

NBER WORKING PAPER SERIES

SELECTING DIRECTORS USING MACHINE LEARNING

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

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

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue  
Cambridge, MA 02138  
March 2018, Revised June 2019

We thank Renée Adams and Reena Aggarwal (who graciously shared data), Lucian Bebchuk, Philip Bond, Lisa Cook, Ran Duchin, Daniel Ferreira (discussant), Fabrizio Ferri, Shan Ge, Jarrad Harford, Ben Hermalin, Joan MacLeod Heminway, Joshua Lee (discussant), Nadya Malenko (discussant), Jordan Nickerson (discussant) Miriam Schwartz-Ziv, Anil Shivdasani, Tracy Yue Wang (discussant), Ayako Yasuda, Luigi Zingales (discussant) and conference and seminar participants at North Carolina, Northeastern, Ohio State, Singapore, Tennessee, Washington, 2017 Pacific Northwest Finance Conference, 2017 WAPFIN Conference at NYU Stern, 2017 NABE TEC Conference, 2018 University of Miami-AFFECT conference, 2018 Drexel Corporate Governance Conference, 2018 ICWSM BOD workshop, 2018 NBER Economics of AI Conference, 2018 Wash. U. Olin Corporate Finance Conference, 2019 AFA Annual Meetings, 2019 NBER Big Data Conference, 2019 Conference on Emerging Technologies in Accounting and Financial Economics at USC and 2019 Wine Country Finance Conference. Special thanks to Ronan Le Bras for providing invaluable help throughout the project. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Isil Erel, Léa H. Stern, Chenhao Tan, and Michael S. Weisbach. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Selecting Directors Using Machine Learning

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

NBER Working Paper No. 24435

March 2018, Revised June 2019

JEL No. G34,M12,M51

**ABSTRACT**

Can algorithms assist firms in their decisions on nominating corporate directors? We construct algorithms to make out-of-sample predictions of director performance. Tests of the quality of these predictions show that directors predicted to do poorly indeed do poorly compared to a realistic pool of candidates. Predictably poor performing directors are more likely to be male, have more past and current directorships, fewer qualifications, and larger networks than the directors the algorithm would recommend in their place. Machine learning holds promise for understanding the process by which governance structures are chosen, and has potential to help real-world firms improve their governance.

Isil Erel  
Fisher College of Business  
Ohio State University  
2100 Neil Avenue  
Columbus, OH 43210  
and NBER  
erel@fisher.osu.edu

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

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

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

## 1. Introduction

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

Many of the concerns about boards come from the director selection process, which has been a source of debate since at least Berle and Means (1932).<sup>2</sup> The selection process for selecting directors is one of the most important yet least studied questions in corporate governance. Despite the checks and balances built into a public corporation's governance system, the CEO often controls the selection of new directors.<sup>3</sup> In practice, appointed directors are almost always supporters of the CEO and his policies. Aside from occasional proxy contests, shareholders have virtually no control over the choice of the directors whose mandate is to represent their interests.

Can machine learning algorithms help address this first order issue in corporate governance? We outline a potential way in which they can. We argue that algorithms can shed light on the decision-making process that governs the nomination of corporate directors.

Key to our approach is the idea that the nomination of a corporate director can be thought of as a prediction problem (Kleinberg et al., 2015), in contrast to a parameter estimation problem, as the literature on boards has traditionally approached it. We consider a potential alternative approach to select directors: one that uses algorithms that rely on data on firms and on their current board members, as well as on potential directors and their attributes, to identify the quality of directors being considered for a given firm's board. We take advantage of advances in machine learning that have revolutionized many fields and have

---

<sup>1</sup> See Hermalin and Weisbach (2003), Adams, Hermalin and Weisbach (2010), and Adams (2017) for surveys.

<sup>2</sup> 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.

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

led to innovations ranging from self-driving cars to facial recognition. In the social sciences, machine learning has great potential for prediction problems such as the one we consider here, the way in which one determines which potential director would be the best for a particular firm. While “traditional” econometrics is typically designed for estimating structural parameters and drawing causal inferences, machine learning is substantially better at making predictions, in part because it does not impose unnecessary structure on the data.<sup>4</sup>

We construct a large database of publicly traded U.S. firms and independent directors appointed between 2000 and 2014. We build several machine learning algorithms designed to predict director performance using director, board and firm level data available to the nominating committee at the time of the nominating decision. We compare the algorithms’ selections of directors to the ones actually chosen by firms. The discrepancies between firms’ actual choices of directors and the choices based on the predictions from our algorithms allow us to characterize which individual features are overrated by decision makers. As such, the algorithm’s predictions can be leveraged as a diagnostic tool to shed light on the decision-making process that governs the selection of corporate directors.

A crucial element of any algorithm designed to select valuable independent directors is a process for assessing a director’s performance in a particular firm. The task of measuring the performance of an individual director is challenging since directors generally act collectively on the board and it is usually impossible for a researcher to ascertain the actions of any particular director. Nevertheless, as Hart and Zingales (2017) emphasize, directors’ fiduciary duty is to represent the interests of the firm’s shareholders. Their popularity among shareholders is thus a natural metric for evaluating them. For that reason, our main measure of director performance is based on levels of shareholder support in annual director re-elections.

An important issue in interpreting the results is whether shareholder votes reflect directors’ quality in addition to their popularity with shareholders. While these notions are to some extent the same since a director’s duty is to serve the interests of shareholders, we recognize that investors often have limited

---

<sup>4</sup> See Athey and Imbens (2017) and Mullainathan and Spiess (2017).

resources and sometimes vote based on simple, check-the-box criteria in routine director elections. To measure the extent to which the algorithm's predictions using shareholder votes apply to other measures of director performance, we also consider the model's ability to predict firm profitability and announcement returns of director appointments. We find that the algorithm's predictions of shareholder votes to re-elect directors are also strongly related to firm profitability and to announcement returns of director appointments.

We construct machine learning algorithms to predict the performance of any potential director at any particular company, taking into consideration who is currently sitting on the board. Using our sample of public firms, we train each algorithm (i.e. fit a model) on a "training" set (directors appointed between 2000 and 2011), and then compare the predictions to the observed data out-of-sample using a "test" set (directors appointed between 2012 and 2014).

We find that these algorithms make accurate out-of-sample predictions of the distribution of outcomes, whether predicting the level of shareholder support or the excess support relative to the slate. The directors the algorithms predicted would do poorly did much worse on average than the directors the algorithm predicted would do well. In comparison, the directors predicted to do poorly by an OLS model do not actually have worse performance out of sample than those the OLS model predicted would do well. Machine learning algorithms, by letting the data speak about the underlying relationships among all candidate predictors, end up fitting the data much better and consequently do better at predicting future outcomes out of sample.

We only observe the votes (i.e. the label) for directors who were actually nominated to the board but do not observe them for potential candidates who were *not* nominated. This "*selective labels*" problem of having voting data at the company in question only for directors who were actually selected is a common issue in prediction problems (see Kleinberg et al. (2017)). In other words, we are only able to evaluate the algorithm's predictive ability for nominated directors. However, if decision makers consider features that are not observable to our algorithm in their nominating decisions of directors, the distribution of outcomes in the set with observed labels (nominated directors) could differ from that in the set with missing labels

(not nominated directors), even if they share exactly the same observable characteristics. In other words, if boards are skilled at using unobservables in their nominating decisions, nominated directors could have higher expected performance than otherwise similar (based on observables) passed-over directors.

Most prediction problems in the social sciences are subject to the issue of selective labels and reliance on unobservables, formalized in Kleinberg et al. (2017). These issues make the evaluation of the algorithms' predictions challenging. We design the quasi-labels approach to overcome this challenge. Quasi-labels represent substitute labels for observations with unobserved outcomes. Importantly, in the procedure we design, quasi-labels do not need to be *perfect* substitutes to labels in order to meet their goal: to assess the algorithms' predictions when labels are missing and the decision maker relies on unobservables.

For each board appointment in our test set, we construct a realistic pool of potential candidates: directors who joined the board of a smaller neighboring company within a year. Presumably these potential candidates would have found the opportunity to be on the board of a larger nearby company attractive, since directorships at larger companies tend to be better paying and more prestigious than directorships at smaller companies. They also signaled that they were available and willing to travel to this specific location for board meetings. Although we do not observe the performance (*i.e.* the *label*) of those potential candidates (the selective labels problem), the design of our candidate pools allows us to observe what we refer to as their "*quasi-label*": their performance on the board they effectively joined.

We find that directors the algorithm predicted would perform poorly (well) indeed do perform poorly (well) when compared to potential available alternatives. Directors in the bottom decile of predicted performance rank at the 27<sup>th</sup> percentile in the distribution of quasi-labels. In contrast, those in the top decile rank at the 78<sup>th</sup> percentile. OLS models are unable to predict *ex ante* who will perform well compared to alternatives and who will not.

While machine learning models do not generate estimates of the underlying structural parameters of a model, we can use the algorithm's predictions to understand the features that are overvalued and undervalued by firms in the director selection process. A striking result in this paper is that machine learning models consistently suggest directors who would have been likely both to accept the directorship and to

outperform the directors that are actually chosen by firms. Relative to algorithm-selected directors, management-selected directors who receive predictably low shareholder approval are more likely to be male, have larger networks, sit on more boards, and are more likely to have a finance background. These attributes characterize the stereotypical director in most large companies. A plausible interpretation of our results is that firms that nominate predictably unpopular directors tend to choose directors who are like existing directors, while the algorithm suggests that adding diversity would be a better idea.

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

## **2. Machine Learning Algorithms to Predict Director Performance**

We build algorithms designed to make an ex ante prediction of directors' level of relative shareholder support, averaged over the first three years of their tenure. The algorithms use a set of observable director, board, and firm features that are available to the nominating committee at the time of the nominating decision. The algorithms are commonly used in the supervised machine learning literature: *lasso*, *ridge*, *neural networks* and *gradient boosting trees*. We train each of these algorithms, i.e. estimate model

---

<sup>5</sup> Li et al. (2018) use machine learning (word embedding) to measure corporate culture.

