Selecting Directors Using Machine Learning

By ISIL EREL, LÉA H. STERN, CHENHAO TAN, AND MICHAEL S. WEISBACH[*]

Can an algorithm assist firms in their nominating decisions of corporate directors? We construct algorithms tasked with making out-of-sample predictions of director performance. We run tests of the quality of these predictions and show that directors predicted to do poorly indeed do poorly compared to a realistic pool of candidates. Predictably unpopular directors are more likely to be male, have held more directorships, have fewer qualifications, and larger networks than the directors the algorithm recommends. Machine learning holds promise for understanding the process by which governance structures are chosen, and has potential to help firms improve their governance.

(JEL C10, C45, G30, M12, M51)

1. Introduction

A company's board of directors is legally responsible for managing the company. In principle, the board of directors reports to the shareholders and represents their interests. In practice, however, there is much variation in director quality and the extent to which they serve shareholders' interests.[1]

Many of the concerns about boards come from the director selection process, which has been a source of debate since at least Berle and Means (1932).[2] Despite

[*] Stern: Foster School of Business, University of Washington (leastern@uw.edu); Erel: Fisher College of Business, Ohio State University (erel@fisher.osu.edu); Tan: Department of Computer Science, University of Colorado, Boulder (chenhao.tan@colorado.edu); Weisbach: Fisher College of Business, Ohio State University (weisbach.2@osu.edu). We thank Renée Adams (who also graciously shared data), Lucian Bebchuk, Philip Bond, Lisa Cook, Ran Duchin, Shan Ge, Jarrad Harford, Ben Hermalin, Joan MacLeod Heminway, Nadya Malenko, Miriam Schwartz-Ziv, Anil Shivdasani, Tracy Yue Wang, and conference and seminar participants at North Carolina, Northeastern, Ohio State, Singapore, Tennessee, Washington, 2017 Pacific Northwest Finance Conference, 2017 WAPFIN Conference at NYU Stern, 2017 NABE TEC Conference, 2018 University of Miami-AFFECT conference, 2018 Drexel Corporate Governance Conference, 2018 ICWSM BOD workshop, the 2018 NBER Economics of AI Conference and the 2018 Wash. U Corporate Finance Conference. Special thanks to Ronan Le Bras for providing invaluable help throughout the project.

[1] See Hermalin and Weisbach (2003), Adams, Hermalin and Weisbach (2010), and Adams (2017) for surveys.

[2] Berle and Means (1932) wrote: "Control will tend to be in the hands of those who select the proxy committee and by whom the election of directors for the ensuing period will be made. Since the proxy committee is appointed by the existing management, the latter can virtually dictate their own successors" (p. 87). Hermalin and Weisbach (1998) present a formal model of this process in which boards vary in their independence from the CEO in equilibrium.

the checks and balances built into a public corporation's governance system, the CEO often controls the selection of new directors.[3] In practice, appointed directors are almost always supporters of the CEO and his policies. Aside from occasional proxy contests, shareholders have virtually no control over the choice of the directors whose mandate is to represent their interests.

We consider a potential alternative approach to select directors: one that uses algorithms that rely on data on firms, potential directors, and their attributes, to identify the quality of directors being considered for a given firm's board. We take advantage of advances in machine learning that have revolutionized many fields and have led to innovations ranging from self-driving cars to facial recognition. In the social sciences, machine learning has great potential for prediction problems such as the one we consider here, the way in which one determines which potential director would be the best for a particular firm. While "traditional" econometrics is typically designed for estimating structural parameters and drawing causal inferences, machine learning is substantially better at making predictions, in part because it does not impose unnecessary structure on the data.[4]

We construct a large database of publicly traded U.S. firms and directors appointed between 2000 and 2014. We build several machine learning algorithms designed to predict director performance using director and firm level data available to the nominating committee at the time of the nominating decision. We compare the algorithms' selections to the directors who were actually chosen by firms. The discrepancies between firms' actual choices of directors and the choices based on the predictions from our algorithms allow us to characterize which individual features are overrated by decision makers. Our goal in this paper is to use the

[3] See Shivdasani and Yermack (1999) and Kramarz and Thesmar (2013) for anecdotal evidence suggesting that the CEO typically holds a veto power over the choice of directors. See also Cai, Nguyen, and Walkling (2017), who document that more complex firms and firms in more competitive environments are more likely to appoint directors who are connected to the CEO or the existing board.

[4] See Athey and Imbens (2017) and Mullainathan and Spiess (2017).

algorithm's predictions as a diagnostic tool to shed light on the decision-making process behind the selection of directors.

A crucial element of any algorithm designed to select valuable directors is a process for assessing a director's performance in a particular firm. The task of measuring the performance of an individual director is challenging since directors generally act collectively on the board and it is usually impossible for a researcher to ascertain the actions of any particular director. Nevertheless, as Hart and Zingales (2017) emphasize, directors' fiduciary duty is to represent the interests of the firm's shareholders and therefore their popularity among shareholders is a natural metric for evaluating them. Our main results are based on two measures that reflect shareholders' support, which reflect her quality and performance. Our results however are similar when we use alternative performance measures, such as firm profitability or announcement returns around the director's appointment.

We recognize that investors often have limited resources and vote based on simple, check-the-box criteria in routine director elections. However, an emerging literature on director elections documents that the level of shareholder support received by a director is positively related to measures of director performance and shareholder votes do appear to capture shareholders' discontent when present (see Cai et al. (2009), Fischer et al. (2009), Iliev et al. (2015), Aggarwal et al. (2017), and Ertimur et al. (2017)). To adjust for firm-specific factors and to measure shareholder support of an individual director, we use the excess votes (relative to the slate of directors) a *new* independent director receives in *subsequent* elections as a market-based measure of an individual director's performance. We call this measure "excess votes". All results are qualitatively unchanged if we use the fraction of votes in favor and do not subtract the slate's average.

We construct algorithms tasked with predicting the performance of any potential director at any particular company. These algorithms are common in the computer science literature (i.e., lasso, ridge, neural networks, and gradient boosting trees). On our sample of public firms, we train each algorithm (i.e. fit a model) on a "training" set (directors appointed between 2000 and 2011), and then compare the predictions to the observed data out-of-sample using a "test" set (directors appointed between 2012 and 2014).

We find that these algorithms make accurate out-of-sample predictions of shareholder support in director elections, whether predicting the level of shareholder support or the excess support relative to the slate The directors the algorithm predicted would do poorly did much worse on average than the directors the algorithm predicted would do well. In comparison, the directors predicted to do poorly by an OLS model do not actually have worse performance out of sample than those the OLS model predicted would do well. The machine learning algorithms can predict the level of shareholder support new directors will receive out of sample, while the OLS model cannot. Machine learning algorithms, by letting the data speak about the underlying relationships among the variables, ends up fitting the data much better and consequently does better at predicting future outcomes out of sample.

From a decision maker's perspective, it is also necessary to predict the performance of potential alternative directors. However, we only observe the subsequent votes for directors who were actually nominated to the board but do not observe subsequent votes for potential candidates who were *not* nominated. This *"selective labels"* problem of having voting data at the company in question only for directors who were actually selected is a common issue in prediction problems (see Kleinberg et al. (2017)). In addition, if decision makers consider features that are not observable to our algorithm in their nominating decisions of directors, the distribution of outcomes in the set with observed labels (nominated directors) could differ from that in the set with missing labels (not nominated directors), even if they share exactly the same observable characteristics. In other words, if boards are skilled at using unobservables in their nominating decisions, nominated directors could have higher expected performance than otherwise similar (based on observables) passed-over directors.

To determine whether our algorithm could be of use to decision makers, we wish to compare the performance of potential alternative directors to that of the individuals who actually did join the company's board. Our empirical strategy to deal with the selective labels and unobservables issues involves matching each new board appointment to its own realistic pool of potential candidates. Each pool consists of directors who joined the board of a smaller neighboring company in the past year or the following year. Presumably these potential candidates would have found the opportunity to be on the board of a larger nearby company attractive, since directorships at larger companies tend to be better paying and more prestigious than directorships at smaller companies. Directors in this pool were willing to travel to this specific location for board meetings. In addition, our results are similar when we further restrict realistic candidates to those who joined the board of a firm in the same industry. We use the machine learning model to predict the performance of each potential director the firm could nominate using the firm and board characteristics (including committee assignments) of the focal firm.

Although we do not observe the performance (*label*) of potential candidates that were not nominated to the focal board, the design of our candidate pools allows us to observe what we refer to as their *"quasi-label"*: their performance on the board they actually joined.

Many (if not most) prediction problems in the social sciences are subject to the issue of selective labels and reliance on unobservables, introduced and formalized in Kleinberg et al. (2017). This makes the evaluation of a prediction algorithm challenging. In settings in which quasi-labels are available though (i.e. a wide range of prediction problems in the social sciences), they provide a powerful tool to evaluate prediction algorithms. Quasi-labels represent intuitive, best estimate labels for observations with unobserved outcomes. Importantly, quasi-labels do not need to be *perfect* substitutes to labels in order to fulfil their role of assessing how well the algorithms make predictions when labels are missing and the decision maker relies on unobservables.

As highlighted in Kleinberg et al. (2017), computer scientists typically disregard the challenges raised in prediction problems characterized by selective labels and unobservables as their focus is exclusively on prediction accuracy. In contrast, social scientists are likely interested in the extent to which an algorithmic decision aid could improve on human decisions. In the context of judges' bail decisions, the authors design a very clever methodology to circumvent those challenges. Our setting differs significantly from theirs, which leads us to contribute to the literature with a different empirical strategy, one based on quasi-labels. We provide a detailed discussion of this strategy in Section 4.3.

The field of computer science has developed a set of tools, commonly referred to as machine learning, that expand the scope of research questions in the social sciences (see Mullainathan and Spiess, 2017). We expect that as social scientists tackle more prediction problems, the use of quasi-labels designed to evaluate prediction algorithms will increase, as they can be applied in many settings. We hope our present work encourages researchers in that direction.

An important issue in interpreting the results is whether shareholder votes reflect directors' quality rather than their popularity with shareholders. While these notions are to some extent the same since a director's duty is to serve the interests of shareholders, it is possible that the algorithm is picking up the recommendations of shareholder services like ISS or shareholders' biases and views on certain director characteristics (e.g. gender), rather than the true interests of shareholders.

Several pieces of evidence reassure us that this is not the case. First, ISS introduced guidelines to vote in favor of proposals aimed at increasing gender diversity in 2010. Less than 20% of appointments in our training set occur when ISS had those guidelines in place. Second, Iliev and Lowry (2014) show that institutional investors with larger size of ownership vote independently from ISS recommendations. We repeat all our tests by focusing on a subsample of firms with larger-than-median (26%) ownership by the top-5 institutional owners. Our results are unchanged. Third, we compute the mean cumulative abnormal returns (CAR) around the announcement of director appointments in our test set. We find that directors predicted to do poorly by the algorithm have a mean CAR that is significantly lower than that of directors predicted to do well. We also train our algorithms on a sample for which the CAR around the appointment announcement is available, with similar results. Finally, we train the algorithm using firm profitability following director appointments as an alternative measure of performance. While this measure reflects the collective decisions of all management rather than the individual directors, it still reflects the ability of the directors to some extent. Importantly, we find that selecting directors based on predictions of the level of shareholder support does not come at the expense of lower profitability. On the contrary, we show that directors whose subsequent shareholder votes are predictably poor are associated with significantly lower subsequent firm profitability than that for directors with predictably high shareholder votes. Put together, those results strongly suggest that unpopular directors are indeed worse directors.

While machine learning models do not generate estimates of the underlying structural parameters of a model, we can use the algorithm's predictions to understand the characteristics that the model indicates are overvalued and undervalued by firms when they choose directors. A striking result in this paper is that machine-learning models consistently suggest directors who would have been likely both to accept the directorship and to outperform the directors that are actually chosen by firms. Relative to algorithm-selected directors, firm-selected directors who receive predictably low shareholder approval are more likely to be male, have larger networks, sit on more boards, and are more likely to have a finance background. These attributes characterize the stereotypical director in most large companies. A plausible interpretation of our results is that firms that nominate predictably unpopular directors tend to choose directors who are like existing directors, while the algorithm suggests that adding diversity would be a better idea.

Machine-learning tools have the potential to help answer many unanswered questions in the social sciences, both by academics wishing to understand the way the world actually works[5], and by practitioners and policy makers wishing to make better real-world decisions. In terms of boards of directors, an algorithmic decision aid could allow firms to choose better among existing candidates, without stripping decision makers of their judgement. We emphasize that algorithms complement human judgement rather than provide a substitute for judgement. As such, simple economics predict that the economic value of the "output" of board decisions increases with the use of algorithmic decision aids (see Autor, 2015 and Agrawal. Gans and Goldfarb, 2017) In addition, algorithmic decision aids could help firms identify alternative choices of potential directors, thereby opening up board seats to a broader set of candidates with more diverse backgrounds and experiences, who would have otherwise been overlooked.[6]

This paper is organized as follows. The next section describes the machine-learning algorithms we use and develops a framework that helps us assess the performance of these algorithms. In the third section, we present our data and summary statistics. In the fourth section, we present the performance of our prediction models and compare firms' actual choices of directors with potential alternatives. Section 5 compares the characteristics of model-selected directors with those of directors actually chosen by firms to evaluate the types of directors that tend to be overvalued in the decision-making process. The final section includes an

[5] Li et al. (2018) use machine learning (word embedding) to measure corporate culture.

[6] We thank Oren Etzioni, CEO of the Allen Institute for Artificial Intelligence, for originally pointing this out to us.

extensive discussion of our approach and findings, puts them in perspective, discusses possible extensions, and concludes.

