Predicting forced CEO turnover using machine learning

Juebin Zeng, Helen Lu and Henk Berkman

Abstract

This study predicts forced CEO turnover with machine learning. In out-of-sample tests, our machine learning model substantially outperforms traditional models across different performance metrics. Machine learning's predictions show evidence to support the strong-form RPE and reject the weak-form RPE in CEO dismissal process. Globally, performance related, incentive related, and risk-taking related features contribute the most in predicting forced CEO turnover. Locally, machine learning can deal with sophisticated interactions and nonlinearity among the features, especially detecting the skill-matching between CEOs and firms. Finally, this study reveals that CEO entrenchment leads to undervaluation of CEOs' poor performance explaining why some CEOs are rarely fired. In addition, this study suggests that directors may misattributes financial distress to CEOs who should not be fired, to preserve their reputation and positions.

1.0 Introduction

The topic of CEO dismissals has gained significant attention in financial academic research. Understanding the process and reasons for CEOs' removal is crucial, given the potential contribution of CEO skills to firm performance. Any inefficiencies in the removal process could negatively impact the success of the firm. Previous studies have extensively examined the mechanism and determinants of CEO dismissals, involving various dimensions such as firm characteristics, CEO attributes, corporate governance, and relative performance. The varied dimensions are not surprising, given the different underlying theories and variable selections. However, the limited number of variables and the frequent use of the logistic regression method[1] in prior literature may not fully capture the intricate and non-linear interactions among variables when predicting CEO dismissals. Consequently, the resulting outcomes are plausible and cannot distinguish which dimension dominates the board's consideration when removing CEOs.

Recent advances have demonstrated that machine learning algorithms are better alternatives to traditional models in predicting uncommon firm-level phenomena, such as accounting fraud (Bao et al., 2020), future earnings change (Chen et al., 2022) and financial misstatement (Bertomeu et al., 2021). In this paper, we present a machine learning approach for identifying the critical features (equivalent to *variables* in econometrics) associated with forced CEO turnover and predicting such turnover events. The specific machine learning approach used in this paper is Gradient Boost Machines (GBMs), especially LightGBM, which is a tree-based machine learning model and belongs to ensemble learning family[2].

There are two major objectives for this study. For the first objective, this paper wishes to develop a model that can accurately predict as many forced CEO turnovers as possible. However, given the rarity of forced CEO turnover events, the dataset is likely to be imbalanced. While aiming for a high number of correctly predicted forced CEO turnovers (true positives), it is expected that there will be a high number of misclassifications (false positives). Therefore, it is important to carefully consider the trade-off between the number of true positives and the number of false positives when dealing with imbalanced classification problems.

[1] Logistic regression is widely used in forced turnover determination. Hazard model and probit regression are also applied in previous studies.

[2] This will be briefly discussed in our research design section.

For the second objective, this paper aims to validate and reconcile the determinants of forced CEO turnover and the related theory (i.e., relative performance evaluation theory) . First, in the LightGBM model with a large number of input variables, decision trees are iteratively added to improve performance. The model evaluates the contribution of each feature by analysing the split points on that feature. If the split point does not improve model performance, it suggests that the feature has low importance in predicting forced CEO turnovers. This helps to identify the most relevant features[3], which can be used to confirm or refute the determinants of forced CEO turnover found in previous studies. Second, if the machine learning model is valid and outperforms traditional models, it is expected that the machine learning model's predictions are more reliable than traditional models' predictions. This enables an examination of previous relative performance evaluation evidence regarding CEO turnover, and help to either support or reject it.

This study employs the LightGBM model to predict the binary target variable of forced CEO turnover, which indicates whether a CEO was fired in a given fiscal year. To evaluate the performance of the ML model, two high-profile traditional models are selected as benchmarks. Additionally, regularized logistic regression (LASSO[4]) is also applied using all features as input to compare with the LightGBM model. The first traditional model is developed by Bushman, Dai and Wang (2010). They adopt probit regressions to mainly examine the effects of individual and industrial volatilities on forced CEO turnover. The second traditional model is from Jenter and Kanaan (2015). They employ logit regressions to examine the impact of relative performance on forced CEO turnover.

In out-of-sample results, this paper shows that the LightGBM model outperforms two traditional models as well as the LASSO model across three different performance metrics. Specifically, the LightGBM model achieves a significantly larger area under the Receiver Operating Characteristics curve (ROC AUC) and larger area under the Precision-Recall curve (PR AUC). The LightGBM model also outperforms the other models with the highest average Normalized Discounted Cumulative Gain at the top 3% of predictions (NDCG@3%), and this outperformance persists over years in test sample. Another way to evaluate the quality of our

[3] We use SHAP values (SHapley Additive exPlanations), which is introduced by (Lundberg & Lee, 2017), to measure the contribution of each feature.

[4] LASSO (Least Absolute Shrinkage and Selection Operator) is a simple machine learning model and a variant of Logistic regression model. It can regularize coefficients of variables by imposing penalties to shrink coefficients. This helps in improving model interpretability, reducing overfitting, and enhancing prediction accuracy, especially in situations where there are many potential features available.

machine learning model is to evaluate the way in which actual forced CEO turnover rate (the percentage of forced CEO turnover) increases related to predicted turnover probability. We divide the predicted turnover probabilities in deciles for all models. We find a similar trend across all models, whereby the actual turnover rate increases as the predicted probability decile increases. Our machine learning model outperforms the other models in predicting CEO turnover, with the lowest turnover rate of 1.33% in the bottom decile (decile 1) and the highest turnover rate of 16.96% in the top decile (decile 10).

Our machine learning model has demonstrated superior predictive power across multiple evaluation methods. Given its accurate predictions, it is intriguing to investigate how these predictions can contribute to a better validation of previous CEO turnover related hypotheses and theories. Prior studies have indicated that firm performance, especially CEO-induced performance (idiosyncratic return) and industry-induced performance (peer return), plays an important role in predicting CEO turnover. In out-of-sample results, the average turnover probability predicted by our ML model successfully captures a significant declining trend across idiosyncratic return deciles, with an average predicted probability of 18.05% in the bottom decile (decile 1) and 8.48% in the top decile (decile 10). The two traditional models also capture such trend but it is not as pronounced. For industry-induced performance, the traditional models exhibit a slight declining trend (insignificant) in capturing industry-induced performance. Furthermore, our machine learning model's predictions demonstrate the ineffectiveness of industry-induced performance in predicting forced CEO turnovers. The results challenge the weak-form relative performance evaluation theory raised by Jenter and Kanaan (2015), while providing support for the strong-form relative performance evaluation theory.