<sup>6</sup> We thank Oren Etzioni, CEO of the Allen Institute for Artificial Intelligence, for originally pointing out this benefit of our approach to us.

parameters, on directors appointed between 2000 and 2011 and test them on directors appointed between 2012 and 2014. Following the terminology in machine learning, we call the data from 2000-2011 the “training set” (in-sample data) and the data from 2012 to 2014 the “test set” (out-of-sample data).

The algorithms combine candidate predictors in highly flexible and nonlinear ways to produce their best prediction of the specified outcome variable for new observations in the test set. There are a number of well-known machine learning algorithms that can be used for our prediction exercise. We use four of these algorithms to predict director performance, and give a brief summary of each in this section.

### 2.1. *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 the bias-variance tradeoff in the machine learning literature: if an algorithm fits in-sample data too well (low bias), it has high variance and thus does not perform as well on out-of-sample data. *Lasso* and *ridge* are both linear models that use a regularization term to achieve a balance between bias and variance. They do so by minimizing a loss function that includes in-sample fit and a penalty term that favors simple models, thereby reducing variance (see online appendix for more details).

### 2.2. *Gradient Boosting Trees*

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



of observations in the leaf it ends up in. The random forest algorithm takes an average of the predictions from all the decision trees.

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

### 2.3. Neural Networks

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

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

### 3.1. Measuring Director Performance through Re-Election Results

A challenging yet essential part of designing an algorithm to select directors is the way in which the algorithm measures director performance. We use the relative shareholder support that directors receive in

annual director re-elections as a market-based measure of *individual* directors' performance. Specifically, our main outcome variable is the average level of shareholder support over the first three years of director tenure, adjusted each year by the average support for the entire slate of directors up for re-election on that board that year.<sup>7</sup> All our results are qualitatively unchanged if we task the algorithms with predicting the absolute level of shareholder support, *i.e.* if we do not subtract the average for the slate.

The literature on director re-elections is large. See, for example, Boone, Field, and Karpoff (2007), Linck, Netter, and Yang (2008), Cai, Garner and Walkling (2009), Linck, Netter, and Yang (2009), Fischer et al. (2009), Coles, Daniel and Naveen (2014), Iliev, Lins, Miller, Roth (2015), Aggarwal, Dahiya and Prabhala (2017), Ertimur, Ferri and Oesch (2017), Cai, Nguyen and Walkling (2017), Fedaseyev, Linck, and Wagner (2017), Fos, Li and Tsoutsoura (2017).

One potential concern with using shareholder support as our measure of director performance is that in the vast majority of cases, directors receive overwhelming majority. Most studies report a mean shareholder support around 95%. Therefore, there is virtually no variation in the *outcome* of the re-elections. If the results reflect the market's perception of a director's quality, it must be that variation among winning votes contains meaningful differences in the market's assessment. Consistent with this notion, Cai et al. (2009), Fischer et al. (2009), and Iliev et al. (2015) suggest that cross-sectional variation in shareholder support does in fact reflect market perceptions of director quality. These papers find that vote totals predict stock price reactions to subsequent turnover. In addition, vote totals are negatively related to CEO turnover, board turnover, management compensation levels, and the probabilities of removing poison pills and classified boards. Moreover, director re-elections appear to have real consequences, even if the elections are not contested and the nominated directors end up being re-elected. Fos et al. (2017) find that when directors are closer to getting re-elected, they are more likely to fire CEOs, presumably to persuade shareholders that they are being more diligent. Aggarwal et al. (2017) suggest that directors with low relative support are

---

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

more likely to leave the board, and if they stay, tend to move to less prominent positions. Ertimur et al. (2018) find that when votes are withheld from directors, boards explicitly attempt to address shareholders' concerns.

A second potential concern with using shareholder support as our measure of director performance is that votes could reflect arbitrary recommendations by proxy advisors such as ISS. Ertimur et al. (2018) report that since 2003 large institutional investors take an active role in developing the guidelines that are the basis of ISS recommendations, which, as such, reflect its clients' aggregated preferences. This confirms findings in Aggarwal, Erel and Starks (2016) who show that institutional investors and proxy advisors pay attention to the changing opinions of their beneficiaries and shareholders. The recent literature shows however that institutional investors do *not* follow proxy advisors' recommendations blindly. Aggarwal et al. (2016) find that shareholders are less likely to follow the recommendations of either management or proxy advisory firms as shareholders are forming their own views due to changes in public opinion. Iliev and Lowry (2014) show that institutional investors with larger size of ownership tend to vote more independently from ISS recommendations. Our results are unchanged when we repeat our tests by focusing on a subsample of firms with larger-than-median (26%) ownership by the top-5 institutional owners. In addition, using detailed voting data from 2003-2017, Heath et al. (2019) show that when ISS recommends voting against management, index (active) funds vote with management 54% (42%) of the time. This recent stream of the literature strongly suggests that shareholder votes are not simply the reflection of an arbitrary recommendation issued by proxy advisors.

Overall, the literature finds that shareholder support does reflect perceptions of director quality, that directors care about these perceptions, and that they take actions to influence them. The question then is: are algorithms able to pick up variations in these perceptions of director quality despite the fact that most directors receive extremely high support? Put differently, the highly skewed distribution of our specified outcome variable may be viewed as an additional challenge for the algorithms. Are they still able to identify *ex ante* who will end up in the left tail?

### *3.2. Sample Selection*

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

We obtain data on the level of shareholder support for individual directors from *ISS Voting Analytics* and focus on directors appointed during our sample period. To account for firm level effects, we use *excess votes* as our specified outcome variable in most of our reported results. To construct *excess votes*, we use the number of votes in favor over all votes cast (yes, no, withheld). We then subtract the average for the slate of directors up for re-election on that board and take the average of this variable over the first three years of tenure. Our sample contains the voting outcome, i.e. *excess votes*, for 24,054 new director appointments. All our results are similar when we use shareholder support and do not subtract the support for other directors on the slate.

### 3.3. Summary Statistics

Table 1 presents summary statistics for average shareholder support (mean total votes) and for excess votes over the first three years of tenure. As previously documented in the literature on uncontested director elections, the overall level of shareholder support is typically very high. Given that the mean level of support is .95 and the median is .975 (with a standard deviation of .07), a voting outcome below 95% is a relatively poor outcome. Consequently, a voting outcome below 95% likely reflects a perception of poor performance by the director. Starting in Column 4, we report shareholder support after subtracting the average support for the entire slate of directors in that year. Although shareholder support in uncontested elections is typically very high, shareholders do on occasion oppose newly nominated directors (see figure in online appendix).

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

TABLE 1: SHAREHOLDER SUPPORT SUMMARY STATISTICS

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

Table 2 illustrates that the frequency of shareholder discontent varies by director and board characteristics. For example, the fraction “poor outcomes”, representing the bottom 10% of the sample in terms of excess votes, is 10.6% for male directors and 7.9% for female directors.<sup>8</sup> Similarly, busy directors (serving on three or more boards) experience low shareholder support more frequently than non-busy directors. However, theory provides little guidance regarding the particular variables and functional forms of the relation between the various director, board and firm characteristics and the performance of directors. For example, we do not know whether we should expect female busy directors with a Ph.D. serving on the large board of a small firm in the pharmaceutical industry to receive higher or lower shareholder support on average than a male director who serves on a single small board of a large manufacturing corporation. The problem increases in complexity when many more covariates are likely to matter. For this reason, we rely on an estimation procedure that does not impose the specific form for the relationship between potential

<sup>8</sup> A similar table can be constructed using total shareholder support and defining a poor outcome when shareholder support is below 80% for example.

explanatory variables. Machine learning algorithms therefore provide a disciplined and rigorous approach to model selection (Athey, 2017).

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

TABLE 2: AVERAGE FRACTION OF POOR OUTCOME

This table presents the average fraction of “poor outcome” for various director-level and board-level characteristics. A director is considered to experience a poor outcome if her excess votes is < -2%. Poor outcomes represent 10% of the sample.

## 4. Evaluating Machine Learning Predictions of Director Performance

### 4.1. Model Specification

We develop machine learning algorithms that predict the performance of a potential director. We first “train” each algorithm on the 2000-2011 portion of our sample containing 18,476 new independent director appointments, of which 12,815 are unique directors, at 2,407 firms. Training involves having the algorithm determine which combinations of variables best predict future performance.<sup>9</sup> We evaluate the models’ out-of-sample predictions on the held out 2012-2014 portion of our sample containing 5,578 new director appointments, of which 4,019 are unique directors, at 569 firms. We compare those out-of-sample predictions to those from an OLS model. All comparisons are based on predictions for the 2012-2014 subsample of director appointments, which does not overlap with the 2000-2011 subsample on which the algorithms are trained.

<sup>9</sup> The algorithms rely on a regularizer that balances out in-sample fit and out-of-sample overfitting.

## 4.2. Predictions of Director Performance

Table 3 summarizes the ability of the machine learning models, once trained on the earlier portion of the sample, to predict director success in the later part.<sup>10</sup>

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

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

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

A simple test of a model for predicting performance is whether actual performance is an increasing function of predicted performance. Table 3 indicates that average observed shareholder support almost monotonically increases across model predicted performance percentiles for each machine learning model. However, in contrast to the machine learning models, the average observed outcome of directors in the bottom of the predicted performance distribution using the OLS model (.028) is actually higher than that of directors in the top of the predicted performance distribution (.006).

Among the machine learning algorithms, *XGBoost* performs best at predicting the subsequent success of directors using both excess votes and total votes as measures of director performance.<sup>11</sup> Directors predicted to be in the bottom percentile as predicted by *XGBoost* have an average *observed* excess

<sup>10</sup> See Online Appendix for the OLS model used in the reported results. Alternative models, for example without fixed effects and with different covariates, lead to the same conclusion in terms of OLS's performance to predict director performance.

<sup>11</sup> *XGBoost* is an algorithm with a reputation for generating excellent predictions on a variety of problems, and was the most often used algorithm among the winning solutions in the 2015 machine learning Kaggle competition.

shareholder support of -3.1%, whereas the average observed excess support is 1.1% for directors in the top percentile of predicted performance. This pattern highlights the difference between the machine learning model and OLS in their ability to predict future performance.

