^{HML} \cdot HML_t + \beta_p^{SMB} \cdot SMB_t + \\ \beta_p^{WML} \cdot WML_t + \beta_p^{RMW} \cdot RMW_t + \beta_p^{CMA} \cdot CMA_t + \epsilon_{p,t} \end{aligned} \tag{2}$$

$ER_{p,t}$ is the excess return of portfolio p in month t. $RMRF_t$, HML_t, SMB_t, WML_t, RMW_t, and CMA_t are the respective factor returns in month t, with the β's as their respective regression coefficients. $\epsilon_{p,t}$ is the error term. α_p, the abnormal return of portfolio p, is our main variable of interest.

We use monthly observations for returns and factors. For re-balancing the portfolios at the beginning of each year, we use the BDA score calculated with all officers and directors that were members of the board at any time during the prior year (as well as ATT, respectively).[12] So if we build the portfolios for the year 2013, we use the one-year lagged BDA and ATT scores of 2012. If a company is delisted during one year, we include it in the portfolio upon the month of the delisting and keep only the remaining companies in the portfolio for the rest of the year, i.e. no new companies are added to the portfolio during a year.

We also use two other sorting categories instead of BDA (and keep ATT as the second sorting category): The first one we use is gender diversity, i.e. the Blau index of percentage of female board members. This is a classic measure used for board diversity and could show us how much of a possible α when sorting with BDA stems from "just" straight gender diversity.[13] The second sorting category are the residuals of a regression of BDA on the Blau Index of percentage female board members (i.e.,

[12]This means that some members were not on the board for the whole prior year. This is no shortcoming in your analysis, as there is no clear cutoff point when the influence of a new members began, especially on stakeholder perception (e. g. already with an earlier announcement of the new appointment). So the date of the official appointment would probably be at a later point in time and after the actual influence of the new board member began. The same is true for board members that leave the board during a year, as one cannot say for sure when their influence on decisions or stakeholder perception ends.

[13]Unfortunately, due to data constraints we are not able to build double-sorted portfolio using other board diversity measures such as ethnic diversity.

gender diversity), which we got by estimating the following regression equation:

$$BDA_{i,t} = \beta_0 + \beta_1 GenderDiversity_{i,t} + \epsilon_{i,t} \tag{3}$$

By sorting after this residuals we can build portfolios based on what our BDA measure captures beyond pure gender diversity.[14]

5 Sample description

5.1 Data and summary statistics

We analyze a sample free of survivorship bias of all S&P 100 constituents from 2013 to 2022. We select the S&P 100 index because it is one of the most important indices worldwide and the companies included in it are arguably some the most visible companies in the U.S., if not the world, while others are not household names. Despite this, the large size of these companies ensures that pictures of board members are widely available to stakeholders. We therefore expect our suggested channels to work particularly well for them. This results in a total of 133 firms and $1,330$ firm-year observations, not considering missing values.

Pictures of board members are found via Google image search. We are using the following search request: "First name Surname Company name Year". We add the company name to ensure that the picture, in which the officer is depicted, has a business background. The first colored "portrait picture" is used, i. e. one where the officer looks frontal into the camera whilst his whole face can be seen. When there is no starting date for being officer or director provided, then the officer or director is only considered in the year when his or her tenure ended.[15] If for a certain year no picture of a board member can be found we use a picture from up to three years before or after the missing one.

We calculate the returns of a firm using LSEG Datastream's Total Return Index

[14]This approach is similar to Huang et al. (2014) and their $ABTONE$ measure.

[15]We retrieve start and end dates for board member tenures from LSEG Workspace.

which captures both changes in stock prices as well as dividends. Monthly factor returns and the risk free rate for the U.S. are downloaded from Kenneth French's website.[16] The data points for further analyses including firm performance measures, specifically Tobin's Q and Return on Assets (ROA), as well as control variables and firm characteristics, e.g. total assets, market capitalization, and firm age, are gathered from LSEG Datastream (see section 6.2). Furthermore, we use traditional board diversity metrics, previously introduced in Table 1, provided by LSEG to control whether our new diversity score is correlated with or can be explained them.

The descriptive statistics of our sample can be found in Table 2. Of particular interest is our BDA score, which has an average value of 74.09, close to its median of 74.55. The maximum and minimum values amount to 81.77 and 36.65, respectively. This indicates some variation across the sample with a standard deviation of 3.39.

[Table 2 about here.]

Figures 3 and 4 illustrate some additional information about the BDA score. Given the increased attention to diversity issues in recent years, it may be insightful to analyze the change in the average BDA scores over time. Figure 3 shows a slow but steady increase over the sample period from 2013 to 2022, which is consistent with the growing focus on diversity in recent years.

[Figure 3 about here.]

Furthermore, both the demographic diversity within boards and financial performance may be driven by industry-related dynamics. Figure 4 shows boxplots of the BDA score across GICS industries over the sample period. Most industries have a median score in the range of 72 to 76, with Health Care and Consumer Discretionary at the top. Furthermore, the variation within industries appears to vary, with greater (lower) variation in industries such as Materials and Consumer Staples (Utilities and Health Care).[17]

[16] See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[17] However, it is important to note that the number of firms also varies across industries, as can be seen in Table 17 in the Appendix.

[Figure 4 about here.]

Considering the potential impact of both industries and time on firm performance measures and the BDA scores, they are controlled for in the respective analyses.

5.2 Determinants of board diversity appearance

To gain a deeper understanding of how the diversity aspects reflected in the BDA scores are associated with firm characteristics and to see which types of firms tend to have more or less diverse appearing boards, the BDA scores are regressed on various firm characteristics. Thus, it might be the case that financially successful firms may attract a larger pool of potential board members, which increases the likelihood of a higher level of demographic diversity within the board. We use the following regression equation to determine the association between diversity and certain firm characteristics.

$$BDA_{i,t} = \chi' \boldsymbol{X_{i,t-1}} + \delta_{jt} + \epsilon_{i,t} \tag{4}$$

where i indexes firms, j indexes industries and t indexes years. δ_{jt} are industry times year fixed effects and ϵ represents the error term. The vector $\boldsymbol{X}$ contains commonly used control variables and financial characteristics. These include firm size, as captured by total assets and number of employees, as well as book-to-market ratio, firm age, and board size. Furthermore, the inclusion of returns, cash reserves, dividend yield, and financial leverage allows to assess whether firms with stronger financial positions tend to have more diverse appearing boards.

The results in Table 3 show that growth firms (lower book-to-market ratio) tend to have higher BDA scores. This finding is consistent with the assumption that growth firms may demonstrate higher levels of diversity, as they are likely to be more incentivized to address diversity issues due to the importance of human capital and innovation within such firms (Edmans et al., 2023). Notably, the regression results do not indicate associations between firm size or recent financial performance and the visual appearance of the board.

[Table 3 about here.]

Next, following Edmans et al. (2023), we model our BDA Score with "classical" demographic diversity indicators established in prior literature. For our analyses, we use the following regression equation:

$$BDA_{i,t} = \theta_0 + \theta' \boldsymbol{Y_{i,t}} + \epsilon_{i,t} \tag{5}$$

The vector $\boldsymbol{Y}$ includes various diversity measures. Specifically, we use gender diversity, ethnic diversity, cultural diversity, and skills diversity. The latter one is closer to a measure of cognitive diversity. As ethnic diversity is only filled for about 10% of the observations where gender diversity is available, we use cultural diversity as a proxy in some of our regressions. Table 4 displays the results using the Blau index from Blau (1977) of each diversity measure as explanatory variables.[18] The regressions allow us to determine whether we are actually measuring something different or unseen, respectively, or if we are measuring the same easily observable dimensions of diversity in a different way.

The results in Table 4 show positive and significant coefficients for both gender and ethnic diversity. There are no significant coefficients for cultural and skills diversity. The R^2 for each regression ranges from 9% to 13%. Overall, the results suggest that our score can be partially explained by classical diversity measures, though a large part remains unexplained. This suggests that our BDA score provides additional information beyond these traditionally used measures, while still (partially) incorporating these demographic dimensions of diversity.[19]

[Table 4 about here.]

[18]The Blau index is a measure of heterogeneity that has a maximum value of $(n-1)/n$, where n is the number of categories, i.e. two in the case of gender diversity. The Blau index is then calculated as $1 - \sum_{i=1}^{n} p_i^2$ where p is the percentage value of each category i. This ensures that the index is at its maximum when each category has exactly the same percentage value. This could be assumed as maximum diversity according to the respective dimension.

[19]The results are qualitatively unchanged when using the raw percentages instead of the Blau indices of the diversity measures as explanatory variables.

As a further robustness test, we regress the BDA scores on commonly used proxies for diversity. We use the Diversity & Inclusion Scores by LSEG as such an indicator. The score combines 24 metrics that are categorized into four pillar scores related to diversity and inclusion: diversity (e.g., gender diversity), inclusion (e.g., flexible working hours), people development (e.g., training opportunities), and news and controversies (e.g., wage disputes) (LSEG, 2024). However, it must be noted that the LSEG scores are only available from 2016 on, while we calculated the BDA scores from 2013 on, which reduces the number of observations in this analysis.

[Table 5 about here.]

