NBER WORKING PAPER SERIES

SELECTING DIRECTORS USING MACHINE LEARNING

Isil Erel
Léa H. Stern
Chenhao Tan
Michael S. Weisbach

Working Paper 24435
http://www.nber.org/papers/w24435

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2018

We thank Shan Ge, Ben Hermalin, Nadya Malenko, Tracy Yue Wang, Miriam Schwartz-Ziv, and conference and seminar participants at the University of Washington, the 2017 Pacific Northwest Finance Conference, the 2017 Women Professor in Finance Conference at NYU Stern, the 2017 NABE TEC Conference and the 2018 University of Miami-AFFECT conference. Special thanks to Ronan Le Bras for providing invaluable help throughout the project. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Selecting Directors Using Machine Learning
Isil Erel, Léa H. Stern, Chenhao Tan, and Michael S. Weisbach
NBER Working Paper No. 24435
March 2018
JEL No. G34,M12,M51

ABSTRACT

Can an algorithm assist firms in their hiring decisions of corporate directors? This paper proposes a method of selecting boards of directors that relies on machine learning. We develop algorithms with the goal of selecting directors that would be preferred by the shareholders of a particular firm. Using shareholder support for individual directors in subsequent elections and firm profitability as performance measures, we construct algorithms to make out-of-sample predictions of these measures of director performance. We then run tests of the quality of these predictions and show that, when compared with a realistic pool of potential candidates, directors predicted to do poorly by our algorithms indeed rank much lower in performance than directors who were predicted to do well. Deviations from the benchmark provided by the algorithms suggest that firm-selected directors are more likely to be male, have previously held more directorships, have fewer qualifications and larger networks. Machine learning holds promise for understanding the process by which existing governance structures are chosen, and has potential to help real world firms improve their governance.

Isil Erel
Department of Finance
Ohio State University
832 Fisher Hall
2100 Neil Avenue
Columbus, OH 43210
erel@fisher.osu.edu

Léa H. Stern
Department of Finance and
Business Economics
Foster School of Business
University of Washington
Seattle, WA 98195
leastern@uw.edu

Chenhao Tan
Department of Computer Science
University of Colorado
Boulder, CO 80309
chenhao@chenhaot.com

Michael S. Weisbach
Department of Finance
Fisher College of Business
Ohio State University
2100 Neil Ave.
Columbus, OH 43210
and NBER
weisbach.2@osu.edu

1. Introduction

A company's board of directors is elected by its shareholders and is legally responsible for managing the company. In principle, the board of directors reports to the shareholders and will maximize the firm's value. In practice, however, there is much variation in director quality and the extent to which they serve the shareholders' interest.[1]

Many of the problems with boards occur because of the process by which directors are selected. The board selection process has been discussed since at least Berle and Means (1932) and is still a major source of debate.[2] Ultimately, the issue stems from the fact that despite the checks and balances built into a public corporation's governance system, the CEO nonetheless often effectively controls the board's decisions, including the selection of new directors. In practice, appointed directors are almost always supporters of the CEO and his policies.[3] Aside from occasional proxy contests, shareholders have virtually no control over the choice of the directors whose mandate is to represent their interests.

In this paper, we consider a potential alternative approach to select directors, using algorithms that rely on data on firms, potential directors, and their attributes, to identify the quality of directors being considered for a for a given firm's board. In this paper, the quality and the performance of directors refers to their ability to gather shareholder support. The best possible director is therefore the director who receives the highest approval from shareholders.

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,

[1] The literature on boards' performance and the circumstances under which they do and do not make value-maximizing decisions is enormous. 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.

[3] There is ample anecdotal evidence that the decision does not ultimately rests with the nominating committee as the CEO typically holds a veto power. See Shivdasani and Yermack (1999). See also Cain, 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.

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 hiring 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.

A crucial element of any algorithm designed to select the directors who would be most valuable to a particular firm is a process for assessing a director's performance. 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. One measure of an individual director's performance is the level of shareholder support that a particular director receives during shareholder elections. The recent literature on director elections documents that the level of shareholder support received by a director is positively related to measures of director performance (see Cai et al. (2009), Fischer et al. (2009), Iliev et al. (2015), Aggarwal et al. (2017), and Ertimur et al. (2017)). We therefore use the level of shareholder support a *new* independent director receives in *subsequent* elections as a market-based measure of an individual director's performance. Since directors' fiduciary duty is to serve shareholders' interests, their popularity among shareholders is an appropriate metric for evaluating them.

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

Using voting totals in subsequent elections as the market's assessment of a new director's performance, we construct algorithms to select the directors who are most likely to receive the highest approval from shareholders in the future. These algorithms rely on methodological approaches now common in the computer science literature (i.e., lasso, ridge, random forest, neural networks, and *XGBoost*). On our sample of public firms, we fit each model on a "training" subsample (directors appointed between 2000 and 2011), and then compare the predictions to the observed data on a "testing" subsample (directors appointed between 2012 and 2014).

We find that these algorithms make accurate out-of-sample predictions of shareholder support in director elections. For example, our model based on *XGBoost*, predicts the average fraction of votes in favor of a director with a mean absolute error of about 4% out of sample. More importantly, the directors the algorithm predicted would do poorly subsequently 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 receive out of sample, while the OLS model cannot.

While it is important to be able to make accurate out of sample predictions if the goal is to assist decision makers, it is not sufficient as we are able to assess how well our algorithm predicts only on cases for which we observe the actual outcome. However, the set of observations for which we observe the actual outcome is not a random subset of observations. We only observe the performance of directors who were actually hired by the board and do not observe such performance for potential candidates who were not hired. 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 take into account features that are not observable to our algorithm in their decision process, the distribution of outcomes in the set with observed labels (hired directors) may differ from that in the set with missing labels (not hired directors), even if they share exactly the same observable characteristics.

To determine whether our algorithm can assist decision makers, we compare the performance of potential candidate directors to that of the individuals who did join the company's board. Each new board appointment is matched to its own realistic pool of potential candidate directors which consists of directors who joined a smaller neighboring company in the past year or the following year. Presumably these potential candidates were available and would have found the opportunity to be on the board of a larger nearby company to be attractive, since larger companies usually pay their directors better than smaller companies, and their directorships are more prestigious. The idea is to check for example whether the potential candidates our algorithm identified as promising candidates would indeed be good directors.[5] Because the focal board did not hire them, we do know whether they indeed would be (that is the crux of the selective labels problem). It could be that the focal firm did not hire them because it relied on unobservables that would effectively make them poor directors.

Although we do not observe the performance of potential candidates that were not hired, 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. We exploit the observation that there is on average little variation in director performance -- the fraction of votes received in director elections -- across the different boards they joined during our sample period. If they performed well on the board they ended up joining for example, this reassures us that there is not something fundamentally wrong that our algorithm could not observe and that would prevent them from being successful directors.

Of course, quasi-labels are not a perfect substitute for the level of support a director would have gathered on the focal board. However, it is important to keep in mind that the main purpose of constructing candidate pools and using quasi-labels is not to identify the absolute best director, but rather to *evaluate* the predictive accuracy of our algorithm in the presence of the selective labels problem and

[5] The algorithm considers the firm, board and committee assignments of the focal firm when making a performance prediction for a potential candidate who joined a different board.

unobservables.[6] In particular, the results from this exercise are not affected by the endogenous matching between boards and directors as they would be if the purpose of the exercise was to draw causal inference. Of course, it is possible that due to the endogenous nature of the board-director match, some quasi-labels might be inflated, i.e. the performance of the available candidate may not be as high on the focal board. However, our results go both ways: whereas the observed performance of directors our machine-learning models predicted would do poorly ranks low in the distribution of quasi-labels, that of directors predicted to do well ranks high.[7] The symmetry of our results provides reassurance that our quasi-labels are not systematically inflated due to the endogenous nature of the board-director match. In our setting, quasi-labels are simply a noisier benchmark than the ideal benchmark. The results from this exercise confirm that our algorithms are able to identify the directors who will gather high shareholder approval rates at a particular board and those who will not.