Figure 1 shows the average observed level of shareholder support for directors across the ten deciles of predicted performance for OLS and for the machine learning algorithms in the 2012-14 test period. The figure shows how the mean shareholder support for a director is an increasing function of the predicted one for all the machine learning algorithms, but not for the OLS model. The difference in the predictive ability of various models illustrates the difference between standard econometric approaches and machine learning. OLS fits the data well in sample but poorly out of sample. In contrast, machine learning algorithms are specifically designed to predict well out of sample.

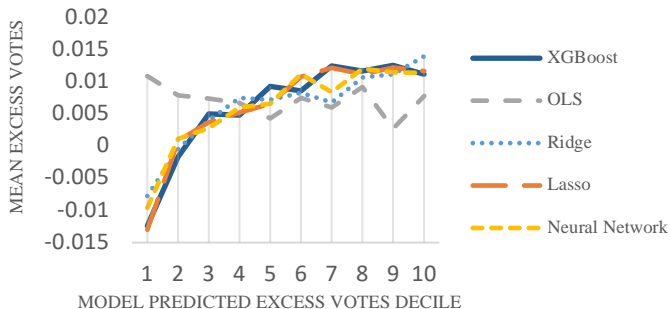


FIGURE 1: MEAN OBSERVED *EXCESS VOTES* VS. PREDICTED *EXCESS VOTES*

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

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



it. Machine learning, by letting the data speak about the underlying relationships among the variables, ends up fitting the data better and consequently does better at predicting outcomes out of sample.

#### 4.3. *Excluding Poorly Performing Firms*

A possible concern with this analysis is that the relation between predicted performance and subsequent performance could occur only because of poorly performing firms. A poorly performing firm would likely be less attractive to a director, so it could be that only low ability directors are attracted to poorly performing firms, even if the firms are relatively large and otherwise prestigious. Because of their low ability, these directors would tend to do worse *ex post*. We repeat our analyses omitting firms that experience negative abnormal returns in the year prior to the nomination. We find similar results without poorly performing firms in the sample. 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.

### 5. **Designing the Quasi-Labels Procedure to Evaluate the Algorithm**

The results so far suggest that directors identified by our algorithm as likely to have low (high) future shareholder support, are in fact on average more likely to have low (high) support in subsequent elections. Accurate out of sample predictions, however, are not sufficient to imply that algorithms could assist firms in their nominating decisions of corporate directors. Specifically, there are two important and related challenges in assessing whether the algorithmic predictions can actually lead to better outcomes. First, we can only observe how well our algorithm's predictions do for directors who are actually appointed to that position (the *selective labels* problem). Second, when deciding on their choice of directors, decision makers presumably take factors into account that are not observable to the algorithm. Therefore, directors who were nominated, although they could share the same observable features as potential alternatives, could differ in terms of unobservables. In particular, they could have been chosen because they have a set of skills that are valuable to the firm, or because they have a personal relationship with the CEO or existing directors. A firm could also have decided not to nominate a candidate based on some characteristics unobservable to the

algorithm that would make this candidate a poor choice. We cannot observe these factors, yet they could lead to different average outcomes for nominated vs. not nominated, even if both are identical on the basis of observable characteristics.

To formalize these concepts, we develop a framework in the spirit of Kleinberg et al. (2017) and present it in the online appendix. Our empirical strategy to address these concerns involves the design of a pool of realistic potential candidates for each vacant board position. Using this pool of potential candidates, we wish to evaluate the algorithm’s predictions of the performance of the directors who firms actually nominated. In cases where our algorithm predicted low performance for these directors, we are interested in whether there were plausible alternatives available, how they would have performed, and how the director who was nominated actually performed compared to those alternatives.

Each new board appointment in the test set is associated with a candidate pool, comprised of directors who, within one year of the appointment, joined the board of a smaller neighboring firm.<sup>12</sup> By revealed preference, we know that these directors were available to join a board at that time and were willing to travel to that specific location for board meetings. We restrict the pool of potential candidates to directors who joined a *smaller* neighboring firm since the prestige and remuneration of being a director tends to increase with company size (see Masulis and Mobbs, 2014). There are on average 147 candidates in a candidate pool. Our results are similar if we modify how we construct pools of candidates, for example by further restricting the set of candidates in candidate pools to directors who joined the board of a firm in the same industry.<sup>13</sup>

To generate predictions for potential candidates at the focal firm, our algorithms use the board and firm characteristics as well as the committee assignments of the appointment at the focal firm with the individual potential candidate’s features. We do not observe the performance of these potential candidates (the essence of the selective labels problem). We do however observe what we refer to as their “*quasi-label*”, which is

---

<sup>12</sup> A neighboring firm is defined as a firm whose headquarters is within 100 miles of the focal firm’s headquarters. The average distance with the focal board is 35 miles (median distance is 26 miles).

<sup>13</sup> There are on average 33 candidates in these more restrictive candidate pools.

an informative signal that serves as a substitute for a direct measure of performance. In our setting, a quasi-label is the director’s performance on the board he or she actually joined. This measure represents an indication of how potential candidates would have performed on the focal board.

We provide a schematic representation of the quasi-label procedure that we design in Figure 2. We first rank all nominated directors in our test set according to their performance as predicted by the algorithm. For all nominated directors in the bottom decile of predicted performance, we consider their associated candidate pool and rank candidates in this candidate pool according to their predicted performance *on the focal board*. We re-rank promising candidates according to their quasi-labels. The question we consider is: how does the observed performance of hired directors predicted to do poorly compare to the performance of available alternative candidates?

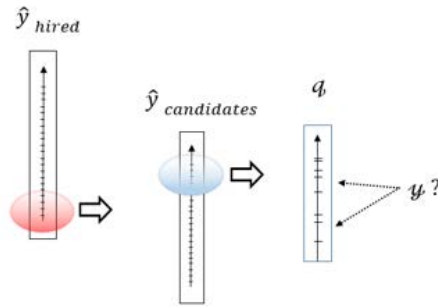


FIGURE 2: ASSESSING THE ALGORITHM’S PREDICTIONS USING QUASI-LABELS

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

If the observed performance of the nominated director ranks high in the distribution of quasi-labels, this would suggest that even though our algorithm predicted this particular director would do poorly, she ended up doing well relative to available alternatives. The focal board might have relied on unobservables in the nomination process, and the high rank in the distribution of quasi-labels would suggest that unobservables were used as *signal*. On the other hand, if the observed performance ranks low in the distribution of quasi-labels, then our algorithm would have identified *ex ante* that this director would

perform poorly, and relative to alternatives, she indeed did perform poorly. This pattern would suggest that any unobservables used in the nomination decision process was not a signal of performance, but was noise, bias, or related to agency problems.

Table 4 presents the median rank in the distribution of quasi-labels for directors in the bottom and top deciles of predicted performance for several machine learning algorithms, as well as for an OLS model. For all machine learning models, nominated directors predicted to do well performed noticeably better than available alternative candidates, while nominated directors predicted to do poorly performed worse than available alternative candidates. *XGBoost* and *lasso* again appear to be the preferred algorithms. They can best discriminate *ex ante* the directors who will do well from those who will not. In the rest of the paper, we focus on results with *XGBoost* to simplify the discussion.<sup>14</sup> The median director predicted by the *XGBoost* algorithm to be in the *bottom* decile of performance ranks at the 27<sup>th</sup> percentile in the distribution of quasi-labels. The median director predicted to be in the *top* decile ranks at the 78<sup>th</sup> percentile in the distribution of quasi-labels. In contrast, the predictions from the OLS model are uninformative about subsequent performance; directors rank around the 75<sup>th</sup> percentile in the distribution of quasi-labels, regardless of whether OLS predicted they would perform well or not.

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

TABLE 4: EVALUATING THE PREDICTIONS USING QUASI-LABELS

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

<sup>14</sup> *XGBoost* is the algorithm that systematically performs very well on the various specified outcomes that we ask the algorithms to predict. All results are similar across models.

Table 4 includes results for the top and bottom deciles of predicted performance. However, our results are similar across all deciles of performance and when compared to *all* potential candidates (i.e. not conditioning on the most promising candidates). Figure 3 shows that the mean and median rank (percentile) in the distribution of quasi-labels almost monotonically increases across deciles of predicted performance. The observed performance of hired directors in the test set is compared to the performance of *all* potential candidates in their respective candidate pool. These results suggest that machine learning models can be helpful to predict whether an individual will be successful as a director in a particular firm.

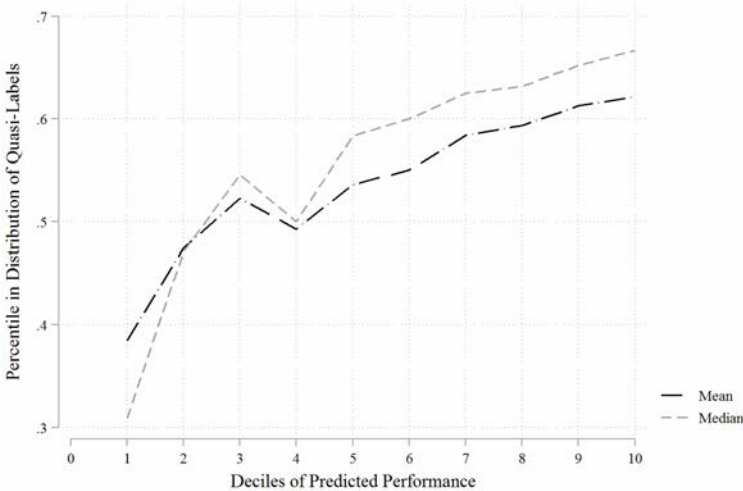


FIGURE 3: MEAN AND MEDIAN RANK IN QUASI-LABEL DISTRIBUTION ACROSS DECILES OF PREDICTED PERFORMANCE

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

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

There potentially exist many settings in which quasi-labels may be used to assess the algorithms’ predictions if they represent a plausible substitute for missing labels. Under the assumption that the difference between the unobserved missing label and its quasi-label does not vary in a predictable way

across the distribution of the predicted outcomes, the quasi-label procedure potentially offers a useful approach in various contexts.<sup>15</sup>

In our setting, the endogenous nature of the board-director match could lead to systematically inflated quasi-labels, i.e. by revealed preference, the performance of the available candidate would not be as high on the focal board.<sup>16</sup> If quasi-labels are inflated due to the endogenous board-director match, then the difference between the quasi-label and the missing label is expected to be positive. However, we do not expect this difference to vary in any predictable way across deciles of predicted performance. A closer look at our quasi-label procedure results indicates that the rank in the distribution of quasi-labels increases almost monotonically across model prediction deciles (Figure 3). Therefore, whereas nominated directors predicted to do poorly rank *low* compared to alternative candidates, nominated directors predicted to do well rank *high*. The symmetry of this result suggests that our quasi-labels are not systematically inflated.