The results in Table 5 show highly significant and mostly positive associations between the LSEG scores and our BDA score. This indicates that firms with more diverse appearing boards actually tend to show better performance in various diversity aspects captured by the LSEG scores, supporting the assumption that demographic diversity as captured by the BDA scores can be a valuable and informative signal for stakeholders. However, low values for R^2, ranging from 4% to 8%, indicate that the diversity aspects captured by LSEG only partially explain our measure for diversity, which is plausible due to inherently different approaches in the construction of the variables.

6 Results

6.1 Factor Model Analysis

In this section, we analyze the performance of the double-sorted portfolios to test our hypothesis that a diverse appearing board can serve as a signal to stakeholders, particularly for firms that receive less attention from stakeholders. We begin by constructing long-short portfolios for the different sorting categories to identify any significant differences between the two groups. Specifically, for each sorting category, i.e. overall BDA score, ATT, gender diversity based on Blau Index of percentage of female board members, and residuals of a regression of BDA on gender diversity, we build a long-short portfolio with a long position in the companies with a high BDA score (ATT, gender diversity, residuals) and a short position in the companies with a low BDA score (ATT, gender diversity, residuals). The excess returns of these portfolios are analyzed using the six-factor model from Equation (2). Any significant differences within the categories could indicate that there are other effects at play if we do indeed find an outperformance of a portfolio. The results are displayed in Table 6.

[Table 6 about here.]

The variable of interest is the α of each long-short portfolio. Using the overall BDA score for the tercile sorting in the first column shows an insignificant coefficient, i.e. no indication of significant performance differences can be detected between the high BDA firms and the low BDA firms. The same result holds when ATT, gender diversity, and residuals are used for the portfolio sorting, as shown second to fourth column. So overall, neither the high/low BDA portfolio nor the high/low ATT portfolio has performed better than their counterpart in the respective long-short portfolio. This means that if we indeed find a statistically significant α, this effect is probably not driven by a outperformance of high/low BDA firms or high/low ATT firms alone. The same holds true for the gender diversity and high residuals portfolios.[20]

Next, we analyze the performance of the portfolios from the independent double-sorting. We start with double-sorting using one-year lagged values. The results are displayed in Table 7. Panel A shows the results when using BDA and ATT as sorting criteria. To get a better sense of where the BDA results could stem from, Panel B shows the results when using gender diversity[21] instead of BDA, and Panel C shows the results when using the residuals from Equation (3). For brevity reasons we only report the six-factor α's in the table.[22]

[20] As we conduct the double-sorting based on the two-year lagged scores, too, we also built long-short portfolios based on those scores. The results basically stay the same. Additionally we checked whether the S&P 100 and the equal-weighted S&P 100 have statistically significant α's to see if the index itself outperforms the market index from Kenneth French's homepage (in addition to the other five factors). Both indices do not have a significant α. These results are all available on request.

[21] Gender diversity is hereafter defined as the Blau index calculated on percentage of female board members.

[22] The complete regression results for all nine portfolios as well as the number of firms in each

[Table 7 about here.]

High portfolios in the respective sorting category are indicated by (1), while low portfolios in the respective sorting category are indicated by (3). The results show that the high-BDA-low-ATT portfolio outperforms, indicated by a positive monthly α, which is statistically significant at the 5% level. Furthermore, the coefficient of 0.48% indicates also economic significance. The effect seems to persist in a lower manner for the middle BDA firms that receive low stakeholder attention. This is indicated by the portfolio's α of 0.30%, which is also statistically significant at the 5% level. It still suggests economic significance, but not as much as the α of the high-BDA-low-ATT portfolio. This implies that for firms with lower stakeholder attention, a diverse appearing board can indeed be a positive signal to stakeholders. However, the signal decreases for firms with less diverse appearing board. Consistent with this, we do not find an alpha for the low-BDA-low-ATT portfolio, i.e. the portfolio consisting of firms with low stakeholder attention and the least diverse appearing boards. We also find a significant α for the middle-BDA-high-ATT portfolio, i.e., the firms that receive high stakeholder attention and are in the middle tercile in terms of board diversity appearance. We cannot explain these results with our impact channel.

To check whether our result are driven by pure gender diversity or something that the BDA scores measures additionally, i.e. the residuals, we turn our attention to the results in Panels B and C. Looking first at the results of using gender diversity and ATT as sorting criteria in Panel B, we find a statistically significant α of 0.48% for the high-gender diversity-low-ATT portfolio. So again the α is statistically and economically significant, which could indicate that stakeholders also think of pure board gender diversity as a valuable signals for firms that do not receive high stakeholder attention and therefore could change their behavior towards those firms favorably. In contrast to the results from Panel A, i.e. when using the BDA score, we cannot find a decreasing effect when the board gets "less gender diverse". Again, we have a statistically significant α in a portfolio where our impact channel does not deliver an explanation,

portfolio are available on request.

namely the low-gender diversity-high-ATT portfolio.

Lastly, if we use the information that our BDA scores adds on top to gender diversity, i.e. the residuals of Equation (3), next to ATT as sorting criteria in Panel C, we find that only the high-residuals-high-ATT portfolio has a statistically significant α. So it seems to be the case that investors indeed take the appearance of the board beyond just raw gender diversity into account and change their behavior towards the firm favorably. Unfortunately, due to data constraints we cannot check whether this results is driven by other observable dimensions such as ethnic diversity or age dispersion in the board, so we do not claim a causal link between board appearance and a firm's abnormal returns. But overall, our results indicate that a diverse appearing board could indeed serve as a positive signal to stakeholders of firms that receive less stakeholder attention.[23]

As it is difficult to say for sure when a potentially favorable change in stakeholder behavior towards a high-BDA firm materializes in terms of an α, we next use the two-year lagged scores to build the portfolio (e.g., the high-BDA-low-ATT portfolio of 2022 is built using the firms that were in the highest BDA tercile and the lowest ATT tercile in 2020). The results are displayed in Table 8. As before when using the one-year lagged scores to build the portfolio, we distinguish between using BDA score as sorting criteria in Panel A, gender diversity in Panel B, and the residuals in Panel C (next to ATT, respectively).

[Table 8 about here.]

The results seem even more convincing than when we use the one-year lagged values to determine the portfolio sorts. First looking at the double-sorting using the two-year lagged BDA score and the two-year lagged ATT score in Panel A, the high-BDA-

[23] In another analysis, we also used the one-year lagged officer BDA score, calculated using just the officers of the board, as second sorting dimension next to the one-year lagged ATT score. Stakeholders could look only at the officers, i.e. the board members who are in charge operationally (rather than having "only" supervisory duties in case of the directors in the board), and change the behavior towards the firm based on their appearance. We find an α of 0.41%, which is significant at the 5%-level, for the high-officer BDA-low-ATT portfolio. The only other portfolio with a significant α is the low-officer BDA-low-ATT portfolio. So the results are similar to those when using the overall board BDA score and are available on request.

low-ATT portfolio has a highly significant and economically very large α of 0.73%, which is even higher when using the respective one-year lagged scores. Additionally, when we use board gender diversity instead of BDA in Panel B, we do not find an outperformance for the high-gender diversity-low-ATT portfolio. But when we look at the residuals, i.e., what our BDA score measures in addition to just gender diversity, we find a statistically significant α for the high-residuals-low-ATT portfolio. With a coefficient of 0.47% it is also economically large. So our results suggest that the outperformance materializes not only over a one-year period, but in an even larger manner over a two-year period. The latter effect cannot be explained by pure board gender diversity. By this, our results suggest that a positive change in stakeholder behavior materializes to a positive α not only over the short-term, but also over the (near-)mid-term.[24] Moreover, the effect for the residuals in Panel C persists in the surrounding portfolios, i.e., for the high-residuals-middle-ATT portfolio and middle-residuals-low-ATT portfolio, suggesting that the effect of a positive change of stakeholder behavior could persist even when stakeholder attention or the board diversity appearance beyond pure gender diversity is in the middle tercile.[25]

Again, our results could be driven by other demographical dimensions of board diversity that we cannot measure. But overall, our results at least indicate that a more diverse appearing board of firms with low public attention is indeed a valuable and positive signal to stakeholders. Using both BDA and residuals, we find a statistically significant and economically large α for the double-sorted portfolios consisting of high-

[24]When expanding the holding period of each stock in the respective portfolio to two years instead of using the two-year lagged portfolio sort, we get a statistically significant α that is between the one for the one-year lagged portfolio sort and the two-year lagged portfolio sort. This can be expected, as the this approach combines both sorts into one portfolio. We also checked whether we can detect an α in the high-BDA-low-ATT portfolio using portfolio sorts with an even larger lag. We cannot find a significant coefficient when using the three-year lagged and five-year lagged scores for building the portfolios, but we do indeed find a economically large α of 0.67% when using the four-year lagged portfolio sort, which is statistically significant at the 1% level. But overall, our results (available on request) suggest that the potentially positive change in behavior of stakeholders materializes to a positive α over the short- to (near-)mid-term rather than over a longer time horizon.

[25]We also calculated the α's using the two-year lagged officer BDA score and ATT to sort the portfolios. As the high-officer BDA-low-ATT portfolio does not have a significant α, the effect of a potentially positive change in stakeholder behavior seems to be more short-term for the officer appearance.

BDA (residuals)-low-ATT firms. The results are even stronger when we use the two-year lagged scores to build the portfolios, suggesting that the results of a potentially positive change in the behavior of stakeholders materialized in terms of α not only in the short-term, but also in the (near-)mid-term.