As an alternative measure of director performance, we use firm profitability following director appointments. 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. Using firm profitability as the measure of director ability, the machine-learning models also predict well out of sample, while OLS does not. 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 subsequent firm profitability that is significantly lower than that for directors with high predicted shareholder votes.

Machine-learning tools have the potential to help answer many unanswered questions in the social sciences. 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

[6] The algorithm could give a prediction for *any* candidate, not only those in the potential candidate pool. In addition, decision makers could still rely on unobservables to hire among candidates predicted to do well.

[7] In contrast, directors rank around the 50th percentile in the distribution of quasi-labels, regardless of whether OLS predicted they would perform well or not.

are actually chosen by firms. We find that when compared 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. Therefore, a plausible interpretation of our results is that firms that hire predictably unpopular directors tend to choose directors who are like existing directors, while the algorithm suggests that adding diversity would be a better idea. The fact that our algorithm can generate a prediction for any potential candidate has the additional benefit of broadening the set of potential directors that companies could consider, thereby opening up board seats to a new set of candidates with more diverse backgrounds and experiences who would have otherwise been overlooked.[8]

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 compare the performance of our prediction models and provide various characteristics that affect directors' performance. The fifth section includes an 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 several algorithms that are designed to make an ex ante prediction of directors' level of shareholder support, *Y*, over the first three years of their tenure. The algorithms use a set of observable director, board and firm features, *W,* that are available to the nominating committee at the time of the hiring decision. The algorithms are among the most commonly used in the machine learning literature: *lasso*, *ridge*, *random forest*, *neural networks* and *XGBoost*. We train each of these algorithms, i.e.

[8] Field, Souther, and Yore (2017) show that diverse directors are significantly less likely to hold key positions on the board than their non-diverse counterparts with similar professional qualifications.

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 training set (in-sample data) and the data from 2012 to 2014 testing set (out-of-sample data). The variable the algorithms try to predict is the average level of shareholder support over the first three years of director tenure.[9]

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

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. In contrast, *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 Appendix 1 for more details).

2.1.2. *Random Forest*

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, we estimate multiple trees 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

[9] The data on director elections in ISS Voting Analytics is not always available for all election years. Using the average of available election outcomes over the first three years of tenure allows us to expand the number of usable observations. The distribution of shareholder support is essentially the same, whether it is one year or three years after initial appointment. In other words, there is no trend in shareholder votes over the first years of tenure. We obtain similar results using voting data at year one, year two or year three instead of using the average over the first three years.

data point can then 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 takes an average of the predictions from all the decision trees.

2.1.3. *Gradient Boosting Trees*

Similar to random forest, *gradient boosting trees* 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.

2.1.4. *Neural Networks*

Artificial neural networks are designed to mimic the way the brain processes information. A neural network 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. Figure 1 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 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). Similar to the neurons in the hidden layers, the output neuron judges its input via an activation function and decides from which neurons it will accept the triggered values. The output is the weighted sum of the activated neurons in the last hidden layer. Training a network involves modifying the weights on the synapses so as to minimize a cost function (typically 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: the problem of having performance data at the company in question for only directors who were actually selected, as well as the fact that real world decision makers often make their decisions based on variables that are unobservable to us. To formalize these concepts, we develop a framework similar to that presented by Kleinberg et al. (2017) and present it in Appendix 2.

In our setting, the selective labels problem refers to the limitation that we can only observe how well our algorithm performs out of sample on instances for which we observe the outcome. However, the subset of instances for which we observe the outcome is not a random subset of instances but instead the result of the selection made by decision makers. Presumably, decision makers take into account features that are not observable to the algorithm. Therefore, directors who were hired, although they might share the same exact observable features as other directors not hired, 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 hire a candidate based on some characteristics unobservable to the algorithm that would make this candidate a poor choice. Since we cannot observe these factors, they could lead to different average outcomes for hired vs. not hired, even if both are identical on the basis of observable characteristics.

Our empirical strategy to address these concerns involves designing a pool of potential candidates for each vacant board position. The goal is to evaluate the algorithm's predictions of the performance of directors who were actually hired. 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 hired actually performed compared to those alternatives.

To construct candidate pools, we consider directors who joined the board of a neighboring company around the same time.[10] 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 and remuneration of being a director tends to increase with company size (see Masulis and Mobbs, 2014).[11]

There were 19,464 directors in our sample who joined multiple boards of public companies over our sample period. The median standard deviation of shareholder support over the first three years of their tenure across the different boards they join is .024. This low standard deviation suggests that on average directors tend to receive similar support from shareholders on the different boards they join. Therefore, although we do not observe the performance of potential candidates that were not hired, we observe what we refer to as their *"quasi-label"*: their performance on the board they actually joined.

We find that directors our algorithm predicted would receive low shareholder support do poorly when compared to available alternative candidates. Specifically, we show that the actual performance of directors our machine-learning models predicted would do poorly (well) ranks low (high) in the distribution of quasi-labels. In contrast, directors rank around the 50th percentile in the distribution of quasi-labels, regardless of whether OLS predicted they would perform well or not.

Obviously, quasi-labels are not a perfect substitute for the level of support a director would have gathered on the focal board. However, our empirical strategy only uses candidate pools and quasi-labels to deal with the selective labels problem and evaluate the performance of the algorithm. For our purpose, quasi-labels are a noisier benchmark than our ideal benchmark. The basis for using quasi-labels is to check whether the potential available candidates our algorithm identified as promising candidates for example would indeed be good directors. It could be that the focal board did not hire them because it relied on unobservables that would effectively make them poor directors. If they performed well on the

[10] A neighboring firm is a defined as a firm whose headquarters is within 100 miles of the focal firm's headquarters.
[11] There are on average 192 candidates in a candidate pool.

board they ended up joining, however, this rules out the possibility that there is something fundamentally wrong that our algorithm could not observe and that would prevent them from being successful directors. Of course, it is still possible that due to the endogenous nature of the board-director match, the quasi-labels might be inflated, i.e. the performance of the available candidate may not be as high on the focal board. However, our results go both ways: hired directors predicted to do poorly rank low compared to how other candidates would have performed, but we also find that hired directors predicted to do well rank *high* compared to alternatives. The symmetry of our results provides reassurance that our quasi-labels are not systematically inflated due to the endogenous nature of the board-director match.

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 must rely on a market-based measure that incorporates the information of market participants. Therefore, we use the proportion of votes that an individual receives in director elections, 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 variation in vote outcomes 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, the results of director elections appear to have real consequences, even if the elections are not contested and the nominated directors are 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 vote totals 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 the concerns of the shareholders. Overall, the recent literature strongly suggests that vote totals do reflect perceptions of director quality, that directors care about these perceptions, and take actions designed to influence them. Therefore, in this paper, the quality and the performance of directors refers to their ability to gather shareholder support and we assume that the best possible director is the director who receives the highest approval from shareholders.

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. Given these requirements, it is natural to consider a sample of boards from large, publicly-traded, U.S. firms. We identify 41,051 new independent directors appointed to 4,887 unique corporate boards between 2000 and 2014 using *BoardEx*. *BoardEx* is also our main data source for director as well as board-level characteristics. We obtain firm-level characteristics from *Compustat* and *CRSP*. Average market capitalization of firms in our sample is $6.6 billion.

We obtain data on the level of shareholder support for individual directors from *ISS Voting Analytics*. Again we concentrate on new directors only and use the average fraction of votes in favor of a given director over all votes cast for reelection over the first three years of her/his director's tenure. We have the voting outcome (average over the first three years of tenure) for 26,024 new director appointments.

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.[12] However, some recent research provides evidence on the decline in this influence. 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.[13]

3.3. Summary Statistics

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 (Cai et al., 2009, Fischer et al., 2009, Iliev et al., 2015, Aggarwal et al., 2017 and Ertimur et al., 2017). 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. Figures 2 and 3 show that although shareholder support in uncontested elections is typically very high, shareholders do on occasion oppose newly hired directors. 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.