While traditional econometric modelling produces coefficient estimates for each feature, machine learning algorithms do not. However, SHAP values (<u>SH</u>apley <u>A</u>dditive ex<u>P</u>lanations) introduced by Lundberg and Lee (2017) can help to quantify the contribution of each feature to predicting forced CEO turnover. In terms of global contribution, we find that incentive-related features, market-based performance features, and risk-related features contribute most (in top 10 important features) to predicting forced CEO turnovers. Accounting-based performance features, governance-related features and CEO background features do not hold significant importance in global contribution. The insignificance of certain features in global contribution does not necessarily imply their lack of importance locally. We take Microsoft and Computer Task Group as two examples to demonstrate how machine learning algorithms determine CEO dismissals locally. We find that high (low) incentive payment, high (low) idiosyncratic return, low (high) risk-taking, and high (low) CEO power push down (up) the probability of CEO dismissal. Moreover, we find skill-matching between CEOs and firms reduce the probability of CEO dismissal. Specifically in the examples above, the CEO who works in computer software industry with a technology background is less likely to be fired.

Finally, this paper partially reconciles two questions in firing process: why CEOs who are predicted to be dismissed are still in position and why CEOs who are not predicted to be dismissed but are fired. We find firms that have CEOs who have higher ownership and who are also the chairman of the board are likely to retain CEOs who are predicted to be dismissed. This pattern is consistent with the view that poor corporate governance reduces the turnover-performance sensitivity. For CEOs who are not predicted to be fired but are fired, we find that those CEOs are less powerful and are in relatively large firms. However, those firms are facing financial distress. One potential explanation can be that the directors of poorly performing large companies may resort to firing CEOs to shield themselves from accountability and safeguard their positions and standing.

The applications of machine learning techniques to finance and accounting research topics are under-exploited. This paper complements to this strand of papers, including but not limited to accounting fraud and financial misstatement detection (Bao et al., 2020; Bertomeu et al., 2021), asset pricing (Geertsema & Lu, 2023; Gu et al., 2020), director selection (Erel et al., 2021) and future earnings change (Chen et al., 2022). This paper utilizes the novel machine learning approach to deliver a comprehensive picture about the determinants of forced CEO turnovers. The results contribute to the literature on relative performance evaluation in CEO turnover decisions (e.g., Bushman et al., 2010; Eisfeldt & Kuhnen, 2013; Jenter & Kanaan, 2015; Jenter & Lewellen, 2021; Parrino, 1997) and literature on factors affecting turnover-performance sensitivity (e.g., Bhagat et al., 2010; Dikolli et al., 2014; Goyal & Park, 2002; Jagannathan & Pritchard, 2017; Kaplan & Minton, 2006, 2006; Liu, 2014, 2014; Weisbach, 1988). The partial reconciliation on firing process by using machine learning predictions in this paper also contributes to the literature on corporate governance and CEO entrenchment (e.g., Ertugrul & Krishnan, 2011; Hazarika et al., 2012; Taylor, 2010)

2.0 Related literature

A significant portion of the extant literature has related firm performance[5] to forced CEO turnover decisions. The key assumption on this relation is that boards regularly revise their evaluation of a CEO's suitability in office based on performance signals that unfold over time. Gibbons and Murphy (1990) are the first to apply relative performance evaluation (RPE) theory in the context of CEO turnover decisions. They propose that boards efficiently incorporate performance information into the assessment of a CEO's ability and follow an optimal dismissal decision rule based on this assessment. Their findings suggest that the optimal turnover decisions are solely associated with the performance relative to a benchmark[6] (i.e., industry performance) that filters out the components of firm performance unrelated to managerial effort or ability. In other words, CEO dismissal decisions are solely influenced by the part of firm performance that reflects CEO efforts and ability (abnormal or idiosyncratic performance), rather than full industry performance (peer performance). This type of RPE is referred to as strong-form RPE. Similar results for other samples and periods that support strong-form RPE have been reported by Parrino (1997), Huson et al. (2001), and Hazarika et al. (2012).

Contrary to strong-form RPE, Jenter and Kanaan (2015) (hereafter, JK) argue that CEO turnover decisions actually follow a weak-form RPE approach. JK perform two-stage regressions to examine turnover-performance relation and decompose firm performance into two components in the first stage: idiosyncratic return (residual) and industry-induced return (expected return). According to JK's findings, CEO dismissals conform to weak-form RPE because poor industry performance (e.g., industry downturn) provides more informative insights into the CEO's ability. Specifically, CEOs with poor performance are more likely to be dismissed in bad times, as underperformers is more revealing about deficiencies. Similar results related to weak-form RPE have also been documented by Eisfeldt and Kuhnen (2013), Gopalan, Milbourn and Song (2010), and Kaplan and Minton (2006).

Similar findings regarding the negative relation between the likelihood of CEO dismissal and industry performance are also reported by Bushman, Dai and Wang (2010) (hereafter, BDW). Different from JK's viewpoint, BDW focus on the impact of market turbulence (e.g., noise,

[5] Most common performance measures in CEO turnover literature are based on stock returns. Some studies also consider using accounting-based performance measures (e.g., Engel et al., 2003; Huson et al., 2001; Weisbach, 1988)

[6] The common benchmark used to obtain abnormal performance is industry return. Some studies also consider marked-based benchmark, such as equal-/value-weighted market return or S&P 500 index return (e.g. Bushman et al., 2010; Kaplan & Minton, 2006)

economy-wide effects, etc.) on the ability of boards to identify underperforming CEOs. They incorporate additional two risk proxies to evaluate informativeness about boards learning CEOs' talent. Specifically, BDW demonstrate that firm performance becomes more diagnostic to replace low-quality CEOs when the risk deriving from the uncertainty about the CEO's talent (idiosyncratic risk) increases. On the other hand, boards become difficult to distinguish the talent of incumbents when the risk deriving from industry (peer risk) increases, leading to a decreased likelihood of CEO dismissal.

The extant turnover-performance literature has reached a consensus that the performance attributed to the CEO's ability and skills (idiosyncratic performance[7]) have robust impact on CEO turnover decisions[8]. However, the role of industry performance in CEO dismissal remains ambiguous. While BDW have replicated JK's findings about the negative impact of industry performance on the likelihood of CEO dismissal, they consider this relation as a "conundrum" and indicate that "…… their tests do not provide convincing support for any of the proposed explanations for the industry effect on CEO turnover". Furthermore, Fee et al. (2018) cast doubt on the relation between industry performance and CEO dismissal and find that the relation is unrobust for different timing conventions and modelling choices. Given the ambiguous evidence about the impact of industry performance on CEO dismissal (i.e., whether CEO dismissals follow strong-form RPE or weak-form RPE), this paper will construct machine learning model to predict forced CEO turnover and utilize the predictions from the model to examine the RPE theory. Additionally, the machine learning model will be benchmarked against the traditional models proposed by JK and BDW to demonstrate the accuracy and reliability of the predictions generated by the machine learning model.