6.2 Firm valuation and accounting performance

Next, we aim to explain where the outperformance of firms with more diverse appearing boards and less public attention could stem from. Thus, we explore the relationship between the BDA score and firm performance measures using the following equation.

$$Firm\ performance_{i,t} = \gamma BDA_{i,t-1} + \zeta' \boldsymbol{Z_{i,t-1}} + \delta_{jt} + \epsilon_{i,t} \tag{6}$$

The coefficient of interest is γ, which quantifies the impact of board diversity appearance, as measured by the BDA score of the previous year, on firm performance in the following year. We use both the raw BDA score and dummies representing the BDA-ATT portfolio to which the firms belong, in accordance with our methodology in section 6.1. This allows us to examine two key aspects: (i) whether firms with an overall more diverse appearing board demonstrate better firm performance, and (ii) whether a more diverse appearance (high BDA) of a board serves particularly well as a signal for firms with low overall attention (low ATT). Following our impact channel, positive results in either case could indicate that greater board diversity appearance may encourage stakeholders to engage with these firms on more favorable terms or increase their willingness to engage with them at all. To mitigate the potential influence of omitted variables, we include control variables in the vector $\boldsymbol{Z}$. These include total assets as a proxy for firm size, book-to-market ratio, and firm age. Furthermore, we repeat the analysis including gender diversity, defined as Blau index of the percentage of female board members, to determine whether a potential association between firm performance and the BDA score can be attributed solely to gender diversity rather than the broader visual appearance of the board. In line with our analysis in section 6.1, we replace the BDA score with gender diversity and the BDA residuals obtained from Equation (3) as independent variables to examine whether the observed results are mainly driven by gender diversity alone or by additional aspects captured by the BDA score.

We use two different measures as proxies for firm performance. First, we employ Tobin's Q as a market-based measure, which is defined as the ratio of the market value of a firm to its replacement value (Kaldor, 1966). It has become common practice in the finance and accounting literature to measure the ratio by comparing the market value of the firm's equity and liabilities to their corresponding book values, as determining the replacement values of a firm's assets can be challenging. Second, we use return on assets (ROA) as an accounting-based measure, which is defined as the ratio of a firm's profitability, as measured by its net income, to its total assets.

The results of the regression with the raw BDA score as the explanatory variable are shown in Table 9. They show no significant association between the overall BDA score and measures of firm performance. Thus, more diverse appearing boards alone do not seem to be associated with better valuation or accounting measures and, following our impact channel, more favorable decisions of stakeholders towards the firm. On the other hand, the results show a significant and positive association between gender diversity and firm performance.

[Table 9 about here.]

To account for the possibility that a potential change in stakeholder behavior may take longer to materialize, we perform an additional regression using two-year lagged BDA scores. The results in Table 10 show a significant and positive association between the BDA score and Tobin's Q.[26] Notably, the positive and significant association between gender diversity and Tobin's Q does not persist over a two-year (and three-year) time horizon, suggesting that the broader appearance of board diversity plays a more important role over longer time horizons. On the other hand, the results do not indicate an association between the BDA score and the accounting-based measure

[26]The significant and positive association between the BDA score and Tobin's Q persists when using three-year lagged BDA scores. Results are available upon request.

ROA. This suggests that potential effects of a more diverse appearing board may be incorporated in the market value of firms, but not (yet) in their accounting figures.

[Table 10 about here.]

In a further step, we replace the BDA scores as independent variables in Equation (6) with dummies, indicating the respective levels of BDA and ATT of the firms.[27] The results are shown in Table 11. Following the assumption that the described impact channel is particularly pronounced for firms with a high level of observable diversity (high BDA score) and a low level of public attention as measured by the Google Search Volume (low ATT), we would expect a positive association of the portfolio consisting of these firms with the employed valuation and accounting measures. While the results do not show significant and positive associations for the high-BDA-low-ATT firms, they do indicate partly significant and negative associations with firm performance measures for firms with lower levels of observable diversity.

[Table 11 about here.]

Table 12 shows qualitatively similar results when constructing the portfolios based on two-year lagged BDA scores and ATT indices.[28]

[Table 12 about here.]

In the next step, we use gender diversity and the BDA residuals as independent variables to examine whether the observed results are mainly driven by gender diversity alone or by additional aspects captured by the BDA score. Table 13 indicates that both gender diversity and the BDA residuals are positively and significantly associated with Tobin's Q. This suggests that the BDA score captures factors beyond gender diversity that are associated with improved firm performance. Furthermore, when gender diversity and the BDA residuals are lagged by two years, the coefficient for the

[27] To avoid multicollinearity in the regression, the middle-BDA middle-ATT portfolio is not included in this analysis.

[28] However, the association seems to be weaker when using a three-year lagged BDA scores and ATT indices. Results are available upon request.

BDA residuals becomes significant at a higher level, implying that the potential positive effects of a more diverse-looking board may take longer to materialize. Notably, the regression results show that the raw BDA residuals seem to be significantly associated with the market-based measure Tobin's Q, as opposed to the accounting-based measure ROA.[29]

[Table 13 about here.]

In the next step, we construct portfolios using gender diversity and the BDA residuals introduced in section 6.1 instead of our BDA scores. This approach allows us to isolate the aspects of our BDA measure that go beyond capturing pure gender diversity. Table 14 presents regressions using portfolios sorted by gender diversity, while Table 15 shows regressions based on portfolios sorted by BDA residuals. Table 14 shows similar results to prior analyses, revealing partly negative and significant associations between firms with lower gender diversity and firm performance measures. However, the results in Table 15 indicate that these negative associations seem to be explained by gender diversity rather than BDA residuals.[30]

[Table 14 about here.]

[Table 15 about here.]

Overall, these analyses support the previously shown results, namely that stakeholders tend to reward higher levels of diversity appearance as more diverse appearing boards seem to be a valuable and positive signal to stakeholders for firms that receive less public attention. Moreover, the findings of this section indicate that less diverse appearing boards tend to send a negative signal to stakeholders as they even seem to penalize lower levels of diversity appearance. However, the negative association can be largely explained by gender diversity. Additional regressions suggest that the broader appearance of board diversity plays a more important role over longer time horizons.

[29]The results are qualitatively unchanged when gender diversity and BDA residuals are lagged by three years. Results are available upon request.

[30]The results are qualitatively unchanged when the portfolios are constructed based on three-year lagged gender diversity and BDA residuals. Results are available upon request.

7 Conclusion

We introduce a new measure for board diversity appearance using a novel approach based on machine learning methods. We use portrait pictures of board members of S&P 100 firms to calculate our BDA score, which quantifies the level of visually observable diversity within boards. Drawing from concepts in social and behavioral psychology, we suggest that stakeholders may use board diversity appearance to approximate unobservable attributes of board members that affect firm performance. Thus, board diversity appearance may serve as a valuable signal to stakeholders. We further propose that this signal may be particularly pronounced for firms receiving low overall attention, where stakeholders face a general lack of information. If stakeholders do indeed use the appearance of boards as a proxy for unobservable information, such as the corporate culture, they may change their behavior towards these firms by engaging with them on more favorable terms or at all, ultimately affecting firm performance.

Using a portfolio approach we double-sort the firms in our sample independently into portfolios based on their board diversity appearance, i.e. our BDA score, and overall stakeholder attention, i.e. ATT. We find that the high-BDA-low-ATT portfolio indeed has a statistically significant and positive α using the six-factor model, among other portfolios. The abnormal return gets even larger when we use the two-year lagged scores to build the portfolios. These findings can be attributed to information in our BDA score beyond pure board gender diversity. Following our impact channel, stakeholders seem to tend to observe the board and change their behavior especially for firms that receive low public attention, i.e. where they do not have a clear picture in their heads. The effects seem to materialize into an α over the short- to mid-term. Building on this, we also find some indication that firms with lower BDA scores exhibit lower firm performance over the mid-term, as measured by Tobin's Q.

Overall, our results provide some indication that stakeholders do indeed value the signal of a diverse appearing board and are more likely to engage with a firm on more favorable terms when the firm receives limited public attention. This highlights the complexity of the "diversity puzzle" in the context of corporate boards. Future research could further refine our measure of board diversity appearance into an even more holistic metric by combining visual and textual data to create a multimodal measure.[31] Incorporating textual data could help to capture additional observable dimensions of diversity, such as the educational background of board members, that are beyond the visual appearance of the board.

[31] See e.g. https://arxiv.org/abs/2301.12597

References

Adams, R. B., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, *94*(2), 291–309.

Adams, S. (2013). The Best and Worst Companies For Women And Minorities [Available at forbes.com/sites/susanadams/2013/03/07/the-best-and-worst-companies-for-women-and-minorities, Last accessed May 2, 2025]. *Forbes*.

Ahmed, S., Ranta, M., Vähämaa, E., & Vähämaa, S. (2023). Facial attractiveness and CEO compensation: Evidence from the banking industry. *Journal of Economics and Business*, *123*, 106095.

Albinger, H. S., & Freeman, S. J. (2000). Corporate Social Performance and Attractiveness as an Employer to Different Job Seeking Populations. *Journal of Business Ethics*, *28*, 243–253.

Anderson, R. C., Reeb, D. M., Upadhyay, A., & Zhao, W. (2011). The Economics of Director Heterogeneity. *Financial Management*, *40*, 5–38.

Becker, G. (1957). *The Economics of Discrimination*. University of Chicago Press.