2. Using Machine Learning to Predict Director Performance

2.1. Algorithms to Predict Performance

We build algorithms designed to make an ex ante prediction of directors' level of shareholder support, over the first three years of their tenure. The algorithms use a set of observable director, board, and firm features that are available to the nominating committee at the time of the nominating decision. The algorithms are among the most commonly used in the machine learning literature: *lasso*, *ridge*, *neural networks* and *gradient boosting trees.* We train each of these algorithms, i.e. estimate model parameters, on directors appointed between 2000 and 2011 and test them on directors appointed between 2012 and 2014. Following the terminology in machine learning, we call the data from 2000-2011 the "training set" (in-sample data) and the data from 2012 to 2014 the "test set" (out-of-sample data).

The variable the algorithms try to predict in the reported results is the average level of shareholder support over the first three years of director tenure, adjusted each year by the average support for the entire slate of directors. The data on director elections is from ISS Voting Analytics. We use the average of available shareholder support over the first three years of tenure.[7]

There are a number of well-known machine learning algorithms that can be used for our prediction exercise. We use four of these algorithms to predict director performance, and give a brief summary of each in this section.

2.1.1. *Lasso* and *Ridge*

[7] The distribution of shareholder support does not change over the first few years of a director's tenure. We obtain similar results using shareholder support at year one, year two or year three instead of using the average over the first three years.

OLS regressions tend to generate poor out-of-sample predictions as they are designed to minimize the in-sample residual sum of squares. This observation is known as bias-variance tradeoff in the machine learning literature: if an algorithm fits in-sample data too well (low bias), it has high variance and thus does not perform as well on out-of-sample data. *Lasso* and *ridge* are both linear models that use a regularization term to achieve a balance between bias and variance. They do so by minimizing a loss function that includes in-sample fit and a penalty term that favors simple models, thereby reducing variance (see online appendix for more details).

2.1.2. *Gradient Boosting Trees*

Gradient Boosting Trees are similar to random forest algorithms. A *random forest* algorithm is an ensemble method that combines multiple decision trees. Intuitively, a single decision tree presents a flow chart where a data point can follow the flow starting from the root to a leaf node associated with its final prediction. The selection of attributes at each node in decision trees is inspired by information theory to maximize information gain. In the *random forest* algorithm, multiple trees are estimated by using a random subset of covariates in each tree. Among those, the covariate that provides the best binary split based on information gain is used to split the data into two partitions and functions as the root of the tree. The algorithm repeats this process until it reaches the bottom of the tree, where each "leaf" or terminal node is comprised of similar observations. Then, a new data point can start at the top of each tree and follow the splits at each node all the way to a leaf node. The prediction for this new data point is the average outcome of observations in the leaf it ends up in. The random forest algorithm takes an average of the predictions from all the decision trees.

Similar to the random forest algorithm, the *gradient boosting trees* algorithm is an ensemble method that combines multiple trees. The key difference lies in that the final prediction is a linear sum of all trees and the goal of each tree is to minimize the residual error of previous trees. The *XGBoost* algorithm provides an efficient implementation of this algorithm that is scalable in all scenarios (Chen and Guestrin, 2016). In the rest of the paper, we use *XGBoost* and *gradient boosting trees* interchangeably.

2.1.3. *Neural Networks*

Loosely speaking, artificial neural networks are inspired by the way the brain processes information. It is structured in layers of neurons connected by synapses. The first layer comprises the input neurons and the final layer represents the output. Layers of neurons between the first and final layers are hidden layers. The figure in the online appendix depicts the structure of a basic neural network with two hidden layers. Neurons x_i are input neurons connected to the next layer of neurons by synapses which carry weights w^1. Each synapse carries its own weight. An activation function (usually a sigmoid to allow for non-linear patterns) is embedded in each neuron in the hidden layers to evaluate its inputs. The set of weights carried by the synapses that reach a neuron are fed into its activation function, which will determine whether or not that neuron is activated. If activated, it triggers the next layer of neurons with the value it was assigned, with weight w^2 (again with each synapse carrying its own weight). Similar to the neurons in the hidden layers, the output neuron judges its input via an activation function and decides from which neurons to accept the triggered values. The output is a weighted sum of the activated neurons in the last hidden layer. Training a network involves modifying the weights on the synapses to minimize a cost function (e.g. the sum of squared errors).

2.2. Assessing Algorithms' Predictions

Assessing whether the algorithmic predictions can actually lead to better outcomes is not a straightforward task. We cannot simply compare the predictions to the actual outcomes in the test set as is typically done in most machine learning applications because of two important challenges. First, we can only observe our algorithm's predictions for directors who are actually appointed to that position (the *selective labels* problem). Second, when deciding on their choice of directors, decision makers presumably take factors into account that are not observable to the algorithm. Therefore, directors who were nominated, although they could share the same observable features as potential alternatives, could differ in terms of unobservables. In particular, they could have been chosen because they have a set of skills that are valuable to the firm, or because they have a personal relationship with the CEO or existing directors. A firm could also have decided not to nominate a candidate based on some characteristics unobservable to the algorithm that would make this candidate a poor choice. We cannot observe these factors, yet they could lead to different average outcomes for nominated vs. not nominated, even if both are identical on the basis of observable characteristics.

To formalize these concepts, we develop a framework similar to that presented by Kleinberg et al. (2017) and present it in the online appendix. Our empirical strategy to address these concerns involves designing a pool of realistic potential candidates for each vacant board position. We wish to evaluate the algorithm's predictions of the performance of directors who were actually nominated. In cases where our algorithm predicted low performance, for example, we are interested in whether there were plausible alternatives available, how they would have performed, and how the director who was nominated actually performed compared to those alternatives. In Section 4.3 we explain how we construct the candidate pools and present results of this comparison between actual choices of directors and potential candidates. We first describe below our sample of directors of focal firms and our algorithms' predictions of director performance.

3. Constructing a Sample on which Algorithms Can Select Directors

3.1. Measuring Director Performance Through Election Results

A challenging part of designing an algorithm to select directors is the way in which the algorithm measures director performance. Most actions that directors take are done collectively with other directors in the privacy of the boardroom, making it harder to assess the performance of a given director. Also, for an outside observer or an algorithm to assess the performance of an individual director, it should rely on a measure that incorporates the information available to market participants. We use the shareholder support that an individual receives in director elections as a market-based measure of individual directors' performance.

An important feature of director elections is that the vast majority of the time, directors receive overwhelming majorities of the vote, with most studies reporting a mean vote of around 95% in favor of the directors. Therefore, there is virtually no variation in the *outcome* of the elections. If the election results reflect the market's perception of a director's quality, it must be that variation among winning votes contains meaningful differences in the market's assessment. Consistent with this notion, Cai et al. (2009), Fischer et al. (2009), and Iliev et al. (2015) suggest that cross-sectional variation in shareholder support does in fact reflect market perceptions of director quality. These papers find that vote totals predict stock price reactions to subsequent turnover. In addition, vote totals are negatively related to CEO turnover, board turnover, management compensation levels, and the probabilities of removing poison pills and classified boards.

In addition, director re-elections appear to have real consequences, even if the elections are not contested and the nominated directors end up being re-elected. Fos et al. (2017) find that when directors are closer to elections, they are more likely to fire CEOs, presumably to persuade shareholders that they are being more diligent. Aggarwal et al. (2017) suggest that directors with low relative support are more likely to leave the board, and if they stay, tend to move to less prominent positions. Finally, Ertimur et al. (2017) find that when votes are withheld from directors, boards explicitly attempt to address shareholders' concerns. Overall, the recent literature strongly suggests that shareholder support does reflect perceptions of director quality, that directors care about these perceptions, and take actions designed to influence them. Furthermore, while many papers show the influence of recommendations by proxy advisory firms (e.g., ISS) on voting by institutional investors on various governance proposals, including director elections,[8] recent research provides evidence on the decline in this influence.[9]

3.2. *Sample Selection*

To evaluate the performance of an algorithm to select directors, we must gather a sample in which we can observe the attributes of firms and boards, and also for which we can measure the performance of directors. Because of these requirements, we analyze a sample of boards from large, publicly-traded, U.S. firms with an average market capitalization of $6.6 billion. We identify 41,015 new independent directors appointed to 4,887 unique corporate boards of these firms between 2000 and 2014 using *BoardEx,* which is our main data source for director and board-level characteristics.

We obtain data on the level of shareholder support for individual directors from *ISS Voting Analytics* and focus on new directors only. To construct *excess votes*, the measure of performance used in reported results, we use the number of votes in favor over all votes cast (yes, no, withheld). We then subtract the average for the entire slate of directors and take the average over the first three years of tenure. Our sample contains the voting outcome, i.e. the *excess votes*, for 24,054 new director appointments. All our results are similar when we use shareholder support and do not subtract the support for other directors on the slate.

3.3. *Summary Statistics*

---

[8] See, for example, Cai, Garner, and Walkling (2009), Daines, Gow and Larcker (2010), Alexander, Chen, Seppi, and Spatt (2010), Ertimur, Ferri and Oesch (2015), Larcker, McCall and Ormazabal (2015), Malenko and Shen (2016), and Ertimur, Ferri and Oesch (2017).

[9] For example, Iliev and Lowry (2014) find that mutual funds vary greatly in their reliance on ISS recommendations. Aggarwal, Erel, and Starks (2016) show that investor voting has become more independent of ISS recommendations in shareholder proposals where ISS recommends a vote against the proposal. A recent striking example of investors choosing to dismiss the recommendation of the lead proxy advisors is when ADP shareholders voted to reelect all incumbent board members in a proxy fight against activist investor William Ackman. All three main proxy advisors had recommended shareholders to oppose the reelection of ADP's directors. See https://www.nytimes.com/2017/11/07/business/dealbook/adp-ackman.html.

Table 1 presents summary statistics for average shareholder support over the first three years of tenure. As previously documented in the literature on uncontested director elections, the overall level of shareholder support is typically very high. Given that the mean level of support is .95 and the median is .975 (with a standard deviation of .07), a voting outcome below 95% is a relatively poor outcome. Consequently, a voting outcome below 95% likely reflects a perception of poor performance by the director. Starting in Column 4, we report shareholder support after subtracting the average support for the entire slate of directors in that year. Although shareholder support in uncontested elections is typically very high, shareholders do on occasion oppose newly nominated directors (see figure in online appendix). The question is whether an algorithm can pick up signals in the data that can reliably predict which directors will ultimately fall in that left tail. Our results show that the answer is yes.

	n	mean total votes	median total votes	mean excess votes	std excess votes	25th ptcl excess votes	median excess votes	75th pctl excess votes
2000	331	0.950	0.974	0.0008	0.0300	-0.0058	0.0004	0.0082
2001	772	0.944	0.970	-0.0001	0.0455	-0.0050	0.0017	0.0134
2002	1,057	0.946	0.970	0.0022	0.0387	-0.0038	0.0015	0.0115
2003	1,774	0.951	0.974	0.0064	0.0359	-0.0014	0.0028	0.0149
2004	2,019	0.953	0.977	0.0069	0.0442	-0.0008	0.0033	0.0153
2005	1,893	0.948	0.974	0.0049	0.0369	-0.0011	0.0033	0.0136
2006	1,789	0.941	0.969	0.0051	0.0412	-0.0016	0.0036	0.0153
2007	1,942	0.940	0.971	0.0045	0.0434	-0.0023	0.0026	0.0157
2008	1,691	0.944	0.973	0.0067	0.0431	-0.0032	0.0034	0.0180
2009	1,541	0.948	0.976	0.0072	0.0435	-0.0020	0.0045	0.0187
2010	1,842	0.948	0.977	0.0039	0.0431	-0.0044	0.0027	0.0152
2011	1,825	0.954	0.981	0.0038	0.0462	-0.0019	0.0035	0.0160
2012	1,862	0.952	0.981	0.0045	0.0422	-0.0007	0.0038	0.0162
2013	2,148	0.948	0.980	0.0027	0.0444	-0.0021	0.0032	0.0139
2014	1,568	0.959	0.985	0.0063	0.0408	-0.0004	0.0045	0.0149
	24,054	0.9484	0.9755	0.0044	0.0413	-0.0024	0.0030	0.0147

TABLE 1: SHAREHOLDER SUPPORT SUMMARY STATISTICS

Notes: This table presents summary statistics for total (columns labeled as *mean/median total votes*) and excess shareholder support over time. Shareholder support is defined as the fraction of votes in favor of a given director over all votes cast for the director's reelection within three years of her tenure. To compute *Excess Votes*, we subtract the average of that variable for the slate of directors up for reelection that year on the focal board. Then we take the average of this relative vote measure over the first three years of the new director's tenure. The data is from ISS Voting Analytics.

Table 2 illustrates that the frequency of shareholder discontent varies by director and board characteristics. For example, the fraction "bad outcomes", representing the bottom 10% of the sample in terms of excess votes, is 10.6% for male directors and 7.9% for female directors. Similarly, busy directors (serving on three or more boards) experience low shareholder support more frequently than non-busy directors. However, theory provides little guidance about the particular variables and functional forms of the relation between the various director, board and firm characteristics and the performance of directors. For example, we do not know whether we should expect female busy directors serving on a large board to receive higher or lower shareholder support on average than a male director who serves on a single small board. The problem increases in complexity when many more covariates are likely to matter. For this reason, we rely on an estimation procedure that does not impose the specific form for the relationship between potential explanatory variables. Machine learning algorithms therefore provide a disciplined and rigorous approach to model selection (Athey, 2017).