In addition to RPE theory, previous studies have looked into various factors in CEO dismissal decisions. These studies include the literature on corporate governance and the ways of CEO entrenchment investigate the factors such as CEO-founder status (e.g., Beneish et al., 2017; Bushman et al., 2010), CEO ownership (e.g., Denis et al., 1997; Goyal & Park, 2002; Hazarika et al., 2012), CEO duality (e.g., Goyal & Park, 2002; Hazarika et al., 2012) and board structure (e.g., Coles et al., 2014; Guo & Masulis, 2015; Kaplan & Minton, 2006; Weisbach, 1988); the literature on investigating the impact of CEO and firm characteristics on CEO turnover such as

[7] Early studies refer to industry-/market-adjusted performance as the performance attributed to the CEO's ability and skills.
[8] See Fee et al. (2018) regarding the discussion about robust models of CEO turnover.

CEO age and tenue (e.g., Allgood & Farrell, 2000; Dikolli et al., 2014; Goyal & Park, 2002; Murphy & Zimmerman, 1993), CEO education background (e.g., Bhagat et al., 2010), CEO compensation (e.g., Campbell et al., 2011; Gao et al., 2012; Jochem et al., 2018; Laux, 2012), firm competition within industry (e.g. DeFond & Park, 1999; Goyal & Park, 2002), firm location (e.g., Jagannathan & Pritchard, 2017), and accounting performance (e.g., Engel et al., 2003; Farrell & Whidbee, 2003; Huson et al., 2001).

Previous studies primarily concentrated on investigating and explaining CEO dismissals within sample under some specific assumptions and often emphasising on causal inferencing – explanatory modelling. Explanatory modelling aims to minimise the bias resulting from model misspecification to obtain the most accurate representation of the underlying theory. In contrast, predictive modelling, such as machine learning algorithms, aims to minimize out-of-sample prediction error, which encompasses the combined effect of bias and estimation variance resulting from using a sample to estimate model parameters. Given the underlying assumptions and the limited number of covariates in explanatory modelling, its predictive power might be limited, leading to predictions that lack practical insights[9]. This paper approaches CEO dismissals as a prediction problem, wherein boards must review CEOs' suitability and make decisions to fire or not. Considering the potentially complex non-linearities and interactions among covariates in CEO dismissal decisions, a data-driven approach like machine learning algorithms is better suited for making accurate predictions. Therefore, in contrast to extant literature, the interest of this paper is to examine the relative importance of features associated with CEO dismissals and utilize the predictions produced by machine learning to provide practical insights in CEO dismissal process.

3.0 Research design

In this section, we divide our research design into two parts. The first part introduces the main machine learning model used in this study and discuss how we deploy our machine learning model. The second part describes the construction of our sample.

[9] For example, within the framework of efficient markets theory in finance, the inability to make accurate predictions about future outcomes is considered a fundamental premise. However, Moritz and Zimmermann (2016) employ machine learning to demonstrate that the historical returns of US firms possess substantial predictive capabilities concerning their future stock prices.

3.1 Machine learning model

The primary machine learning algorithm utilized in this paper is the Gradient Boosting Machine (GBM), which falls under the ensemble learning framework. Ensemble learning is a technique in machine learning that combines multiple individual models, known as base estimators, to enhance prediction accuracy. Boosting is a widely adopted approach within ensemble learning. In boosting, the base estimators are trained sequentially, with each subsequent estimator focusing on the samples that were misclassified by the previous estimator, thereby improving the accuracy. Previous studies (e.g., Chen & Guestrin, 2016; Zhou, 2021) have also shown that ensembles usually outperform any single base estimator. LightGBM is one of the popular implementations utilizing boosting approach, with the properties of high speed and high accuracy. In this paper, the LightGBM model is employed to directly predict forced CEO turnover.

Given the infrequent nature of forced CEO turnover events in each year, it is important to highlight the potential challenge of class imbalance in prediction tasks. The class imbalance may lead to machine learning models misclassifying a large amount of non-turnover instances as turnovers. Previous studies (e.g. Bao et al., 2020; Liu & Zhou, 2013) have proposed several methods to address this issue. These methods include resampling the dataset using sampling algorithms such as Synthetic Minority Oversampling Technique (SMOTE) and utilizing ensemble learning algorithms with variations like RUSBoost or AdaBoost. These methods essentially oversample or undersample the dataset to artificially balance the classes. However, these methods have a common premise that the input dataset must be complete without any missing values.

In corporate finance research, where missing values are common, the implementation of these methods may be challenging. One plausible way is to drop observations with missing values, then implement resampling algorithms. However, dropping observations with missing values (list-wise deletion) means a substantial loss of data availability (losing about 1/3 of our sample size), leading to the loss of valuable information input and causing bias[10]. Another plausible way is to implement resampling algorithms after imputing missing values. However, simple imputation, such as zero, mean or median, may produce bias or unrealistic results on a high-

[10] See Emmanuel et al. (2021) about a survey on missing data in machine learning. They indicate that if the sample size is not large or the missing values do not satisfy Missing Completely at Random (MCAR) assumption, then list-wise deletion is not appropriate. There are three types of missingness. To understand the types of missingness, please refer to Emmanuel et al. (2021) and the course notes from the University of Michigan by Dr. Josh Errickson about multiple imputation (https://dept.stat.lsa.umich.edu/~jerrick/courses/stat701/notes/mi.html#types-of-missing-data).

dimensional dataset (Emmanuel et al., 2021). Moreover, the advanced imputation techniques, such as multiple imputation, strongly rely on the assumption of Missing at Random (MAR) that the missingness is only related to the observable data. Considering some financial variables are missing due to Missing not at Random (MNAR)[11], the resulting imputed dataset may not be representative of the true population. Consequently, performing resampling on an unrepresentative dataset can lead to unexpected bias in the analysis.

This study intends not to drop any observations with missing values, as GBMs can handle missing values by ignoring them at a split point and then assigning them to a child node that reduces the loss. To address the class imbalance issue, this study plans to adjust hyperparameters of LightGBM, specifically the *scale_pos_weight* parameter, to assign a higher weight to the positive class (forced turnover events) during the training process. This weighting scheme helps the model to focus more on accurately predicting the minority class. Additionally, this study intends to control the updates made during the training by limiting the step size through the *max_delta_step*. This helps to prevent large and unstable changes and enhance convergence to predict more forced turnover events.

