# Selecting Directors Using Machine Learning

Léa H. Stern
University of Washington

Isil Erel
Ohio State University

Chenhao Tan
University of Colorado, Boulder

Michael S. Weisbach
Ohio State University and NBER

December 22, 2017

**Abstract**

This paper proposes an alternative method of selecting boards of directors that relies on machine learning. We develop algorithms with the goal of selecting directors that would be most valuable to 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 rank much lower in performance than directors who were predicted to do well. Deviations from the benchmark provided by the algorithms suggest that the suboptimal director choices are selected are more likely to be male, older, and connected to exiting directors. 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.

**Key Words**: Corporate Governance, Boards of Directors, Machine Learning

Contact Information: Léa Stern: Department of Finance and Business Economics, Foster School of Business, University of Washington, Seattle, WA, 98195, email: leastern@uw.edu; Isil Erel: Department of Finance, Fisher College of Business, Ohio State University, Columbus, OH 43210, email: erel@fisher.osu.edu; Chenhao Tan: Department of Computer Science, University of Colorado, Boulder, CO, 80309, email: chenhao.tan@colorado.edu; Michael Weisbach: Department of Finance, Fisher College of Business, Ohio State University, Columbus, OH 43210, email: weisbach.2@osu.edu. We thank 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, and the 2017 NABE TEC Conference. Special thanks to Ronan Le Bras for providing invaluable help throughout the project.

# 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 best possible directors for a given firm. 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

[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.

recognition. In the social sciences, machine learning has great potential for prediction problems such as the one we consider here, the way in which one determines which potential director would add the most value to a firm. While traditional econometrics is useful for estimating structural parameters and drawing causal inferences, Athey and Imbens (2017) argue that machine learning is substantially better at making predictions, largely because it does not impose unnecessary structure on the data.

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 quality of these selections by the algorithms to that of the directors who are 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 the nature and source of errors made by decision makers in selecting directors.

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 an outsider 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 related to firm and 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. This measure reflects the views of shareholders about the quality of the director, which presumably aggregates the publicly available information about his or her quality and incorporates how he/she matches to the company.[4]

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 will subsequently receive the highest approval from shareholders. These algorithms rely on methodological approaches common in the computer science literature (i.e., lasso, ridge, random forest, neural networks, and gradient boosting trees). 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 *gradient boosting trees* predict 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 out of sample do not actually perform worse than those the OLS model predicted would do well. In sum, the machine learning algorithms we develop provide superior predictions on the quality of directors.

To determine whether an algorithm can improve on boards' decisions, we also wish to compare the algorithm's predictions to the choices of directors that firms actually picked. This

[4] Our choice of director performance measure is also guided by the fact that the mandate of board directors is to represent shareholders. Even if not formalized, the nominating committee should in principle form a prediction of how well a potential candidate would perform on that dimension.

task is complicated by the fact that we only observe the performance of directors who were actually hired by the board and do not observe such performance for potential candidates that 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)). We address it by constructing a realistic pool of potential candidate directors: directors who joined the board of a neighboring company around the same time.

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 in the past year or the following year. Presumably these potential candidates would have found the opportunity to be on the board of a larger nearby company to be attractive. Each new board appointment is therefore associated to its own pool of potential candidates. Remember that in evaluating the algorithm’s predictions of the performance of directors who were actually hired, 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 join. Therefore, although we do not observe the performance of potential candidates that were not hired, we observe their *quasi-label*: their performance on the competing board they joined.

We show that directors machine-learning models predicted would do poorly indeed rank lower when compared to candidates’ performance than those they predicted would do well. In contrast, directors rank around the 50$^{th}$ percentile in the distribution of quasi-labels, regardless of whether OLS predicted they would perform well or not. Note that although the board-director match is endogenous, our empirical strategy only uses the candidate pools to *evaluate* the performance of the algorithm. In practice, 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 a candidate pool. We find that directors the algorithm predict would perform poorly rank much lower in the distribution of quasi-labels than those predicted to do well, which provides reassurance that our quasi-labels are not systematically inflated due to the endogenous nature of the board-director match.

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 fit accurately and outperform OLS. 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 predictably poor performing directors based on shareholder votes are associated with firm profitability that is significantly lower in the years following the initial appointment than that for directors with high predicted performance.