Our results suggest that the algorithm is able to identify who will perform well and who will not. Algorithms have the potential to improve on real world boards' nominating decisions by affecting the rank ordering of potential directors.

## 6. Director Popularity or Performance?

An important interpretational issue is understanding exactly what the algorithm is predicting. The fiduciary responsibility of directors is to maximize shareholders' welfare, so choosing directors who will receive the most subsequent votes would seem to be a natural approach. One concern, however, is that many institutional shareholders decide on their votes through recommendations of shareholder services companies such as ISS. ISS introduced guidelines in the latter part of our training period. For example, explicit guidelines to support proposals aimed at increasing female board representation were introduced in 2010. However, our training sample covers data from 2000-2011. Less than 20% of appointments in our training set take place when ISS had those specific guidelines in place.

---

<sup>15</sup> For instance, suppose a researcher wanted to evaluate algorithmic predictions of loan performance. Quasi-labels for denied loans may be the loan performance for the firm (or individual) offered a loan by a different institution.

<sup>16</sup> Note that this would assume that boards and directors are skilled at using unobservables to match.

Following Iliev and Lowry (2014) who argue that institutional investors with larger size of ownership vote independently from ISS recommendations, we repeat all our tests by focusing on a subsample of firms with larger-than-median (26%) ownership by the top-5 institutional owners. Our results (available upon request) remain very similar.

Whereas the recent literature on routine director re-election does report that votes indeed capture the performance of directors, we test whether we find evidence for this in our data. Specifically, we compare the cumulative abnormal returns (CARs) around the announcement of director appointments in our test set for directors predicted to do well to those for directors predicted to do poorly.<sup>17</sup>

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

TABLE 5: CUMULATIVE ABNORMAL RETURNS AROUND APPOINTMENT ANNOUNCEMENTS

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

Table 5 reports the mean CARs using a (-1; +1) window around announcements. The same pattern emerges using longer windows as well. Using our *XGBoost* algorithm to predict excess votes, we find that the mean CAR for directors predicted to do poorly (decile 1) in our test set is -1.94% whereas it is +0.75% for directors predicted to do well (decile 10). The difference is statistically significant at the 1% level. Directors predicted to be unpopular also tend to be viewed by the market as worse directors. We also used the algorithm to predict announcement CARs using a smaller sample for which announcement dates are available, with similar results.

Finally, we train an *XGBoost* algorithm to predict a measure of firm profitability, *EBITDA/Total Assets*, three years post appointment. We then sort directors in our test set into deciles based on predicted

<sup>17</sup> We collect announcement dates from BoardEx, CapitalIQ and Lexis-Nexis.

profitability. We report the actual profitability as well as the shareholder support in the first two rows of Table 6.<sup>18</sup> The model trained to predict profitability in the subsequent period indeed does predict future profitability well. The actual profits for the firms sorted into deciles based on expected profits increase monotonically, with average profits increasing with the model’s expectation of profitability.

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

TABLE 6: COMPARING SHAREHOLDER SUPPORT MODELS WITH PROFITABILITY MODELS

This table reports the actual performance for each decile of *XGBoost*-predicted performance. *XGBoost* is trained to predict 1) firm profitability three years after the director has been appointed (EBITDA/Total Assets) 2) total votes and 3) excess votes. The results are for our test set only (out-of-sample performance for directors appointed between 2012-2014).

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

These results suggest that the choice of training the algorithm on shareholder support in director elections is not crucial for the algorithm to be able to predict director quality. When the model is trained

<sup>18</sup> The correlation of EBITDA/Total Assets with the shareholder support measure is 0.12 (p-value: 0.000).



using profitability instead, the pattern of predictions is similar. The algorithm predicts future subsequent support. Since this support is based on the market's perception of a director's contribution to quality, the results are similar when the algorithm is trained on profitability directly. In addition, for the algorithm trained on shareholder support that we discussed above, we consider whether it can also predict future profitability in addition to future shareholder support. We break the sample into deciles based on the algorithm's predictions of *excess votes* and *total votes*, and present average observed *excess votes*, *total votes* as well as the average profitability for each decile. We present these averages in the bottom four rows of Table 6.

As discussed above, *XGBoost* is successful in predicting future shareholder support (i.e. total votes) and excess votes: average shareholder support in the lowest decile is 92% (-1.2% for excess votes), compared to 97.7% in the top decile (1.1% for excess votes). In addition, it also predicts future profitability. Firms that nominated directors in the bottom decile of predicted shareholder support have an average profitability of -0.3%, whereas firms that nominated directors in the top decile of predicted shareholder support have an average profitability of 10%. When *XGBoost* predicts excess votes, the average profitability of firms in the bottom decile is .6% and it is 11.2% for the top decile (Figure 4). This finding has two important and related implications: first, shareholder votes do appear to be closely related to firm performance, thereby supporting their use as a metric to evaluate director performance and second, nominating directors on the recommendation of an algorithm trained to predict shareholder votes would not come at the expense of poor firm performance.<sup>19</sup>

---

<sup>19</sup> This result alleviates concerns related to the omitted payoff bias articulated in Kleinberg et al. (2017), which in our setting refers to the concern that the decision-maker could have alternative objectives other than satisfying shareholders when making the nominating decision.

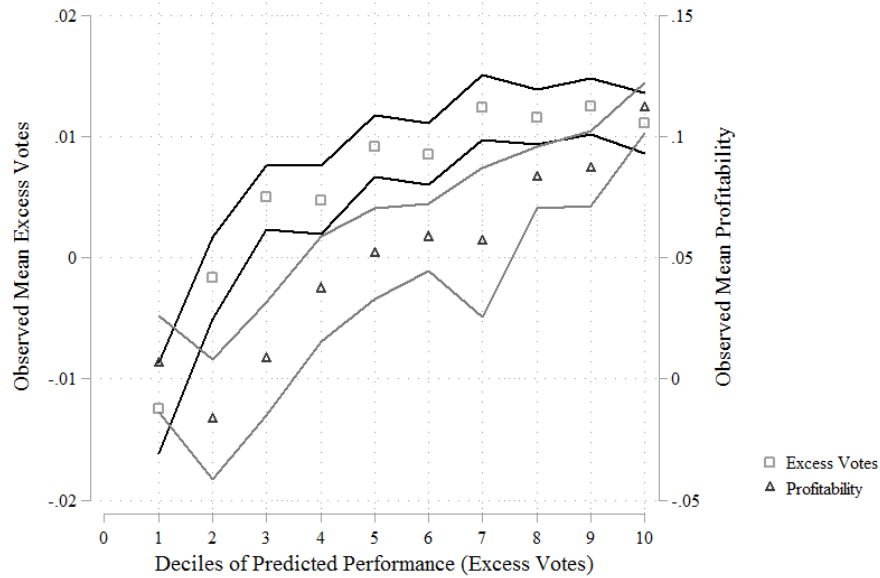


FIGURE 4: OBSERVED PERFORMANCE (EXCESS VOTES AND PROFITABILITY) ACROSS DECILES OF PREDICTED PERFORMANCE

This figure reports the actual mean *Excess Votes* (left y-axis) and mean firm profitability (right y-axis), with their respective 95% confidence interval, for each decile of *XGBoost*-predicted director performance (Excess Votes) for directors in the test set.

## 7. Characteristics that Affect Director Performance

The machine learning models appear to be able to predict which directors are likely to receive more shareholder support in subsequent elections and this support appears to be related to firm level performance. One of the differences with traditional econometric modeling is that the machine learning algorithms do not provide a formula that can be used to infer the influence of any particular independent variable on performance. To understand which characteristics affect director performance, we consider the predictions from the machine learning models and evaluate the extent to which director and firm characteristics are associated with high and low predicted performance.

### 7.1. Univariate Comparisons

Table 7 provides some guidance about which director features are valued by the algorithm in its assessment of directors. This table reports the averages of a number of characteristics of potential directors, boards, and firms that are associated with low and high expected future voting. In particular, it presents the means of these characteristics for the bottom and top deciles of predicted shareholder support predicted by the

*XGBoost* model. There are notable differences between directors in the top and bottom deciles. In particular, directors in the bottom decile are more likely to be male, sit on more current boards, have sat on more boards in the past, have received lower shareholder support in previous elections for other boards they sat on, and have a larger network. These differences suggest that male directors who are on a number of boards tend to be less desirable directors on average, either because they are too busy to do a good job or because they are less likely to monitor the CEO.<sup>20</sup>

Board-level variables that affect predicted excess votes likely reflect perceptions of the quality of governance in a particular firm. Note that the outcome variable presented is the *Excess votes*, which is adjusted for the average support for the slate, and therefore, some of the statistics are harder to interpret. For instance, the average tenure of incumbent board members is about three years for directors in the bottom decile of predicted performance, whereas it is about ten years for those in decile ten. This pattern occurs because a new director is more likely to receive more votes relative to other directors up for reelection if the incumbents have been on the board for a very long time, most likely because of a recent push for board refreshment.<sup>21</sup> In unreported results where the top and bottom deciles refer to unadjusted (total) votes, we see that longer average director tenure, which is likely to reflect an entrenched board, is associated with lower predicted shareholder support.

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

---

<sup>20</sup> Fich and Shivdasani (2006) present evidence suggesting that a director being overly busy can meaningfully affect their monitoring of management.

<sup>21</sup> See for example <https://corpgov.law.harvard.edu/2016/07/29/refreshing-the-board/>

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

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

TABLE 7: TOP VS. BOTTOM DECILE OF PREDICTED PERFORMANCE

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

## 7.2. Multivariate Comparisons

Because director and firm characteristics are not independent from one another, we estimate regressions of predicted performance. As independent variables, we include firm, board and director variables. The coefficients reflect the characteristics that *XGBoost* tends to associate with higher performance. We report estimates of these regressions in Table 8. The relatively low  $R^2$  of these regressions is not surprising and speaks to the importance of feature interactions and non-linearities that *XGBoost* relies on to generate its predictions of subsequent performance.