Following the conventional approach used in prior studies (e.g., Bao et al., 2020; Erel et al., 2021), this study employs a time-series based split for the sample, dividing it into three sets: training, validation, and test sets. The time periods for each set are as follows: the training set covers the years 1992-2008, the validation set covers 2009-2013, and the test set covers 2014-2019. The proportions of the sample size allocated to each set are approximately 60% for the training set, 20% for the validation set, and 20% for the test set. The LightGBM model is initially trained on the training set using the specified hyperparameters. Subsequently, the model's performance is evaluated on the validation set to identify the optimal hyperparameters, ensuring the best performance for the given task. Finally, the trained model is implemented on the test set to assess its out-of-sample performance[12]. We select two traditional models proposed by JK and

[11] For example, Koh and Reeb (2015) indicate firms choose not to disclose R&D expenditures for strategic considerations so that they do not separate R&D expenses from other reported expenses, such as expenses shifting. The blank fields in some financial variables may not be due to random missingness, as they are actually not reported at discretion (MNAR). Therefore, imputation techniques cannot predict the missing values based on observable data.

[12] The number of estimators is determined by early stopping on validation set. A list of learning rate (*learning_rate*) including 0.01, 0.03, 0.05, 0.1, 0.3, a list of maximum number of leaves (*num_leaves*) including 29, 30, 31, 32, a list of class weight (*scale_pos_weight*) including 8, 10, 12, and the limit of step size (*max_delta_step*) of 10 are input in the training process. The model performance for each combination of hyperparameters is evaluated on validation set to determine the optimal combination. The default settings of LightGBM are used for all other hyperparameters.

BDW as well as another simple machine learning model (LASSO) as benchmarks. To make results comparable across models, the benchmark models are evaluated using the same data splits to generate out-of-sample predictions[13]. All reported results in this study are based on the evaluation of the model on the test set.

3.2 Sample selection

This study focuses on predicting forced CEO turnovers using machine learning. The widely used CEO turnover classification strategy is based on the algorithm by Parrino (1997). However, the extant studies (e.g., Fee et al., 2018; Jenter & Lewellen, 2021; Taylor, 2010) raise concerns about the effectiveness of this classification strategy. Parrino's strategy heavily rely on age cut-off, that is, categorizing the departures of CEOs above the age of 60 as voluntary CEO turnovers. However, it is possible for CEOs above the age of 60 to also be forced out due to factors such as lack of innovation or declining (or stagnant) performance, while firms possibly take "retirement" as an excuse for their CEOs' departures to avoid negative impact on stock prices. This study, instead, uses an open-source CEO turnover dataset with available CEO classifications by Gentry, Harrison, Quigley and Boivie (2021). They classify all CEO turnovers from Execusomp based solely on SEC filings and media coverage without any age cut-off. It is less subjective and a more natural way to classify CEO turnovers, and less likely to cause any factitious results in later analysis. Appendix A shows classifications and definitions for CEO departures[14].

CEO and board information in this study is primarily obtained from two sources: Execucomp and BoardEx. The compensation-related data of CEOs, such as total compensation, salary, and bonuses, is sourced from Execucomp. CEOs' attributes (e.g., age, gender, and nationality), CEO work history and education background, board size, and board independence are sourced from BoardEx. Firm fundamentals are sourced from Compustat. Appendix B provides a detailed description of the variables used in this study. A variable name with a suffix of "_r" means this

[13] LightGBM can handle missing values, but traditional models and LASSO cannot. To make results more comparable, we use **IterativeImputer()** in **ScikitLearn** to impute the dataset so that the number of observations trained, validated and tested are the same across all models. Although we have discussed that imputation may not be suitable for financial data, in unpresented results, there is no difference between list-wise deletion and imputation in terms of model performance. Therefore, the decision to use imputation allows for a fair and comparable evaluation of all models, even though it may not be the optimal approach for financial data.

[14] Since this paper studies forced CEO turnovers, we focus on classifications of code 3 (Involuntary – CEO dismissed for job performance) and code 4 (Involuntary - CEO dismissed for legal violations or concerns) and assign a dummy value of 1 to them. For non-turnovers, we assign a dummy value of 0 to them. Following BDW and JK, we exclude the firm-years of the rest types of CEO turnovers. In addition, we carefully and manually deal with code 7, 8, and 9 by either deleting them or assigning 0 to them.

variable has been pre-processed with percentile-ranking[15]. A variable name with a suffix of "_l1" means this variable has been lagged by 1 year.

Following Peters and Wagner (2014), we assign a value of one to the Forced dummy in the last fiscal year in which a dismissed CEO is in office for the greater part of the fiscal year. We use this timing convention, because, in the case of a transition within a fiscal year, ExecuComp records the compensation of the CEO who was in office for the greater part of that fiscal year (Peters and Wagner, 2014). Therefore, we can attribute the compensation data to a CEO in the same year. After carefully merging the data from different sources, the final sample of this study consists of 40,185 firm-year observations, spanning from 1992 to 2019[16]. The distribution of CEO turnovers is shown in Table 1.

Table 1: CEO turnover distribution

Year	Forced turnover (N)	Voluntary turnover (N)	Number of firms	% Forced
1992	3	16	855	0.35%
1993	10	73	986	1.01%
1994	13	101	1057	1.23%
1995	13	98	1149	1.13%
1996	22	148	1276	1.72%
1997	20	143	1347	1.48%
1998	23	184	1431	1.61%
1999	38	207	1415	2.69%
2000	55	185	1412	3.90%
2001	36	118	1456	2.47%
2002	53	130	1501	3.53%
2003	36	162	1540	2.34%
2004	50	206	1515	3.30%
2005	49	153	1436	3.41%
2006	44	217	1543	2.85%
2007	60	189	1756	3.42%
2008	60	152	1721	3.49%
2009	51	134	1702	3.00%
2010	42	176	1671	2.51%
2011	40	189	1635	2.45%
2012	44	160	1609	2.73%
2013	41	167	1594	2.57%
2014	50	192	1578	3.17%

[15] For variables such as firm size, sales and compensation-related variables, they are naturally growing even if the inflation is removed (e.g., the firm size of ten years ago is much smaller and totally different from the firm size now). Therefore, there exists a problem that the decision rules derived from training set may not apply to validation and test sets as we split the sample in time-series way. To alleviate this problem, we percentile-rank these variables year-by-year so that the values of these variables are shrunk to a range of 0 to 1, representing the relative position of a specific feature in the market for a specific year.