The striking result about the machine learning models is that they consistently suggest directors who would have been likely both to accept the directorship and to outperform the directors that are actually chosen by firms. It is possible that the machine-learning model could be suggesting potential directors that existing directors did not consider. Alternatively, as has been argued by many observers, the observed choices of directors in actual firms are a result of numerous agency conflicts. A reason for the superior performance of the algorithms could be that algorithms are not subject to these agency problems. Consistent with this idea, directors chosen by firms are too often male, have too large networks, are on too many other boards, and are too likely to have a finance background, relative to the directors the algorithm suggests that boards select. In other words, boards tend to choose directors who are like themselves, while the algorithm suggests that adding diversity would be a better idea.

Machine-learning tools have the potential to help answer many questions left unanswered in the social sciences. Our setting lends itself well to comparing machine predictions to human decisions for two main reasons. First, our setting allows us to design realistic pools of candidates for whom we are able to observe performance. Second, although performance is a multi-faceted concept in many hiring contexts, the mandate of corporate directors is clear: represent shareholders' interests. The combination of those two characteristics, which are specific to our application, effectively allows us to draw conclusions on the quality of boards' hiring decisions.

In addition, the fact that our algorithm can generate a prediction for any potential candidate could have 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.[5]

This paper is organized as follows. The next section describes machine-learning algorithms we use and develops a framework that help us assess the performance of these algorithms. In the third section, we present our data and summary statistics. In the forth section, we compare performance of our prediction models and provide various characteristics that affect directors' performance. The firth section includes an extensive discussion of our approach and findings, put them in perspective, discuss possible extensions, and then concludes.

## 2. Using Machine Learning to Predict Director Performance

### *2.1. Algorithms to Predict Performance*

[5] 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.

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 *gradient boosting trees*. We train each of these algorithms, i.e. estimate model parameters, on directors appointed between 2000 and 2011 and test them on directors appointed between 2012 and 2014. Following the terminology in machine learning, we call the data from 2000-2011 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.[6] As is typical in the machine learning literature, the way the various algorithms handle missing observations is by a simple imputation procedure: a missing value will be set equal to the median value of all other observations for that variable.

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

[6] 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.

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 $\ell_1$-norm (least absolute deviation). The penalty weight ($\lambda$) 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 $\ell_2$-norm (least squares), therefore shrinking estimates smoothly towards zero, as opposed to setting some estimates to zero as Lasso does.[7]

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

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

"leaf" or terminal node is comprised of similar observations. Then, a new 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, due to two important challenges: the selective labels problem-having performance data at the company in question for only directors who were actually selected-and the problem of the reliance on unobservables by decision makers. To formalize these concepts, we develop a framework similar to that presented by Kleinberg et al. (2017).

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 $\not{f}$. For the purpose of predictive modeling, we are interested in finding a function that closely matches the function $\not{f}$ 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 $\not{f}$. In other words, our algorithm seeks a functional form that maps features W into predictions $\hat{\not{f}}(W)$ 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 $\pi_j$ 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.

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

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

The 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 $\alpha$ 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 $\alpha_j$ 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. Note that although the board-director match is endogenous, our empirical strategy only uses the candidate pools to *evaluate* the performance of the algorithm.

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. In other words, we are interested in whether there were other available candidates that would have performed better and that the algorithm would have chosen over the director actually hired. 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 poor performers would suggest that boards put disproportionate weight on $v_j$.

## 3. Constructing a Sample on which Algorithms Can Select Directors

### *3.1. Measuring Director Performance Through Election Results*

A challenging part of selecting directors is measuring 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 wining 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 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.

*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.[8] 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.[9] Note that all of our results pertain to our "testing" subsample, which includes directors appointed between 2012 and 2014, when the influence of proxy advisory firms is weaker than it was earlier in time.

### *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

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

[9] 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.

newly hired directors. The question we ask in this paper 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. Whereas knowledge of previous research on boards of directors can guide us to select characteristics which we think 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, an estimation procedure that has as much flexibility as possible and allows the data to dictate the form of the relationship between potential explanatory variables is desirable. This flexibility is a unique advantage of our machine-learning algorithms.

## 4. Evaluating Machine Learning Predictions of Director Choice

### *4.1. Comparison of Various Prediction Models*

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 predicts 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.

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 portion. 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. 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.[10] 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.[11] Figure 4 shows the average observed level of shareholder support for directors across the ten

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

[11] *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.

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 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.

*4.2. Assessing the Algorithm's Predictions: Exploiting Quasi-Labels*

Our finding that directors identified by our algorithm as having low (high) future shareholder support, are more likely to have low (high) support in subsequent elections, raises the issue of why firms selected the poor-performing directors who actually did join their boards.