Table 2 further shows that the frequency of shareholder discontent varies by director and board characteristics. For example, directors receive less than 95% of the votes in 28% of the cases, but that number is 29% for male directors and 23% for female directors. Similarly, busy directors (serving on three or more boards) experience lower shareholder support more frequently than non-busy directors.

---

[12] 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).

[13] 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.

Whereas knowledge of previous research on boards of directors can guide us to select characteristics which are likely to matter (the variables in Table 2 were hand curated), theory is unhelpful at guiding us how to interact those variables in order to make accurate predictions of director performance. 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. In addition, there is not one ideal governance structure for all firms. Instead, each firm faces its own governance optimization problem (see Coles, Daniel, and Naveen (2008)). Consequently, it is appropriate to use an estimation procedure such as machine learning that does not impose the specific form for the relationship between potential explanatory variables.

4. Evaluating Machine Learning Predictions of Director Choice

4.1. Model Specification

Using this sample, 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. We first "train" each algorithm on the 2000-2011 portion of our sample consisting of 20,969 new director appointments (14,374 unique directors) at 2,628 firms. Training involves having the algorithm determine which combinations of the sample variables best predict future performance. We then evaluate the models' predictions on the 2012-2014 portion of our sample (5,667 new director appointments -- 4,004 unique directors at 526 firms) and compare the predictions to those from an OLS model. We emphasize that all the results presented below are for the 2012-2014 subsample of director appointments, which is not overlapping with the 2000-2011 subsample on which the models 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. This table presents the average observed shareholder support for directors in various percentiles of predicted shareholder support for five machine-learning algorithms and for an OLS model as a benchmark. A characteristic of a good model for predicting performance is that actual performance is an increasing function of predicted performance. Table 3 indicates that average observed shareholder support does increase across predicted percentiles of shareholder support for each model except for the OLS one. The average observed outcome of directors in the bottom of the predicted performance distribution using the OLS model is actually higher than that of directors in the top of the predicted performance distribution.[14] This pattern highlights the difference between the machine learning model and OLS in their ability to predict future performance.

Among the alternative machine learning algorithms, *XGBoost* performs best at predicting the subsequent success of directors. It is also the one that yields the lowest mean absolute error.[15] Figure 4 shows the average observed level of shareholder support for directors across the ten deciles of predicted performance for OLS and *XGBoost* in the 2012-14 testing period. The figure clearly shows how the mean shareholder support for a director is a monotonic function of the predicted one for the *XGBoost*, but not for the OLS model. The difference in 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: directors predicted to be in the bottom decile as predicted by *XGBoost* have an average observed shareholder support of 93%, whereas the average observed support is 98% for directors in the top decile of predicted performance.

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 Mullanaithan and Spiess (2017), who claim that machine learning is the preferred approach for prediction problems such as this one. One possible reason why the machine learning models do so much better is because they let the data decide which transformations of which variables are relevant, while in OLS (or any other standard econometric technique), the researcher has to specify the structure of the equation

---

[14] See the appendix for details on the OLS model used.

[15] *XGBoost* is an algorithm known for generating state-of-the-art results on a variety of problems. It was the most often used algorithm among the winning solutions in the 2015 machine learning Kaggle competition.

before estimating it. Machine learning, by letting the data speak about the underlying relationships among the variables, ends up fitting the data much better and also does better at predicting future 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. However, as discussed in Section 2.2., accurate out of sample predictions are not sufficient to argue that our algorithm could assist firms in their hiring decisions of corporate directors. In addition, the algorithm should be able to assess how the hired directors performed compared to the way alternatives would have performed. In this section, we present results comparing the performance of directors actually chosen to that of potential alternative choices.

We rely on the following procedure: we rank all hired directors in our test set according to their predicted performance. 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.[16] 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, 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 to the quasi-labels of promising candidates. Figure 5 provides an illustration of this procedure.

Table 4 presents the median ranks in the distribution of quasi-labels for directors in the bottom and top deciles of predicted levels of shareholder support for several machine-learning algorithms, as well as for an OLS model. All machine-learning models find that the directors they predicted would do poorly

[16] To arrive at predictions for potential candidates, our algorithms use the board and firm characteristics of the focal firm and the committee assignments of the director who was hired at the focal firm but the individual director characteristics of the potential candidate.

indeed rank lower when compared to candidates' performance than those they predicted would do well. In contrast, the predictions from the OLS model are uninformative about subsequent performance; directors rank around the 50th percentile in the distribution of quasi-labels, regardless of whether OLS predicted they would perform well or not. *XGBoost* again appears to be the preferred algorithm, because it can best discriminate *ex ante* the directors who will do well from those who will not.

The median director predicted by the *XGBoost* algorithm to be in the bottom decile of shareholder support ranks at the 38th percentile in the distribution of quasi-labels. In contrast, the median director predicted to be in the top decile ranks at the 65th percentile in the distribution of quasi-labels. Figure 6 illustrates these results and contrasts them with the results from an OLS model. There is virtually the same performance for the top and bottom deciles of predicted performance when performance is forecasted by OLS. In contrast, the candidates identified by *XGBoost* as having high performance performed noticeably better than the candidates predicted to have low performance. Machine learning models can predict, at least to some extent, whether a given individual will be successful as a director in a particular firm.

We emphasize that board-director matches are not exogenous, and are likely chosen to maximize the "fit" between directors and firms. However, our empirical strategy uses the candidate pools to *evaluate* the performance of the algorithm and this pool is limited to directors who within the prior or subsequent year joined the board of a smaller nearby company. In practice, though, the algorithm could give a prediction for any candidate, not only those in the candidate pool. On average, there are about two hundred candidates with quasi-labels in candidate pools defined in this manner. The fact that our quasi-labels do not systematically outperform most of the director performance provides reassurance that our quasi-labels are not systematically inflated due to the endogenous nature of the board-director match, alleviating endogeneity concerns. We find similar results to the ones reported in the tables, when we restrict potential candidates to directors who joined a board in the same industry (still on a smaller nearby company, within 100 miles, around the same time). There are on average 25 candidates for which we have quasi-labels in these restricted candidate pools.

To summarize, 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. These results suggest that our algorithm has the potential to improve on real world boards' hiring decisions. It is important to note that this work is a first pass exercise to show the potential of machine learning algorithms in shedding light on the quality of boards' hiring decisions. A more powerful algorithm and/or better data would likely predict future performance even more accurately.

4.4. Robustness on Predicted Performance

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. The median appointed directors appointed who is predicted to be in the bottom decile of performance by the *XGBoost* algorithm turned out to have poor performance ex post, in the 37th percentile of all quasi-labels. In contrast, the median director predicted to be in the top decile of performance subsequently is in the 67th percentile. 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. Characteristics that Affect Director Performance

The machine learning models are able to predict which directors are likely to receive more votes in subsequent elections. These voting totals presumably reflect shareholder satisfaction with director performance, which is a market-based measure of director performance. The predictions come from a sophisticated algorithm based on the variables in our database. One of the differences between this machine learning approach and 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 subsequent performance. To understand which characteristics affect director performance, we consider the predictions coming from the machine learning models and evaluate the extent to which particular variables are associated with high and low predicted performance.

4.5.1. Univariate Comparisons

Table 5 provides some guidance about which director features are valued by the 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 5 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 have fewer qualifications, be male, sit on more current boards, have sat on more boards in the past, and have received lower shareholder support in previous elections for other boards they sat on. 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.[17]

Board-level variables that affect predicted voting totals likely reflect perceptions of the quality of governance in a particular firm. Longer average director tenure, which is likely to reflect an entrenched board, is associated with lower predicted shareholder support. In contrast to the arguments of Yermack (1996), *larger* boards tend to be predicted to have higher support from shareholders.

Firm level variables affecting voting tend to reflect the performance of the firm, with better performance leading to higher predicted shareholder support. Both average ROE and prior stock returns for the bottom predicted decile of shareholder support are substantially worse than for the top decile of

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

predicted shareholder support. In addition, firms in the top decile of predicted support tend to pay much larger dividends, which could reflect that higher dividends are associated with both higher profitability and stability, both of which appear to lead to higher shareholder support in elections.