[16] We only drop observations that cannot be matched in all three sources of Compustat, Execucomp and BoardEx. We do not winsorize our sample as outliers is not a concern for a tree-based machine learning model. The target variable is a dummy variable (0 and 1) which does not include any outliers. In addition, the input variables are split based on a decision rule with a threshold, which can handle outliers robustly.

2015	50	205	1522	3.29%
2016	50	195	1458	3.43%
2017	42	181	1397	3.01%
2018	59	171	1340	4.40%
2019	73	173	1283	5.69%
Total	1127	4425	40185	2.80%

Since this study does not apply widely used approach to encode the types of CEO turnovers, this study replicates the models proposed by BDW and JK to show the robustness of using Gentry et al. (2021)'s turnover dataset. Table 2 provides summary statistics for our sample based on BDW's model[17]. Relative to voluntary turnover sample, on average, forced turnover sample has lower idiosyncratic return (ret_idio_l1), lower ROA (roa_ind_adj_l1), lower CEO tenure (tenure), higher competition (competition_l1) and higher risk (risk_roa_l1). These findings are consistent with Bushman et al. (2010). There are also some differences. For example, in Bushman et al. (2010), the average of firm age for forced turnover sample is significant lower than voluntary turnover sample. However, in our sample, we do not find a significant difference in firm age. The might be due to using different CEO turnover classification scheme to classify CEO turnover.

Table 2: Summary statistics - Bushman et al. (2010)

	Voluntary turnover (N=3895)			Forced turnover (N=1008)			Control sample (N=30222)		
VarName	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
ret_idio_l1	-0.03	0.35	-0.04	-0.16	0.41	-0.15	-0.01	0.37	-0.02
ret_peer_l1	0.21	0.26	0.19	0.18	0.28	0.16	0.20	0.27	0.20
risk_idio_l1	0.34	0.19	0.30	0.39	0.20	0.35	0.35	0.20	0.30
risk_peer_l1	0.17	0.08	0.15	0.18	0.08	0.15	0.17	0.09	0.15
roa_ind_adj_l1	0.02	0.14	0.01	-0.01	0.17	0.00	0.02	0.14	0.01
risk_roa_l1	2.48	5.34	1.12	3.17	4.78	1.67	2.29	4.68	1.09
firmsize_l1	7.65	1.75	7.53	7.73	1.92	7.51	7.62	1.79	7.51
ceoage	60.04	7.54	60.80	56.24	7.22	56.00	56.57	7.27	56.74
tenure	9.84	7.95	7.84	6.66	6.05	5.00	7.91	7.49	5.67
ceofounder	0.12	0.32	0.00	0.07	0.26	0.00	0.11	0.32	0.00
competition_l1	251.76	275.85	172.00	283.25	355.17	178.00	254.01	293.05	171.00
firmage	27.86	16.91	24.00	27.11	17.22	22.00	27.85	16.92	24.00

Table 3 replicates the main regression results of BDW (Table 2 in their paper) and JK (Table 3 in their paper) using our sample. The findings in BDW's model indicate a negative relation

[17] Since JK and BDW use similar variables, we only present the summary statistics for variables used in BDW's model. The difference in sample size between our full sample and the sample in Table 2 is due to the exclusion of observations with missing values.

between idiosyncratic return, peer return, and peer risk with forced CEO turnover, while idiosyncratic return shows a positive relationship with forced CEO turnover, consistent with BDW's main findings. In JK's model, both idiosyncratic return and peer return exhibit a negative relationship with forced CEO turnover, aligning with the weak-form RPE theory proposed by JK. Furthermore, the significance levels of the key independent variables and control variables in Table 3 are similar to those reported in the original studies. By successfully replicating the results from two high-profile studies on forced CEO turnover, this study has demonstrated the robustness of the sample and the validity of using a different turnover classification scheme.

Table 3: Replicated regression results

This table replicates the main results of BDW (Table 2 in their paper) and JK (Table 3 in their paper). Dependent variable is a dummy indicator that equals 1 if a CEO is forced out, otherwise 0. The idiosyncratic return (*ret_idio_l1*), peer return (*ret_peer_l1*), idiosyncratic risk (*risk_idio_l1*) and peer risk (*risk_peer_l1*) are constructed based on BDW's method. We employ BDW's idiosyncratic returns and peer returns in JK's model, as JK do not clearly specify how they construct their RPE variables. The definitions of the remaining variables are shown in Appendix B.

	(1) Bushman et al. (2010)	(2) Jenter and Kanaan (2014)
ret_idio_l1	-0.456***	-1.184***
	(-11.28)	(-13.01)
ret_peer_l1	-0.228***	-0.617***
	(-4.01)	(-5.11)
risk_idio_l1	0.488***	
	(7.22)	
risk_peer_l1	-0.307*	
	(-1.85)	
roa_ind_adj_l1	-0.065	
	(-0.83)	
risk_roa_l1	0.008***	
	(3.65)	
firmsize_l1	0.034***	
	(3.66)	
ceoage	0.003	
	(1.27)	
tenure	-0.009***	-0.021***
	(-3.34)	(-4.28)
ceofounder	-0.140**	
	(-2.43)	
competition_l1	0.000*	
	(1.77)	
firmage	-0.002	
	(-1.58)	
ret_idio_l2		-0.297***
		(-3.04)

ret_peer_l2		-0.266**
		(-2.14)
retire_age		0.138
		(1.14)
high_ownership		-0.240
		(-1.27)
N	31230	31628
N_forced	1008	1009
N_control	30222	30619
Model	Probit	Logit

4.0 Results

In this section, we present machine learning results on forced CEO turnover. First, we examine the model performance based on different performance metrics, benchmarked against two traditional models (from BDW and JK) and LASSO model (another machine learning model). Then, we utilize the machine generated predictions to examine previous CEO turnover hypotheses and theories, with a particular focus on RPE theory. Furthermore, this study quantifies global and local contribution of variables in predicting forced CEO turnovers. Finally, this study interprets the difference between machine predicted turnovers and actual turnovers through two questions: Why do CEOs often remain in their positions despite being predicted to be fired, and conversely, why are CEOs sometimes fired even when the predictions suggest they should not be.