Poor director performance does not necessarily imply that firms were wrong to hire them initially. Firms' hiring decisions could have been reasonable ones that did not work out because of poor luck. However, the model suggests that the situation is worse than that. Since the model's predictions are done using information publicly available at the time when the new directors are chosen, boards could have, and from the point of view of shareholders, *should have* known that these directors' performance would ultimately be poor.

There are several possible explanations why boards would choose suboptimal directors when information indicating the likelihood of poor subsequent performance is available. One possible reason is that boards maximize their own utility rather than shareholders. Consequently, new directors are selected because existing boards and CEOs like them personally but they do not best serve shareholders' interest (see Hermalin and Weisbach (1998) for a formal development of this idea). A second possible explanation of this result is that the firms that hired predictably bad performers as directors did so because there were no better alternatives available. In other words, it could be that in the cases where poor performing directors are chosen, all other possible directors who would been willing to serve would have performed even worse than the ones firms actually did select as directors. Finally, it is possible that since machine learning models are new and boards in our sample did not have access to these, the boards did not know that they were making poor choices of new directors.

To evaluate whether this result occurs because of a shortage of potential alternative directors we consider whether there were alternative directors who likely would have been willing to serve, and would have performed better than the directors who were selected. We start from the assumption that individuals who just joined, or are about to join a board, would have been attracted to the possibility of joining a larger board in the same geographical region around the same time. Directors of nearby smaller companies are plausible alternative choices for directorships, since it is generally more prestigious (and lucrative) to join boards of larger companies than of smaller ones. A large fraction of these directors likely would have been willing to join the board of a nearby larger company if asked.

For this reason, we construct pools of potential candidates for each new board position by looking at directors who joined the board of a smaller company within 100 miles of the firm headquarter in the past one or following one year of the new director's election. There are on average 192 such individuals for each director appointment in our sample for whom we have shareholder support data available. We refer to potential alternative directors' performance on the boards they actually joined as their "quasi-labels". This quasi-label is an indication of the performance these directors would have achieved had they been asked to be directors of the larger firms whose choices we are considering.

It is possible that the directors' performance would have been different had they joined the board of the firms we are considering. However, individual directors' support tends to be consistent across the companies they serve. There were 19,464 directors in our sample who joined multiple boards over our sample period. The average (median) standard deviation of shareholder support over the first three years of their tenure across the different boards they join is only .045 (.024). This low standard deviation suggests that on average directors tend to receive similar support from shareholders on the different boards they join.

We are interested in cases in which directors are predicted to perform poorly, and whether there were plausible alternatives available to the boards that selected them. To measure the expected performance of potential alternative choices, we calculate how the directors' predicted and actual performance compare to that from the alternative pool the firm could have hired. Because firms are in different locations and of different sizes, each director opening has its own candidate pool, for which there is a different distribution of possible outcomes. We calculate the predicted performance of each possible alternative choice for each directorship that was filled, in addition to the director who was actually chosen. We rank these predicted performances and then compare them to the actual subsequent performance of each director, measured as his or her voting totals at the company for which he or she was chosen to be a director.

Table 4 shows the median ranks in the quasi-labels distributions 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 indeed rank lower when compared to candidates' performance than those they predicted would do well. In contrast, directors rank around the 50$^{th}$ 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 is the one that is able to best discriminate *ex ante* which directors will do well from those who will not.

The median director who is predicted by the *XGBoost* algorithm to be in the bottom decile of shareholder support is the 38$^{th}$ percentile of actual observed performance. In contrast, the median director predicted to be in the top decile ranks in the 65$^{th}$ percentile of the observed performance. Figure 5 illustrates the median rank of predicted shareholder support among potential alternative candidates using quasi-labels in Deciles 1 and 10 with *XGBoost* and OLS models. This figure illustrates the difference between the *XGBoost* and OLS approaches. There is virtually the same subsequent 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 in subsequent elections 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.

Note that although the board-director match is endogenous, our empirical strategy only uses the candidate pools to *evaluate* the performance of the algorithm. In practice, the algorithm could give a prediction for any candidate, not only those in the candidate pool, which is limited to directors who within the past year or the following year joined the board of a smaller nearby company. On average, there are about two hundred candidates with quasi-labels in a candidate pool. We find that directors the algorithm predict would perform poorly rank much lower in the distribution of quasi-labels than those predicted to do well, which provides reassurance that our quasi-labels are not systematically inflated due to the endogenous nature of the board-director match. 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.