Director variables related to predicted subsequent shareholder support are gender, a dummy variable that indicates whether the director is “busy” and the number of listed boards a director serves on. In particular, the algorithm suggests that male directors and directors who are on at least three boards (“busy” directors) tend to receive less support from shareholders, which we found was related to firm performance as well.

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

*t*-statistics in parentheses

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

TABLE 8: THE DETERMINANTS OF PREDICTIONS: OLS REGRESSIONS

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

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

Board level variables that are significantly related to the predicted shareholder support are the size of the board, the average tenure of incumbent board members, and the average number of independent directors. These variables again are likely to reflect the independence of the board from management. Firm-level variables that appear to be associated with subsequent performance are size (total assets), operating performance, and whether the firm pays dividends.

### *7.3. Overvalued Director Characteristics*

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

In Table 9, we report characteristics of directors who were nominated, but whom the algorithm predicted would do poorly and who indeed subsequently did poorly. Compared to promising candidates identified by the algorithm, predictably unpopular directors are on average more likely to be male, have fewer degrees post undergraduate, a larger professional network, more current and past directorships, and are more likely to have a background in finance.<sup>22</sup>

---

<sup>22</sup> Predictions for candidates assume the same committee assignments as the nominated director. We find very similar results for all alternative specifications mentioned in previous sections.

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

TABLE 9: OVERVALUED DIRECTOR CHARACTERISTICS

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

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



## 8. Summary and Discussion

We present a new approach based on algorithms to select the directors of publicly traded companies. In developing the machine learning algorithms, we contribute to our understanding of governance, specifically boards of directors, in at least four ways. First, we evaluate whether it is possible to construct an algorithm that can predict 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 provide evidence that director popularity is related to their expected value and that shareholder support is thus a meaningful proxy for their performance. Finally, we use the selections from the algorithms as benchmarks to understand the process through which directors are actually chosen and identify the types of individuals who are more likely to be chosen as directors *counter* to the interests of shareholders.

There are a number of methodological issues we must address before we can design such an algorithm. We must be able to measure the performance of a director to predict which potential directors will be of highest quality. Our approach is based on levels of shareholder support. We ensure the validity of our specified outcome variable in several ways. 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 vote-based performance measure is an *individual* measure which reflects the support the director personally has from the shareholders she represents and which should incorporate all publicly available information about her performance. We show that the vote-based performance measure is informative about directors' quality as it is related to firm profitability and firm value. Importantly, selecting directors based upon predictions of how popular they will be with shareholders does not come at the expense of other dimensions management likely cares about.

Using publicly available data on firm, board, and director characteristics, our algorithm can predict the success of directors. In comparison to the machine learning models, standard econometric models fit the data poorly out of sample. Specifically, the observed performance of individual directors is not related to

the predictions of performance of an OLS model. The fact that the machine learning models dramatically outperform econometric approaches is consistent with the arguments of Athey and Imbens (2017) and Mullainathan and Spiess (2017): machine learning is a valuable approach for prediction problems in the social sciences.

There is an additional methodological issue we need to address before we can conclude that algorithms can help us understand the director nomination process. We observe the predictive accuracy of our algorithm only for directors who were nominated. We design a quasi-labels procedure which exploits the fraction of votes plausible candidates received at the company whose board they joined as an indication of their performance. We find that directors the algorithm predicted would do poorly (well) indeed do poorly (well) when compared to realistic alternatives.

The differences between the directors suggested by the algorithm and those actually selected by firms allow us to assess the features that are overrated in the director nomination process. Comparing predictably poor performing directors to promising candidates suggested by the algorithm, it appears that firms choose directors who are more likely to be male, have a large network, have many past and current directorships, and have a finance background. In a sense, the algorithm is saying exactly what institutional shareholders have been saying for a long time: that directors who are not old friends of management and come from different backgrounds are more likely to monitor management. In addition, less connected directors potentially provide different and potentially more useful opinions about policy. For example, TIAA-CREF (now TIAA) has had a corporate governance policy aimed in large part at diversifying boards of directors since the 1990s for this reason (see Biggs (1996) and Carleton et al. (1998)).<sup>23</sup>

Our finding on the predictability of which directors will or will not be popular with shareholders has important implications for corporate governance. Observers since Smith (1776) and Berle and Means

---

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

(1932) have been concerned about whether managers intentionally select boards that maximize their own interests rather than those of the shareholders. The psychology literature started by Meehl (1954) has found that due to behavioral biases, even simple algorithms can outperform humans in deciding on personnel decisions. We believe that machine learning algorithms, with their powerful predictive ability, present an opportunity for firms to improve their selection process.

A natural question concerns the applicability of algorithms such as the ones we developed in practice. We view our work as a “first pass” approach, aimed at bringing the predictive power of machine learning tools to the issue of director selection. More sophisticated models with richer data would undoubtedly predict individual director performance better than the models presented here. If algorithms such as these are used in practice in the future as we suspect they will be, practitioners will undoubtedly have access to much better data than we have and should be able to predict director performance more accurately than we do in this paper. An important benefit of algorithms is that they are not prone to the agency conflicts that occur when boards and CEOs together select new directors.

Algorithmic bias and algorithmic fairness represent concerns of growing importance. Our data, as is almost always the case in the social sciences, was generated by human decisions. As a result, machine learning algorithms can generate bias amplification (Zhao et al., 2017). While our choice of specified outcome helps mitigate the concern of “historical bias” (the decision maker and the evaluator are separate entities), special attention should be paid to various properties of any dataset prior to the implementation of an algorithmic decision aid. As Miller (2018) argues however, the perils of human bias are arguably worse than the perils of algorithmic bias. The benchmark against which algorithms should be evaluated should not necessarily be zero bias, but rather prevailing biases given that humans make decisions.

Institutional investors are likely to find the algorithm’s independence from agency conflicts particularly appealing and are likely to use their influence to encourage boards to rely on an algorithmic decision aid such as the one presented here for director selections in the future. An important advantage of an algorithm over the way in which directors have been chosen historically is that “algorithms can overcome the harmful effects of cognitive biases” (Sunstein, 2018). Rivera (2012) studies the hiring practices of top investment

banks, consulting and law firms and concludes that recruiters overvalue personal fit which is not necessarily a function of expected performance. In the context of lower skill workers, Hoffman et al. (2017) find that managers who hire against test recommendations end up with worse average hires. Cowgill (2018) shows that the job-screening algorithm at a software company prefers “nontraditional” candidates. Our results suggest that the same idea applies to the nominating of corporate directors. Including algorithmic input to limit (but not strip) discretion and reliance on soft information in these decisions could help minimize agency problems, and thus lead to a modified rank ordering of candidates that could in turn lead to better directors than the current process.

On the other hand, if the algorithm omits attributes of potential directors that are valuable to management, such as specialized knowledge of an industry or government connections, then it potentially could lead to suboptimal solutions. This is why we advocate for tools built on algorithms as decision aids, not substitutes for human judgement. Humans and machines both have limits and make different kinds of mistakes, i.e. they tend to have uncorrelated errors. Achieving the right balance in the division of labor between humans and machines to take advantage of their relative strengths is key.<sup>24</sup>

In this paper, we use 21<sup>st</sup> century technology to confirm an observation that dates back over two hundred years: the board selection process leads to directors who often those nearest at hand and are not necessarily the best choices to serve shareholders’ interests. This technology can, however, in addition to confirming this observation, provide us with the tools to change it. By providing a prediction of performance for *any* potential candidate, a machine learning algorithm could expand the set of potential directors and identify individuals with the skills necessary to become successful directors, who would have otherwise been overlooked. We expect that in the not too distant future, algorithms will fundamentally change the way corporate governance structures are chosen, and that shareholders will be the beneficiaries.

---

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

## References

- Adams, R. B (2017) Boards and the Directors Who Sit on Them. Chapter 6 of *Handbook of the Economics of Corporate Governance*, edited by B. E. Hermalin and M. S. Weisbach, Elsevier.
- Adams, R., B. Hermalin and M. Weisbach (2010) The Role of Boards of Directors in Corporate Governance: A Conceptual Framework and Survey. *Journal of Economic Literature* 48, 58–107.
- Alexander, C., M. Chen, D. Seppi, and C. Spatt (2010) Interim News and the Role of Proxy Voting Advice. *Review of Financial Studies* 23: 4419-4454.
- Aggarwal, R., S. Dahiya and N. Prabhala (2017) The Power of Shareholder Votes: Evidence from Uncontested Director Elections. *Georgetown McDonough School of Business Research Paper*.
- Aggarwal, R., I. Erel, and L. Starks (2016) Influence of Public Opinion on Investor Voting and Proxy Advisors, Working Paper, Ohio State University.
- Agrawal, A., J. Gans and A. Goldfarb (2018) Prediction Machines. *Harvard Business Review Press*, Boston, Massachusetts.
- Athey, S. (2017) Beyond Predictions: Using Big Data for Policy Problems. *Science* 355: 483-385.
- Athey, S. and G.W. Imbens (2017) The State of Applied Econometrics: Causality and Policy Evaluation, *Journal of Economic Perspectives*, 31, 3-32.
- Autor, D. (2015) Why Are There Still So Many Jobs? The History and Future of Workplace Automation, *Journal of Economic Perspectives*, 29, 3-30.
- Biggs, John (1996) Corporate Governance Assessment: A TIAA-CREF Initiative, *Directors Monthly*, 20 (10) 1-8.
- Boone, A., L. Field, and J. Karpoff (2007) The Determinants of Corporate Board Size and Composition: An Empirical Analysis, *Journal of Financial Economics* 85, 66-101.
- 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.L., N. Daniel and L. Naveen (2014) Co-opted Boards. *Review of Financial Studies* 27: 1751-1796.
- Coles, J., N. Daniel and L. Naveen (2015) Board Groupthink. *Working Paper*.
- Cowgill, B. (2018) Automating Judgement and Decision-making: Theory and Evidence from Resumé Screening. *Working Paper*.
- Daines, R., I. Gow, and D. Larcker (2010) Rating the ratings: How Good are Commercial Governance Ratings? *Journal of Financial Economics* 98: 439-461.
- Edmans, A. (2012) The Link Between Job Satisfaction and Firm Value, With Implications for Corporate Social Responsibility. *Academy of Management Perspectives*: 26, 1-19.
- Ertimur, Y. F. Ferri, and D. Oesch (2013) Shareholder Votes and Proxy Advisors: Evidence from Say on Pay. *Journal of Accounting Research* 51 (5): 951-996.
- Ertimur, Y. F. Ferri, and D. Oesch (2018) Understanding Uncontested Director Elections. *Management Science* 64(7): 3400-3420.
- Fedaseyev, V., J. Linck, and H. Wagner (2017) Do Qualifications Matter? New Evidence on Board Functions and Director Compensation, *Journal of Corporate Finance*, forthcoming.
- Fich, E. and A. Shivdasani (2006) Are Busy Boards Effective Monitors? *The Journal of Finance* 61: 689-724.
- Field, L., M. Souther, and A. Yore (2017) Does diversity pay in the boardroom? University of Delaware Working Paper.
- Fischer, P., J. Gramlich, B. Miller and H. White (2009) Investor Perceptions of Board Performance: Evidence from Uncontested Director Elections. *Journal of Accounting and Economics* 48: 172–189.
- Hart O. and L. Zingales (2017) Companies Should Maximize Shareholder Welfare Not Market Value. *Journal of Law, Finance, and Accounting*, 2: 247–274.
- Hastie, T., R. Tibsharani and M. Wainwright (2015) Statistical Learning with Sparsity: The Lasso and Generalizations. *Chapman and Hall/CRC*.
- Heath, D., D. Macciocchi, R. Michaely and M. Ringgenberg (2019) Passive Investors are Passive Monitors. Working Paper.