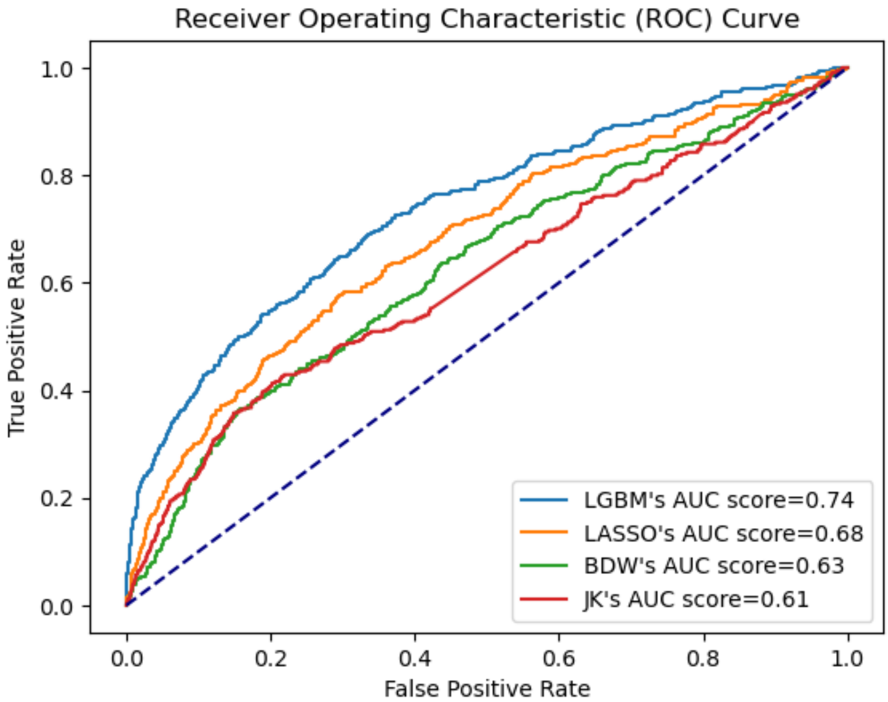
4.1 Model performance

This study adopts four different performance metrics to evaluate model performance on test set. The first performance metric is widely used in binary classification problems – Receiver Operating Characteristic Area Under the Curve (ROC AUC). A ROC curve is a line representing the model's diagnostic performance in terms of true positive rate and false positive rate at different prediction thresholds. AUC score is a metric to measure the area under ROC curve, representing the overall performance of a model. A larger AUC score, ranging from 0 to 1, indicates better predictive performance of the model. Figure 1.A shows that the overall performance of LightGBM outperforms traditional models with the largest ROC AUC score of 0.74. The simple machine learning model, LASSO, also performs better than the two traditional models with a ROC AUC score of 0.68. JK's model performs the worst with a ROC AUC score of 0.61, but still better than the random guessing (no-skill) score of 0.50. However, a higher ROC

AUC score may overestimate the predictive power or discrimination power for imbalanced data, as ROC AUC score mainly evaluates the overall performance of a model[18].

Figure 1.A: Receiver Operating Characteristic (ROC) curves

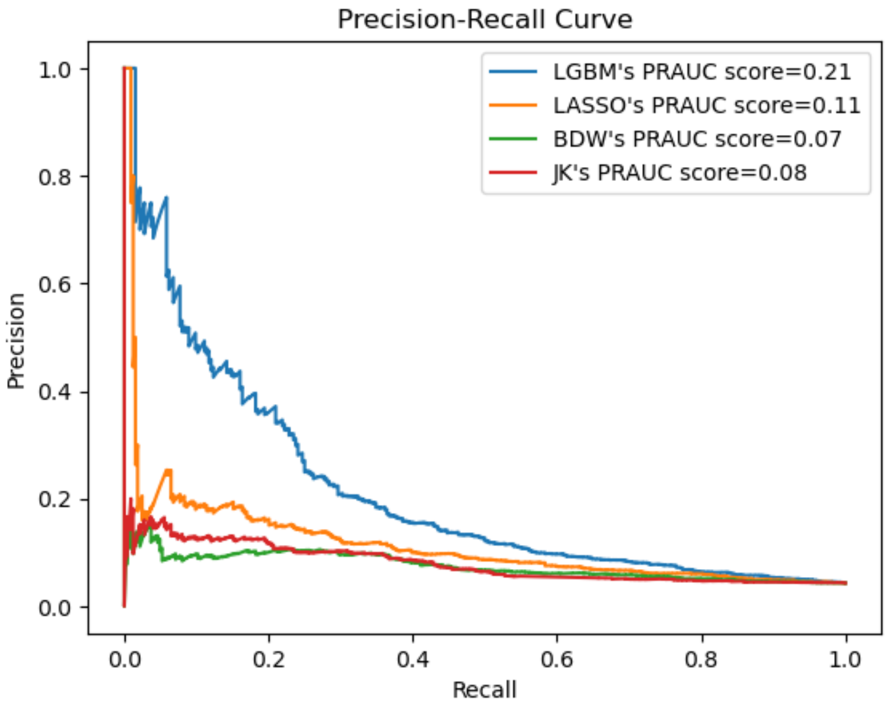
This figure shows ROC AUC curves for different models and their respective AUC scores. False positive rate is the ratio of false positives (the number of no-turnovers that are falsely predicted as turnovers) to true negatives (the number of no-turnovers that are correctly predicted) plus false positives. True positive rate is the ratio of true positives (the number of forced turnovers that are correctly predicted) to true positives plus false negatives (the number of forced turnovers that are falsely predicted as non-turnovers). ROC AUC scores represent the areas under ROC curves, ranging from 0 to 1. A higher ROC AUC score means a better model performance.



[18] Appendix C presents the confusion matrices for the four models employed in this study. Although the two traditional models exhibit better performance than random guessing based on ROC AUC scores, they fail to predict any true positives, indicating limited predictive capability. Furthermore, while the LASSO model predicts more true positives compared to the LightGBM model and outperforms the traditional models, this increased number of true positives comes at a high expense of misclassifying non-turnovers as turnovers.

Figure 1.B Precision-Recall curves

This figure shows PR AUC curves for different models and their respective AUC scores. Recall, also known as true positive rate, is the ratio of true positives (the number of forced turnovers that are correctly predicted) to true positives plus false negatives (the number of forced turnovers that are falsely predicted as non-turnovers). Precision is the ratio of false positives (the number of no-turnovers that are falsely predicted as turnovers) to false positive plus true positives. PR AUC scores represent the areas under ROC curves, ranging from 0 to 1. A higher PR AUC score means a better discrimination and predictive power on minority class (forced CEO turnover).