Becker, G. (1964). *Human Capital*. University of Chicago Press.

Bernile, G., Bhagwat, V., & Yonker, S. (2018). Board diversity, firm risk, and corporate policies. *Journal of Financial Economics*, *127*(3), 588–612.

Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis*, *45*, 150–156.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure* (Vol. 7). Free Press New York.

Campbell, K., & Mínguez-Vera, A. (2008). Gender Diversity in the Boardroom and Firm Financial Performance. *Journal of Business Ethics*, *83*, 435–451.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82.

Carter, D. A., D'Souza, F., Simkins, B. J., & Simpson, W. G. (2010). The Gender and Ethnic Diversity of US Boards and Board Committees and Firm Financial Performance. *Corporate Governance: An International Review*, *18*(5), 396–414.

Carter, D. A., Simkins, B. J., & Simpson, W. G. (2003). Corporate Governance, Board Diversity, and Firm Value. *Financial Review*, *38*, 33–53.

Chai, D., Dai, M., Gharghori, P., & Hong, B. (2021). Internet search intensity and its relation with trading activity and stock returns. *International Review of Finance*, *21*(1), 282–311.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The journal of finance*, *66*(5), 1461–1499.

Dodd, O., Frijns, B., Gong, K., & Liao, S. (2022). *Board Cultural Diversity and Firm Performance Under Competitive Pressures* (Working paper). Available at SSRN 4212481.

Edmans, A., Flammer, C., & Glossner, S. (2023). *Diversity, Equity, and Inclusion* (Working Paper No. 31215). National Bureau of Economic Research.

Ekinci, C., & Bulut, A. E. (2021). Google search and stock returns: A study on BIST 100 stocks. *Global Finance Journal*, *47*, 100518.

Erhardt, N. L., Werbel, J. D., & Shrader, C. B. (2003). Board of Director Diversity and Firm Financial Performance. *Corporate Governance: An International Review*, *11*, 102–111.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, *116*(1), 1–22.

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of financial economics*, *128*(2), 234–252.

Fombrun, C., & Shanley, M. (1990). What's in a Name? Reputation Building and Corporate Strategy. *Academy of Management Journal*, *3*(2), 233–258.

Gardenswartz, L., & Rowe, A. (2003). *Diverse Teams at Work: Capitalizing on the Power of Diversity*. Society for Human Resource Management.

Giannetti, M., & Wang, T. Y. (2023). Public attention to gender equality and board gender diversity. *Journal of Financial and Quantitative Analysis*, *58*(2), 485–511.

Green, J., & Hand, J. R. M. (2021). *Diversity Matters/Delivers/Wins Revisited in S&P 500 Firms* (Working paper). Available at SSRN 3849562.

Hambrick, D. C., Cho, T. S., & Chen, M.-J. (1996). The Influence of Top Management Team Heterogeneity on Firms' Competitive Moves. *Administrative Science Quarterly*, *41*, 659–684.

Herring, C. (2009). Does Diversity Pay?: Race, Gender, and the Business Case for Diversity. *American Sociological Review*, *74*(2), 208–224.

Hillman, A. J., & Dalziel, T. (2003). Boards of Directors and Firm Performance: Integrating Agency and Resource Dependence Perspectives. *Academy of Management Review*, *28*(3), 383–396.

Hoffman, L. R., & Maier, N. R. F. (1961). Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *The Journal of Abnormal and Social Psychology*, *62*, 401–407.

Hoffman, L. R. (1959). Homogeneity of member personality and its effect on group problem-solving. *The Journal of Abnormal and Social Psychology*, *58*, 27–32.

Hsieh, T.-S., Kim, J.-B., Wang, R. R., & Wang, Z. (2020). Seeing is believing? Executives' facial trustworthiness, auditor tenure, and audit fees. *Journal of Accounting and Economics*, *69*(1), 101260.

Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone management. *The accounting review*, *89*(3), 1083–1113.

Hunt, V., Prince, S., Dixon-Fyle, S., & Dolan, K. (2020). Diversity wins — How inclusion matters [McKinsey & Company, Last accessed May 2, 2025]. https://www.mckinsey.com/~/media/mckinsey/featured%5C%20insights/diversity%5C%20and%5C%20inclusion/diversity%5C%20win%5C%20how%5C%20inclusion%5C%20matters/diversity-wins-how-inclusion-matters-vf.pdf

Jia, Y., Lent, L. v., & Zeng, Y. (2014). Masculinity, Testosterone, and Financial Misreporting. *Journal of Accounting Research*, *52*(5), 1195–1246.

Kachur, A., Osin, E., Davydov, D., Shutilov, K., & Novokshonov, A. (2020). Assessing the Big Five personality traits using real-life static facial images. *Scientific Reports*, *10*(1), 8487.

Kaldor, N. (1966). Marginal Productivity and the Macro-Economic Theories of Distribution: Comment on Samuelson and Modigliani. *The Review of Economic Studies*, *33*(4), 309–319.

Kamiya, S., Kim, Y. H. (, & Park, S. (2019). The face of risk: CEO facial masculinity and firm risk. *European Financial Management*, *25*(2), 239–270.

King, D. E. (2009). Dlib-Ml: A Machine Learning Toolkit. *J. Mach. Learn. Res.*, *10*, 1755–1758.

King, D. E. (2017). High Quality Face Recognition with Deep Metric Learning [Last accessed May 2, 2025]. http://blog.dlib.net/2017/02/high-quality-face-recognition-with-deep.html

Kolla, M., & Savadamuthu, A. (2023). The Impact of Racial Distribution in Training Data on Face Recognition Bias: A Closer Look. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, 313–322. https://doi.org/10.1109/WACVW58289.2023.00035

Levi, G., & Hassner, T. (2015). Age and Gender Classification Using Convolutional Neural Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops*. %5Curl%7Bhttps://osnathassner.github.io/talhassner/projects/cnn_agegender%7D

LSEG. (2024). *Diversity and Inclusion Scores from LSEG – Methodology*. Retrieved August 16, 2024, from https://www.lseg.com/content/dam/ftse-russell/en_us/documents/methodology/diversity-inclusion-rating-methodology.pdf

Lu, Y., & Teo, M. (2022). Do Alpha Males Deliver Alpha? Facial Width-to-Height Ratio and Hedge Funds. *Journal of Financial and Quantitative Analysis*, *57*(5), 1727–1770.

Mandasari, R., Mardiana, R., & Dewinda, M. C. (2023). Firm Performance and Stock Returns: The Moderating Role of Google Search Volume Index: Evidence from Companies Listed in Indonesian Sharia Stock Index. *Journal of Applied Business Administration*, *7*(2), 255–264.

Mazières, A., Menezes, T., & Roth, C. (2021). Computational appraisal of gender representativeness in popular movies. *Humanities and Social Sciences Communications*, *8*(1), 137.

Minton, B. A., Taillard, J. P., & Williamson, R. (2014). Financial Expertise of the Board, Risk Taking, and Performance: Evidence from Bank Holding Companies. *Journal of Financial and Quantitative Analysis*, *49*(2), 351–380.

Nemeth, C. J. (1986). Differential contributions of majority and minority influence. *Psychological Review*, *93*(1), 23–32.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica*, *55*, 703–708.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256.

Obaid, K., & Pukthuanthong, K. (2022). A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics*, *144*(1), 273–297.

Petersen, M. A. (2008). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, *22*(1), 435–480.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific reports*, *3*(1), 1–6.

Radford, A., Józefowicz, R., & Sutskever, I. (2017). *Learning to Generate Reviews and Discovering Sentiment* (Working Paper). arXiv preprint 1704.01444.

Rindova, V. P. (1999). What Corporate Boards have to do with Strategy: A Cognitive Perspective. *Journal of Management Studies*, *36*(7), 953–975.

Schakel, A. M. J., & Wilson, B. J. (2015). *Measuring Word Significance using Distributed Representations of Words* (Working Paper). arXiv preprint 1508.02297.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 815–823.

Spence, M. (1978). Job Market Signaling. In *Uncertainty in economics* (pp. 281–306). Academic Press.

Thorndike, E. (1920). A Constant Error in Psychological Ratings. *Journal of Applied Psychology*, *4*, 25–29.

Tkachenko, Y., & Jedidi, K. (2020). *What Personal Information Can a Consumer Facial Image Reveal? Implications for Marketing ROI and Consumer Privacy* (Working paper). Available at SSRN 3616470.

Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking & Finance*, *41*, 17–35.

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*, *114*(2), 246–257.

Westphal, J. D., & Milton, L. P. (2000). How Experience and Network Ties Affect the Influence of Demographic Minorities on Corporate Boards. *Administrative Science Quarterly*, *45*(2), 366–398.

Wingard, J. (2021). No More Excuses: How 4 Leading Companies Diversified Their Boards [Available at forbes.com/sites/jasonwingard/2021/08/12/no-more-excuses-how-4-leading-companies-diversified-their-boards, Last accessed May 2, 2025]. *Forbes*.

Xu, Q., Fernando, G. D., & Schneible, R. A. (2022). Age diversity, firm performance and managerial ability. *Review of Accounting and Finance*, *21*(4), 276–298.

Yoshinaga, C., & Rocco, F. (2020). Investor attention: can google search volumes predict stock returns? *BBR. Brazilian Business Review*, *17*, 523–539.