4.5.2. Multivariate Comparisons

Because the director and firm characteristics are not independent from one another, it is important to understand which characteristics are driving the predictions of the machine learning algorithms. To do so, we estimate regressions of the predicted values of subsequent votes from the *XGBoost* model. As independent variables, we include both firm level and director variables. The coefficients from these regressions reflect the characteristics that the algorithm associates with higher subsequent performance.

We present estimates of this regression in Table 6. Director level variables that are related to predicted subsequent shareholder support are gender, whether the director is busy and the number of listed boards. In particular, our algorithm suggests that male, busy directors who are on a number of boards are likely to be less popular with shareholders. This pattern could reflect the commonly stated concern of shareholders that directors are too often in a "old boys club", in which 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)).

Board level variables that are significantly related to the predictions of shareholder support for a director are the size of the board, the average tenure of incumbent board members, and the fraction of women on the board. 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.

We emphasize that the purpose of this exercise is simply to provide some information on which characteristics affect the predictions more. We note however, that in these regressions, the R^2 is fairly low (below 40%), which speaks to the importance of feature interactions and non-linearities that *XGBoost* relies on to generate its predictions.[18]

4.5.3. Director Characteristics That Are Overvalued

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 hired 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, our algorithm provides a diagnostic tool that can help evaluate the way in which directors are chosen.

In Table 7, we report characteristics of directors who were hired, but whom the algorithm predicted would do poorly (and subsequently did poorly). Compared to more promising candidates as identified by our algorithms, predictably unpopular directors are more likely to be male, have fewer degrees post undergraduate, have a larger professional network, have sat on more boards in the past, to be sitting on more current boards, have a background in finance, and have received lower shareholder support in the past. This comparison is for each new board seat, holding committee assignments constant for hired directors and candidates. In addition, these are averages across all new board positions.

These results highlight the features that are 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 traditional literature on boards going back to Smith (1776) and Berle and Means (1932), managers and existing directors could implicitly collude to hire 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

[18] Kleinberg et al. (2017) make a similar argument when they report results of linear projections of predicted crime rates in their bail application.

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 sophisticated algorithms we present. Understanding why firm-selected directors differ from algorithm-selected directors is likely to be an important topic of future research.

4.6. Firm Profitability as a Measure of Performance

The algorithms we have presented are trained to predict shareholder support in subsequent elections. The advantage of this approach is that shareholder support is a market-based measure of each director's performance, so reflects perceptions of how a particular director contributes to performance. An alternative approach would be to train the algorithms to evaluate the effect of a particular director on expected profitability. In principle, the two approaches would yield similar results, since shareholders vote on directors based on their perceptions of how that director would affect profitability.

To evaluate the extent to which models trained on profitability differ from those trained on shareholder approval of directors, we train an *XGBoost* algorithm to predict a firm level profitability measure (EBITDA/Total Assets) three years after the director appointment on the 2000-2011 "training" subsample. We sort the directors into deciles based on expected profitability. We report the actual profitability as well as the shareholder voting support in the subsequent 2012-2014 period in the first two rows of Table 8.[19]

The model trained to predict expected profitability in the subsequent period actually 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 hired directors in the 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

[19] The correlation of EBITDA/Total Assets with the shareholder support measure is .14 (p-value: 0.000).

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.

These results suggest that the choice of training the algorithm on shareholder support is not crucial for the algorithm to be able to select high quality directors. 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.[20] In addition, for the algorithm trained on shareholder support that we discussed above, we consider whether it can predict future profitability in addition to future shareholder support. We break the sample into deciles based on the algorithm's predictions of shareholder support during the training period (2000-2011), and present average subsequent shareholder support in the subsequent three years (2012-2014) as well as the average profitability during this period for each decile. We present these averages in the third and fourth columns of Table 8.

As discussed above, this algorithm is successful in predicting future shareholder support: average shareholder support in the lowest decile is 92%, compared to 97.7% in the top decile. In addition, it also predicts future profitability. Firms that hired directors in the bottom decile of predicted performance have an average profitability of -0.3%, whereas firms that hired directors in the top decile of predicted performance have an average profitability of 10%.

[20] A possible concern is that differences in predicted performance could be primarily driven by differences in performance of the firms hiring the directors, especially those firms that are doing extremely poorly. To evaluate whether the algorithm's ability to select directors well occurs only because of the poorly performing firms in our sample, we replicate the results using only firms that experience positive excess returns leading up to the nomination of the director. The results are similar to those reported above: on this subsample, the algorithm can successfully predict future director performance.

5. Summary and Discussion

In this paper, 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 three 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, we use the selections from the algorithms as benchmarks to understand the process through which directors are actually chosen and 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. First, 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 the shareholder elections. This vote, which is shown to be highly 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, 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.

We find that our machine-learning algorithms fit the data well. The realized performance following the appointment of a director is a monotonic function of the predicted performance. Using publicly available data on firm, board, and director characteristics, our *XGBoost* algorithm can accurately predict the success of individual 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 Mullanaithan and Spiess (2017) that machine learning is a promising approach for prediction problems in social sciences.

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 to diversify boards of directors since the 1990s for this reason (see Biggs (1996) and Carleton et al. (1998)).[21]

[21] 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

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. Consequently, it is easy to imagine that a machine-learning algorithm, which is much more sophisticated than the algorithms relied on by psychologists, would allow firms to improve their board selection process.

A natural question concerns the applicability of algorithms such as the one we developed in practice. The algorithms we present should be treated as "first pass" 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.[22] Institutional investors are likely to find this attribute particularly appealing and are likely to use their influence to encourage boards to rely on an algorithm such as the one presented here for director selections in the future.

---

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

[22] 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). 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 we believe our algorithm is not prone to propagating biases.

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. We expect (hope?) that in the future, more sophisticated versions of this algorithm will be used by boards to improve their choices of directors and the way in which corporate governance serves shareholders' interests.

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. In the context of lower skill workers, Hoffman, Kahn and Li (2017) find that managers who hire against test recommendations end up with worse average hires. Our results suggest that the same applies to the hiring of corporate directors. This lack of discretion and reliance on soft information in hiring could potentially minimize agency problems, and thus algorithmic selection could be a desirable process from the shareholders' perspective. 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.

Machine learning has revolutionized many fields. We believe that it is likely that corporate governance will be affected in the future as well. By providing a prediction of performance for *any* potential candidate, a machine-learning algorithm could *de facto* expand the set of potential directors and identify individuals with the skills necessary to become successful directors, who would have otherwise been overlooked. Consequently, we expect that in the not too distant future, machine-learning techniques will fundamentally change the way corporate governance structures are chosen, and shareholders will be the beneficiaries.

References

Adams, R. (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.

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

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.*

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.

Hastie, T., R. Tibsharani and M. Wainwright (2015) Statistical Learning with Sparsity: The Lasso and Generalizations. *Chapman and Hall/CRC.*

Hermalin, Benjamin E., and Michael S. Weisbach (1998) Endogenously Chosen Boards of Directors and Their Monitoring of the CEO. *American Economic Review*, 88(1): 96–118.

Hermalin, Benjamin E., and Michael 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, Gerard and Gordon Phillips (2010) Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies* 23 (10), 3773-3811.

Hoberg, Gerard and Gordon Phillips (2016) Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy* 124 (5) 1423-1465.

Hoffman, Mitchell, 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, Daniel (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*

Larcker, D., A. McCall, and G. Ormazabal (2015) Outsourcing Shareholder Voting to Proxy Advisory Firms. *Journal of Law and Economics* 58, 173-204.

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, Paul E. (1954) *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, University of Minnesota Press: Minneapolis.

Mitchell, Tom (1997) Machine Learning. *McGraw Hill*.

Mullanaithan, Sendhil and Jann Spiess (2017) Machine learning: An Applied Econometric Approach, *Journal of Economic Perspectives*, 31, Number 4, 87-106.

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.