Since the aim of using machine learning in this study is to predict as many forced CEO turnover firm-years as possible with misclassifying fewer non-CEO turnover firm-years, we utilize another performance metric to evaluate the trade-off between recall (true positive rate) and precision (true positives /(true positives + false positives)) – Precision-Recall Curve Area Under the Curve (PR AUC). Ideally, the study would like to have high recall and precision (successfully predicting positives with fewer false negatives and fewer false positives). However, this is extremely hard for imbalanced sample. PR curves focus on the model performance on predicting minority class (here, the forced CEO turnover class) by looking into the sensitivity between recall and precision at different prediction thresholds. As Figure 1.B shows, a high recall is accompanied by a low precision for traditional models as well as LASSO model, indicating that a high recall (predicting more true positives) brings a large number of false positives. However, ML model has a significant larger area under its PR curve (0.21) relative to the other models (0.11 for LASSO, 0.07 for BDW and 0.08 for JK), showing that the trade-off between recall and precision for LightGBM model is less sensitive than for other models. The evidence suggests that LightGBM model has higher discrimination and predictive power than traditional models as well as LASSO.

Following the approach of Bao et al. (2020), this study utilizes Normalized Discounted Cumulative Gain at the top k positions (NDCG@k) as the third performance metric to assess out-of-sample performance. NDCG@k is commonly employed to evaluate the quality or ranking performance of search engine algorithms or recommendation systems up to a specified cut-off point (k). Specifically, in this study, the task of predicting CEO turnovers can be viewed as a ranking problem, wherein the aim is to determine whether a CEO dismissal decision should be recommended or not. The evaluation of out-of-sample performance can be restricted to a limited number of observations with the highest predicted probability of forced CEO turnover, as the focus of this study lies on the predictive power concerning the minority class. Therefore, NDCG@k evaluates how well the dismissal recommendations provided by a model align with the actual turnovers[19].

We choose a cut-off of 3%, which represents the top 3% of out-of-sample firm-years with the highest predicted probability of forced CEO turnovers, to calculate the NDCG@3% scores. The reason for selecting 3% is that it aligns with the approximate percentage of forced turnovers observed annually. Following Bao et al. (2020), this study calculates NDCG@3% score annually during the test period. These individual NDCG@3% scores are then averaged to obtain an overall NDCG@3% score, which is used to evaluate the performance of a model. Panel A of Table 4 indicates that the LightGBM model achieves the highest NDCG@3% score throughout the test sample period, suggesting that the LightGBM model consistently outperforms the other three models in every year of the out-of-sample period. Panel B of Table 4 provide a summary of the performance for different models used in this study. The average NDCG@3% score for each model is reported. Panel B shows the evidence that LightGBM model completely outperforms the other models across different performance metrics in predicting forced CEO turnovers, with the highest ROC AUC score, highest PR AUC score, and highest average NDCG@3% score[20].

[19] Please refer to Bao et al., (2020)'s paper for a detailed explanation of NDCG@k.

[20] Appendix D provides the average NDCG scores under different cut-offs. LightGBM model still perform the best among the four models.

Table 4: NDCG@3% scores

Panel A of this table shows the NDCG@3% scores across different years for different models. NDCG score ranges from 0 to 1. A higher NDCG score represent a better model's ranking performance, i.e., a greater alignment between the predicted dismissals and the actual dismissals. Panel B of this table summarize the performance of different models based on ROC AUC, PR AUC, and average NDCG@3%.

Panel A: NDCG@3% scores across different test years				
Year	LightGBM	LASSO	BDW	JK
2014	0.44	0.23	0.06	0.07
2015	0.43	0.22	0.12	0.10
2016	0.35	0.23	0.10	0.14
2017	0.20	0.00	0.02	0.07
2018	0.45	0.30	0.15	0.18
2019	0.53	0.27	0.15	0.16

Panel B: A summary of performance metrics			
Model	ROC AUC	PR AUC	NDCG@3%
LightGBM	0.74	0.21	0.40
LASSO	0.68	0.11	0.21
BDW	0.63	0.07	0.10
JK	0.61	0.08	0.12

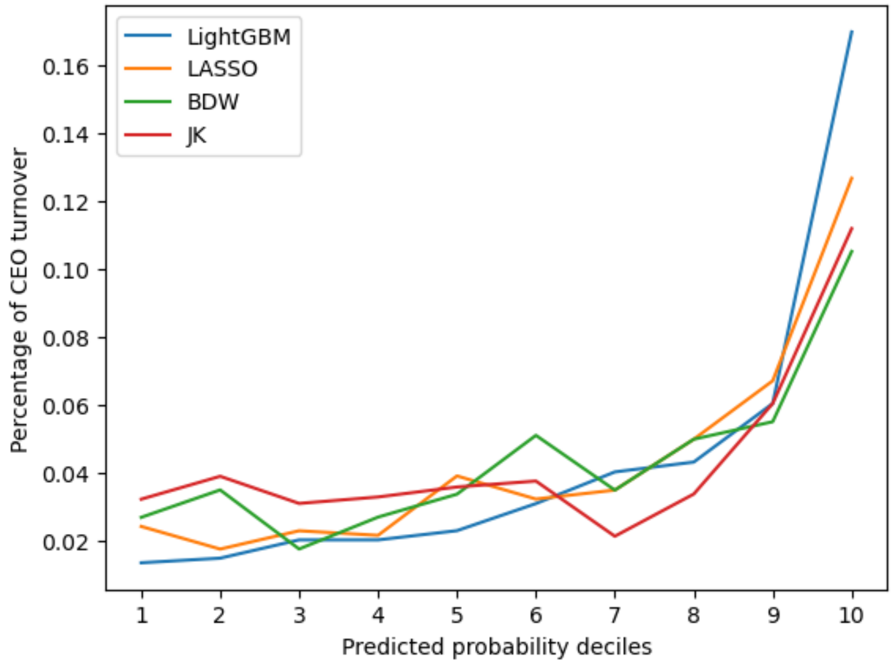
So far, this study has evaluated the out-of-sample performance for different models based on commonly used performance metrics in machine learning. However, another important aspect to consider is the relation between actual forced turnover rate and predicted forced CEO turnover probability. This can serve as an additional "performance metric" to evaluate the effectiveness of a model. Specifically, if a model can successfully classify forced CEO turnovers, it is expected that the forced turnover rate is extremely low (high) in the bottom (top) predicted probability decile. Deviation from this pattern would indicate an erroneous model, regardless of its performance in different performance metrics.

Figure 2 illustrates the relation between predicted probability deciles and actual turnover rates for all models. It shows a pattern that the actual turnover rate increases as the predicted probability decile increases. However, JK's model performs poorly compared to other models, particularly in the lower deciles. The forced CEO turnover rate does not consistently rise as the predicted probability decile progresses from decile 1 to decile 7.