Zhang, L. (2020). An Institutional Approach to Gender Diversity and Firm Performance. *Organization Science*, *31*(2), 439–457.

A Appendix

Definitions of variables

[Table 16 about here.]

Firm-year observations across GICS industries

[Table 17 about here.]

Figures

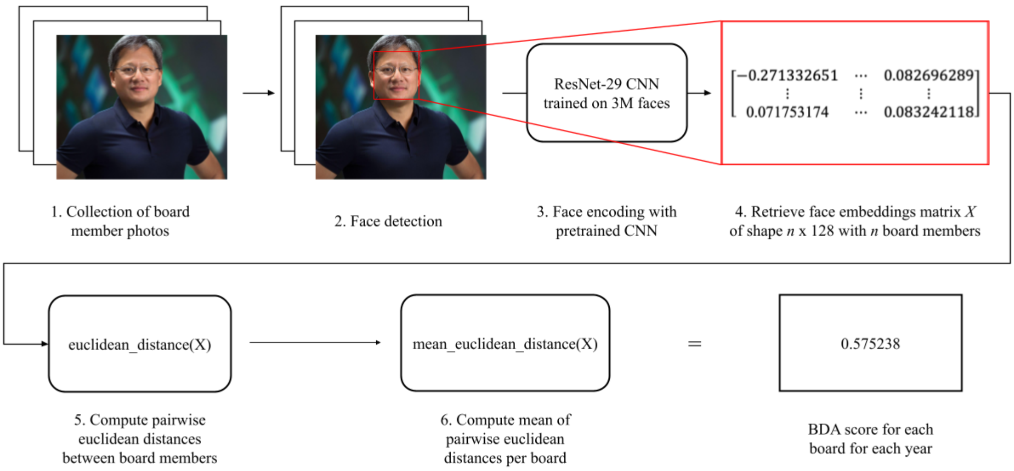

Figure 1: Our deep-learning-based computer vision approach to measuring firm-level board diversity appearance

This figure illustrates our approach to measuring firm-level board diversity appearance. Source of portrait image: https://www.nvidia.com/en-us/about-nvidia/board-of-directors/jensen-huang/

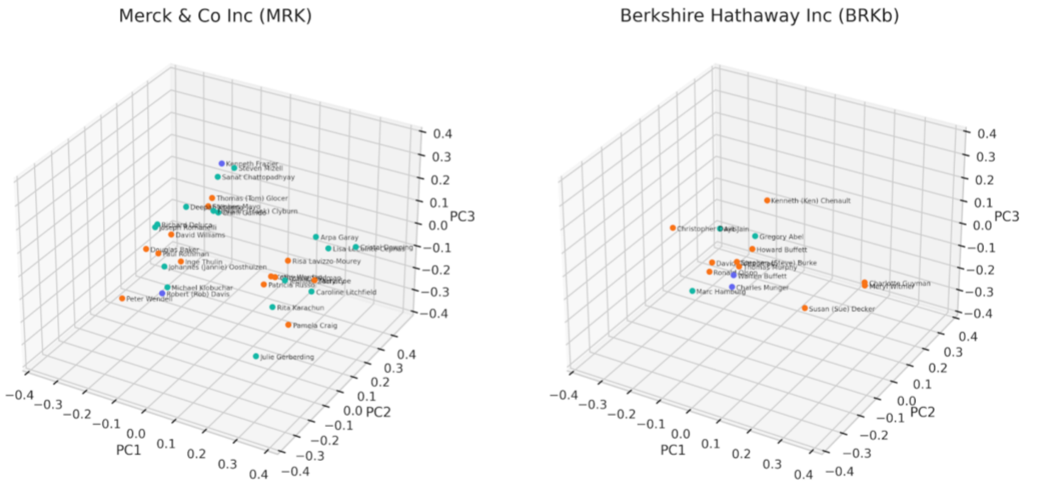

Figure 2: Visualization of the board diversity appearance
This figure illustrates the board diversity appearance for the two companies "Merck & Co., Inc." and "Berkshire Hathaway Inc.". We performed Principle Component Analysis (PCA) to reduce the number of dimensions from 128 to 3. All directors are indicated by orange dots, officers are indicated by teal dots, and members of the board who are both officers and directors are indicated by blue dots.

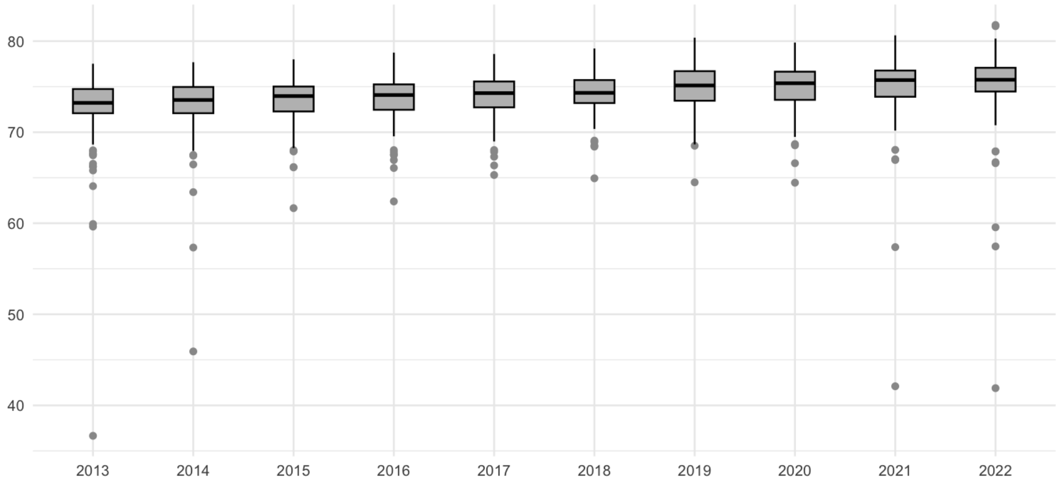

Figure 3: BDA scores over time
This figure shows boxplots of the BDA scores over time. The sample period is from 2013 to 2022.

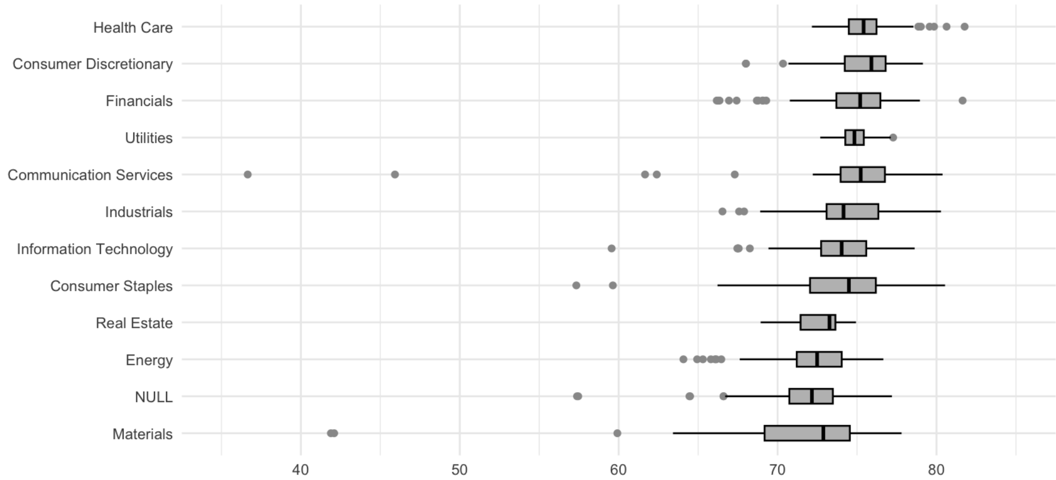

Figure 4: BDA scores across industries
This figure shows boxplots of the BDA scores across GICS industries. The sample period is from 2013 to 2022.

Tables

Table 1
Regressing Principal Components on standard diversity measures

	PC1	PC2	PC3
Gender Diversity	0.1799^{***}	0.0168	-0.0037
	(0.0213)	(0.0384)	(0.0392)
Ethnic Diversity	0.0782^{***}	0.0995^{***}	0.0794^{**}
	(0.0166)	(0.0223)	(0.0354)
Skills Diversity	-0.0054	0.0218	0.0399^{**}
	(0.0103)	(0.0276)	(0.0178)
Constant	-0.0559^{***}	-0.0300^{***}	-0.0371
	(0.0104)	(0.0111)	(0.0230)
Observations	115	115	115
R^2	0.3678	0.1233	0.1233
Adjusted R^2	0.3507	0.0996	0.0996

This table displays results of the regressions of different Principal Components from the PCA on the BDA scores on classical diversity measures, expressed as percentages. The first column has PC1 as dependent variable. The second column has PC2 as dependent variable. The third column has PC3 as dependent variable. The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 2
Descriptive statistics

	Mean	Median	Standard deviation	Minimum	Maximum	Observa-tions
BDA Score	74.09	74.55	3.39	36.65	81.77	1,266
ATT	3.18	0.77	8.17	0.00	72.75	1,198
Monthly Returns	0.01	0.01	0.08	-0.83	2.14	15,870
Yearly Returns	0.16	0.12	0.39	-0.70	7.43	1,432
Tobin's Q	2.55	1.96	1.80	0.95	10.27	1,354
ROA	0.07	0.06	0.07	-0.18	0.29	1,369
Total Assets	199.88	71.88	400.46	7.33	2,419.51	1,369
Market Capitalization	162.50	92.77	261.39	2.07	2,993.85	1,354
Net Income	6.41	3.90	9.50	-7.44	59.97	1,463
Book to Market	0.36	0.27	0.33	-0.10	1.47	1,354
Firm Age	45.09	33.00	33.73	1.00	140.00	1,269
Cash	0.14	0.11	0.13	0.00	0.56	1,248
Dividend Yield	0.02	0.02	0.02	0.00	0.08	1,353
Leverage	0.68	0.68	0.20	0.19	1.22	1,369
Number of Employees	127.78	64.00	235.30	1.40	2,300.00	1,351
Board Size	12.00	12.00	2.08	1.00	19.00	1,238

This table reports descriptive statistics for the variables used in this analysis. The variables total assets, market capitalization, and net income are reported in millions and the number of employees is reported in thousands. Accounting measures are winsorized at the 1% and 99% levels. All variables are retrieved from LSEG Datastream. The sample period is from 2013 to 2022.