#### *4.2.1. Robustness*

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 $37^{th}$ percentile of all quasi-labels. In contrast, the median director predicted to be in the top decile of performance subsequently is in the $67^{th}$ 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.

### *4.3. Characteristics that Affect Director Performance*

The machine learning models are able to predict which directors are likely to receive more votes in subsequent elections. Presumably, voting totals reflect shareholder satisfaction with director performance, as a market-based measure of director performance. The predictions come from a complicated algorithm based on the variables in our database. Presumably, some variables in the database are more important than others in their impact on subsequent performance. However, unlike a regression model, machine learning models do not produce estimated functions with parameters one can easily examine. Instead, to understand the characteristics affecting director performance, we consider the prediction values from the machine learning models and try to measure which variables appear to be associated with high and low predicted performance.

Table 5 provides some guidance about which director features are overrated by boards and management when making their hiring decisions. This table reports some differences in director, board and firm level characteristics between the bottom and top deciles of predicted shareholder support using the *XGBoost* model. Notable differences between directors in the top and bottom decile are that 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.

Because these qualifications are not independent from one another, we wish to understand which observed characteristics of directors lead to higher subsequent votes. We estimate regressions of the predicted values of subsequent votes from the *XGBoost* model. These regressions include both firm-level and director variables. The coefficients from these regressions will reflect the characteristics that the algorithm associates with higher subsequent performance.

We present estimates of this regression in Table 6. Firm-level variables that appear to be associated with subsequent performance are size (total assets), operating performance and whether the firm pays dividends. 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. At the director level, gender, whether the director is busy and the number of listed boards the director sat on appear to impact the predictions of our algorithm. Note that the purpose of this exercise is simply to provide some information on which characteristics affect the predictions more. In these regressions, the $R^2$ is fairly low (below 40%). The relatively low $R^2$ speaks to the importance of feature interactions and non-linearities that *XGBoost* relies on to generate its predictions.[12]

The algorithm's predictions also help identify the individual director features that tend to be overvalued or undervalued by management when selecting a new director to serve on the board. 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 to evaluate the way in which directors are chosen.

In Table 7, we report on the 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 bad 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. And they suggest that management makes predictably poor director hires.

*4.4. Firm Profitability as a Measure of Performance*

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

The mandate of board directors is to represent shareholders and thus shareholder support at annual elections is a natural measure of director performance. Nevertheless, one might be concerned that hiring on the basis of a prediction of the level of shareholder support might come at the cost of what management ultimately cares about: firm profitability. In this subsection, we show that this is not the case in two ways.

First, we train an *XGBoost* algorithm to predict a firm level profitability measure (EBITDA/Total Assets) three years after the director appointment. We rank hired directors based on this alternative predicted performance and report results in the first two rows of Table 8. [13]

Selecting directors on the basis of predicted performance using one performance measure does not lead to hiring directors who will underperform along the alternative criteria. On the contrary, both measures are consistent. Firms that hired directors in the bottom (top) decile of predicted performance have an average profitability of -49.8% (20.5%). When our algorithm is trained to predict future profitability, the hired directors at firms predicted to have low (high) profitability three years following their appointment have an average shareholder support of 94% (96%). The difference is significant at the 1% level.

Next, we train our algorithm to predict shareholder support, the 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%. Average shareholder support in the lowest decile is 92% and it becomes 97.7% in the top decile.

Importantly, these results are very similar if we focus on firms that experience positive excess returns leading up to the nomination of the director.

[13] The correlation with the shareholder support measure is .14 (p-value: 0.000).

## 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, the extent to which suboptimal choices are knowable ex ante, 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. 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. A limitation of the analysis is that we cannot know whether a potential director would have received the same support from shareholders at another company than at the company where he was actually a director. We do note, however, that most directors who accept multiple directorships during our sample period tend to receive the same support level in all the companies in which they serve.

The machine-learning algorithms we construct in this way fit the data well. Using publicly available data on firm, board, and director characteristics, our *XGBoost* algorithm predicts shareholder support with a mean absolute error of about 4% out of sample. In addition, the realized performance following the appointment of a director is a monotonic function of the predicted performance. This model 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, econometric models fit the data poorly out of sample. 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 paper's finding that it is possible to predict, with a high degree of accuracy, 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 who maximize their own interests rather than those of the shareholders. The notion that boards are aware of directors who are expected to perform better and nonetheless select suboptimal directors to join is consistent with this idea.

The differences between the directors suggested by the algorithm and those actually selected by firms allow us to assess whether boards knowledgeably pick suboptimal directors. Comparing predictably bad 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 other words, directors and managers select new directors who are more like themselves even if they are not as good as other possible choices at maximizing value.