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.

Appendix 1: Details of Some Algorithms used to Predict Performance

A1.1. 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 chosen via cross-validation to ensure generalization, i.e., accurate out-of-sample predictions.

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.[23]

A1.2. Random Forests

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.

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

A key difference between decision tree learning and Ridge and Lasso regression lies in 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 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 learn a random forest regressor, a number of decision tree regressors are fitted by randomly sample data from the training instances with replacement and also randomly sample a subset of features. The average values of all decision tree regressors is used to predict the value of an instance.

A1.3. Gradient Boosting Trees

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:

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

   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. Specifically, 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), $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda||w||^2$ 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.

Appendix 2: 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' hiring 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 that 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 hiring a director with specific characteristics; for example, they could derive private benefits from hiring 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 hiring a director with certain features, respectively.

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

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

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

The hiring 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' hired 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 hiring rule α that produces a higher payoff than the baseline: the outcome of board *j*'s actual hiring decision?

The difference in the expected payoffs between the two hiring 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 hired under the alternative hiring 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 hired director, their performance would be identical. This is the equivalent of assuming that unobservables, *z*, play no role in hiring 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 hired, although they might share the same exact observable features as other directors not hired, might differ in terms of unobservables. These unobservables could lead to different average outcomes for hired vs. not hired, 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 hires generated by the algorithm's hiring 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' hiring decisions. Our approach is to contrast those decisions to an alternative hiring 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 hiring 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 hiring decisions, we add the assumption that among those 150 alternative hires, 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.

Appendix 3: Data Definitions

A.3.1. 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
Avgtimeothco	The average time that a director sits on the board of quoted companies
Bkgd 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"
Bkgd CEO	Equals to one if job history includes CEO title
Bkgd 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"
Bkgd hr	Equals to one if job history includes in title one of the following: "hr " "recruitment" "human resource"
Bkgd law	Equals to one if job history includes in title one of the following: "lawyer" "legal" "attorney" "judge" "judicial"
Bkgd 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"
Bkgd 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"
Bkgd 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"
Bkgd politician	Equals to one if job history includes in title one of the following: "politician" "senator" "political" "deputy" "governor"
Bkgd 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"
Bkgd 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
Connected to CEO	Equals to one if director was connected to CEO before joining the board
Connected to nominating com	Equals to one if director was connected to a member of the nominating committee before joining the board
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
Has_connections	Equals to one if director had at least one established connection to an incumbent board member prior to joining the board
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	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.3.2. 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 independent	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 time in company	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.3.3. Firm level features

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

Variable	**Definition**
act	current assets - Total
aqc	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
ceq	ordinary equity - Total
ch	cash -
che	cash and short term investments -
cogs	cost of good sold -
csho	common shares outstanding -
div_payer	dichotomous variable equal to 1 if the total amount of dividends to ordinary equity > 0
dltt	long term debt - Total - *Source* : Compustat
dp	depreciation and amortization -
dvc	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
invt	Inventories - Total
lct	Current liabilities - Total
leverage	Total long term debt / total assets
ln_numinstblockowners	Logarithm of one plus the number of institutional blockholders.
ln_numinstowners	Logarithm of one plus the number of institutional investors.
MB	(common shares outstanding * stock price)/ ordinary equity
mib	Minority interest
mkvalt	Market value
ni	Net income
prcc_c	Price Close - Annual - Calendar
prcc_f	Price Close - Annual - Fiscal
prodmktfluid	Product market fluidity. Hoberg and Phillips
profitability	ebitda/total assets
ratio_instblockown	Fraction owned by blockholders.
ratio_instown	Fraction owned by institutional investors.
ratio_maxinstown	Fraction owned by largest institutional investor.
ratio_top10instown	Fraction owned by top ten institutional investors.
ratio_top5instown	Fraction owned by top five institutional investors.
re	Retained earnings
rea	Retained earnings restatements
ret12	Cumulative stock return in the twelve months leading up to the appointment.
ret3	Cumulative stock return in the three months leading up to the appointment.

ret6	Cumulative stock return in the six months leading up to the appointment.
revt	Revenue - Total
roa	Net income / total assets
roe	Net income / ordinary equity
sale	Net sales - Total
seq	Stockholders' equity - Total
seta	Settlement (Litigation/Insurance) After-tax
Total assets	total assets -
wcap	Working capital
xi	extraordinary items
xrd	R&D expenses

A.3.4. Industry and market level features

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

Variable	**Definition**
excess_returns12	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
excess_returns3	cumulative stock return in the three months leading up to the appointment minus cumulative returns on the S&P500 in the three months leading up to the appointment
excess_returns6	cumulative stock return in the six months leading up to the appointment minus cumulative returns on the S&P500 in the six 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
mkt3	cumulative returns on the S&P500 in the three months leading up to the appointment
mkt6	cumulative returns on the S&P500 in the six months leading up to the appointment
tnic3*	3-digit, text-base industry classifications from Hoberg and Phillips (2010, 2016)

Figure 1: Basic Structure of a Neural Network with Two Hidden Layers.

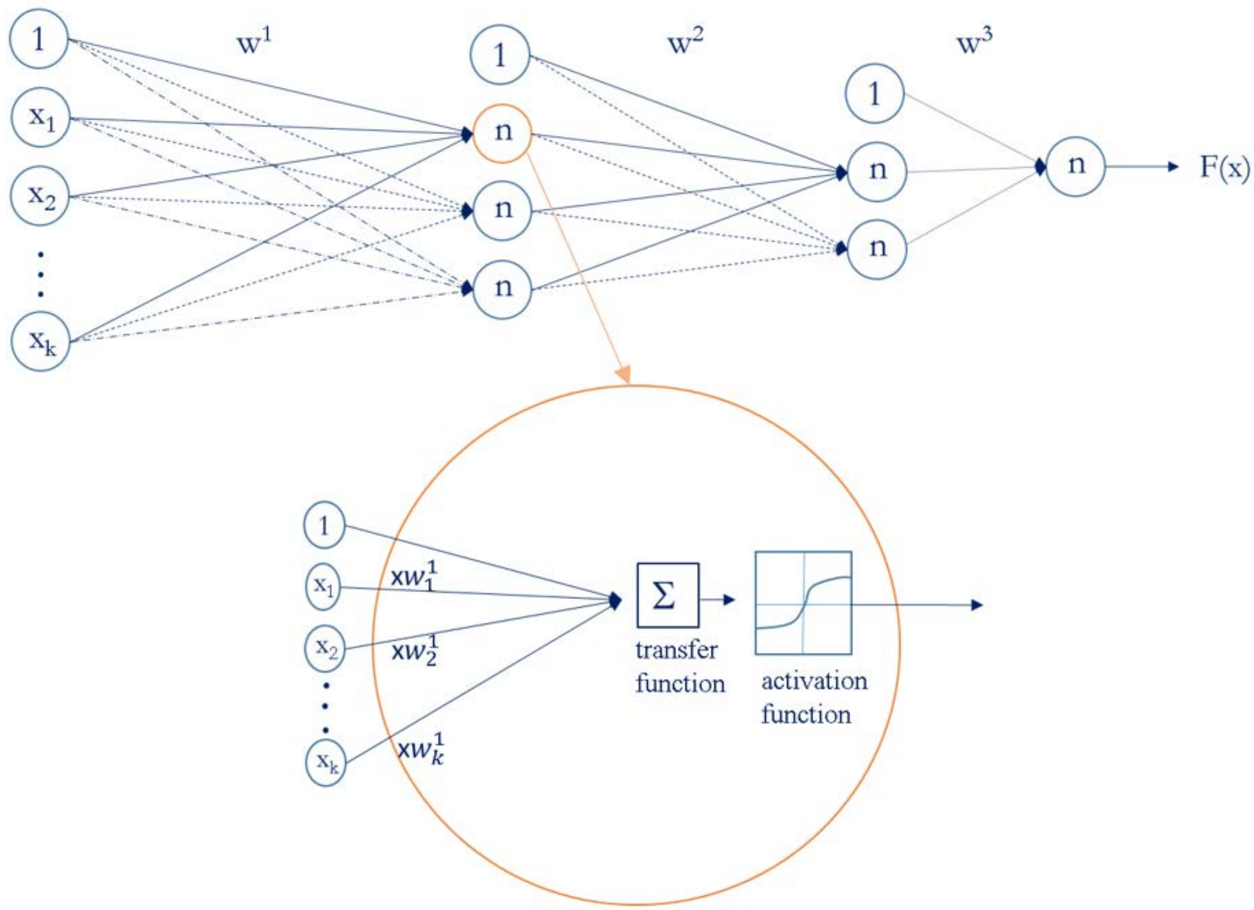
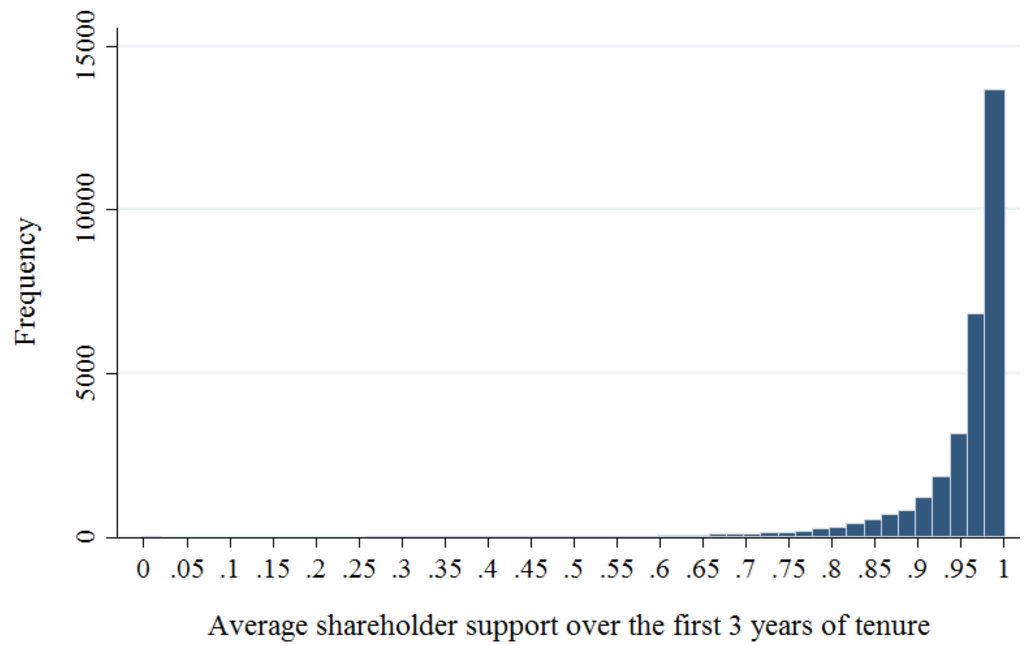