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

Hermalin, B. E., and M. S. Weisbach (2003) Boards of Directors as an Endogenously Determined Institution: A Survey of the Economic Literature. *Federal Reserve Bank of New York Economic Policy Review*, 9(1): 7–26.

Hoberg, G. and G. Phillips (2010) Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies* 23 (10), 3773-3811.

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

Hoffman, M., L. B. Kahn and D. Li (2017) Discretion in Hiring. *The Quarterly Journal of Economics*. qjx042, <https://doi.org/10.1093/qje/qjx042>.

Iliev, P., K. Lins, D. Miller and L. Roth (2015) Shareholder Voting and Corporate Governance Around the World. *The Review of Financial Studies* 28(8): 2167–2202.

Iliev, P. and M. Lowry (2014) Are Mutual Funds Active Voters? *Review of Financial Studies* 28: 446-485.

Kahneman, D. (2011) *Thinking, Fast and Slow*, Farrar, Strauss and Giroux: New York.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan (2017) Human Decisions and Machine Predictions. *The Quarterly Journal of Economics* 133 (1) 237–293.

Kleinberg J., Liang A. and S. Mullainathan (2017) The Theory Is Predictive, but Is It Complete? An Application to Human Perception of Randomness. *Working Paper*.

Kleinberg, J., J. Ludwig, S. Mullainathan and Z. Obermeyer (2015) Prediction Policy Problems. *American Economic Review Papers & Proceedings* 105(5): 491–495.

Kleinberg, J., J. Ludwig, S. Mullainathan and C. Sunstein (2019) Discrimination in the Age of Algorithms. NBER Working Paper No. 25548.

Kramarz, F. and D. Thesmar (2013) Social Networks in the Boardroom. *Journal of the European Economic Association* 11(4): 780-807.

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

Li, K., M. Feng, R. Shen, and X. Yan (2018) Measuring Corporate Culture Using Machine Learning. *Working Paper*.

Linck, J, J. Netter, and T. Yang (2008) Determinants of Board Structure, *Journal of Financial Economics* 87, 308-328.

Linck, J, J. Netter, and T. Yang (2009) The effects and unintended consequences of the Sarbanes-Oxley Act on the supply and demand for directors, *Review of Financial Studies* 8, 3287–3328.

Malenko, N., and Y. Shen (2016) The Role of Proxy Advisory Firms: Evidence from a Regression- Discontinuity Design, *The Review of Financial Studies* 29: 3394–3427.

Masulis, R. and S. Mobbs (2014) Independent Director Incentives: Where do Talented Directors Spend their Limited Time and Energy, *Journal of Financial Economics* 111, 406-429.

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

Miller, A. (2018) Want Less-Biased Decisions? Use Algorithms. *Harvard Business Review*.

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

Mullainathan, S. and J. Spiess (2017) Machine learning: An Applied Econometric Approach, *Journal of Economic Perspectives*, 31, 4: 87-106.

Pearl J. and D. Mackensie (2018) *The Book of Why*. Basic Books.

Peters, R. H. and L. A. Taylor (2017) Intangible capital and the investment-q relation. *Journal of Financial Economics* 123 (2): 251-272.

Rivera, L.(2012) Hiring as Cultural Matching: The Case of Elite Professional Service Firms, *American Sociological Review*, 77: 999-1022.

Shivdasani, A. and D. Yermack (1999) CEO Involvement in the Selection of New Board Members: An Empirical Analysis, *The Journal of Finance*, 54, 1829-1853.

Shmueli, G. (2010) To Explain or to Predict? *Statistical Science* 25(3): 289-310.

Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, Indianapolis, Liberty Press.

Sunstein, C. (2018) Algorithms, Correcting Biases. Working paper.

Suresh, H. and J. V. Gutttag (2019) A Framework for Understanding Unintended Consequences of Machine Learning. arXiv:1901.10002. In the Proceedings of *Association for the Advancement of Artificial Intelligence (AAAI) 2019*.

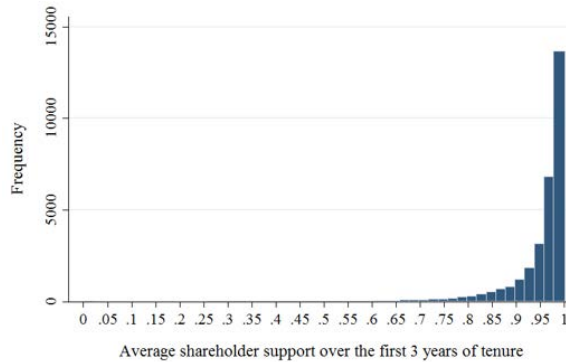
Yermack, D. (1996) Higher Market Valuation of Firms with a Small Board of Directors, *Journal of Financial Economics* 40: 185-211.

Zhao Jieyu, T. Wang, M. Yatskar, V. Ordonez, K. Chang (2017) Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints, *EMNLP 2017*.

## For Online Publication

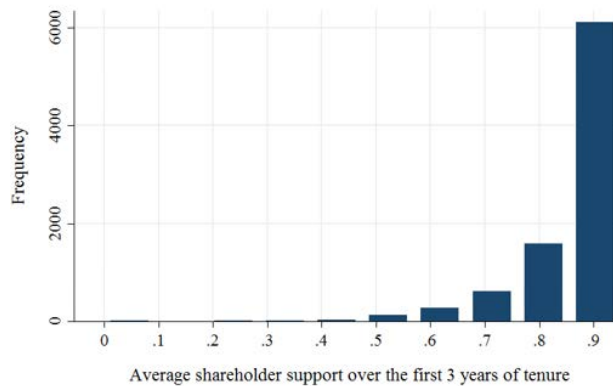
### VOTES DISTRIBUTION

Shareholder Support: Fraction of Votes “For”



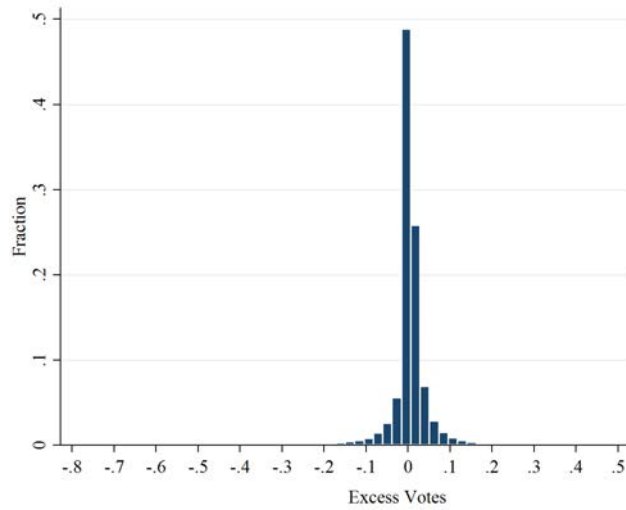
This figure shows the distribution of average shareholder support, defined as the fraction of votes in favor of a given director over all votes cast for the director’s reelection within three years of her tenure. The data is from ISS Voting Analytics.

Distribution of Poor Outcomes: Fraction of Votes “For” Below 95%



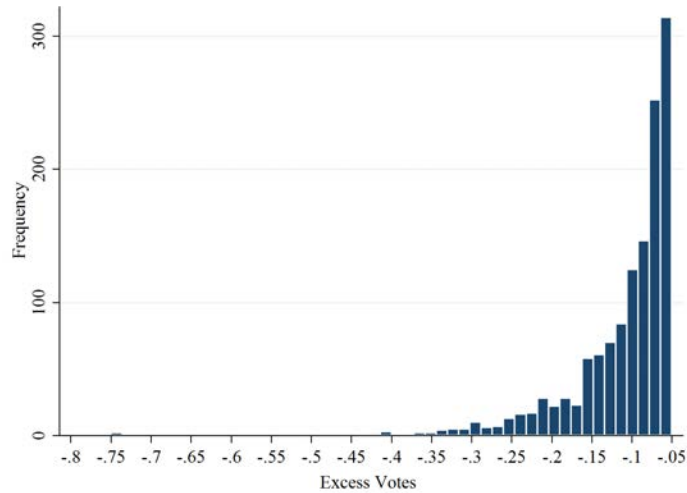
This figure shows the distribution of average shareholder support for values under its mean value of 95%. Shareholder support is defined as the fraction of votes in favor of a given director over all votes cast for the director’s reelection within three years of her tenure. The data is from ISS Voting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

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



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

Distribution of Poor Outcomes: Excess Votes below -5%



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



## OLS MODEL

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

t-statistics in parentheses

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

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

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

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

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

where  $\|\beta\|_1$  is the  $\ell_1$ -norm (least absolute deviation). The penalty weight ( $\lambda$ ) on the sum of the absolute values of coefficients is set using the default parameter in scikit-learn<sup>25</sup>.