	Full sample	yes	no	Difference p-value
Director level				
Male	0.102	0.106	0.079	0.000
Foreign	0.101	0.115	0.100	0.138
Qualifications > median	0.102	0.094	0.106	0.005
Network size > median	0.102	0.108	0.096	0.002
Generation BBB	0.101	0.093	0.118	0.000
Generation X	0.101	0.151	0.096	0.000
Busy director	0.102	0.145	0.090	0.000
Finance background	0.102	0.106	0.101	0.328
Board level				
Fraction male > median	0.102	0.116	0.091	0.000
Board size > median	0.102	0.089	0.114	0.000
Nationality mix > median	0.102	0.108	0.100	0.064
Attrition rate > median	0.098	0.106	0.086	0.000

TABLE 2: AVERAGE FRACTION OF BAD OUTCOME

Notes: This table presents the average fraction of "bad outcome". A director is considered to experience a bad outcome if her excess votes is < -2%. Bad outcomes represent 10% of the sample. Shareholder support is defined as the fraction of votes for a director over all votes cast for that director. Shareholder discontent is presented for various director-level and board-level characteristics.

4. Evaluating Machine Learning Predictions of Director Performance

4.1. Model Specification

We develop machine-learning algorithms that predict the quality of a potential director, using the subsequent voting as a measure of a director's quality. There are 24,054 appointments for which we observe the vote outcome and can compute *excess votes*. We first "train" each algorithm on the 2000-2011 portion of our sample containing 18,476 new independent director appointments, of which 12,815 are unique, at 2,407 firms. Training involves having the algorithm determine which combinations of variables best predict future performance.[10] We evaluate the models' out-of-sample predictions on the held out 2012-2014 portion of our sample containing 5,578 new director appointments, of which 4,019 are unique directors, at 569 firms. We compare those out-of-sample predictions to those from an OLS model. All comparisons are based on predictions for the 2012-2014 subsample of director appointments, which does not overlap with the 2000-2011 subsample on which the algorithms are trained.

4.2. Predictions of Director Performance

Table 3 summarizes the ability of the machine learning models, once trained on the earlier portion of the sample, to predict director success in the later part.[11]

		Average Observed Performance for Directors in a Given Percentile of Predicted Performance as Predicted by:				
	Predicted Percentile of Excess Votes	OLS	XGBoost	Ridge	Lasso	Neural Network
Directors predicted to perform **poorly**	1%	0.028	-0.031	-0.012	-0.024	-0.014
	5%	-0.018	-0.014	-0.013	-0.015	-0.010
	10%	0.014	-0.008	0.000	-0.008	-0.001
Directors predicted to perform **well**	90%	0.013	0.013	0.011	0.011	0.011
	95%	0.007	0.012	0.014	0.013	0.016
	100%	0.006	0.011	0.009	0.016	0.015

TABLE 3: OLS VS. MACHINE LEARNING TO PREDICT DIRECTOR PERFORMANCE

[10] The algorithms rely on a regularizer that balances out in-sample fit and out-of-sample overfitting.
[11] See online appendix for details on the OLS model used. Using alternative models, for example, without fixed effects, lead to similar results in terms of OLS's performance in predicting director performance.

Notes: This table reports the average observed level of excess shareholder support over the first three years of a new director's tenure for directors who were ranked by their predicted level of shareholder support by an OLS model and several machine-learning algorithms (XGBoost, Ridge, Lasso and Neural Network). Shareholder support is defined as the fraction of votes in favor of a given director over all votes cast for the director's reelection within three years of her tenure. To compute *Excess Votes*, we subtract the average of that variable for the slate of directors up for reelection that year on the focal board. Then we take the average of this relative vote measure over the first three years of the new director's tenure.

A simple test of a model for predicting performance is whether actual performance is an increasing function of predicted performance. Table 3 indicates that average observed shareholder support almost monotonically increases across model predicted performance percentiles for each machine learning model but not for the OLS one. In contrast to the machine learning models, the average observed outcome of directors in the bottom of the predicted performance distribution using the OLS model (.028) is actually higher than that of directors in the top of the predicted performance distribution (.006).

Among the machine learning algorithms, *XGBoost* and *lasso* perform best at predicting the subsequent success of directors using excess votes as a measure of director performance. Using total votes (not adjusted for the mean vote in the slate) as a performance measure, *XGBoost* performs better.[12, 13] Directors predicted to be in the bottom percentile as predicted by *XGBoost* have an average observed excess shareholder support of -3.1%, whereas the average observed support is 1.1% for directors in the top percentile of predicted performance. This pattern highlights the difference between the machine learning model and OLS in their ability to predict future performance.

Figure 1 shows the average observed level of shareholder support for directors across the ten deciles of predicted performance for OLS and for the machine learning algorithms in the 2012-14 test period. The figure shows how the mean shareholder support for a director is an increasing function of the predicted one for all the machine learning algorithms, but not for the OLS model. The difference in

[12] These results are not reported but are available from the authors upon request.
[13] *XGBoost* is an algorithm with a reputation for generating excellent predictions on a variety of problems, and was the most often used algorithm among the winning solutions in the 2015 machine learning Kaggle competition.

the predictive ability of various models illustrates the difference between standard econometric approaches and machine learning. OLS fits the data well in sample but poorly out of sample. In contrast, machine learning algorithms are specifically designed to predict well out of sample.

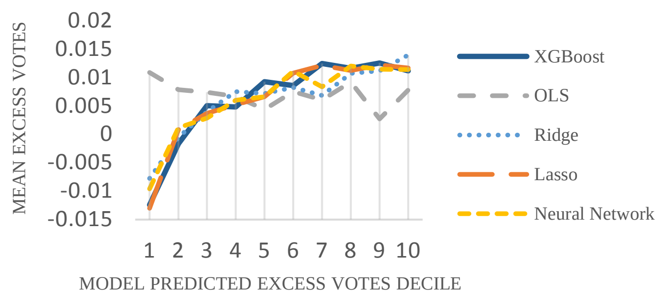


FIGURE 1: MEAN OBSERVED *EXCESS VOTES* VS. PREDICTED *EXCESS VOTES*

Notes: This figure shows the average observed level of excess shareholder support for directors across the ten deciles of predicted performance for OLS and *XGBoost* in the 2012-14 test set. To compute *excess votes*, we first compute the fraction of votes in favor of a given director over all votes cast for the director. Next, we subtract the average of that variable for the slate of directors up for reelection that year on the focal board. Finally, we take the average of this relative vote measure over the first three years of the new director's tenure.

The fact that machine learning models perform substantially better than OLS at predicting director performance out of sample is consistent with the arguments of Athey and Imbens (2017) and Mullainathan and Spiess (2017), who emphasize that machine learning should be the preferred approach for prediction problems such as this one. One possible reason why the machine learning models do much better is because they let the data decide which transformations of which variables are relevant, while in OLS (or other standard econometric technique), the researcher must specify the structure of the equation before estimating it. Machine learning, by letting the data speak about the underlying relationships among the variables, ends up fitting the data better and consequently does better at predicting outcomes out of sample.

4.3. Comparing Firms' Actual Choices of Directors with Potential Alternatives

The results suggest that directors identified by our algorithm as likely to have low (high) future shareholder support, are in fact more likely to have low (high) support in subsequent elections. Accurate out of sample predictions however are not sufficient to argue that our algorithm could assist firms in their nominating decisions of corporate directors. The selective labels problem and reliance on unobservables by the decision maker make it challenging to evaluate the prediction algorithms. Such an evaluation is necessary however in order to assess whether an algorithmic decision aid could improve on the current decision making process. The algorithm should be able to assess how the nominated directors performed compared to how alternatives would have performed. Our empirical strategy revolves around the creation of realistic candidate pools and the introduction of the quasi-labels method of evaluation.

To construct realistic candidate pools, we consider directors who joined the board of a neighboring company around the same time.[14] These directors were available to join a board at that time and were willing to travel to that specific location for board meetings. Furthermore, to alleviate concerns related to the ability of a particular firm to attract promising directors, we restrict the pool of potential candidates to directors who joined a *smaller* neighboring company within a year around the focal appointment, since the prestige and remuneration of being a director tends to increase with company size (see Masulis and Mobbs, 2014). There are on average 147 candidates in a candidate pool. All our results are similar if we further restrict the set of candidates in candidate pools to directors who join the board of a firm in the same industry.[15]

Although we do not observe the performance of these potential candidates, we do observe what we refer to as a *"quasi-label"*, which is an informative estimate that substitutes for labels. In our setting, a quasi-label is a director's performance on the board he or she actually joined. Quasi-labels represent an *indication* of how

[14] A neighboring firm is a defined as a firm whose headquarters is within 100 miles of the focal firm's headquarters. The average distance with the focal board is 35 miles (median distance is 26 miles).
[15] There are on average 33 candidates in these more restrictive candidate pools.

potential candidates would have performed on the focal board. Recall that the algorithm can generate a prediction of performance for *any* candidate, not only those in candidate pools. Candidate pools and quasi-labels are simply part of our strategy to assess the quality of the predictions when labels are missing. To generate predictions for potential candidates at the focal firm, our algorithms use the board and firm characteristics as well as the committee assignments of the appointment at the focal firm with the individual potential candidate's features.

We provide a schematic representation of the quasi-label procedure in Figure 2. First we rank all nominated directors in our test set according to their predicted performance. Then, for each of the nominated directors in the bottom decile of predicted performance, we consider their associated candidate pool and rank candidates in this candidate pool according to their predicted performance *on the focal board*. We retain the top decile of candidates, who are the most promising candidates based on our algorithms' predictions. We then re-rank these promising candidates according to their quasi-labels. We then compare the observed performance of the nominated director on the focal board to the quasi-labels of promising candidates.

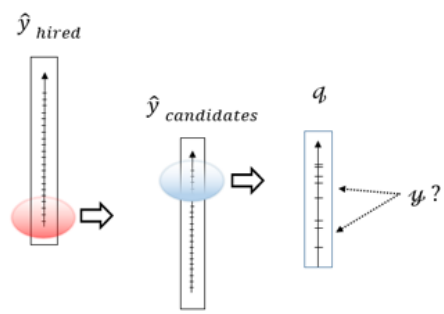


FIGURE 2: ASSESSING THE ALGORITHM'S PREDICTIONS USING QUASI-LABELS

Notes: This figure shows the procedure to evaluate our algorithmic predictions using quasi-labels. We rank all hired directors in our test set according to their predicted performance ($\hat{y}_{hired}$). The bottom decile represents directors who were predicted to receive low shareholder approval. For each of these hired directors, whom our algorithm predicted would be unpopular, we consider their associated candidate pool and rank candidates in this candidate pool according to their predicted performance on the focal board ($\hat{y}_{candidates}$). We retain the top decile of candidates, who are the most promising candidates based on our algorithms' predictions. We then re-rank these promising candidates according to their quasi-labels q, i.e. their performance on the board they actually joined. The goal is then to compare the observed performance of the hired director on the focal board (y) to the quasi-labels of promising candidates.

We are interested in where the actual performance of directors predicted to do poorly actually sits in the distribution of quasi-labels. If the observed performance of the nominated director ranks high in the distribution of quasi-labels, this would suggest that even though our algorithm predicted this particular director would do poorly, she ended up doing well relative to available alternatives. The focal board might have relied on unobservables in the nomination process, and the high rank in the distribution of quasi-labels would suggest that unobservables were used as signal. On the other hand, if the observed performance ranks low in the distribution of quasi-labels, then our algorithm identified ex ante that this director would perform poorly, and relative to alternatives, she indeed did perform poorly. What this would suggest, is that any unobservables used in the nomination decision process was not signal, but either noise, bias, or related to agency problems.

Table 4 presents the median rank in the distribution of quasi-labels for directors in the bottom and top deciles of predicted performance for several machine-learning algorithms, as well as for an OLS model. For all machine-learning models, nominated directors predicted to do well performed noticeably better than available alternative candidates, while nominated directors predicted to do poorly performed worse than available alternative candidates. *XGBoost* and *lasso* again appear to be the preferred algorithms. They can best discriminate *ex ante* the directors who will do well from those who will not. In the rest of the paper, we focus on results with *XGBoost* to simplify the discussion. The median director predicted by the *XGBoost* algorithm to be in the bottom decile of performance ranks at the 27^{th} percentile in the distribution of quasi-labels. The median director predicted to be in the top decile ranks at the 78^{th} percentile in the distribution of quasi-labels. In contrast, the predictions from the OLS model are uninformative about subsequent performance; directors rank around the 75^{th} percentile in the distribution of quasi-labels, regardless of whether OLS predicted they would perform well or not.

	Median percentile of observed performance in the distribution of quasi-labels (candidate pools)				
	OLS	XGBoost	Ridge	Lasso	Neural Network
Bottom decile of predicted performance	77th	27th	37th	23rd	29th
Top decile of predicted performance	75th	78th	82nd	79th	69th

TABLE 4: EVALUATING THE PREDICTIONS USING QUASI-LABELS

Notes: This table reports how nominated directors rank in the distribution of quasi-labels of their candidate pool. For each nominated director in our test set, we construct a pool of potential candidates who could have been considered for the position. Those candidates are directors who accepted to serve on the board of a smaller nearby company within a year before or after the nominated director was appointed. The quasi-label for each of these candidates is how she performed on the competing board she chose to sit on. The first (second) row shows the median percentile of observed performance in the distribution of quasi-labels for directors the model predicted to be in the bottom (top) decile of predicted performance. Each column presents the results from a different model.