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 since the 1990s had a corporate governance policy aimed in large part to diversify boards of directors for this reason (see Biggs (1996) and Carleton et al. (1998)).

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.[14] 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.

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 judgement on the part of directors and

[14] 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.

current management. This lack of discretion 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 as well in the future. 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.

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.

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.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan (2017) Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*

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.

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.

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.

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

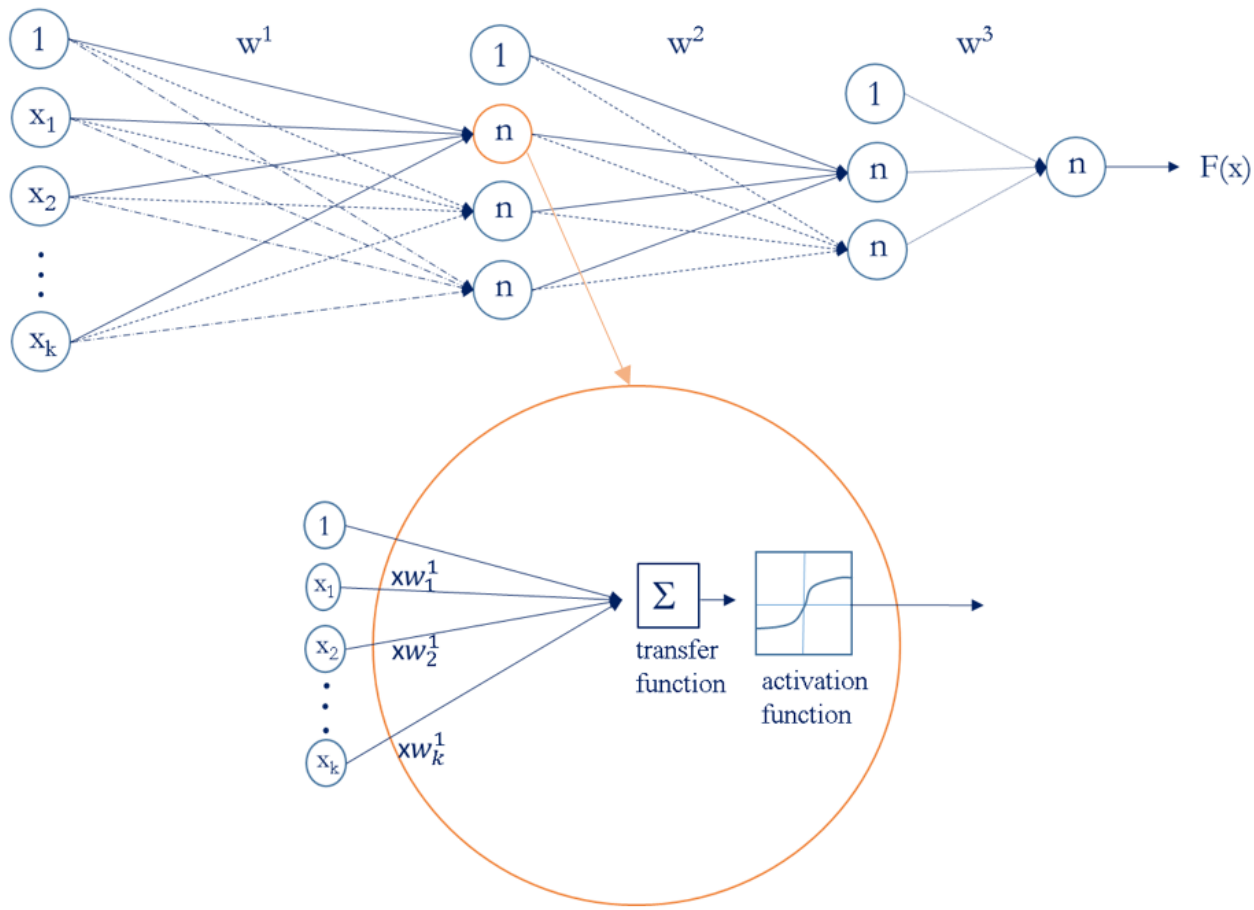
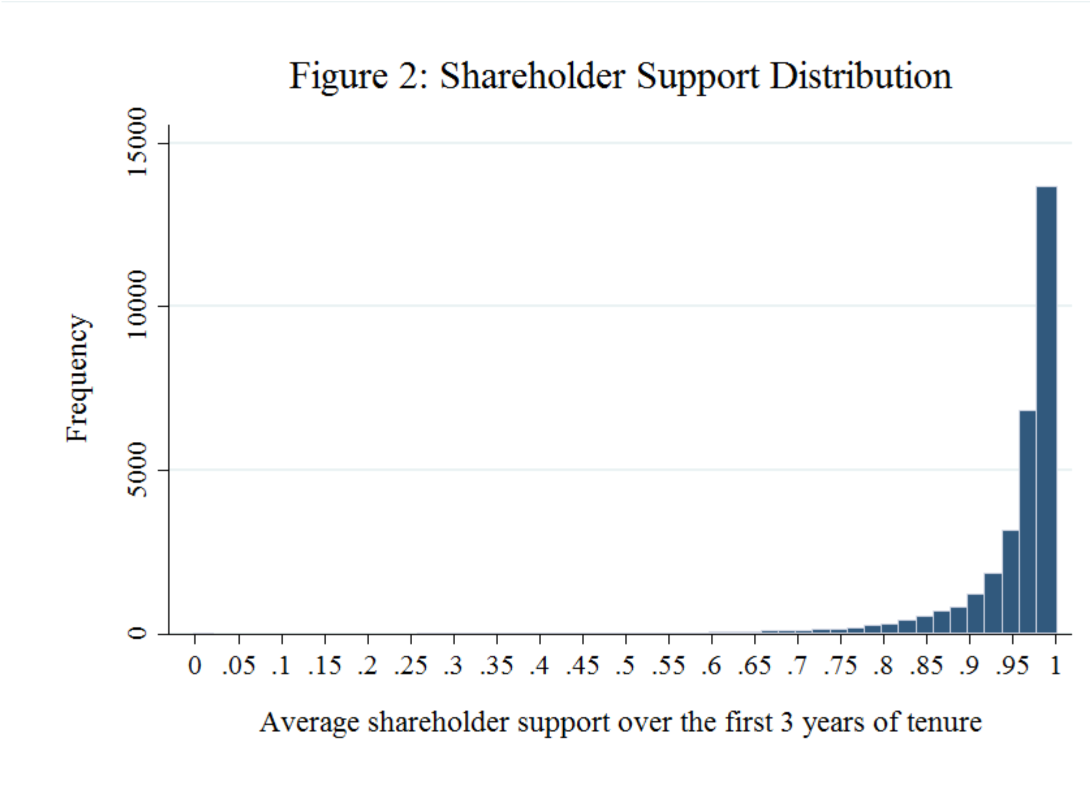
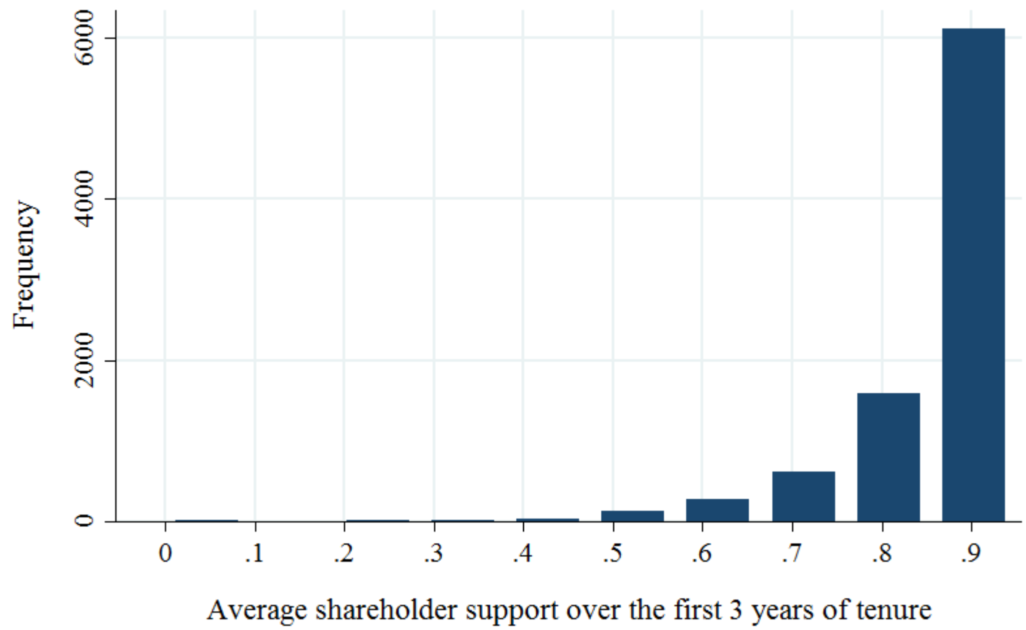