Table 3
Regressing BDA scores on firm characteristics

	BDA	
	(1)	(2)
Returns	-0.4254	-0.1780
	(0.2508)	(0.3663)
ln Total Assets	0.2474	-0.0957
	(0.1662)	(0.2794)
ln Book to Market	-0.3611*	-0.5109**
	(0.1881)	(0.2067)
ln Firm Age	0.0002	-0.1237
	(0.2074)	(0.2224)
ln Cash		0.0891
		(0.1377)
Dividend Yield		17.1828
		(12.1558)
ln Leverage		-0.7987
		(0.7096)
ln Number of Employees		0.1857
		(0.2387)
ln Board Size		2.9399
		(2.0051)
Industry x Year FE	Yes	Yes
Observations	1,013	897
R^2	0.2863	0.3253
Adjusted R^2	0.2001	0.2279

This table presents results of the regressions of the BDA scores on firm characteristics over the period 2013 to 2022. Both models include industry times year fixed effects. All independent variables are lagged by one year. All independent variables except for returns and dividend yield are logarithmized. Accounting measures are winsorized at the 1% and 99% levels. The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 4
Regressing BDA scores on standard diversity measures calculated using the Blau Index

	BDA				
	(1)	(2)	(3)	(4)	(5)
Gender Diversity$_{Blau}$	0.1125***	0.1139***	0.0956***	0.1139**	0.0970***
	(0.0268)	(0.0431)	(0.0234)	(0.0446)	(0.0232)
Ethnic Diversity$_{Blau}$		0.0398*		0.0397**	
		(0.0209)		(0.0176)	
Cultural Diversity$_{Blau}$			−0.0177		−0.0171
			(0.0120)		(0.0114)
Skills Diversity$_{Blau}$				0.0005	−0.0149
				(0.0294)	(0.0372)
Constant	0.7024***	0.6917***	0.7109***	0.6916***	0.7169***
	(0.0107)	(0.0160)	(0.0092)	(0.0164)	(0.0195)
Observations	1,058	115	450	115	450
R^2	0.1238	0.1231	0.0970	0.1231	0.0986
Adjusted R^2	0.1230	0.1074	0.0930	0.0994	0.0926

This table displays the results of the regressions of the BDA scores on classical diversity measures, expressed as Blau Indices. The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 5
Regressing BDA scores on Diversity & Inclusion scores by LSEG

	BDA				
	(1)	(2)	(3)	(4)	(5)
DI Score	0.0547***				
	(0.0187)				
DIV Score		0.0538***			
		(0.0136)			
INC Score			0.0244***		
			(0.0072)		
PPL Score				0.0327***	
				(0.0105)	
CON Score					-0.0401***
					(0.0112)
Constant	71.7877***	73.0730***	73.5700***	73.3730***	78.6632***
	(1.1649)	(0.5388)	(0.5478)	(0.6398)	(0.9408)
Observations	737	773	773	773	773
R^2	0.0442	0.0772	0.0624	0.0462	0.0473
Adjusted R^2	0.0429	0.0760	0.0612	0.0449	0.0460

This table presents results of the regressions of the BDA scores on the Diversity & Inclusion Score (DI) from LSEG and its four subcategories (DIV, INC, PPL, CON). The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 6
Factor model analysis of the returns of long-short portfolios built using different sorting criteria and tercile cut-off points

Portfolio sort based on:	BDA	ATT	Gender diversity	Residuals
RMRF	-0.1748^{***}	0.0856	-0.0351	-0.2017^{***}
	(0.0439)	(0.0586)	(0.0430)	(0.0491)
SMB	-0.0552	-0.0265	0.0000	-0.0402
	(0.0693)	(0.0650)	(0.0860)	(0.0614)
HML	-0.1696^{**}	-0.1523^{*}	-0.0160	-0.1749^{***}
	(0.0655)	(0.0819)	(0.0630)	(0.0651)
WML	0.0288	0.0087	0.0167	0.0163
	(0.0409)	(0.0374)	(0.0350)	(0.0507)
RMW	0.1544	0.1687	0.0208	0.1680
	(0.1103)	(0.1201)	(0.1162)	(0.1227)
CMA	-0.1126	-0.4355^{***}	-0.2400^{**}	-0.1486
	(0.1621)	(0.1471)	(0.1048)	(0.1179)
α	-0.0016	-0.0020	-0.0026	-0.0003
	(0.0026)	(0.0019)	(0.0022)	(0.0018)
Observations	120	120	120	120
R^2	0.2417	0.3955	0.0848	0.3345
Adjusted R^2	0.2014	0.3634	0.0362	0.2992

This table shows the results for long-short portfolios using the three different sorting categories. The portfolios are each long in the firms of the highest tercile of the respective sorting category and low in the firms of the lowest tercile in the respective sorting category. In the first column the overall BDA score is used as sorting category. In the seconds column ATT is used as sorting category. In the third column the Blau index of percentage of female board members is used as sorting category. In the fourth column the residuals of a regression of BDA on Blau index of percentage of female board members is used as sorting category. The corresponding standard errors are calculated following Newey and West (1987) and reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 7
Six-factor α's for portfolios independently double-sorted using different sorting criteria based on one-year lagged portfolio sort

Panel A. Double-sorting on BDA and ATT

		Sorting on ATT		
		(1)	(2)	(3)
	(1)	0.0005	0.0001	0.0048**
		(0.0017)	(0.0013)	(0.0019)
Sorting on BDA	(2)	0.0031*	−0.0015	0.0030**
		(0.0016)	(0.0013)	(0.0014)
	(3)	0.0016	0.0032	0.0002
		(0.0020)	(0.0021)	(0.0024)

Panel B. Double-sorting on gender diversity and ATT

		Sorting on ATT		
		(1)	(2)	(3)
	(1)	0.0021	−0.0024	0.0048**
		(0.0015)	(0.0016)	(0.0024)
Sorting on Gender Diversity (Blau Index)	(2)	−0.00002	0.0012	0.0023
		(0.0013)	(0.0012)	(0.0021)
	(3)	0.0043**	0.0002	0.0026
		(0.0019)	(0.0022)	(0.0021)

Panel C. Double-sorting on residuals and ATT

		Sorting on ATT		
		(1)	(2)	(3)
	(1)	0.0003	0.0015	0.0042*
		(0.0018)	(0.0013)	(0.0022)
Sorting on Residuals	(2)	0.0026	−0.0008	0.0030
		(0.0019)	(0.0012)	(0.0019)
	(3)	0.0009	−0.0001	0.0003
		(0.0017)	(0.0012)	(0.0025)

This table shows the six-factor α's for portfolios sorted based on one-year lagged ATT and a second category. In Panel A, the one-year lagged BDA score is used as second sorting category. In Panel B, one-year lagged gender diversity, measured as the Blau index of percentage of female board members, is used as second sorting category. In Panel C, the one-year lagged residuals of a regression of BDA on gender diversity are used as second sorting category. For all categories terciles are used as cut-off points for the sorting. (1) indicates the high tercile for the respective sorting category. (2) indicates the middle tercile for the respective sorting category. (3) indicates the low tercile for the respective sorting category. The corresponding standard errors are calculated following Newey and West (1987) and reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 8
Six-factor α's for portfolios independently double-sorted using different sorting criteria based on two-year lagged portfolio sort

Panel A. Double-sorting on BDA and ATT

		Sorting on ATT (1)	(2)	(3)
	(1)	-0.0006	0.0012	0.0073^{***}
		(0.0020)	(0.0022)	(0.0022)
Sorting on BDA	(2)	0.0029^{**}	0.0008	0.0003
		(0.0017)	(0.0010)	(0.0015)
	(3)	0.0039	-0.0009	0.0022
		(0.0026)	(0.0014)	(0.0024)

Panel B. Double-sorting on gender diversity and ATT

		Sorting on ATT (1)	(2)	(3)
	(1)	0.0013	-0.0007	0.0033
		(0.0019)	(0.0019)	(0.0026)
Sorting on Gender Diversity	(2)	0.0031^{*}	-0.0012	0.0024
		(0.0017)	(0.0016)	(0.0021)
	(3)	-0.0010	0.0020	0.0033
		(0.0020)	(0.0016)	(0.0027)