So far, we presented results for the top and bottom deciles of predicted performance. However, our results are similar across all deciles of performance and when compared to *all* potential candidates (i.e. not conditioning on the most promising candidates). Figure 3 shows that the median rank (percentile) in the distribution of quasi-labels almost monotonically increases across deciles of predicted performance, These results suggest that machine-learning models can predict, at least to some extent, whether an individual will be successful as a director in a particular firm.

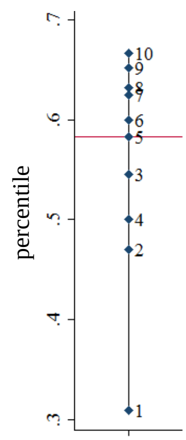


FIGURE 3: MEDIAN RANK IN QUASI-LABEL DISTRIBUTION FOR THE TEN DECILES OF PREDICTED PERFORMANCE

Notes: This figure shows the median rank in the distribution of quasi-labels for directors in each of the ten deciles of *XGBoost* predicted performance (using *Excess votes*). The observed performance of nominated directors in our test set is compared to the quasi-labels of all potential candidates in their respective candidate pool.

We emphasize that board-director matches are not exogenous, and are likely chosen with the intent of maximizing the “fit” between directors and firms. Quasi-labels are not perfect substitutes for labels (the level of support a director would have gathered on the focal board). However, note that the empirical strategy only uses candidate pools and quasi-labels to evaluate the algorithm’s performance in the presence of selective labels.

There potentially exist many settings in which quasi-labels may be used to assess the algorithms’ predictions if they represent a reasonable substitute for missing labels. We posit that under the assumption that the difference between the unobserved missing label and its quasi-label is noise (i.e. is unpredictable), the quasi-label procedure offers a useful approach in various contexts.[16] If there is a concern that there exists a systematic bias in those quasi-labels, i.e. if the difference between the missing label and the quasi-label is systematically positive or systematically negative for all observations, a closer examination of the results may be warranted.[17]

In our setting, the endogenous nature of the board-director match might lead to systematically inflated quasi-labels, i.e. by revealed preferences, the performance of the available candidate would not be as high on the focal board.[18] If quasi-labels are inflated due to the endogenous board-director match, then the difference between the quasi-label and the missing label is expected to be positive. However, we do not expect this difference to differ in any predictable way across deciles of predicted performance. A closer look at our quasi-label procedure results shows that the rank in the distribution of quasi-labels increases almost monotonically across model prediction deciles. Therefore, nominated directors predicted to do

[16] For instance, suppose a researcher wants to evaluate algorithmic predictions of loan performance. Quasi-labels for denied loans may be the loan performance for the firm (or individual) offered a loan by a different institution.
[17] Deriving a formal framework that details the conditions under which the quasi-label procedure is applicable is an interesting avenue for future research.
[18] Note that this would assume that boards and directors are skilled at using unobservables to match.

poorly rank low compared to how other candidates would have performed, but we also find that nominated directors predicted to do well rank *high* compared to alternatives. This symmetry provides reassurance that our results are not affected by inflated quasi-labels.

In sum, focusing on realistic potential candidates for each new board position, our algorithm is able to identify, with reasonable precision, those who will perform well and those who will not. This matters to the extent that this changes the rank ordering of candidate directors. It suggests that our algorithm has the potential to improve on real world boards' nominating decisions. Importantly, this work is a first pass exercise to show the potential of machine learning in shedding light on the quality of boards' nominating decisions.

4.4. Excluding Poorly Performing Firms

A possible concern with this analysis is that the relation between predicted performance and subsequent performance could occur only because of poorly performing firms. A poorly performing firm would likely be less attractive to a director, so it could be that only low ability directors are attracted to poorly performing firms, even if the firms are relatively large and otherwise prestigious. Because of their low ability, these directors would tend to do worse *ex post*.

For this reason, we repeat our analyses omitting firms that experience negative abnormal returns in the year prior to the election. Even without poorly performing firms in the sample, the results are very similar to those reported above. For this reason, it does not appear that the relation between subsequent performance and predicted performance compared to alternative potential directors is driven by poorly performing firms with disgruntled shareholders.

4.5. Director Popularity or Performance?

An important interpretational issue is understanding exactly what the algorithm is predicting. The fiduciary responsibility of directors is to maximize shareholders' welfare, so choosing directors who will receive the most subsequent votes would seem to be a natural approach. One concern, however, is that many institutional shareholders decide on their votes through recommendations of shareholder services companies such as ISS. While in principle the goal of these services is to maximize shareholder welfare, it is possible that they instead follow mechanical rules and that our algorithm simply suggests directors who will do well by these rules. ISS's influence is unlikely to drive our results for several reasons.

First, ISS introduced guidelines in the latter part of our training period. For example, guidelines to support proposals aimed at increasing female board representation were first introduced in 2010. However, our training sample covers data from 2000-2011. Less than 20% of appointments in our training set take place when ISS had those guidelines in place. Second, following Iliev and Lowry (2014) who argue that institutional investors with larger size of ownership vote independently from ISS recommendations, we repeat all our tests by focusing on a subsample of firms with larger-than-median (26%) ownership by the top-5 institutional owners. Our results (available upon request) remain very similar. Third, we compare the cumulative abnormal returns (CARs) around the announcement of director appointments in our test set for directors predicted to do well to those for directors predicted to do poorly.[19] Table 5 reports the mean CARs

	N	Mean	Median
Directors in Decile 1 of predicted performance (excess votes)	292	-1.94%	-0.64%
Directors in Decile 10 of predicted performance (excess votes)	575	0.75%	0.34%
Difference in means (p-value)		**0.0043**	

TABLE 5: CUMULATIVE ABNORMAL RETURNS AROUND APPOINTMENT ANNOUNCEMENTS

Notes: This table reports the mean and median cumulative abnormal returns for directors predicted to do poorly and for directors predicted to do well. Directors predicted to do poorly (well) are directors in decile 1 (decile 10) of predicted performance (excess votes) as predicted by the XGBoost algorithm. The results are shown for appointments in the test set only. The cumulative abnormal returns reported are computed using a (-1; +1) window.

[19] We collect announcement dates from *BoardEx*, *CapitalIQ* and *Lexis-Nexis*.

using a (-1; +1) window around announcements. The same pattern emerges using longer windows as well. Using our *XGBoost* algorithm to predict excess votes, we find that the mean CAR for directors predicted to do poorly (decile 1) in our test set is -1.94% whereas it is +0.75% for directors predicted to do well (decile 10). The difference is statistically significant at the 1% level. Directors predicted to be unpopular also tend to be viewed by the market as worse directors. We also used the algorithm to predict announcement CARs using a smaller sample for which announcement dates are available, with similar results.

Finally, we train an *XGBoost* algorithm to predict a measure of firm profitability, *EBITDA/Total Assets*, three years post appointment. We then sort directors in our test set into deciles of predicted profitability. We report the actual profitability as well as the shareholder support in the first two rows of Table 6.[20]

		1	2	3	4	5	6	7	8	9	10	Decile 10 - 1 *p-value*
Algorithm trained on **profitability**	Average observed profitability	-0.498	-0.064	-0.017	0.017	0.078	0.083	0.113	0.114	0.144	0.205	0.0000
	Average observed shareholder support	0.942	0.946	0.956	0.937	0.957	0.961	0.953	0.954	0.960	0.961	0.0002
	Average observed excess votes	-0.0004	0.002	0.006	0.002	0.006	0.004	0.003	0.005	0.006	0.004	0.0668
Algorithm trained on **excess votes**	Average observed profitability	0.006	-0.017	0.008	0.037	0.052	0.058	0.057	0.083	0.087	0.112	0.0000
	Average observed excess votes	-0.012	-0.002	0.005	0.005	0.009	0.009	0.012	0.012	0.013	0.011	0.0000
Algorithm trained on **total votes**	Average observed profitability	-0.003	-0.032	-0.031	-0.018	0.024	0.029	0.058	0.075	0.086	0.100	0.0000
	Average observed shareholder support	0.920	0.937	0.946	0.948	0.950	0.957	0.957	0.966	0.972	0.977	0.0000

TABLE 6: COMPARING SHAREHOLDER SUPPORT MODELS WITH PROFITABILITY MODELS

Notes: This table reports the actual performance for each decile of *XGBoost*-predicted performance. *XGBoost* is trained to predict firm profitability three years after the director has been appointed (EBITDA/Total Assets), total votes as well as excess votes. The results are for our test set only (out-of-sample performance for directors appointed between 2012-2014).

The model trained to predict profitability in the subsequent period indeed does predict future profitability. The actual profits for the firms sorted into deciles based on expected profits increase monotonically, with average profits increasing with the model's expectation of profitability. Firms that nominated directors in the

[20] The correlation of EBITDA/Total Assets with the shareholder support measure is 0.12 (p-value: 0.000).

bottom decile of predicted performance have an average profitability of -49.8% and in the top decile is 20.5%. What is perhaps more surprising is that even though the model is trained to predict profitability, it also does reasonably well at predicting future shareholder support. Directors predicted to be in the bottom decile of profitability have shareholder support of 94% three years subsequent to the model's training, and directors predicted to be in the top decile have shareholder support of 96%. The difference between the two is statistically significantly different from zero at the 1% level. The model trained on profitability also does reasonably well at predicting *excess votes*. The average *excess votes* is -.0004 for directors in the bottom decile of predicted profitability and it is .004 for those in the top decile (the p-value of the difference is 6.68%).

These results suggest that the choice of training the algorithm on shareholder support in director elections is not crucial for the algorithm to be able to predict director quality. When the model is trained using profitability instead, the pattern of predictions is similar. The algorithm predicts future subsequent support. Since this support is based on the market's perception of a director's contribution to quality, the results are similar when the algorithm is trained on profitability directly. In addition, for the algorithm trained on shareholder support that we discussed above, we consider whether it can also predict future profitability in addition to future shareholder support. We break the sample into deciles based on the algorithm's predictions of *excess votes* and *total votes*, and present average observed *excess votes*, *total votes* as well as the average profitability for each decile. We present these averages in the bottom four rows of Table 6.

As discussed above, *XGBoost* is successful in predicting future shareholder support (i.e. total votes) and excess votes: average shareholder support in the lowest decile is 92% (-1.2% for excess votes), compared to 97.7% in the top decile (1.1% for excess votes). In addition, it also predicts future profitability. Firms that nominated directors in the bottom decile of predicted shareholder support have an average profitability of -0.3%, whereas firms that nominated directors in the top decile of predicted shareholder support have an average profitability of 10%. When *XGBoost* predicts excess votes, the average profitability of firms in the bottom decile is .6% and it is 11.2% for the top decile. This finding suggests that nominating directors on the recommendation of an algorithm trained to predict shareholder votes would not come at the expense of poor firm performance.[21]

5. Characteristics that Affect Director Performance

The machine learning models appear to be able to predict which directors are likely to receive more votes than average in subsequent elections. These excess votes presumably reflect shareholder satisfaction with individual director performance, which is a market-based measure of director performance.

One of the differences with traditional econometric modeling is that the machine learning algorithms do not provide a formula that can be used to infer the influence of any particular independent variable on performance. To understand which characteristics affect director performance, we consider the predictions from the machine learning models and evaluate the extent to which director and firm characteristics are associated with high and low predicted performance.

5.1. Univariate Comparisons

[21] This result alleviates concerns related to the omitted payoff bias articulated in Kleinberg et al. (2017), which in our setting refers to the concern that the decision-maker could have alternative objectives other than satisfying shareholders when making the nominating decision.

Table 7 provides some guidance about which director features are valued by the

	Mean		
	Bottom decile of predicted performance	**Top** decile of predicted performance	Difference p-value
Director level			
Age	56.3	57.0	0.083
Audit committee	0.236	0.818	0.000
Audit committee chair	0.039	0.077	0.001
Background academic	0.060	0.049	0.330
Background finance	0.190	0.122	0.000
Background lawyer	0.026	0.017	0.233
Background manager	0.335	0.318	0.471
Background marketing	0.084	0.026	0.000
Background military	0.010	0.006	0.405
Background politician	0.029	0.011	0.008
Background science	0.040	0.011	0.000
Background technology	0.021	0.007	0.021
Busy	0.520	0.120	0.000
Chairman	0.098	0.001	0.000
Compensation committee	0.624	0.059	0.000
Compensation committee chair	0.175	0.024	0.000
Foreign	0.156	0.088	0.005
Governance chair	0.045	0.011	0.000
Governance committee	0.168	0.122	0.008
International work experience	0.109	0.037	0.000
Male	0.897	0.746	0.000
Network size	1540	1327	0.000
Nomination chair	0.004	0.001	0.318
Nomination committee	0.023	0.011	0.057
Number of qualifications	2.208	2.282	0.180
Total current number of boards sitting on	2.848	1.545	0.000
Total number of listed boards sat on	5.814	2.289	0.000
Ivy league	0.217	0.109	0.000
MBA	0.466	0.410	0.064
Nb previous jobs same FF48 industry	0.105	0.037	0.000
Nb previous directorships same FF48 industry	0.342	0.037	0.000
Board level			
Gender ratio	0.105	0.153	0.000
Nationality mix	0.128	0.084	0.000
Board attrition	0.102	0.054	0.000
Average tenure of incumbent directors	3.443	9.731	0.000
Average tot. nb of boards incumbent directors sit on	1.672	1.809	0.000
Board size	8.5	10.2	0.000
CEO SOX certified	0.539	0.995	0.000
Chairman is CEO	0.357	0.496	0.001
Chairman is CEO with tenure ≥ 5	0.600	0.983	0.000
Indep. directors compensation over CEO tot. compensation	0.912	1.172	0.280
Mean past voting shareholder support	-0.012	0.011	0.000
Number of female directors	1.007	1.611	0.000
Incumbent directors with finance background	0.117	0.221	0.000
Busy incumbent directors	0.173	0.210	0.000
Average age of incumbent directors	57.5	63.0	0.000
Average network size of incumbent directors	1239	1347	0.007

Firm level			
Dividend payer	0.298	0.630	0.000
Excess returns 12 months leading up to appointment	0.028	-0.018	0.126
Firm age	10	30	0.000
Hoberg-Phillips product market fluidity	7.446	6.237	0.000
Institutional ownership %	0.586	0.711	0.000
Largest 10 institutional shareholders %	0.367	0.421	0.000
Largest 5 institutional shareholders %	0.275	0.303	0.001
Largest institutional shareholder %	0.106	0.102	0.492
Leverage	0.266	0.191	0.000
Log (number of institutional blockholders)	1.010	1.250	0.000
Log (number of institutional owners)	4.971	5.279	0.000
Ownership by blockholders %	0.193	0.226	0.002
ROE	-0.110	0.194	0.353
Stock returns prior 12 months	0.158	0.116	0.188
Total assets	17600	30435	0.087
Number of analysts	8.4	12.0	0.000
Short interest (%)	0.036	0.053	0.000
Peter & Taylor Total Q	4.291	0.990	0.000

TABLE 7: TOP VS. BOTTOM DECILE OF PREDICTED PERFORMANCE

Notes: This table reports the mean of firm and director level features for directors in the bottom decile of predicted excess votes and compares it to the mean for directors in the top decile of predicted excess votes. These results are for directors in our test set. Because we do not need the actual vote outcomes for this exercise but only the predictions, this test set covers appointments up to 2016. The algorithm used to predict performance is XGBoost.

algorithm in its assessment of directors. This table reports the averages of a number of characteristics of potential directors, boards, and firms that are associated with low and high expected future voting.