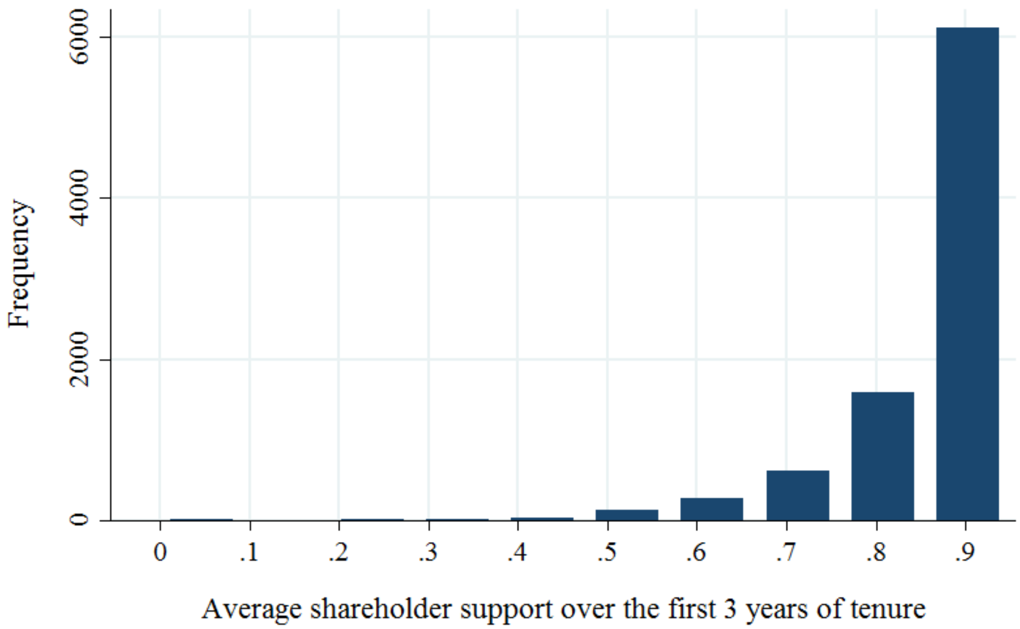
Figure 1 depicts the structure of a basic neural network with two hidden layers. Neurons x_i are input neurons connected to the next layer of neurons by synapses which carry weights w^1. Each synapse carries its own weight. An activation function (usually a sigmoid to allow for non-linear patterns) is 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 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).

Figure 2: Shareholder Support Distribution



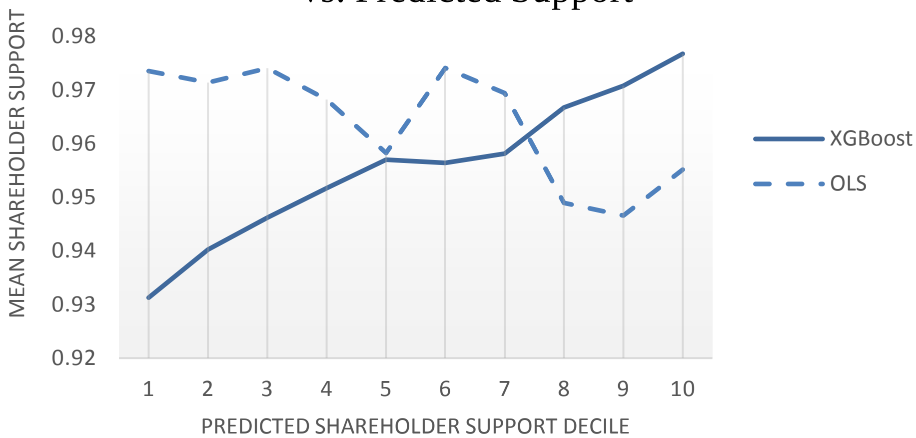
This figure shows the distribution of average shareholder support, defined as the fraction of votes in favor of a given director over all votes cast for the director's reelection within three years of her tenure. The data is from ISS Voting Analytics.