Panel C. Double-sorting on residuals and ATT

		Sorting on ATT (1)	(2)	(3)
	(1)	-0.0003	0.0038^{**}	0.0047^{*}
		(0.0015)	(0.0015)	(0.0025)
Sorting on residuals	(2)	0.0043^{*}	0.0010	0.0040^{**}
		(0.0023)	(0.0012)	(0.0015)
	(3)	0.0026	-0.0038^{**}	0.0021
		(0.0023)	(0.0018)	(0.0027)

This table shows the six-factor α's for portfolios sorted based on two-year lagged ATT and a second category. In Panel A, the two-year lagged BDA score is used as second sorting category. In Panel B, two-year lagged gender diversity, measured as the Blau index of percentage of female board members, is used as second sorting category. In Panel C, the two-year lagged residuals of a regression of BDA on gender diversity are used as second sorting category. For all categories terciles are used as cut-off points for the sorting. (1) indicates the high tercile for the respective sorting category. (2) indicates the middle tercile for the respective sorting category. (3) indicates the low tercile for the respective sorting category. The corresponding standard errors are calculated following Newey and West (1987) and reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 9
Regressing firm value and accounting performance measures on BDA scores

	Tobin's Q		ROA	
	(1)	(2)	(3)	(4)
BDA Score	0.0569	0.0606	0.0011	0.0017
	(0.0318)	(0.0372)	(0.0016)	(0.0016)
Gender Diversity		0.0220*		0.0009**
		(0.0099)		(0.0004)
ln Total Assets	-0.8859***	-0.9106***	-0.0066	-0.0079
	(0.1754)	(0.1742)	(0.0056)	(0.0057)
ln Book to Market			-0.0311***	-0.0313***
			(0.0052)	(0.0053)
ln Firm Age	0.0328	-0.0034	0.0005	-0.0008
	(0.1047)	(0.1078)	(0.0044)	(0.0041)
Industry x Year FE	Yes	Yes	Yes	Yes
Observations	1,174	1,161	1,124	1,114
R^2	0.4999	0.5115	0.4855	0.5079
Adjusted R^2	0.4440	0.4556	0.4245	0.4485

This table presents results of the regressions of accounting performance measures on the BDA scores and controls over the period 2013 to 2022. All models include industry times year fixed effects. Gender Diversity is the Blau index of the percentage of female board members. All independent variables are lagged by one year. The accounting measures are winsorized at the 1% and 99% levels. The control variables are logarithmized. In order to avoid multicollinearity problems, the book-to-market ratio is not used as a control variable in the regressions (1) and (2). The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 10
Regressing firm value and accounting performance measures on 2-year-lagged BDA scores

	Tobin's Q		ROA	
	(1)	(2)	(3)	(4)
BDA Score	0.0777**	0.0952**	0.0012	0.0019
	(0.0328)	(0.0362)	(0.0015)	(0.0016)
Gender Diversity		0.0176		0.0010**
		(0.0108)		(0.0004)
ln Total Assets	-0.9060***	-0.9312***	-0.0073	-0.0089
	(0.1850)	(0.1845)	(0.0056)	(0.0057)
ln Book to Market			-0.0313***	-0.0307***
			(0.0051)	(0.0052)
ln Firm Age	0.0386	0.0089	-0.0002	-0.0021
	(0.1066)	(0.1123)	(0.0045)	(0.0043)
Industry x Year FE	Yes	Yes	Yes	Yes
Observations	1,056	1,041	1,007	995
R^2	0.5018	0.5123	0.5051	0.5306
Adjusted R^2	0.4462	0.4564	0.4462	0.4734

This table presents results of the regressions of accounting performance measures on the BDA scores and controls over the period 2013 to 2022. All models include industry times year fixed effects. Gender Diversity is the Blau index of the percentage of female board members. The variables BDA Score and Gender Diversity are lagged by two years. Control variables are lagged by one year and logarithmized. The accounting measures are winsorized at the 1% and 99% levels. In order to avoid multicollinearity problems, the book-to-market ratio is not used as a control variable in the regressions (1) and (2). The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 11
Regressing firm value and accounting performance measures on BDA/ATT portfolios

	Tobin's Q		ROA	
	(1)	(2)	(3)	(4)
High BDA/High ATT	0.2379	0.1814	-0.0022	-0.0038
	(0.2800)	(0.2725)	(0.0093)	(0.0093)
High BDA/Middle ATT	0.1800	0.0986	0.0072	0.0045
	(0.2651)	(0.2547)	(0.0085)	(0.0083)
High BDA/Low ATT	-0.4266	-0.4979	-0.0021	-0.0045
	(0.3488)	(0.3388)	(0.0167)	(0.0163)
Middle BDA/High ATT	0.3253	0.2818	-0.0215	-0.0230*
	(0.2944)	(0.2924)	(0.0119)	(0.0122)
Middle BDA/Low ATT	-0.2966	-0.3078	-0.0169*	-0.0164*
	(0.2209)	(0.2126)	(0.0087)	(0.0088)
Low BDA/High ATT	0.0014	0.0460	-0.0156	-0.0134
	(0.2835)	(0.2976)	(0.0112)	(0.0110)
Low BDA/Middle ATT	-0.7052**	-0.6650**	-0.0242**	-0.0195*
	(0.2671)	(0.2517)	(0.0098)	(0.0094)
Low BDA/Low ATT	-0.4259	-0.4262	-0.0174**	-0.0165*
	(0.2658)	(0.2653)	(0.0077)	(0.0076)
Gender Diversity		0.0193*		0.0008**
		(0.0097)		(0.0004)
ln Total Assets	-0.9585***	-0.9785***	-0.0065	-0.0073
	(0.1785)	(0.1780)	(0.0057)	(0.0059)
ln Book to Market			-0.0306***	-0.0314***
			(0.0050)	(0.0052)
ln Firm Age	0.0247	-0.0137	0.0005	-0.0007
	(0.1053)	(0.1063)	(0.0044)	(0.0042)
Industry x Year FE	Yes	Yes	Yes	Yes
Observations	1,174	1,161	1,124	1,114
R^2	0.5182	0.5274	0.5010	0.5183
Adjusted R^2	0.4608	0.4698	0.4380	0.4562

This table presents results of the regressions of accounting performance measures on portfolios based on independent double-sorting using overall BDA score and ATT as sorting categories. For both categories terciles are used as cut-off points for the sorting. High, Middle, and Low represent the top, middle, and bottom terciles, respectively, for the given sorting category. All independent variables are lagged by one year. All models include industry times year fixed effects. Gender Diversity is the Blau index of the percentage of female board members. The accounting measures are winsorized at the 1% and 99% levels. The control variables are logarithmized. In order to avoid multicollinearity problems, the book-to-market ratio is not used as a control variable in the regressions (1) and (2). The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 12
Regressing firm value and accounting performance measures on 2-year lagged BDA/ATT portfolios

	Tobin's Q		ROA	
	(1)	(2)	(3)	(4)
High BDA/High ATT	0.2119	0.1569	-0.0034	-0.0055
	(0.3179)	(0.3142)	(0.0088)	(0.0087)
High BDA/Middle ATT	0.2046	0.1284	0.0058	0.0018
	(0.3055)	(0.2896)	(0.0077)	(0.0080)
High BDA/Low ATT	-0.2703	-0.3460	0.0099	0.0063
	(0.4027)	(0.3924)	(0.0146)	(0.0138)
Middle BDA/High ATT	0.2602	0.2166	-0.0221*	-0.0244*
	(0.3424)	(0.3357)	(0.0117)	(0.0120)
Middle BDA/Low ATT	-0.2777	-0.2740	-0.0221**	-0.0219**
	(0.2624)	(0.2517)	(0.0080)	(0.0083)
Low BDA/High ATT	-0.0511	-0.0007	-0.0103	-0.0089
	(0.3252)	(0.3396)	(0.0112)	(0.0115)
Low BDA/Middle ATT	-0.7638**	-0.7210**	-0.0208*	-0.0155
	(0.3002)	(0.3033)	(0.0105)	(0.0100)
Low BDA/Low ATT	-0.3871	-0.3879	-0.0136*	-0.0134
	(0.2950)	(0.2961)	(0.0069)	(0.0074)
Gender Diversity		0.0165		0.0009**
		(0.0103)		(0.0004)
ln Total Assets	-0.9666***	-0.9839***	-0.0063	-0.0074
	(0.1918)	(0.1919)	(0.0058)	(0.0060)
ln Book to Market			-0.0312***	-0.0313***
			(0.0051)	(0.0051)
ln Firm Age	0.0439	-0.0002	-0.0001	-0.0021
	(0.1101)	(0.1128)	(0.0045)	(0.0044)
Industry x Year FE	Yes	Yes	Yes	Yes
Observations	1,056	1,041	1,007	995
R^2	0.5128	0.5197	0.5213	0.5416
Adjusted R^2	0.4544	0.4606	0.4601	0.4816

This table presents results of the regressions of accounting performance measures on portfolios based on independent double-sorting using overall BDA score and ATT as sorting categories. For both categories terciles are used as cut-off points for the sorting. High, Middle, and Low represent the top, middle, and bottom terciles, respectively, for the given sorting category. The portfolios are based on two-year lagged BDA scores and ATT indices. Gender Diversity is lagged by two years. Control variables are lagged by one year. All models include industry times year fixed effects. Gender Diversity is the Blau index of the percentage of female board members. The accounting measures are winsorized at the 1% and 99% levels. The control variables are logarithmized. In order to avoid multicollinearity problems, the book-to-market ratio is not used as a control variable in the regression of Tobin's Q on the BDA scores. The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 13
Regressing firm value and accounting performance measures on one-year and two-year lagged gender diversity and BDA residuals