Figure 1 depicts the structure of a basic neural network with two hidden layers. Neurons $x_i$ are input neurons connected to the next layer of neurons by synapses which carry weights $w^1$. Each synapse carries its own weight. An activation function (usually a sigmoid to allow for non-linear patterns)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



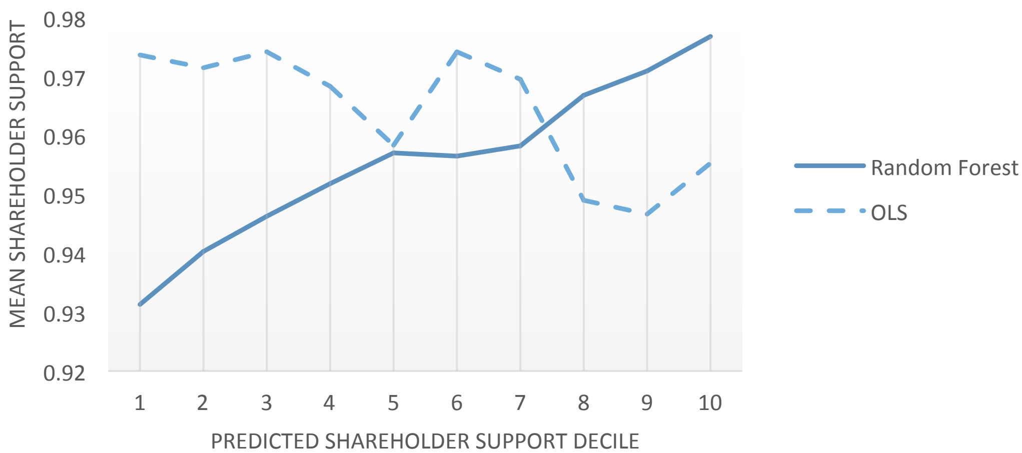

This figure shows the distribution of average shareholder support, defined as the fraction of votes in favor of a given director over all votes cast for the director's reelection within three years of her tenure. The data is from ISS Voting Analytics.

Figure 3: Shareholder Support Below 95%



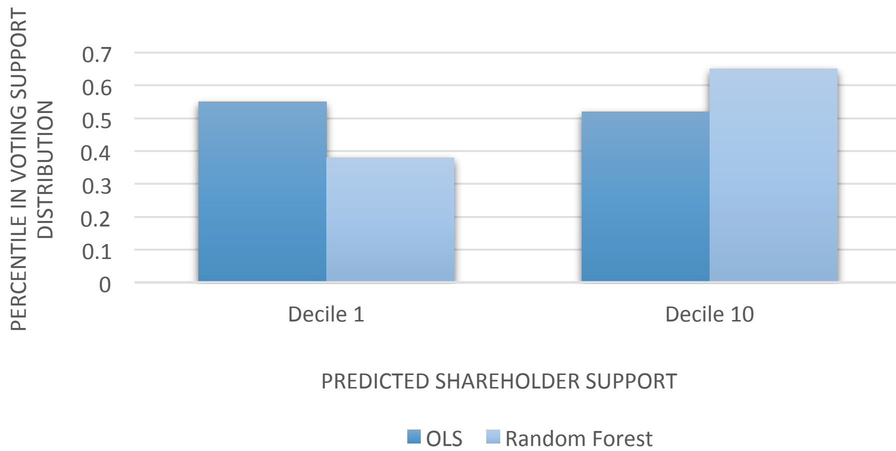

This figure shows the distribution of average shareholder support for values under its mean value of 95%. Shareholder support is defined as the fraction of votes in favor of a given director over all votes cast for the director's reelection within three years of her tenure. The data is from ISS Voting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

Figure 4: Mean Actual Shareholder 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: Median Rank Among Potential Candidates' using Quasi-Labels




This figure illustrates the median rank of predicted shareholder support among potential alternative candidates using quasi-labels in Deciles 1 and 10 with *XGBoost* and OLS models.