In particular, it presents the means of these characteristics for the bottom and top deciles of predicted shareholder support predicted by the *XGBoost* model.

Table 7 indicates that there are notable differences between directors in the top and bottom deciles. In particular, directors in the bottom decile are more likely to be male, sit on more current boards, have sat on more boards in the past, have received lower shareholder support in previous elections for other boards they sat on, and have a larger network. These differences suggest that male directors who are on a number of boards tend to be less desirable directors, either because they are too busy to do a good job or because they are less likely to monitor the CEO.[22]

[22] Fich and Shivdasani (2006) present evidence suggesting that a director being overly busy can meaningfully affect their monitoring of management.

Board-level variables that affect predicted excess votes likely reflect perceptions of the quality of governance in a particular firm. Note that the outcome variable presented is the *excess votes*, which is adjusted for the average vote in the slate, and therefore, some of the statistics are harder to interpret. For instance, the average tenure of incumbent board members is about three years for directors in the bottom decile of predicted performance, whereas it is about ten years for those in decile ten. This pattern occurs because a new director is more likely to receive more votes relative to other directors up for reelection if the others have been there forever. In unreported results where the top and bottom deciles refer to unadjusted (total) votes, we see that longer average director tenure, which is likely to reflect an entrenched board, is associated with lower predicted shareholder support.

Firm level variables affecting voting tend to reflect the performance of the firm, with better performance leading to higher predicted shareholder support. While prior 12-month stock returns for the bottom predicted decile of shareholder support are not different from that for the top decile of predicted shareholder support, average ROE is significantly larger for the top decile.

5.2. Multivariate Comparisons

Because director and firm characteristics are not independent from one another, we estimate regressions of predicted performance. As independent variables, we include both firm and director variables. The coefficients reflect the characteristics that *XGBoost* tends to associate with higher performance. We report estimates of these regressions in Table 8. Director variables related to predicted subsequent shareholder support are gender, a dummy variable that indicates whether the director is "busy" and the number of listed boards a director serves on. In particular, our algorithm suggests that male directors and directors who are on a number of boards ("busy" directors) tend to be less popular with shareholders.

Dependent variable: predicted performance	(1)	(2)	(3)	(4)
Busy	-0.006***	-0.005***	-0.005***	-0.005***
	(-24.332)	(-13.183)	(-12.230)	(-12.087)
Male	-0.001***	-0.001	-0.001	-0.001*
	(-4.623)	(-1.398)	(-1.603)	(-1.688)
Age		-0.000**	-0.000**	-0.000**
		(-2.001)	(-2.079)	(-2.242)
MBA		0.000	0.000	0.000
		(1.074)	(1.108)	(1.150)
Ivy league		-0.001**	-0.001*	-0.001*
		(-2.555)	(-1.869)	(-1.864)
Background lawyer			-0.002	-0.001
			(-1.521)	(-1.394)
Background academic			0.000	0.000
			(0.180)	(0.134)
Background finance			-0.001	-0.001
			(-1.344)	(-1.568)
Network size			-0.000***	-0.000***
			(-2.838)	(-2.691)
Ln (Assets)	0.001***	0.000	0.000	0.000
	(9.054)	(0.276)	(0.016)	(0.160)
ROA	0.001***	0.000	0.000	0.000
	(4.280)	(0.024)	(0.156)	(0.124)
Board size	-0.000***	-0.000**	-0.000*	-0.000*
	(-3.303)	(-2.027)	(-1.751)	(-1.843)
Average nb independent directors	0.009***	0.006***	0.006***	0.005**
	(20.814)	(3.055)	(2.906)	(2.529)
Chairman duality		0.001***	0.001***	0.001***
		(3.399)	(3.481)	(3.319)
Excess returns 12 months leading up to appointment		0.000	0.000	0.000
		(1.049)	(1.058)	(1.156)
Number of female directors			0.000	0.000
			(0.172)	(0.554)
Average tenure of incumbent directors		0.000***	0.000***	0.000***
		(6.441)	(6.099)	(4.907)
Log (number of institutional owners)		0.001**	0.001**	0.001*
		(2.155)	(2.431)	(1.798)
Compensation committee chair				-0.002*
				(-1.803)
Audit committee chair				0.002**
				(2.524)
Governance committee chair				-0.002
				(-1.457)
Nomination committee chair				-0.002
				(-0.599)
Firm Age				0.000***
				(2.925)
Constant	-0.002***	0.003	0.003	0.005*
	(-3.762)	(1.053)	(1.128)	(1.867)
Observations	7,738	1,893	1,883	1,883
R-squared	0.153	0.131	0.136	0.146

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

TABLE 8: THE DETERMINANTS OF PREDICTIONS: OLS REGRESSIONS

Notes: This table reports the results from OLS regression models of the predicted excess votes in our test set on some firm level and director level features. Because we do not need the actual vote outcomes for this exercise but only the predictions, this test set covers appointments up to 2016. The algorithm used to generate the predictions is *XGBoost*.

This pattern could reflect the commonly stated concern of shareholders that directors are too often the same people (almost always men), are on many boards but do not monitor to the extent that shareholders would like (see for example Biggs (1996)). Consistently, network size has a significantly negative coefficient as well.

Board level variables that are significantly related to the predicted shareholder support are the size of the board, the average tenure of incumbent board members, and the average number of independent directors. These variables again are likely to reflect the independence of the board from management. Firm-level variables that appear to be associated with subsequent performance are size (total assets), operating performance, and whether the firm pays dividends. A word of caution in interpreting these coefficients: the R^2 is fairly low (below 20%), which speaks to the importance of feature interactions and non-linearities that *XGBoost* relies on to generate its predictions of subsequent performance.

5.3. Overvalued Director Characteristics

The algorithm's predictions also help identify the individual director features that tend to be overvalued or undervalued by firms when they select new directors. We identify directors who were nominated but are predictably of low quality and compare them to those directors the algorithm would have preferred for that specific board position. The patterns of discrepancies between these two groups recognize the types of directors that tend to be overvalued in the nomination process. In other words, the algorithm provides a diagnostic tool that can help evaluate the way in which directors are chosen.

In Table 9, we report characteristics of directors who were nominated, but whom the algorithm predicted would do poorly (and subsequently did poorly). Compared to promising candidates as identified by our algorithms, predictably unpopular directors are more likely to be male, have fewer degrees post undergraduate, a larger professional network, more current and past directorships, and a background in finance. This comparison is for each new board seat, holding committee assignments constant for nominated directors and candidates. These are averages across all new board positions. We find similar results for all alternative specifications mentioned in previous sections.

	Hired directors with predicted and observed low shareholder support	**Promising candidates for this board position**	
	Mean	Mean	Difference *p-value*
Male	0.984	0.835	0.000
Number of qualifications	2.1	2.4	0.000
Ivy League	0.29	0.26	0.523
MBA	0.57	0.38	0.000
Network size	1673	1428	0.000
Total number of listed boards sat on	6.4	2.3	0.000
Total number of unlisted boards sat on	11.0	2.7	0.000
Total current number of boards sitting on	3.1	1.5	0.000
Number previous jobs same industry	0.11	0.08	0.223
Number previous directorships same industry	0.26	0.07	0.000
Busy	0.64	0.10	0.000
Director age	54.1	54.6	0.353
Background academic	0.021	0.000	0.001
Background finance	0.094	0.046	0.002
International work experience	0.130	0.018	0.000

TABLE 9: OVERVALUED DIRECTOR CHARACTERISTICS

Notes: This table reports the mean of director features for directors in our test set (out of sample predictions) whom our XGBoost algorithm predicted would be in the bottom decile of performance and indeed ended up in the bottom decile of actual performance (predictably unpopular directors) and compares it to the mean for potential candidates the board could have nominated instead, whom our XGBoost algorithm predicted would be in the top decile.

These results highlight the features that are likely overrated by management when nominating directors. They are consistent with the view that directors tend to come from an "old boys club", in which men who have sat on a lot of boards are chosen to be directors, even if they received poor shareholder support at the firms on whose boards they serve. The underlying reason for this pattern, however, is not clear. As suggested by the literature on boards going back to Smith (1776) and Berle and Means (1932), managers and existing directors could implicitly collude to nominate new directors unlikely to rock the boat and upset the rents managers and existing directors receive from their current positions. Alternatively, a long literature in psychology dating to Meehl (1954) and highlighted in Kahneman (2011) has found that even simple algorithms can outperform interviews by trained professionals at predicting subsequent performance in a number of contexts. It is possible that managers and boards could be attempting to find value-maximizing directors but because of behavioral biases, could underperform the algorithms we present. Understanding why firm-selected directors differ from algorithm-selected directors is likely to be an important topic of future research.

6. Summary and Discussion

We present a machine-learning approach to selecting the directors of publicly traded companies. In developing the machine learning algorithms, we contribute to our understanding of governance, specifically boards of directors, in at least four ways. First, we evaluate whether it is possible to construct an algorithm that accurately forecasts whether a particular individual will be successful as a director in a particular firm. Second, we compare alternative approaches to forecasting director performance; in particular, how traditional econometric approaches compare to newer machine learning techniques. Third, there is no consensus in the literature on which metric to use for evaluating director performance. We provide evidence that director popularity is related to their expected value and therefore shareholder support is a reasonable proxy for their performance. Finally, we use the selections from the algorithms as benchmarks to understand the process through which directors are actually chosen and identify the types of individuals who are more likely to be chosen as directors *counter* to the interests of shareholders.

There are a number of methodological issues we must address before we can construct such an algorithm. We must be able to measure the performance of a director to predict which potential directors will be of highest quality. Measurement of directors' performance is complicated by the fact that most directors' actions occur in the privacy of the boardroom where they are not observable to an outside observer. In addition, most of what directors do occurs within the structure of the board, so we cannot isolate their individual contributions. Our approach is based on the fraction of votes a director receives in shareholder elections. This vote, which is shown to be informative about directors' quality in the prior literature, reflects the support the director personally has from the shareholders and should incorporate all publicly available information about the director's performance. In addition, predicted future votes are positively related to CARs around the announcements of director appointments.

Using publicly available data on firm, board, and director characteristics, our *XGBoost* algorithm can predict the success of directors, and in particular, can identify which directors are likely to be unpopular with shareholders. In comparison to the machine-learning models, standard econometric models fit the data poorly out of sample. Specifically, the observed performance of individual directors is not related to the predictions of performance of an OLS model. The fact that the machine learning models dramatically outperform econometric approaches is consistent with the arguments of Athey and Imbens (2017) and Mullainathan and Spiess (2017): machine learning is a valuable approach for prediction problems in social sciences.

While we can observe the fraction of support an existing director has from shareholders, we cannot observe the votes a potential director who was not chosen would have received, nor whether a potential director for a firm would have been willing to accept the directorship. We address this issue by constructing a pool of potential directors from those who around that time accept a directorship at a smaller nearby company, so presumably would have been attracted to a directorship at a larger, neighboring company. To evaluate the performance of our algorithm, we use the fraction of votes he received at the company where he was a director as our measure of this potential director's performance.

The differences between the directors suggested by the algorithm and those actually selected by firms allow us to assess the features that are overrated in the director nomination process. Comparing predictably unpopular directors to promising candidates suggested by the algorithm, it appears that firms choose directors who are much more likely to be male, have a large network, have a lot of board experience, currently serve on more boards, and have a finance background.

In a sense, the algorithm is saying exactly what institutional shareholders have been saying for a long time: that directors who are not old friends of management and come from different backgrounds are more likely to monitor management. In addition, less connected directors potentially provide different and potentially more useful opinions about policy. For example, TIAA-CREF (now TIAA) has had a corporate governance policy aimed in large part at diversifying boards of directors since the 1990s for this reason (see Biggs (1996) and Carleton et al. (1998)).[23]

Our finding on the predictability of which directors will or will not be popular with shareholders has important implications for corporate governance. Observers since Smith (1776) and Berle and Means (1932) have been concerned about whether managers intentionally select boards that maximize their own interests rather than those of the shareholders. In addition, a psychology literature started by Meehl (1954) has found that because of behavioral biases, even simple algorithms can outperform humans in deciding on personnel decisions. It is easy to imagine that a machine learning algorithm, which is more sophisticated than the algorithms relied on by psychologists, would allow firms to improve their selection process.