*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.<sup>26</sup>

### A.2.2. Gradient Boosting Trees

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

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

---

<sup>25</sup> <http://scikit-learn.org/stable/>

<sup>26</sup> For a detailed discussion of sparse estimators, we refer interested readers to Hastie, Tibshirani and Wainwright (2015).

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

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

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

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

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

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

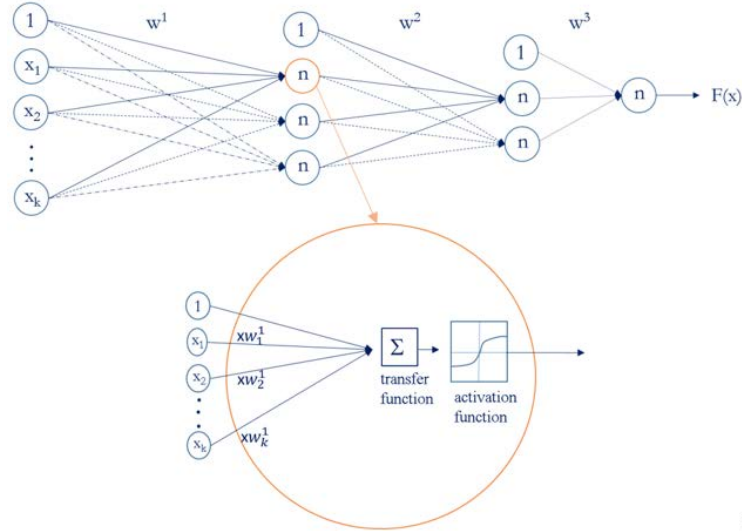
2. Regularized objectives. The split in a decision tree regressor of gradient boosting trees optimizes a regularized global objective that balances the predictive performance and the complexity of decision tree regressors. The loss function is formulated as:

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

where  $l$  refers to a differentiable loss function that measures the difference between the predicted value and the target value (in our case, it is simply squared loss),  $\Omega(f_k) = \gamma T + \frac{1}{2} \lambda ||w||^2$  and measures the complexity of a tree,  $T$  refers to the number of leaves in the tree and  $w$  refers to the score at a leaf. A simple tree has a small number of leaves and each leaf has a small score.  $\gamma$  and  $\lambda$  are parameters to control how these two complexity measures are weighted in the final objective function. The name

gradient boosting trees arise from the fact that a gradient will be computed in the algorithm to optimize the above objective function. Please refer to Chen et al, 2016 for a detailed discussion.

### A.2.3. Neural Networks



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

## A FRAMEWORK TO ASSESS ALGORITHMS' PREDICTIONS

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

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

$\mathcal{F}$  is operationalized by a functional form  $f$ . For the purpose of predictive modeling, we are interested in finding a function that closely matches the function  $f$  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  $f$ . In other words, our supervised machine learning algorithm seeks to learn a functional form that maps features  $W$  into predictions  $\hat{f}(W)$  that generalize well on out-of-sample data (Shmueli, 2010).

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

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

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

For the purpose of the model and similar to Kleinberg et al. (2017), we shrink the dimension of  $\vec{x}$  to a vector with three unidimensional characteristics  $w$ ,  $z$  and  $b$ . In addition, we assume 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:

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

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

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

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

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

The nominating rule  $h_j$  depends on  $k_j(x)$ , the board's *assessment* of future performance for a director with characteristics  $x$ . For a given  $g_j(x)$ , the board chooses the director with the highest  $k_j(x)$ . We do not observe boards' relative weights on director performance,  $u_j$ , and their own preferences for directors

with particular characteristics,  $v_j$ . In a world of perfect corporate governance, boards are only concerned with their mandate (i.e. representing shareholders' interests) and  $v_j = 0$ .

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

The difference in the expected payoffs between the two nominating rules  $\alpha_j$  and  $h_j$  is:

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

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

$$\begin{aligned} E[y|\alpha, w] - E[y|h, w] &= E[w + z|\alpha, w] - E[w + z|h, w] \\ &= E[w|\alpha, w] - E[w|h, w] + E[z|\alpha, w] - E[z|h, w] \\ &= E[z|\alpha, w] - E[z|h, w] \end{aligned}$$

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

We exploit the design of our pool of candidate directors for each board seat in order to compare the performance of our algorithm to board decisions. Although we do not have labels for nominees generated by the algorithm's nominating rule,  $E[y|\alpha]$ , we observe their *quasi-label*: their performance on the smaller neighboring board they joined around the same time.

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



## DATA DEFINITIONS

### A.4.1. Individual Director Features

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

<b><u>Variable</u></b>	<b><u>Definition</u></b>
Age	Director age
Audit chair	Equals one if director is chair of the audit committee
Audit member	Equals one if director is a member of the audit committee
Avgtimeothco	The average time that a director sits on the board of quoted companies
Bkgd academic	Equals one if job history includes in title one of the following: "professor" "academic" "lecturer" "teacher" "instructor" "faculty" "fellow" "dean" "teaching"
Bkgd finance	Equals one if job history includes in title one of the following: "underwriter" "investment" "broker" "banker" "banking" "economist" "finance" "treasure" "audit" "cfo" "financial" "controller" "accounting" "accountant" "actuary" "floor trader" "equity" "general partner" "market maker" "hedge fund"
Bkgd hr	Equals one if job history includes in title one of the following: "hr" "recruitment" "human resource"
Bkgd law	Equals one if job history includes in title one of the following: "lawyer" "legal" "attorney" "judge" "judicial"
Bkgd manager	Equals one if job history includes in title one of the following: "manager" "vp" "president" "director" "administrator" "administrative" "executive" "coo" "chief operating" "operation" "secretary" "founder" "clerk" "division md" "employee" "associate" "head of division"
Bkgd marketing	Equals one if job history includes in title one of the following: "marketing" "publisher" "mktg" "sales" "brand manager" "regional manager" "communication" "merchandising" "comms" "distribution" "media"
Bkgd military	Equals one if job history includes in title one of the following: "captain" "soldier" "lieutenant" "admiral" "military" "commanding" "commander" "commandant" "infantry" "veteran" "sergeant" "army"
Bkgd politician	Equals one if job history includes in title one of the following: "politician" "senator" "political" "deputy" "governor"
Bkgd science	Equals one if job history includes in title one of the following: "researcher" "medical" "doctor" "scientist" "physician" "engineer" "biologist" "geologist" "physicist" "metallurgist" "science" "scientific" "pharmacist"
Bkgd technology	Equals one if job history includes in title one of the following: "technology" "software" "programmer" "it" "chief information officer" "database" "system administrator" "developer"
Bonus	Annual bonus payments (in thousands)
Busy	Equals one if directors sits on three or more boards
CEO	Equals one if director is the company's CEO
Chairman	Equals one if director is chairman of the board
Compensation chair	Equals one if director is chair of the compensation committee
Compensation committee	Equals one if director is a member of the compensation committee

Experience CEO	Equals one if director has experience as CEO of a publicly traded company
Experience CFO	Equals one if director has experience as CFO of a publicly traded company
Experience Chairman	Equals one if director has experience as Chairman of a publicly traded company
Experience exec VP	Equals one if director has experience as executive VP of a publicly traded company
Experience President	Equals one if director has experience as President of a publicly traded company
Foreign	Equals one if director's nationality is not American
GenBBB	Equals one if director was born between 1946 and 1964
GenDepBB	Equals one if director was born in or before 1926
Gender	Equals one if director is male
GenMature	Equals one if director was born between 1927 and 1945
GenX	Equals one if director was born between 1965 and 1980
GenY	Equals one if director was born in 1981 or after
Governance chair	Equals one if director is chair of the governance committee
Governance member	Equals one if director is a member of the governance committee
HistInternational	Equals one if job history includes a position outside the United States
Ivy league	Equals one if director went to an Ivy League college
Lead_independent	Equals one if director is lead independent director
MBA	Equals one if director holds an MBA degree
Mean past voting outcome	Average shareholder support during the first three years of tenure for previous board positions (starting in 2002 <i>Source: ISS Voting Analytics</i> )
Mean_support_3yrs	Average shareholder support over the first three years of tenure. <i>Source: ISS Voting Analytics</i>
Network size	Network size of director (number of overlaps through employment, other activities, and education)
Nomination chair	Equals one if director is chair of the nomination committee
Nomination member	Equals one if director is a member of the nomination committee
Number connections	Number of established connections to incumbent board members prior to joining the board
Number qualifications	Number of qualifications at undergraduate level and above
Nb prev jobs industry	Number of previous jobs in same FF48 industry
Time prev jobs industry	Time spent on jobs in same FF48 industry
Nb prev jobs different industry	Number of previous jobs in different FF48 industry
Time prev jobs different industry	Time spent on jobs in different FF48 industry
Other chair	Equals one if director is chair of a committee other than compensation, audit, governance or nomination
Other member	Equals one if director is a member of a committee other than compensation, audit, governance or nomination
Other compensation	Value of annual ad hoc cash payments such as relocation or fringe benefits awarded during last reporting period (in thousands)
Perf to total compensation	Performance to total - Ratio of Value of LTIPs Held to Total Compensation
Salary	Base annual pay in cash (in thousands)
Timeretirement	Time to retirement (assumed to be 70 years old)
Tot Current Nb Listed Boards sitting on	The number of Boards of publicly listed companies that an individual serves on

Tot Current Nb Other Boards sitting on	The number of Boards for organizations other than publicly listed or private companies that an individual serves on
Tot Current Nb Unlisted Boards sitting on	The number of Boards of private companies that an individual serves on
Tot Nb Listed Boards sat on	The number of Boards of publicly listed companies that an individual has served on
Tot Nb Other Boards sat on	The number of Boards for organizations other than publicly listed or private companies that an individual has served on
Tot Nb unlisted Boards sat on	The number of Boards of private companies that an individual has served on
Total Compensation	Salary + Bonus
Total director compensation	Salary plus Bonus plus Other Compensation plus Employers Defined Retirement/Pension Contribution
Total equity linked wealth	A valuation of total wealth at the end of the period for the individual based on the closing stock price of the last annual report
Value of shares held	Value of shares held at the end of the reporting period for the individual based on the closing stock price of the annual report