**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*** | 0.002*** | 0.002*** | 0.002*** | 0.001*** | 0.002*** |
| | *(19.850)* | *(17.930)* | *(7.911)* | *(6.993)* | *(5.961)* | *(4.663)* |
| Board Size | 0.0001*** | 0.0001*** | 0.001*** | 0.001*** | 0.001*** | 0.001** |
| | *(10.530)* | *(9.110)* | *(6.992)* | *(4.430)* | *(5.942)* | *(2.567)* |
| ROA | -0.001*** | -0.001 | 0.013*** | 0.015*** | 0.017*** | 0.019*** |
| | *(-2.620)* | *(-0.870)* | *(3.539)* | *(3.496)* | *(3.718)* | *(2.710)* |
| Fraction independent | -0.005*** | 0.004 | 0.002 | 0.001 | 0.002 | -0.002 |
| | *(-2.660)* | *(1.450)* | *(0.405)* | *(0.269)* | *(0.348)* | *(-0.261)* |
| Director Age | | 0.000 | -0.000** | -0.000** | -0.000** | 0.000 |
| | | *(-1.23)* | *(-2.027)* | *(-2.305)* | *(-2.218)* | *(0.854)* |
| CEO is chairman | | -0.001*** | -0.001 | -0.001 | -0.001 | -0.002 |
| | | *(-3.120)* | *(-1.176)* | *(-1.512)* | *(-1.494)* | *(-1.287)* |
| Busy | | | -0.004*** | -0.002** | -0.002** | 0.000 |
| | | | *(-5.060)* | *(-2.149)* | *(-2.172)* | *(0.125)* |
| CEO is Chairman with tenure ≥ 5 years | | | 0.000 | 0.000 | -0.001 | -0.002 |
| | | | *(0.336)* | *(0.015)* | *(-0.574)* | *(-0.981)* |
| Average tenure of incumbent directors | | | -0.000*** | -0.000*** | -0.000*** | -0.001*** |
| | | | *(-4.355)* | *(-4.523)* | *(-3.492)* | *(-2.818)* |
| Average time in other companies | | | 0.000 | 0.000 | 0.000 | 0.000 |
| | | | *(0.255)* | *(0.352)* | *(0.130)* | *(0.170)* |
| Male | | | -0.003*** | -0.002*** | -0.002* | -0.001 |
| | | | *(-4.447)* | *(-2.701)* | *(-1.948)* | *(-0.723)* |
| Foreign | | | -0.002* | -0.001 | -0.001 | 0.002 |
| | | | *(-1.731)* | *(-1.405)* | *(-0.630)* | *(1.167)* |
| Number of female directors | | | | 0.002*** | -0.001 | 0.000 |
| | | | | *(5.441)* | *(-1.010)* | *(0.285)* |
| Average network size of incumbent directors | | | | 0.000 | 0.000 | 0.000 |
| | | | | *(0.747)* | *(0.239)* | *(0.327)* |
| Stock returns prior 3 months | | | | 0.002 | 0.001 | 0.000 |
| | | | | *(1.328)* | *(0.658)* | *(0.025)* |
| Average age of incumbent directors | | | | 0.000 | 0.000 | 0.000 |
| | | | | *(1.629)* | *(1.178)* | *(1.540)* |
| Average number of listed boards incumbent directors sat on | | | | -0.002*** | -0.002*** | -0.003*** |
| | | | | *(-5.990)* | *(-5.492)* | *(-5.825)* |
| Firm age | | | | | 0.000 | 0.000 |
| | | | | | *(0.211)* | *(0.767)* |
| Leverage | | | | | -0.002 | 0.001 |
| | | | | | *(-0.971)* | *(0.438)* |
| Dividend payer | | | | | 0.003*** | 0.002* |
| | | | | | *(3.268)* | *(1.667)* |
| Number of qualifications | | | | | 0.000 | -0.001 |
| | | | | | *(0.267)* | *(-1.632)* |
| Gender ratio (1 is all male) | | | | | -0.027*** | -0.029*** |
| | | | | | *(-4.239)* | *(-2.627)* |
| Network size | | | | | 0.000 | 0.000 |
| | | | | | *(0.758)* | *(0.287)* |
| Stock returns prior 12 months | | | | | 0.001* | 0.001 |
| | | | | | *(1.670)* | *(1.232)* |
| Chairman | | | | | -0.005 | -0.007 |
| | | | | | *(-1.644)* | *(-1.640)* |
| Average number of qualifications of incumbent directors | | | | | 0.001 | 0.001 |
| | | | | | *(1.383)* | *(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*** | 0.929*** | 0.939*** | 0.934*** | 0.956*** | 0.954*** |
| | *(650.19)* | *(375.47)* | *(206.186)* | *(141.379)* | *(104.881)* | *(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: Overrated Individual Features

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: Omitted Payoff Bias**

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 |