A natural question concerns the applicability of algorithms such as the ones we developed in practice. The algorithms we present should be treated as "first pass"

[23] Similarly, Glenn Kelman, the CEO of RedFin, recently wrote: "Redfin has recently completed a search for new board directors, […] and we had to change our process, soliciting many different sources for candidates rather than relying exclusively on board members' connections. If you don't pay attention to diversity, you'll end up hiring people who are nearest at hand, who have had similar jobs for decades before. This is how society replicates itself from generation to generation, in a process that seems completely innocuous to those who aren't the ones shut out." https://www.redfin.com/blog/2016/11/how-to-triple-the-number-of-women-appointed-to-boards-in-three-years.html

approaches; presumably more sophisticated models would predict director performance even better than the ones presented in this paper. In addition, our algorithms rely on publicly available data; if one had more detailed private data on director backgrounds, performance, etc., one could improve the algorithm's fit as well. If algorithms such as these are used in practice in the future as we suspect they will be, practitioners will undoubtedly have access to much better data than we have and should be able to predict director performance more accurately than we do in this paper. An important benefit of algorithms is that they are not prone to the agency conflicts that occur when boards and CEOs together select new directors.

Algorithmic bias is a concern of growing importance and algorithms are only as impartial as the data that feed them. If the data is generated by human decisions, machine learning algorithms can generate bias amplification (see Zhao et al., 2017). As Miller (2018) argues however, the perils of human bias are arguably worse than the perils of algorithmic bias. An important feature of our application is that the decision maker and the evaluator are separate entities: the board decides on the identity of the new director while shareholders vote. If we assume that the set of biases and incentives are independent between investors who vote (generate the left-hand side variable in our model) and board members who select new directors (generate the right hand side variables in our model), then this helps mitigate the concern that the algorithm acts as a bias propagator.

Institutional investors are likely to find this independence from agency conflicts particularly appealing and are likely to use their influence to encourage boards to rely on an algorithmic decision aid such as the one presented here for director selections in the future. An important advantage of an algorithm over the way in which directors have been chosen historically is that algorithms do not allow for judgment on the part of directors and current management. As Sunstein (2018) writes: "algorithms can overcome the harmful effects of cognitive biases". Rivera (2012) studies the hiring practices of top investment banks, consulting and law firms and concludes that recruiters overvalue personal fit which is not necessarily a function of expected performance. In the context of lower skill workers, Hoffman et al. (2017) find that managers who hire against test recommendations end up with worse average hires. Cowgill (2018) shows that in the context of hiring at a software company, the job-screening algorithm prefers "nontraditional" candidates. Our results suggest that the same idea applies to the nominating of corporate directors. Including algorithmic input to limit discretion and reliance on soft information in these decisions could help minimize agency problems, and thus lead to a modified rank ordering of candidates that could in turn lead to better directors than the current process. On the other hand, if the algorithm omits attributes of potential directors that are valuable to management, such as specialized knowledge of an industry or government connections, then it potentially could lead to suboptimal solutions. This is why we advocate for tools built on algorithms as decision aids, not substitutes for human judgement. Humans and machines both have limits and make different kinds of mistakes, i.e. they are likely to have uncorrelated errors. Achieving the right balance in the division of labor between humans and machines to take advantage of their relative strengths is key.[24]

In this paper, we use 21st century technology to confirm an observation that dates back over two hundred years: the board selection process leads to directors who often are not the best choices to serve shareholders' interests. This technology can, however, in addition to confirming this observation, provide us with the tools to change it. By providing a prediction of performance for *any* potential candidate, a machine-learning algorithm could expand the set of potential directors and identify individuals with the skills necessary to become successful directors, who would have otherwise been overlooked. We expect that in the not too distant future,

[24] The issues around the consequences of AI-based decisions are exposed in grounded discussions in Agrawal, Gans and Goldfarb (2018)

machine-learning techniques will fundamentally change the way corporate governance structures are chosen, and that shareholders will be the beneficiaries.

References

Adams, R. B (2017) Boards and the Directors Who Sit on Them. Chapter 6 of *Handbook of the Economics of Corporate Governance,* edited by B. E. Hermalin and M. S. Weisbach, Elsevier.

Adams, R., B. Hermalin and M. Weisbach (2010) The Role of Boards of Directors in Corporate Governance: A Conceptual Framework and Survey. *Journal of Economic Literature* 48, 58–107.

Alexander, C., M. Chen, D. Seppi, and C. Spatt (2010) Interim News and the Role of Proxy Voting Advice. *Review of Financial Studies* 23: 4419-4454.

Aggarwal, R., S. Dahiya and N. Prabhala (2017) The Power of Shareholder Votes: Evidence from Uncontested Director Elections. *Georgetown McDonough School of Business Research Paper*.

Aggarwal, R., I. Erel, and L. Starks (2016) Influence of Public Opinion on Investor Voting and Proxy Advisors, Working Paper, Ohio State University.

Agrawal, A., J. Gans and A. Goldfarb (2018) Prediction Machines. *Harvard Business Review Press*, Boston, Massachussetts.

Athey, S. (2017) Beyond Predictions: Using Big Data for Policy Problems. *Science* 355: 483-385.

Athey, S. and G.W. Imbens (2017) The State of Applied Econometrics: Causality and Policy Evaluation, *Journal of Economic Perspectives*, 31, 3-32.

Autor, D. (2015) Why Are There Still So Many Jobs? The History and Future of Workplace Automation, *Journal of Economic Perspectives*, 29, 3-30.

Biggs, John (1996) Corporate Governance Assessment: A TIAA-CREF Initiative, *Directors Monthly,* 20 (10) 1-8.

Cai, J., J. Garner and R. Walkling (2009) Electing Directors. *The Journal of Finance*, 64(5): 2389-2421.

Cai, J., T. Nguyen, and R. Walkling (2017) Director Appointments – It is Who You Know, Drexel University Working Paper.

Carleton, Willard T., James M. Nelson and Michael S. Weisbach (1998) The Influence of Institutions on Corporate Governance through Private Negotiations: Evidence from TIAA-CREF, *The Journal of Finance*, 53, 1335-1362.

Chen, T. and C. Guestrin (2016) XGBoost: A Scalable Tree Boosting System. KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 785-794.

Coles, J., N. Daniel and L. Naveen (2008) Boards: Does One Size Fit All? *Journal of Financial Economics,* 87, 329-356.

Coles, J., N. Daniel and L. Naveen (2015) Board Groupthink. *Working Paper.*

Cowgill, B. (2018) Automating Judgement and Decision-making: Theory and Evidence from Resumé Screening. *Working Paper*.

Daines, R., I. Gow, and D. Larcker (2010) Rating the ratings: How Good are Commercial Governance Ratings? *Journal of Financial Economics* 98: 439-461.

Ertimur, Y. F. Ferri, and D. Oesch (2013) Understanding Uncontested Director Elections. *Management Science*, forthcoming.

Fich, E. and A. Shivdasani (2006) Are Busy Boards Effective Monitors? *The Journal of Finance* 61: 689-724.

Field, L., M. Souther, and A. Yore (2017) Does diversity pay in the boardroom? University of Delaware Working Paper.

Fischer, P., J. Gramlich, B. Miller and H. White (2009) Investor Perceptions of Board Performance: Evidence from Uncontested Director Elections. *Journal of Accounting and Economics* 48: 172–189.
Hart O. and L. Zingales (2017) Companies Should Maximize Shareholder Welfare Not Market Value. *Journal of Law, Finance, and Accounting*, 2: 247–274.
Hastie, T., R. Tibsharani and M. Wainwright (2015) Statistical Learning with Sparsity: The Lasso and Generalizations. *Chapman and Hall/CRC.*
Hermalin, B. E., and M. S. Weisbach (1998) Endogenously Chosen Boards of Directors and Their Monitoring of the CEO. *American Economic Review*, 88(1): 96–118.
Hermalin, B. E., and M. S. Weisbach (2003) Boards of Directors as an Endogenously Determined Institution: A Survey of the Economic Literature. *Federal Reserve Bank of New York Economic Policy Review*, 9(1): 7–26.
Hoberg, G. and G. Phillips (2010) Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies* 23 (10), 3773-3811.
Hoberg, G. and G. Phillips (2016) Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy* 124 (5) 1423-1465.
Hoffman, M., L. B. Kahn and D. Li (2017) Discretion in Hiring. *The Quarterly Journal of Economics. qjx042, https://doi.org/10.1093/qje/qjx042.*
Iliev, P., K. Lins, D. Miller and L. Roth (2015) Shareholder Voting and Corporate Governance Around the World. *The Review of Financial Studies* 28(8): 2167–2202.
Iliev, P. and M. Lowry (2014) Are Mutual Funds Active Voters? *Review of Financial Studies* 28: 446-485.
Kahneman, D. (2011) *Thinking, Fast and Slow*, Farrar, Strauss and Giroux: New York.
Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan (2017) Human Decisions and Machine Predictions. *The Quarterly Journal of Economics* 133 (1) 237–293.
Kleinberg J., Liang A. and S. Mullainathan (2017) The Theory Is Predictive, but Is It Complete? An Application to Human Perception of Randomness. *Working Paper*.
Kleinberg, J., J. Ludwig, S. Mullainathan and Z. Obermeyer (2015) Prediction Policy Problems. *American Economic Review* 105(5): *491–495.*
Kramarz, F. and D. Thesmar (2013) Social Networks in the Boardroom. *Journal of the European Economic Association* 11(4): 780-807.
Larcker, D., A. McCall, and G. Ormazabal (2015) Outsourcing Shareholder Voting to Proxy Advisory Firms. *Journal of Law and Economics* 58, 173-204.
Li, K., M. Feng, R. Shen, and X. Yan (2018) Measuring Corporate Culture Using Machine Learning. *Working Paper.*
Linck, J, J. Netter, and T. Yang (2009) The effects and unintended consequences of the Sarbanes-Oxley Act on the supply and demand for directors, Review of Financial Studies 8, 3287–3328.
Malenko, N., and Y. Shen (2016) The Role of Proxy Advisory Firms: Evidence from a Regression-Discontinuity Design, *The Review of Financial Studies* 29: 3394–3427.
Masulis, R. and S. Mobbs (2014) Independent Director Incentives: Where do Talented Directors Spend their Limited Time and Energy, *Journal of Financial Economics* 111, 406-429.
Meehl, P. E. (1954) *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, University of Minnesota Press: Minneapolis.
Miller, A. (2018) Want Less-Biased Decisions? Use Algorithms. *Harvard Business Review.*
Mitchell, T. (1997) Machine Learning. *McGraw Hill*.
Mullainathan, S. and J. Spiess (2017) Machine learning: An Applied Econometric Approach, *Journal of Economic Perspectives*, 31, 4: 87-106.
Pearl J. and D. Mackensie (2018) The Book of Why. *Basic Books.*
Rivera, L.(2012) Hiring as Cultural Matching: The Case of Elite Professional Service Firms, *American Sociological Review*, 77: 999-1022.

Shivdasani, A. and D. Yermack (1999) CEO Involvement in the Selection of New Board Members: An Empirical Analysis, *The Journal of Finance*, 54, 1829-1853.
Shmueli, G. (2010) To Explain or to Predict? *Statistical Science* 25(3): 289-310.
Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, Indianapolis, Liberty Press.
Sunstein, C. (2018) Algorithms, Correcting Biases. Working paper.
Yermack, D. (1996) Higher Market Valuation of Firms with a Small Board of Directors, *Journal of Financial Economics* 40: 185-211.
Zhao Jieyu, T. Wang, M. Yatskar, V. Ordonez, K. Chang (2017) Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints, *EMNLP* 2017.

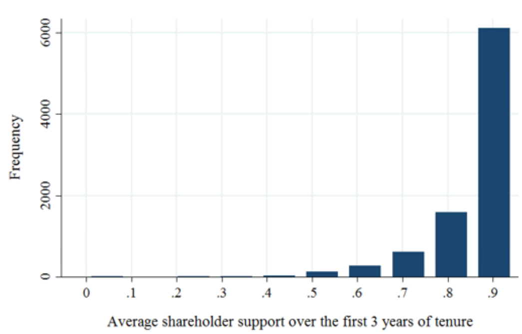
For Online Publication

VOTES DISTRIBUTION

Shareholder Support: Fraction of Votes "For"

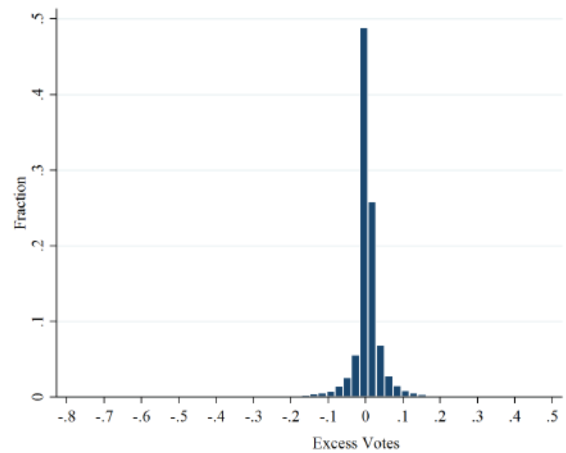
This figure shows the distribution of average shareholder support, defined as the fraction of votes in favor of a given director over all votes cast for the director's reelection within three years of her tenure. The data is from ISS Voting Analytics.