Figure 3: Shareholder Support Below 95%



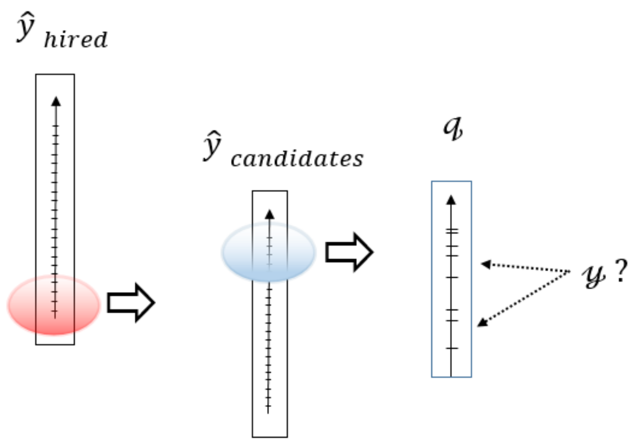
This figure shows the distribution of average shareholder support for values under its mean value of 95%. Shareholder support is defined as the fraction of votes in favor of a given director over all votes cast for the director's reelection within three years of her tenure. The data is from ISS Voting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

Figure 4: Mean Observed Shareholder Support
vs. Predicted Support



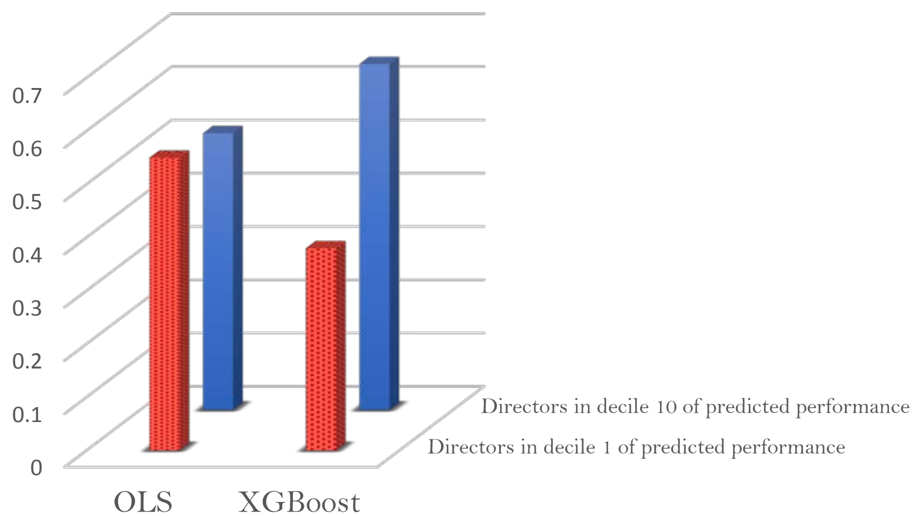
This figure shows the average observed level of shareholder support for directors across the ten deciles of predicted performance for OLS and *XGBoost* in the 2012-14 testing period. Shareholder support is defined as the fraction of votes in favor of a given director over all votes cast for the director's reelection within three years of her tenure.

Figure 5: Assessing the Algorithm's Predictions Using Quasi-Labels



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.

Figure 6: Median Rank in Distribution of Quasi-Labels



This figure illustrates how hired directors in our test set actually performed when compared to potential candidates. Potential candidates are those in the candidate pool associated with each hired director in our test set. We use quasi-labels as an indication of how a potential candidate would have performed on the board of the focal firm. The median director predicted by the *XGBoost* algorithm to be in the bottom decile of shareholder support ranks at the 38th percentile in the distribution of quasi-labels. In contrast, the median director predicted to be in the top decile ranks at the 65th percentile in the distribution of quasi-labels. There is virtually the same performance for the top and bottom deciles of predicted performance when performance is predicted using OLS.

Table 1: Shareholder Support Summary Statistics

This table presents summary statistics for 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. The data is from ISS Voting Analytics.

	N	Mean	Std	25th	50th	75th
2000	748	0.950	0.064	0.944	0.974	0.986
2001	1050	0.944	0.074	0.938	0.970	0.985
2002	1217	0.946	0.074	0.939	0.970	0.986
2003	1949	0.951	0.068	0.945	0.974	0.988
2004	2229	0.953	0.072	0.947	0.977	0.989
2005	2050	0.948	0.072	0.941	0.974	0.989
2006	1959	0.941	0.078	0.927	0.969	0.988
2007	2176	0.940	0.085	0.931	0.971	0.988
2008	1834	0.944	0.075	0.932	0.973	0.988
2009	1690	0.948	0.080	0.945	0.976	0.989
2010	2027	0.948	0.077	0.940	0.977	0.990
2011	1993	0.954	0.069	0.948	0.981	0.992
2012	1914	0.952	0.076	0.951	0.981	0.992
2013	2091	0.948	0.080	0.946	0.980	0.992
2014	1097	0.959	0.071	0.962	0.985	0.993
	26024	0.948	0.074	0.942	0.975	0.989

Table 2: Average Fraction of Bad Outcome

This table presents average fraction of "bad outcome," the shareholder support when it is smaller than its long-term mean 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. Shareholder discontent is presented for various director-level and board-level characteristics.

	Full sample	Yes	No	Difference p-value
Director level				
Male	0.281	0.289	0.227	0.000
Foreign	0.281	0.279	0.281	0.893
Qualifications > median	0.282	0.273	0.286	0.021
Network size > median	0.282	0.279	0.285	0.272
Generation DepBB	0.281	0.308	0.281	0.666
Generation Mature	0.281	0.300	0.275	0.000
Generation BBB	0.281	0.267	0.307	0.000
Generation X	0.281	0.321	0.277	0.000
Generation Y	0.281	0.490	0.280	0.001
Busy director	0.282	0.314	0.275	0.000
Connected to CEO	0.277	0.298	0.261	0.000
Connected to nominating committee	0.277	0.315	0.274	0.012
Board level				
Fraction male > median	0.282	0.250	0.312	0.000
Board size > median	0.282	0.251	0.306	0.000
Nationality mix > median	0.279	0.271	0.282	0.060
Attrition rate > median	0.283	0.303	0.253	0.000

Table 3: OLS Model vs. Random Forest to Predict Director Performance

This table reports the average observed level of 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).

		Average Observed Shareholder Support for Directors in a Given Percentile of Predicted Performance as Predicted by:				
	Predicted Percentile of Shareholder Support	OLS	XGBoost	Ridge	Lasso	Neural Network
Directors predicted to perform **poorly**	1%	0.981	0.883	0.891	0.901	0.904
	5%	0.981	0.925	0.930	0.935	0.939
	10%	0.976	0.947	0.932	0.953	0.942
Directors predicted to perform **well**	90%	0.984	0.982	0.959	0.956	0.966
	95%	0.978	0.980	0.971	0.967	0.968
	100%	0.931	0.983	0.976	0.973	0.967

Table 4: Evaluating the Predictions Using Quasi-Labels

This table reports how hired directors rank in the distribution of quasi-labels of their candidate pool. For each hired 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 in the same industry within a year before or after the hired 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.

	Median percentile of observed performance in the distribution of quasi-labels (candidate pools)				
	OLS	XGBoost	Ridge	Lasso	Neural Network
Bottom decile of predicted performance	55th	38th	43th	44th	50th
Top decile of predicted performance	52th	65th	55th	64th	68th

Table 5: Top vs. Bottom Decile of Predicted Performance

This table reports the mean of firm and director level features for directors in the bottom decile of predicted shareholder support and compares it to the mean for directors in the top decile of predicted shareholder support. These results are for directors in our train set and our test set. The algorithm used to predict performance is XGBoost.

	Mean		
	Bottom decile of predicted performance	**Top** decile of predicted performance	Difference p-value
Director level			
Age	56.5	56.0	0.091
Audit committee	0.285	0.801	0.000
Audit committee chair	0.070	0.146	0.000
Background academic	0.031	0.023	0.129
Background advisor	0.075	0.069	0.476
Background finance	0.055	0.127	0.000
Background human resources	0.002	0.003	0.563
Background lawyer	0.011	0.019	0.041
Background manager	0.202	0.244	0.001
Background marketing	0.055	0.053	0.736
Background military	0.016	0.014	0.610
Background politician	0.015	0.018	0.395
Background science	0.031	0.025	0.264
Background technology	0.017	0.013	0.254
Board attrition	0.089	0.080	0.232
Busy	0.286	0.186	0.000
Chairman	0.031	0.009	0.000
Compensation committee	0.606	0.062	0.000
Compensation committee chair	0.097	0.049	0.000
Connected to CEO	0.422	0.456	0.262
Connected to incumbent director	0.246	0.301	0.000
Connected to nominating committee member	0.101	0.074	0.107
Foreign	0.065	0.058	0.469
Gender ratio (1 is all male)	0.947	0.864	0.000
Governance chair	0.062	0.038	0.000
Governance committee	0.176	0.134	0.000
International work experience	0.069	0.067	0.809
Male	0.929	0.795	0.000
Nationality mix	0.076	0.091	0.004
Network size	1304	1343	0.377
Nomination chair	0.018	0.006	0.000
Nomination committee	0.064	0.016	0.000
Number of qualifications	2.180	2.205	0.504
Total current number of boards sitting on	2.367	1.702	0.000
Total number of listed boards sat on	3.750	2.454	0.000

Cont.