#### A.4.2. Board-level features

*Source: BoardEx except if stated otherwise*

*(as of when the director joins the board)*

<u>Variable</u>	<u>Definition</u>
Attrition rate	Number of Directors that have left a role as a Fraction of average number of Directors for the preceding reporting period
Average age	Average age of directors on the board
Average busy	Fraction of directors currently sitting on three or more boards
Average foreign	Fraction of directors with nationality other than American
Average independent	Fraction of non-executive directors on the board
Average Ivy League	Fraction of directors who went to an Ivy League college
Average MBA	Fraction of directors holding an MBA
Average nb qualifications	Average number of qualifications at undergraduate level and above of directors on the board
Average network size	Average network size of directors on the board (number of overlaps through employment, other activities, and education)
Average tenure	Average board tenure of directors on the board
Average time in company	Average time in company for executive and non-executive directors on the board
Avg tot current nb listed boards	The average number of boards of publicly listed companies directors currently serve on
Avg tot nb listed boards sat on	The average number of boards of publicly listed companies directors have served on
Avg Bkgd academic	Fraction of directors with an academic background (job history)
Avg Bkgd CEO	Fraction of directors with a CEO background (job history)
Avg Bkgd finance	Fraction of directors with a finance background (job history)
Avg Bkgd hr	Fraction of directors with a human resources background (job history)
Avg Bkgd law	Fraction of directors with a law background (job history)
Avg Bkgd manager	Fraction of directors with a manager background (job history)
Avg Bkgd marketing	Fraction of directors with a marketing background (job history)
Avg Bkgd military	Fraction of directors with a military background (job history)
Avg Bkgd politician	Fraction of directors with a political background (job history)
Avg Bkgd science	Fraction of directors with a scientific background (job history)

Avg Bkgd technology	Fraction of directors with a technology background (job history)
Avg Experience CEO	Fraction of directors with experience as CEO of a publicly traded company
Avg Experience CFO	Fraction of directors with experience as CFO of a publicly traded company
Avg Experience Chairman	Fraction of directors with experience as Chairman of a publicly traded company
Avg Experience exec VP	Fraction of directors with experience as executive VP of a publicly traded company
Avg Experience President	Fraction of directors with experience as President of a publicly traded company
Board Pay Slice - salary	Tot indep comp/ CEO salary
Board Pay Slice - total	Tot indep comp/ CEO total compensation
Board size	Number of directors on the board
BOSS	Equals one if the CEO is also the chairman of the board and the President
CEO bonus	CEO's bonus
CEO salary	CEO's salary
CEO total compensation	CEO total compensation (salary plus bonus)
Chairman duality	Equals one if the CEO is chairman of the board
Classified	Equals one if board is classified
Count Female	Number of women on the board
Fracdirafter	Coopted Directors as Fraction of Total Board (Data from Lalitha Naveen's website)
Fracdirafterindep	Coopted Independent Directors as Fraction of Total Board (Data from Lalitha Naveen's website)
Twfracdirafter	Tenure Weighted Coopted Directors as Fraction of Total Board (Data from Lalitha Naveen's website)
Twfracdirafterindep	Tenure-Weighted Coopted Independent Directors as Fraction (Data from Lalitha Naveen's website)
Gender ratio	The Fraction of male directors
Nationality Mix	Fraction of Directors from different countries
Nb independent	Number of independent directors
Nb international experience	Number of directors with international experience
Stdev age	Standard deviation of directors' age
Stdev current listed board	Standard deviation of the number of listed boards each director currently serves on
Stdev listed board sat on	Standard deviation of the number of quoted boards sat on for all directors on the board
Stdev number qualifications	Standard deviation of the number of qualifications at undergraduate level and above for all directors on the board
Stdev Time in Company	Standard deviation of time in the company for all directors on the board
Stdev Time on Board	Standard deviation of time on board for all directors on the board
Succession Factor	Measurement of the Clustering of Directors around retirement age
Tot indep comp	Sum of all independent directors' total compensation
Tot indep comp scaled	Sum of all independent directors' total compensation divided by the number of independent directors

### A.4.3 Firm level features

Source: Compustat /CRSP except if stated otherwise

(as of when the director joins the board)

<u>Variable</u>	<u>Definition</u>
Current assets	Current assets - Total
Acquisitions	Acquisitions
Auditor	Dichotomous variable for each auditing firm
BCW	Equals one if firm was on the Fortune-Best Company to work for list within 10 years preceding the nomination (from Alex Edmans' website)
Blank check	Equals one if firm has a blank check provision (from ISS RiskMetrics)
CAPX	Capital expenditures
CEOSO1	Equals to one if the CEO is exempt from filing Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002
CFOSO1	Equals to one if the CFO is exempt from filing Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002
CEOSO2	Equals to one if the CEO has not filed Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002
CFOSO2	Equals to one if the CFO has not filed Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002
CEOSO3	Equals to one if the CEO has filed Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002
CFOSO3	Equals to one if the CFO has filed Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002
Equity (ordinary)	Ordinary equity - Total
Cash	Cash
Cash and ST investments	cash and short term investments -
COGS	Cost of good sold -
Shares outstanding	Common shares outstanding -
Dividend payer	Dichotomous variable equal to 1 if the total amount of dividends to ordinary equity > 0
LT debt	Long term debt - Total - Source : Compustat
Depreciation	Depreciation and amortization -
Dividends	Total amount of dividends to ordinary equity
EBIT	Earnings Before Interest and Taxes
EBITDA	Earnings Before Interest
Finterms_negative	Loughran-McDonald Negative word proportion (from Wrds SEC Analytics Suite)
Finterms_positive	Loughran-McDonald Positive word proportion (from Wrds SEC Analytics Suite)
Finterms_litigious	Loughran-McDonald litigious word proportion (from Wrds SEC Analytics Suite)
Finterms_uncertainty	Loughran-McDonald uncertainty word proportion (from Wrds SEC Analytics Suite)
Firm age	Time since IPO or first occurrence on CRSP
Fsize	Size of annual report file (from Wrds SEC Analytics Suite)
Golden parachute	Equals one if firm has a golden parachute provision (from ISS RiskMetrics)

Gunnin_fox_index	Gunning Fog Readability Index (from Wrds SEC Analytics Suite)
Harvardiv_negative	Harvard General Inquirer negative word count (from Wrds SEC Analytics Suite)
Incorp	Dichotomous variable for state of incorporation
Inventories	Inventories - Total
Q_int	Peter & Taylor Total Q-Firm's intangible capital estimated replacement cost (from Wrds)
Q_int_know	Peter & Taylor Total Q-Firm's knowledge capital replacement cost (from Wrds)
Q_int_offbs	Peter & Taylor Total Q-Portion of K_int that doesn't appear on firm's balance sheet (from Wrds)
Q_int_org	Peter & Taylor Total Q-Firm's intangible capital estimated replacement cost (from Wrds)
Current liabilities	Current liabilities - Total
Leverage	Total long term debt / total assets
Ln(nb insti blocks)	Logarithm of one plus the number of institutional blockholders.
Ln(nb insti owners)	Logarithm of one plus the number of institutional investors.
Majority vote standard	Equals one if requires a director to receive support from a majority of the shares cast to be elected. (from ISS RiskMetrics)
MB	(common shares outstanding * stock price)/ ordinary equity
Minority interest	Minority interest
Mkt value	Market value
Net income	Net income
NumestYr	Average Annual Number of Analysts (From EPS estimates from IBES)
Plurality vote	Equals one if a director need only receive one vote to be elected. (from ISS RiskMetrics)
Price (calendar)	Price Close - Annual - Calendar
Price (fiscal)	Price Close - Annual - Fiscal
Product Mkt fluidity	Product market fluidity. Hoberg and Phillips
Profitability	ebitda/total assets
Q_tot	Peter & Taylor Total Q-Total q (from Wrds)
Block ownership %	Fraction owned by blockholders.
Institutional ownership %	Fraction owned by institutional investors.
Largest inst. shr. %	Fraction owned by largest institutional investor.
Largest 10 inst. shr. %	Fraction owned by top ten institutional investors.
Largest 5 shr. %	Fraction owned by top five institutional investors.
Retained earnings	Retained earnings
Retained earn. (restated)	Retained earnings restatements
12-month return	Cumulative stock return in the twelve months leading up to the appointment.
3-month return	Cumulative stock return in the three months leading up to the appointment.
6-month return	Cumulative stock return in the six months leading up to the appointment.
Revenue	Revenue - Total
RIX	RIX Readability index (from Wrds SEC Analytics Suite)

ROA	Net income / total assets
ROE	Net income / ordinary equity
Sales	Net sales - Total
Equity (total)	Stockholders' equity - Total
Settlements	Settlement (Litigation/Insurance) After-tax
slpctyr	Average Annual Short Interest as a % of Shares Outstanding
Total assets	total assets -
Working Capital	Working capital
Extraordinary items	extraordinary items
R&D	R&D expenses

#### A.4.4. Industry and market level features

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

<u>Variable</u>	<u>Definition</u>
Industry ROA	Return on assets of firms with same 3-digit SIC code
Market3	Cumulative returns on the S&P500 in the three months leading up to the appointment
Market6	Cumulative returns on the S&P500 in the six months leading up to the appointment
Market12	Cumulative returns on the S&P500 in the twelve months leading up to the appointment
ExcessReturns3	Cumulative stock return in the three months leading up to the appointment minus cumulative returns on the S&P500 in the three months leading up to the appointment
ExcessReturns6	Cumulative stock return in the six months leading up to the appointment minus cumulative returns on the S&P500 in the six months leading up to the appointment
ExcessReturns12	Cumulative stock return in the twelve months leading up to the appointment minus cumulative returns on the S&P500 in the twelve months leading up to the appointment
Tnic3*	3-digit, text-based industry classifications from Hoberg and Phillips (2010, 2016)