Distribution of Bad Outcomes: Fraction of Votes "For" Below 95%



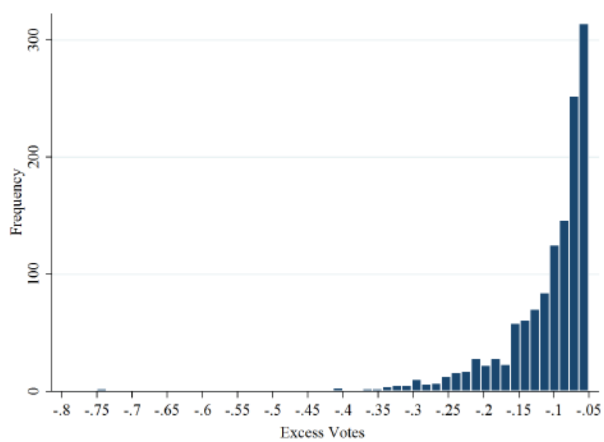
This figure shows the distribution of average shareholder support for values under its mean value of 95%. Shareholder support is defined as the fraction of votes in favor of a given director over all votes cast for the director's reelection within three years of her tenure. The data is from ISS Voting Analytics.

Excess Votes: Fraction of Votes “For” Minus the Slate’s Average



This figure shows the distribution of *excess votes* for our sample. To compute *excess votes*, we compute the fraction of votes in favor of a given director over all votes cast for the director. Next, we subtract the average of that variable for the slate of directors up for reelection that year on the focal board. Finally, we take the average of this relative vote measure over the first three years of the new director’s tenure. The data is from ISS Voting Analytics.

Distribution of Bad Outcomes: Excess Votes below -5%



This figure shows the distribution of *excess votes* below -5%.

OLS MODEL

This table reports coefficients from an OLS regression of excess votes on various director, firm, and board characteristics. OLS model also includes firm fixed effects. Excess vote is defined as the average observed level of shareholder support over the first three years of a new director's tenure, minus the average vote for all directors in the same slate. The regression sample contains director appointments between 2000-2011.

Dependent Variables: Excess Votes	
Compensation committee chair	-0.003
	(-1.281)
Audit committee chair	0.006***
	(2.975)
Governance committee chair	0.003
	(1.338)
Nomination committee chair	0.002
	(0.399)
Nb previous jobs same FF48 industry	-0.002
	(-1.117)
Background finance	0.002
	(1.341)
Background law	-0.006**
	(-2.320)
MBA	0.001
	(1.229)
Ivy league	-0.001
	(-0.615)
Male	0.001
	(0.609)
Age (director)	0.000
	(-0.068)
Number of qualifications	0.000
	(0.133)
Ln (Assets)	0.003**
	(2.323)
Leverage	-0.007
	(-1.285)
M/B	0.000
	(0.185)
Largest 5 institutional shareholders %	0.012
	(1.553)
ROA	0.000
	(0.078)
Product market fluidity	0.000
	(-0.486)
12-month return	0.000
	(-0.452)
Dividend payer	0.003
	(1.147)
Board size	0.000
	(0.375)
Number of female directors	0.000
	(0.371)
Average nb independent directors	-0.009
	(-1.232)
Average age	0.000
	(0.774)
Constant	-0.026*
	(-1.650)
Observations	10,601
Number of firms	2,820
R-squared	0.005

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

ALGORITHMS USED TO PREDICT PERFORMANCE: SOME DETAILS

A.2.1. Less is More: The Case for *Lasso* and *Ridge*

Lasso and *ridge* are both linear models that use a regularization term to achieve a balance between bias and variance. They do so by minimizing a loss function that includes in-sample fit and a penalty term that favors simple models, thereby reducing variance. Prediction accuracy is thus improved by setting some coefficients to zero and shrinking others. To achieve this goal, lasso and ridge combine the minimization of the sum of the squared errors with the norm of parameters. The lasso estimator solves the problem:

$$\min_{\beta} \sum_{j=1}^{k} (y_i - x_i \beta)^2 \ + \ \lambda \ \cdot \ \|\beta\|_1$$

where $\|\beta\|_1$ is the ℓ_1-norm (least absolute deviation). The penalty weight (λ) on the sum of the absolute values of coefficients is set using the default parameter in scikit-learn[25].

Ridge is similar to *lasso* except that the bound on the parameter estimates is the ℓ_2-norm (least squares), therefore shrinking estimates smoothly towards zero, as opposed to setting some estimates to zero as Lasso does.[26]

A.2.2. Gradient Boosting Trees

Gradient Boosting Trees are related to random forests. A decision tree is the basic building block of random forests. A decision tree defines a tree-shape flow graph to support decisions. An instance is classified by starting from the root of the tree, testing the feature specified by the node, moving down the branch corresponding to the feature value in the given instance.

A key difference between decision tree learning and Ridge and Lasso regression lies in the fact that there is no explicit objective function that a decision tree optimizes. Instead, the learning process is a greedy recursive algorithm that finds the best feature to split the current data based on a criterion. In our paper, we_use a decision tree regressor

[25] http://scikit-learn.org/stable/

[26] For a detailed discussion of sparse estimators, we refer interested readers to Hastie, Tibshirani and Wainwright (2015).

where the criterion aims to minimize the mean squared error in each branch. Refer to Mitchell (1997) for more details on decision tree learning.

Random forest is an ensemble method based on decision trees. The main intuition is that a single decision tree can be noisy but is able to function as a weak learner. An ensemble of weak learners makes a strong learner. To train a random forest regressor, a number of decision tree regressors are fitted by randomly sampling data from the training instances with replacement and also randomly sampling a subset of features. The average values of all decision tree regressors is used to predict the value of an instance.

Gradient boosting tree is another ensemble method based on decision trees. It differs from random forests in two aspects:

1. Boosting. To predict the value of an instance, gradient boosting trees uses K additive functions instead of computing the average:

   $$\hat{y_i} = \sum_{k=1}^{K} f_k(x_i),$$

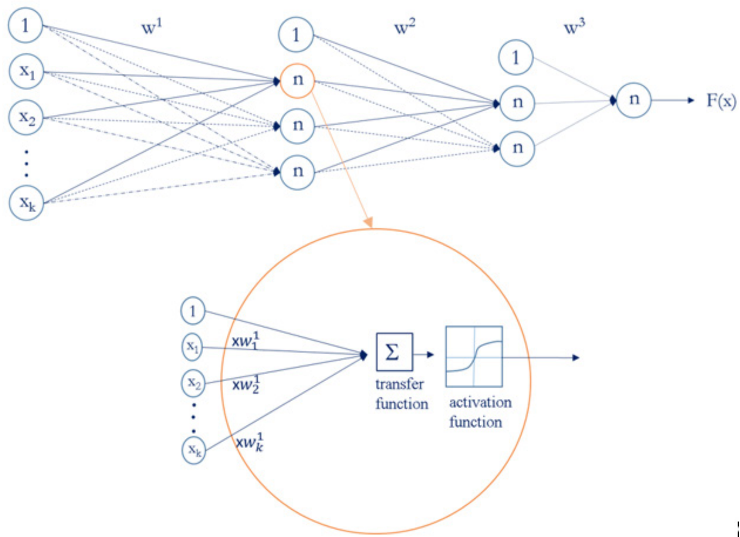
   where f_k is a decision tree regressor. In other words, in boosting, each additional decision tree attempts to fit the residual error, whereas each decision tree in random forest attempts to fit the target value y directly.

2. Regularized objectives. The split in a decision tree regressor of gradient boosting trees optimizes a regularized global objective that balances the predictive performance and the complexity of decision tree regressors. The loss function is formulated as:

   $$L = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k),$$

   where l refers to a differentiable loss function that measures the difference between the predicted value and the target value (in our case, it is simply squared loss), $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\left|\left|w\right|\right|^2$ and measures the complexity of a tree, T refers to the number of leaves in the tree and w refers to the score at a leave. A simple tree has a small number of leaves and each leave has a small score. γ and λ are parameters to control how these two complexity measures are weighted in the final objective function. The name gradient boosting trees arise from the fact that a gradient will be computed in the algorithm to optimize the above objective function. Please refer to Chen et al, 2016 for a detailed discussion.

A.2.3. Neural Networks



The figure above depicts the structure of a basic neural network with two hidden layers. Neurons x_i are input neurons connected to the next layer of neurons by synapses which carry weights w^1. Each synapse carries its own weight. An activation function (usually a sigmoid to allow for non-linear patterns) is embedded in each neuron in the hidden layers to evaluate its inputs. The set of weights carried by the synapses that reach a neuron are fed into its activation function, which will determine whether that neuron is activated. If activated, it then triggers the next layer of neurons with the value it was assigned, with weight w^2 (again with each synapse carrying its own weight).

A Framework to Assess Algorithms' Predictions

Following Kleinberg et al. (2017), we develop a framework to understand the issues faced when assessing the prediction accuracy of our algorithms. Suppose that the true data generating process is given by $\mathcal{Y} = \mathcal{F}(\mathcal{W}, \mathcal{Z})$, where $\mathcal{W}$ and $\mathcal{Y}$ are operationalized by W, our vector of inputs and Y, our outcome variable (i.e., director performance). $\mathcal{Z}$ represents a set of features that affect director performance and that are observable by the board but not by the algorithm. An example of such a feature would be idiosyncratic knowledge of the firm or its industry that would make a potential director more valuable.

In addition, there are features $\mathcal{B}$ that do *not* affect director performance and are unobservable to the algorithm, but could nonetheless affect boards' nominating decisions. Examples of such features could be a candidate's political views, or the neighborhood where he grew up. The board's preferences for certain features in $\mathcal{B}$ could be conscious or even could represent an implicit bias of which they are unaware of. The important point is that these attributes of a potential director can influence boards' decisions even though they are uncorrelated with performance.

$\mathcal{F}$ is operationalized by a functional form $\mathscr{f}$. For the purpose of predictive modeling, we are interested in finding a function that closely matches the function $\mathscr{f}$ in out-of-sample data. Compared to classic causal hypothesis testing, we do not make strong assumptions about the structure of $\mathcal{F}$ and thus do not focus on examining the estimated parameters and claim that these parameters match $\mathscr{f}$. In other words, our algorithm seeks a functional form that maps features W into predictions $\hat{\mathscr{f}}(W)$ that generalize well on out-of-sample data (Shmueli, 2010).

A director is characterized by $\vec{x}$, composed of three vectors of features and outcome *y*:

$$\vec{x} = \begin{bmatrix} W \\ \mathrm{Z} \\ \mathrm{B} \end{bmatrix}$$

Note that x may include not only director characteristics but also firm and board characteristics so that both the board and the algorithm try to assess a director's future performance on a specific board.

For the purpose of the model, we shrink the dimension of $\vec{x}$ to a vector with three unidimensional characteristics w, z and b. In addition, we make the assumption that the sum of *w* and *z* is distributed between 0 and 1 and that their sum equals *y* on average:

$$E[Y = y|W = w, Z = z] = E[y|w, z] = w + z$$

Each board j has a payoff function π_j that is a function of the director's performance as well as of the director's characteristics as defined by $\vec{x}$. For each director (x, y) in the candidate pool $\mathcal{D}$ of size k, the board's payoff is characterized as:

$$\pi_j(x, y) = \underbrace{u_j y}_{\substack{benefits\ from \\ director's\ performance}} + \underbrace{v_j g_j(x)}_{\substack{benefits\ from\ hiring \\ director\ with\ characteristics\ x}}$$

$g_j(x)$ is a board specific function that maps directors' characteristics into a score. We can think of $g_j(x)$ as a measure of the utility the board derives from nominating a director with specific characteristics; for example, they could derive private benefits from nominating someone from their own network. The variables u_j and v_j are the weights that board *j* puts on director performance and on the benefits it derives from nominating a director with certain features, respectively.

We assume that board *j* chooses a nominating rule h_j such that it maximizes its expected payoff.

$$h_j \in \{0,1\}^k \ and\ \|h_j\|_0 = 1$$

$$\Pi_j(h_j) = \sum_{i \in \mathcal{D}} h_{j,i} E[\pi_j(x_i, y_i)]$$

The nominating rule h_j depends on $k_j(x)$, the board's *assessment* of future performance for a director with characteristics x. For a given $g_j(x)$, the board chooses the director with the highest $k_j(x)$. We do not observe boards' relative weights on director performance, u_j, and their own preferences for directors with particular characteristics, v_j. In a world of perfect corporate governance, boards are only concerned with their mandate (i.e. representing shareholders' interests) and $v_j = 0$.

We set $v_j = 0$ not because we believe in a world of perfect governance but because our question is: can an algorithm identify a director x'' with better performance than director x' nominated by board *j*, whom the board will like at least equally well? In other words, conditional on $g_j(x'') \geq g_j(x')$, can an algorithm recommend a nominating rule α that produces a higher payoff than the baseline: the outcome of board *j*'s actual nominating decision?