	Tobin's Q		ROA	
	1yr lag	2yr lag	1yr lag	2yr lag
Gender Diversity	0.0288**	0.0266**	0.0011**	0.0010**
	(0.0105)	(0.0105)	(0.0004)	(0.0004)
BDA Residuals	6.8418*	7.9827**	0.2141	0.2152
	(3.3761)	(3.3735)	(0.1566)	(0.1517)
ln Total Assets	-0.8964***	-0.9025***	-0.0070	-0.0075
	(0.1692)	(0.1740)	(0.0057)	(0.0058)
ln Book to Market			-0.0318***	-0.0320***
			(0.0054)	(0.0054)
ln Firm Age	-0.0500	-0.0455	-0.0018	-0.0028
	(0.1038)	(0.1062)	(0.0040)	(0.0041)
Industry x Year FE	Yes	Yes	Yes	Yes
Observations	1,219	1,173	1,173	1,126
R^2	0.5142	0.5099	0.5039	0.5170
Adjusted R^2	0.4571	0.4545	0.4425	0.4594

This table presents results of the regressions of accounting performance measures on gender diversity (as the Blau index of the percentage of female board members) and BDA residuals (as and controls over the period 2013 to 2022. All models include industry times year fixed effects. The variables Gender Diversity and BDA Residuals are lagged by one year in the first and third regression and by two years in the second and fourth regression. All control variables are lagged by one year and logarithmized. The accounting measures are winsorized at the 1% and 99% levels. In order to avoid multicollinearity problems, the book-to-market ratio is not used as a control variable in the regression with Tobin's Q. The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 14
Regressing firm value and accounting performance measures on one-year and two-year lagged GEN/ATT portfolios

	Tobin's Q		ROA	
	1yr lag	2yr lag	1yr lag	2yr lag
High GEN/High ATT	0.2289	0.1988	-0.0186*	-0.0224*
	(0.2866)	(0.3218)	(0.0101)	(0.0117)
High GEN/Middle ATT	-0.1404	-0.2150	-0.0041	-0.0098
	(0.2397)	(0.2897)	(0.0054)	(0.0081)
High GEN/Low ATT	-0.2998	-0.2038	-0.0052	-0.0007
	(0.2748)	(0.3236)	(0.0147)	(0.0152)
Middle GEN/High ATT	0.2138	0.2663	-0.0155	-0.0193*
	(0.2367)	(0.3028)	(0.0094)	(0.0088)
Middle GEN/Low ATT	-0.4371	-0.3505	-0.0197*	-0.0251**
	(0.2655)	(0.3036)	(0.0105)	(0.0092)
Low GEN/High ATT	-0.1724	-0.2177	-0.0202*	-0.0240**
	(0.2293)	(0.2620)	(0.0095)	(0.0101)
Low GEN/Middle ATT	-0.7021**	-0.5470	-0.0311**	-0.0325***
	(0.3047)	(0.3687)	(0.0103)	(0.0081)
Low GEN/Low ATT	-0.6131*	-0.4898	-0.0340**	-0.0305**
	(0.2996)	(0.3376)	(0.0106)	(0.0125)
ln Total Assets	-0.9541***	-0.9562***	-0.0064	-0.0059
	(0.1808)	(0.1932)	(0.0057)	(0.0060)
ln Book to Market			-0.0323***	-0.0328***
			(0.0049)	(0.0052)
ln Firm Age	-0.0554	-0.0528	-0.0016	-0.0028
	(0.1013)	(0.1084)	(0.0040)	(0.0042)
Industry x Year FE	Yes	Yes	Yes	Yes
Observations	1,184	1,062	1,137	1,016
R^2	0.5151	0.5055	0.5140	0.5355
Adjusted R^2	0.4578	0.4466	0.4534	0.4767

This table presents results of the regressions of accounting performance measures on portfolios based on independent double-sorting using gender diversity (indicated by GEN and defined as the Blau index of the percentage of female board members) and ATT as sorting categories. For both categories terciles are used as cut-off points for the sorting. High, Middle, and Low represent the top, middle, and bottom terciles, respectively, for the given sorting category. The portfolios are based on one-year (two-year) lagged gender diversity percentages and ATT indices in the first and third (second and fourth) regression. Control variables are lagged by one year and logarithmized. All models include industry times year fixed effects. The accounting measures are winsorized at the 1% and 99% levels. In order to avoid multicollinearity problems, the book-to-market ratio is not used as a control variable in the regression of Tobin's Q on the BDA scores. The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 15
Regressing firm value and accounting performance measures on one-year and two-year lagged RES/ATT portfolios

	Tobin's Q		ROA	
	1yr lag	2yr lag	1yr lag	2yr lag
High RES/High ATT	0.2163	0.1896	0.0074	0.0001
	(0.2403)	(0.2693)	(0.0119)	(0.0109)
High RES/Middle ATT	0.0038	0.1351	0.0087	0.0036
	(0.2071)	(0.2490)	(0.0084)	(0.0088)
High RES/Low ATT	-0.5028	-0.3651	-0.0061	-0.0017
	(0.3600)	(0.3901)	(0.0171)	(0.0178)
Middle RES/High ATT	0.1444	0.2180	-0.0134	-0.0181
	(0.2808)	(0.2971)	(0.0092)	(0.0104)
Middle RES/Low ATT	-0.1414	-0.0104	-0.0025	-0.0009
	(0.2148)	(0.2433)	(0.0094)	(0.0113)
Low RES/High ATT	0.1941	0.1381	-0.0050	-0.0044
	(0.2906)	(0.3437)	(0.0106)	(0.0137)
Low RES/Middle ATT	-0.6073**	-0.6884*	0.0016	-0.0032
	(0.2624)	(0.3288)	(0.0122)	(0.0113)
Low RES/Low ATT	-0.4372	-0.4565	-0.0091	-0.0129
	(0.2535)	(0.2591)	(0.0083)	(0.0085)
ln Total Assets	-0.9798***	-0.9758***	-0.0081	-0.0070
	(0.1724)	(0.1911)	(0.0057)	(0.0058)
ln Book to Market			-0.0292***	-0.0313***
			(0.0048)	(0.0051)
ln Firm Age	0.0251	0.0335	0.0019	0.0005
	(0.1026)	(0.1113)	(0.0046)	(0.0048)
Industry x Year FE	Yes	Yes	Yes	Yes
Observations	1,008	1,010	969	965
R^2	0.5467	0.5181	0.4810	0.5035
Adjusted R^2	0.4894	0.4573	0.4117	0.4369

This table presents results of the regressions of accounting performance measures on portfolios based on independent double-sorting using the BDA residuals (indicated by RES and defined as the residuals of a regression of BDA on the Blau Index of percentage female board members) and ATT as sorting categories. For both categories terciles are used as cut-off points for the sorting. High, Middle, and Low represent the top, middle, and bottom terciles, respectively, for the given sorting category. The portfolios are based on one-year (two-year) lagged BDA residuals and ATT indices in the first and third (second and fourth) regression. Control variables are lagged by one year and logarithmized. All models include industry times year fixed effects. The accounting measures are winsorized at the 1% and 99% levels. In order to avoid multicollinearity problems, the book-to-market ratio is not used as a control variable in the regression of Tobin's Q on the BDA scores. The corresponding standard errors clustered by firm and year following Petersen (2008) are reported in parentheses. Significance levels are indicated by *, **, and *** and correspond to the 10%, 5%, and 1% significance levels, respectively.

Table 16
Definitions of variables

Panel A. Traditional diversity measures	
Gender Diversity	Percentage of female board members.
Ethnic Diversity	Percentage of board members classified under racial/ethnicity minority groups.
Cultural Diversity	Percentage of board members with a cultural background different from the location of the corporate headquarters.
Skills Diversity	Percentage of board members with neither an industry-specific background nor a strong financial background.
Panel B. Firm performance measures	
Tobin's Q	Sum of market value of equity and book value of liabilities, divided by the book value of assets.
Return on Assets (ROA)	Net income divided by book value of total assets.
Return on Equity (ROE)	Net income divided by book value of equity.
Panel C. Control variables and firm characteristics	
Total Assets	Book value of total assets.
Market Capitalization	Total market value of the company.
Net Income	Net income after preferred dividends.
Book to Market	Total shareholders' equity divided by market value of the company.
Firm Age	Firm age from the date of incorporation.
Cash	Sum of cash and short term investments divided by book value of total assets.
Dividend Yield	Total common and preferred dividends paid divided by market value of the company.
Leverage	Book value of liabilities (calculated as book value of total assets minus total shareholders' equity) divided by book value of total assets.
Number of Employees	Number of full-time and part-time employees in the company.
Board Size	Number of board members.

This table provides definitions for all variables used. The data is retrieved from LSEG Workspace and Datastream.

Table 17
Firm-year observations across GICS industries

GICS industry	Number of firm-year observations
Financials	209
Health Care	176
Information Technology	165
Consumer Staples	143
Industrials	143
Energy	132
Communication Services	121
Consumer Discretionary	121
NULL	110
Materials	66
Utilities	55
Real Estate	22

This table reports the number of firm-year observations included in our sample for every GICS industry. The sample period is from 2013 to 2022.