Board level			
Average tenure of incumbent directors	6.115	4.661	0.000
Average total number of current boards incumbent directors sit on	1.796	1.935	0.000
Board size	8.5	10.9	0.000
CEO SOX certified	0.947	0.855	0.000
Chairman is CEO	0.331	0.359	0.149
Chairman is CEO with tenure ≥ 5	0.093	0.109	0.187
Independent directors compensation over CEO total compensation	1.026	0.844	0.607
Mean past voting shareholder support	0.946	0.952	0.270
Number of female directors	0.709	1.741	0.000
Firm level			
Dividend payer	0.237	0.605	0.000
Excess returns 12 months leading up to appointment	0.075	0.151	0.002
Firm age	14.804	17.869	0.000
Hoberg-Phillips product market fluidity	7.383	7.625	0.055
Institutional ownership %	0.625	0.482	0.000
Largest 10 institutional shareholders %	0.425	0.272	0.000
Largest 5 institutional shareholders %	0.309	0.195	0.000
Largest institutional shareholder %	0.110	0.070	0.000
Leverage	0.177	0.229	0.000
Log (number of institutional blockholders)	1.195	0.656	0.000
Log (number of institutional owners)	4.444	4.663	0.000
Ownership by blockholders %	0.237	0.107	0.000
ROE	-11.355	0.422	0.294
Stock returns prior 12 months	0.127	0.252	0.000
Stock returns prior 3 months	0.025	0.076	0.000
Total assets	6643	22207	0.000

Table 6: The Determinants of Predictions: Ordinary Least Squares Regressions

This table reports the results from OLS regression models of the predicted levels of shareholder support in our test set (out-of-sample predictions for directors appointed between 2012 and 2014) on some firm level and director level features. The algorithm used to generate the predictions is XGBoost.

Dependent variable: predicted performance	(1)	(2)	(3)	(4)	(5)	(6)
ln(assets)	0.002*** *(19.850)*	0.002*** *(17.930)*	0.002*** *(7.911)*	0.002*** *(6.993)*	0.001*** *(5.961)*	0.002*** *(4.663)*
Board Size	0.0001*** *(10.530)*	0.0001*** *(9.110)*	0.001*** *(6.992)*	0.001*** *(4.430)*	0.001*** *(5.942)*	0.001** *(2.567)*
ROA	-0.001*** *(-2.620)*	-0.001 *(-0.870)*	0.013*** *(3.539)*	0.015*** *(3.496)*	0.017*** *(3.718)*	0.019*** *(2.710)*
Fraction independent	-0.005*** *(-2.660)*	0.004 *(1.450)*	0.002 *(0.405)*	0.001 *(0.269)*	0.002 *(0.348)*	-0.002 *(-0.261)*
Director Age		0.000 *(-1.23)*	-0.000** *(-2.027)*	-0.000** *(-2.305)*	-0.000** *(-2.218)*	0.000 *(0.854)*
CEO is chairman		-0.001*** *(-3.120)*	-0.001 *(-1.176)*	-0.001 *(-1.512)*	-0.001 *(-1.494)*	-0.002 *(-1.287)*
Busy			-0.004*** *(-5.060)*	-0.002** *(-2.149)*	-0.002** *(-2.172)*	0.000 *(0.125)*
CEO is Chairman with tenure ≥ 5 years			0.000 *(0.336)*	0.000 *(0.015)*	-0.001 *(-0.574)*	-0.002 *(-0.981)*
Average tenure of incumbent directors			-0.000*** *(-4.355)*	-0.000*** *(-4.523)*	-0.000*** *(-3.492)*	-0.001*** *(-2.818)*
Average time in other companies			0.000 *(0.255)*	0.000 *(0.352)*	0.000 *(0.130)*	0.000 *(0.170)*
Male			-0.003*** *(-4.447)*	-0.002*** *(-2.701)*	-0.002* *(-1.948)*	-0.001 *(-0.723)*
Foreign			-0.002* *(-1.731)*	-0.001 *(-1.405)*	-0.001 *(-0.630)*	0.002 *(1.167)*
Number of female directors				0.002*** *(5.441)*	-0.001 *(-1.010)*	0.000 *(0.285)*
Average network size of incumbent directors				0.000 *(0.747)*	0.000 *(0.239)*	0.000 *(0.327)*
Stock returns prior 3 months				0.002 *(1.328)*	0.001 *(0.658)*	0.000 *(0.025)*
Average age of incumbent directors				0.000 *(1.629)*	0.000 *(1.178)*	0.000 *(1.540)*
Average number of listed boards incumbent directors sat on				-0.002*** *(-5.990)*	-0.002*** *(-5.492)*	-0.003*** *(-5.825)*
Firm age					0.000 *(0.211)*	0.000 *(0.767)*
Leverage					-0.002 *(-0.971)*	0.001 *(0.438)*
Dividend payer					0.003*** *(3.268)*	0.002* *(1.667)*
Number of qualifications					0.000 *(0.267)*	-0.001 *(-1.632)*
Gender ratio (1 is all male)					-0.027*** *(-4.239)*	-0.029*** *(-2.627)*
Network size					0.000 *(0.758)*	0.000 *(0.287)*
Stock returns prior 12 months					0.001* *(1.670)*	0.001 *(1.232)*
Chairman					-0.005 *(-1.644)*	-0.007 *(-1.640)*
Average number of qualifications of incumbent directors					0.001 *(1.383)*	0.001 *(0.457)*
Compensation chair						-0.001 *(-0.320)*
Audit chair						0.001 *(0.462)*
Governance chair						-0.001 *(-0.286)*
Nomination chair						0.007 *(0.630)*
Industry ROA						0.000 *(0.074)*
ROE						0.000 *(0.428)*
Connected to the CEO dummy						-0.001 *(-0.623)*
Connected to a member of the nominating committee dummy						0.000 *(0.070)*
Number of incumbent directors known						0.000 *(0.346)*
Constant	0.934*** *(650.19)*	0.929*** *(375.47)*	0.939*** *(206.186)*	0.934*** *(141.379)*	0.956*** *(104.881)*	0.954*** *(66.713)*
Observations	5,481	3,227	1,363	1,235	1,183	489
R-squared	0.167	0.247	0.218	0.269	0.303	0.348

Table 7: Overvalued Director Characteristics

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 shareholder support and indeed ended up being in the bottom decile (predictably bad directors) and compares it to the mean for candidates the board could have hired instead, whom our XGBoost algorithm predicted would be in the top decile of shareholder support.

	Hired directors with predicted and observed low shareholder support	**Promising candidates for this board position**	
	Mean	Mean	Difference p-value
Male	0.965	0.774	0.000
Foreign	0.153	0.142	0.634
Number of qualifications	2.1	2.4	0.000
Network size	1529	1298	0.000
Total number of listed boards sat on	6.9	2.7	0.000
Total number of unlisted boards sat on	9.5	5.2	0.000
Total current number of boards sitting on	3.1	1.6	0.000
Busy	0.41	0.14	0.000
Director age	56.8	58.3	0.001
Background academic	0.014	0.015	0.852
Background finance	0.085	0.039	0.000
International work experience	0.038	0.048	0.276
Mean past voting shareholder support	0.909	0.966	0.000

Table 8: Comparing Shareholder Support Models with Profitability Models

This table reports the observed outcome (profitability or shareholder support) for each decile of predicted performance when performance is measured either as the level of shareholder support or firm profitability (EBITDA/Total Assets). We provide the results when an XGBoost algorithm is trained to predict firm level profitability three years after the director has been appointed and when it is trained to predict the level of shareholder support. The results are for our test set only (out of sample performance for directors appointed between 2012 and 2014).

		1	2	3	4	5	6	7	8	9	10	Difference 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
Algorithm trained on **shareholder support**	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