The difference in the expected payoffs between the two nominating rules α_j and h_j is:

$$\Pi_j(\alpha_j) - \Pi_j(h_j) = \sum_{i\in\mathcal{D}} \alpha_{j,i} E[\pi_j(x_i, y_i)] - \sum_{i\in\mathcal{D}} h_{j,i} E[\pi_j(x_i, y_i)]$$
$$= \underbrace{E[y|\ \alpha]}_{missing\ label} - \underbrace{E[y|\ h]}_{observed\ label}$$

We do not observe the performance of directors who would be nominated under the alternative nominating rule produced by the algorithm. As discussed in Kleinberg et al. (2017), missing labels are often dealt with in the machine learning literature by various imputation procedures. However, this approach would assume that if a director shares the same set of observable feature values, *w*, as the nominated director, their performance would be identical. This is the equivalent of assuming that unobservables, *z*, play no role in nominating decisions. For a given *w*, the imputation error would therefore be:

$$E[y|\alpha, w] - E[y|h, w] = E[w + z|\alpha, w] - E[w + z|h, w]$$
$$= E[w|\alpha, w] - E[w|h, w] + E[z|\alpha, w] - E[z|h, w]$$
$$= E[z|\alpha, w] - E[z|h, w]$$

This imputation error points up the *selective labels problem* as described by Kleinberg et al. (2017). In our setting, it refers to the possibility that directors who were nominated, although they might share the same exact observable features as other directors not nominated, might differ in terms of unobservables. These unobservables could lead to different average outcomes for nominated vs. not nominated, even if both are identical on the basis of observable characteristics.

We exploit the design of our pool of candidate directors for each board seat in order to compare the performance of our algorithm to board decisions. We consider directors who joined the board of a neighboring company around the same time. These directors were available to join a board at that time and willing to travel to that specific location for board meetings. Furthermore, to alleviate concerns related to the ability of a particular firm to attract promising directors, we restrict the pool of potential candidates to directors who joined a *smaller* neighboring company around the same time, since the prestige of being a director tends to increase with company size (see Masulis and Mobbs, 2014). In addition, we note that there is on average very little variation in shareholder support for individual director performance across the different boards they join during our sample period. Therefore, although we do not have labels for nominees generated by the algorithm's nominating rule, $E[y|\alpha]$, we observe their *quasi-label*: their performance on the smaller neighboring board they joined around the same time.

We are interested in evaluating the quality of boards' nominating decisions. Our approach is to contrast those decisions to an alternative nominating rule that our algorithm would have chosen. For example, using the notation introduced in this section, if the algorithm predicted a director with characteristics x' would perform very poorly and there were 150 other candidates the algorithm predicted would do better, there are effectively 150 alternative nominating rules α that would yield a higher payoff in terms of benefits derived from director performance. To allow boards to use unobservables to make their nominating decisions, we add the assumption that among those 150 alternative nominees, there exists at least one director with characteristics x'' such that $g_j(x'') \geq g_j(x')$. When we analyze the quasi-labels of those potential candidates, we explore whether they indeed do much better on average than director x' when x' was predicted to do poorly, and worse when x' was predicted to do well.

There are two, not mutually exclusive, reasons why the selections of the algorithm could outperform the actual directors selected by firms: first, the algorithm actually attempts to choose value maximizing directors while actual boards do not ($u_j = 0$), and second, the machine learning approach outperforms the choices firms would have made even if they were attempting to maximize value. In other words, boards are "mispredicting" future performance, i.e. the technology $k_j(x)$ they use to assess the future performance of candidates is inapt. Results related to chosen directors who were predictably unpopular would suggest that boards put disproportionate weight on v_j.

DATA DEFINITIONS

A.4.1. Individual Director Features
Source: BoardEx except if stated otherwise
(as of when the director joins the board)

Variable	**Definition**
Age	Director age
Audit chair	Equals to one if director is chair of the audit committee
Audit member	Equals to one if director is a member of the audit committee
Avg. time on the board	The average time that a director sits on the board of quoted companies
Background academic	Dichotomous variable equal to (*henceforth "Equals to"*) one if job history includes in title one of the following: "professor" "academic" "lecturer" "teacher" "instructor" "faculty" "fellow" "dean" "teaching"
Background CEO	Equals to one if job history includes CEO title
Background finance	Equals to one if job history includes in title one of the following: "underwriter" "investment" "broker" "banker" "banking" "economist" "finance" "treasure" "audit" "cfo" "financial" "controller" "accounting" "accountant" "actuary" "floor trader" "equity" "general partner" "market maker" "hedge fund"
Background hr	Equals to one if job history includes in title one of the following: "hr " "recruitment" "human resource"
Background law	Equals to one if job history includes in title one of the following: "lawyer" "legal" "attorney" "judge" "judicial"
Background manager	Equals to one if job history includes in title one of the following: "manager" "vp" "president" "director" "administrator" "administrative" "executive" "coo" "chief operating" "operation" "secretary" "founder" "clerk" "division md" "employee" "associate" "head of division"
Background marketing	Equals to one if job history includes in title one of the following: "marketing" "publisher" "mktg" "sales" "brand manager" "regional manager" "communication" "merchandising" "comms" "distribution" "media"
Background military	Equals to one if job history includes in title one of the following: "captain" "soldier" "lieutenant" "admiral" "military" "commanding" "commander" "commandant" "infantry" "veteran" "sergeant" "army"
Background politician	Equals to one if job history includes in title one of the following: "politician" "senator" "political" "deputy" "governor"
Background science	Equals to one if job history includes in title one of the following: "researcher" "medical" "doctor" "scientist" "physician" "engineer" "biologist" "geologist" "physicist" "metallurgist" "science" "scientific" "pharmacist"
Background technology	Equals to one if job history includes in title one of the following: "technology" "software" "programmer" " it " "chief information officer" "database" "system administrator" "developer"
Bonus	Annual bonus payments (in thousands)
Busy	Equals to one if directors sits on three or more boards
CEO	Equals to one if director is the company's CEO
Chairman	Equals to one if director is chairman of the board
Compensation chair	Equals to one if director is chair of the compensation committee
Compensation committee	Equals to one if director is a member of the compensation committee
Employer contribution	Employers Defined Retirement/Pension Contribution
Equity linked remuneration ratio	Equity Linked Compensation as a proportion of total compensation for the individual based on the closing stock price of the last annual report
Foreign	Equals to one if director's nationality is not American
GenBBB	Equals to one if director was born between 1946 and 1964
GenDepBB	Equals to one if director was born in or before 1926
Gender	Equals to one if director is male
GenMature	Equals to one if director was born between 1927 and 1945
GenX	Equals to one if director was born between 1965 and 1980

GenY	Equals to one if director was born in 1981 or after
Governance chair	Equals to one if director is chair of the governance committee
Governance member	Equals to one if director is a member of the governance committee
HistInternational	Equals to one if job history includes a position outside the United States
Independent	Equals to one if director is not an executive director
Lead independent director	Equals to one if director is lead independent director
Mean past voting outcome	Average shareholder support during the first three years of tenure for previous board positions (starting in 2002). Source: ISS Voting Analytics
Mean_support_3yrs	Average shareholder support over the first three years of tenure. Source: ISS Voting Analytics
Network size	Network size of director (number of overlaps through employment, other activities, and education)
Nomination chair	Equals to one if director is chair of the nomination committee
Nomination member	Equals to one if director is a member of the nomination committee
Number connections	Number of established connections to incumbent board members prior to joining the board
Number qualifications	Number of qualifications at undergraduate level and above
Other chair	Equals to one if director is chair of a committee other than compensation, audit, governance or nomination
Other member	Equals to one if director is a member of a committee other than compensation, audit, governance or nomination
Other compensation	Value of annual ad hoc cash payments such as relocation or fringe benefits awarded during last reporting period (in thousands)
Perf to total compensation	Performance to total - Ratio of Value of LTIPs Held to Total Compensation
Salary	Base annual pay in cash (in thousands)
Timeretirement	Time to retirement (assumed to be 70 years old)
Tot Current Nb Listed Boards sitting on	The number of Boards of publicly listed companies that an individual serves on
Tot Current Nb Other Boards sitting on	The number of Boards for organizations other than publicly listed or private companies that an individual serves on
Tot Current Nb Unlisted Boards sitting on	The number of Boards of private companies that an individual serves on
Tot Nb Listed Boards sat on	The number of Boards of publicly listed companies that an individual has served on
Tot Nb Other Boards sat on	The number of Boards for organizations other than publicly listed or private companies that an individual has served on
Tot Nb unlisted Boards sat on	The number of Boards of private companies that an individual has served on
Total Compensation	Salary + Bonus
Total director compensation	Salary plus Bonus plus Other Compensation plus Employers Defined Retirement/Pension Contribution
Total equity linked wealth	A valuation of total wealth at the end of the period for the individual based on the closing stock price of the last annual report
Value of shares held	Value of shares held at the end of the reporting period for the individual based on the closing stock price of the annual report

A.4.2. Board-level features

Source: BoardEx except if stated otherwise
(as of when the director joins the board)

Variable	**Definition**
Attrition rate	Number of Directors that have left a role as a proportion of average number of Directors for the preceding reporting period
Average age	Average age of directors on the board
Average nb independent directors	Fraction of non executive directors on the board
Average nb qualifications	Average number of qualifications at undergraduate level and above of directors on the board
Average network size	Average network size of directors on the board (number of overlaps through employment, other activities, and education)
Average tenure	Average board tenure of directors on the board
Average tenure of incumbent directors	Average time in company for executive and non-executive directors on the board
Avg tot current nb listed boards	The average number of boards of publicly listed companies directors currently serve on
Avg tot nb listed boards sat on	The average number of boards of publicly listed companies directors have served on
Board Pay Slice - salary	Tot indep comp/ CEO salary
Board Pay Slice - total	Tot indep comp/ CEO total compensation
Board size	Number of directors on the board
BOSS	Dichotomous variable equal to one if the CEO is also the chairman of the board and the President
CEO bonus	CEO's bonus
CEO salary	CEO's salary
CEO total compensation	CEO total compensation (salary plus bonus)
Chairman duality	Dichotomous variable equal to one if the CEO is chairman of the board
Count Female	Number of women on the board
Gender ratio	The proportion of male directors
Nationality Mix	Proportion of Directors from different countries
Nb independent	Number of independent directors
Stdev age	Standard deviation of directors' age
Stdev current listed board	Standard deviation of the number of listed boards each director currently serves on
Stdev listed board sat on	Standard deviation of the number of quoted boards sat on for all directors on the board
Stdev number qualifications	Standard deviation of the number of qualifications at undergraduate level and above for all directors on the board
Stdev Time in Company	Standard deviation of time in the company for all directors on the board
Stdev Time on Board	Standard deviation of time on board for all directors on the board
Succession Factor	Measurement of the Clustering of Directors around retirement age
Tot indep comp	Sum of all independent directors' total compensation
Tot indep comp scaled	Sum of all independent directors' total compensation divided by the number of independent directors

A.4.3. Firm level features

Source: Compustat /CRSP except if stated otherwise
(as of when the director joins the board)

Variable	Definition
Current assets	current assets - Total
Acquisitions	acquisitions -
Auditor	Dichotomous variable for each auditing firm
CAPX	capital expenditures -
CEOSO1	Equals to one if the CEO is exempt from filing Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002
CFOSO1	Equals to one if the CFO is exempt from these filing Certification Documents
CEOSO2	Equals to one if the CEO has not filed Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002
CFOSO2	Equals to one if the CFO has not filed these Certification Documents
CEOSO3	Equals to one if the CEO has filed Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002
CFOSO3	Equals to one if the CFO has filed these Certification Documents
Equity (ordinary)	ordinary equity - Total
Cash	cash -
Cash and ST investments	cash and short term investments -
COGS	cost of good sold -
Shares outstanding	common shares outstanding -
Dividend payer	dichotomous variable equal to 1 if the total amount of dividends to ordinary equity > 0
LT debt	long term debt - Total - *Source* : Compustat
Depreciation	depreciation and amortization -
Dividends	total amount of dividends to ordinary equity
EBIT	Earnings Before Interest and Taxes
EBITDA	Earnings Before Interest
Firm age	time since IPO or first occurrence on CRSP
Inventories	Inventories - Total
Current liabilities	Current liabilities - Total
Leverage	Total long term debt / total assets
Ln (nb institutional blocks)	Logarithm of one plus the number of institutional blockholders.
Ln (nb institutional owners)	Logarithm of one plus the number of institutional investors.
M/B	(common shares outstanding * stock price)/ ordinary equity
Minority interest	Minority interest
Mkt value	Market value
NI	Net income
Price (calendar)	Price Close - Annual - Calendar
Price (fiscal)	Price Close - Annual - Fiscal
Product market fluidity	Product market fluidity. Hoberg and Phillips
Profitability	ebitda/total assets
Block ownership %	Fraction owned by blockholders.
Institutional ownership %	Fraction owned by institutional investors.
Largest inst. shareholder %	Fraction owned by largest institutional investor.
Largest 10 inst. shrholders %	Fraction owned by top ten institutional investors.
Largest 5 inst. shrholders %	Fraction owned by top five institutional investors.
Retained earnings	Retained earnings
Retained earnings (restated)	Retained earnings restatements
12-month return	Cumulative stock return in the twelve months leading up to the appointment.
Revenue	Revenue - Total
ROA	Net income / total assets
ROE	Net income / ordinary equity
Sales	Net sales - Total
Equity (total)	Stockholders' equity - Total
Settlements	Settlement (Litigation/Insurance) After-tax

Total Assets	Total assets -
Working capital	Working capital
Extraordinary items	Extraordinary items
R&D	R&D expenses

A.4.4. Industry and market level features
Source: Compustat /CRSP except if stated otherwise
(as of when the director joins the board)

Variable	**Definition**
Excess returns 12-month leading up to appointment	cumulative stock return in the twelve months leading up to the appointment minus cumulative returns on the S&P500 in the twelve months leading up to the appointment
Industry ROA	return on assets of firms with same 3-digit SIC code
Mkt12	cumulative returns on the S&P500 in the twelve months leading up to the appointment
Tnic3*	3-digit, text-base industry classifications from Hoberg and Phillips (2010, 2016)