On the other hand, LightGBM model consistently exhibits a rise in forced CEO turnover rates from decile 1 to decile 7, followed by a significant jump from decile 8 to decile 10. It also outperforms other models, with the lowest turnover rate of 1.33% in the first decile and the highest turnover rate of 16.96% in the tenth decile. These findings provide additional evidence supporting the validity and predictive power of LightGBM model in predicting forced CEO turnovers.

Figure 2: Forced CEO turnover rate versus predicted forced turnover probability

This figure shows the forced CEO turnover rate across the 10 deciles of predicted forced turnover probability for LightGBM, LASSO and traditional models in the test period. The predicted probabilities are divided into 10 deciles every test year, ranging from low (decile 1) to high (decile 10). Within each decile, the percentage of forced CEO turnovers is calculated.



4.2 RPE theory examination

This subsection aims to investigate the widely used RPE theory in the context of determining forced CEO turnover. Previous studies have identified two primary variants of the RPE theory: strong-form RPE and weak-form RPE. The main distinction between these two forms lies in the consideration of industry-induced performance's impact on the probability of forced CEO turnover.

Within the strong-form RPE assumption, CEO dismissals are exclusively determined by the CEO's individual performance, as measured by idiosyncratic return. Conversely, the weak-form RPE assumption acknowledges that CEO dismissals are contingent upon both the CEO-induced performance and industry-induced performance (peer return) (JK, 2015). Moreover, the effect of CEO-induced performance on CEO dismissals is influenced by the concurrent industry-induced performance.

Taking the advantage of the enhanced predictive power of LightGBM model, this subsection employs these more reliable and accurate predictions to conduct a two-fold investigation. Firstly, it examines the applicability of the RPE theory. Then, it identifies the specific form of the RPE theory. This study divides idiosyncratic returns into 10 deciles in each year of the test period. If the RPE theory is valid, it is expected that the forced CEO turnover rate and predicted forced CEO probability significantly decreases from the bottom decile to the top decile.

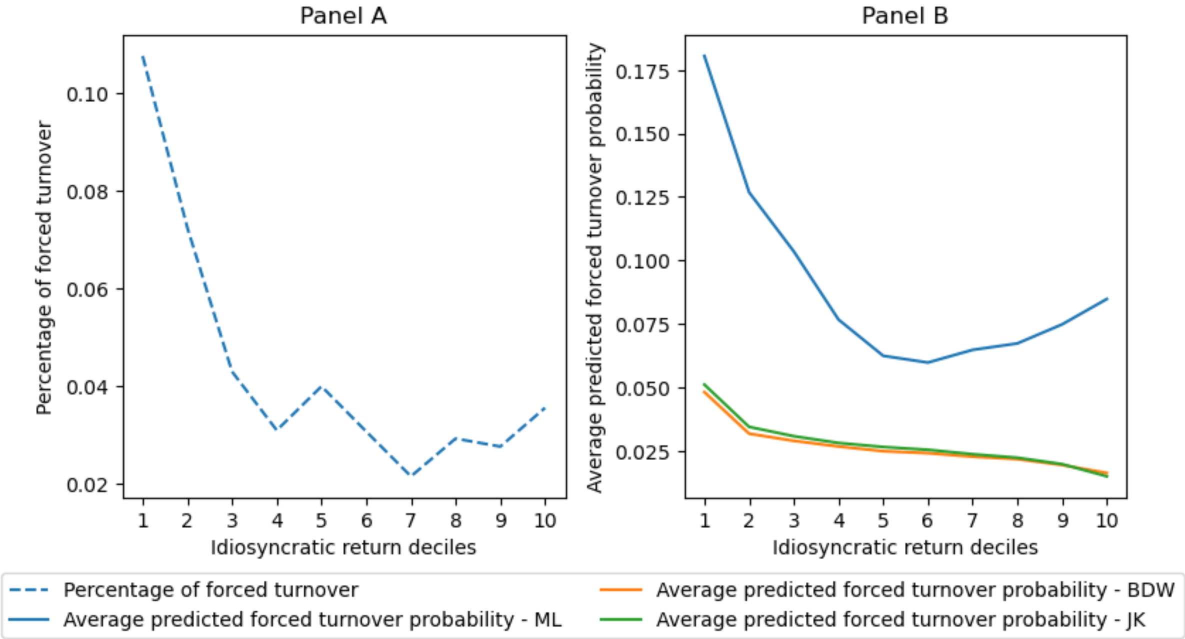
This study will exclusively present the outcomes of LightGBM model. The results of the LASSO model will not be included, given the established superiority of LightGBM observed in previous analyses. Therefore, the results of LightGBM will serve as representative of the machine learning (ML) outcomes. For brevity, LightGBM model will be referred to as ML model in the later analyses.

Since machine learning model does not provide coefficient estimates, this study plots forced CEO turnover rate and predicted forced CEO turnover probabilities across performance deciles to show turnover-performance relation. Panel A of Figure 3 illustrates an overall decline in the forced CEO turnover rate across idiosyncratic return deciles. Specifically, within the test set, there is a strong downward trend in the percentage of forced CEO turnover as CEO-induced performance increases from low (decile 1) to high (decile 6). In addition, a slight increase in the percentage of forced turnover is observed from decile 7 onwards. Given the significant negative relation between forced CEO turnover rate and CEO-induced performance, it is expected that a high (low) percentage of CEO turnover would correspond to a high (low) average predicted forced turnover probability.

In Panel B, LightGBM model's average predicted forced turnover probability mirrors the trend presented in Panel A. Both traditional models exhibit a similar trend, where the average predicted probability decreases as CEO-induced performance increases. However, this trend is not as pronounced (declining from around 5% in decile 1 to around 2.3% in decile 10) compared to the over 7% decrease in the percentage of CEO turnover across performance deciles. Moreover, neither traditional model captures the slight increase in the forced turnover rate observed from decile 7. In summary, the predictions of ML model corroborate the RPE theory, indicating that outperforming CEOs are less likely to face dismissal. Furthermore, the findings also present additional evidence supporting the superior predictive capability of ML model in predicting forced CEO turnover.

Figure 3: CEO-induced performance and forced CEO turnover

This figure shows forced CEO turnover rate and average predicted forced turnover probability across 10 deciles of idiosyncratic return (*ret_idio_l1*). Idiosyncratic returns are divided into 10 deciles in each year of the test period, ranging from low (decile 1) to high (decile 10). Within each decile, the percentage of forced CEO turnovers is calculated in Panel A, and the average of predicted forced CEO turnover probability is calculated in Panel B.

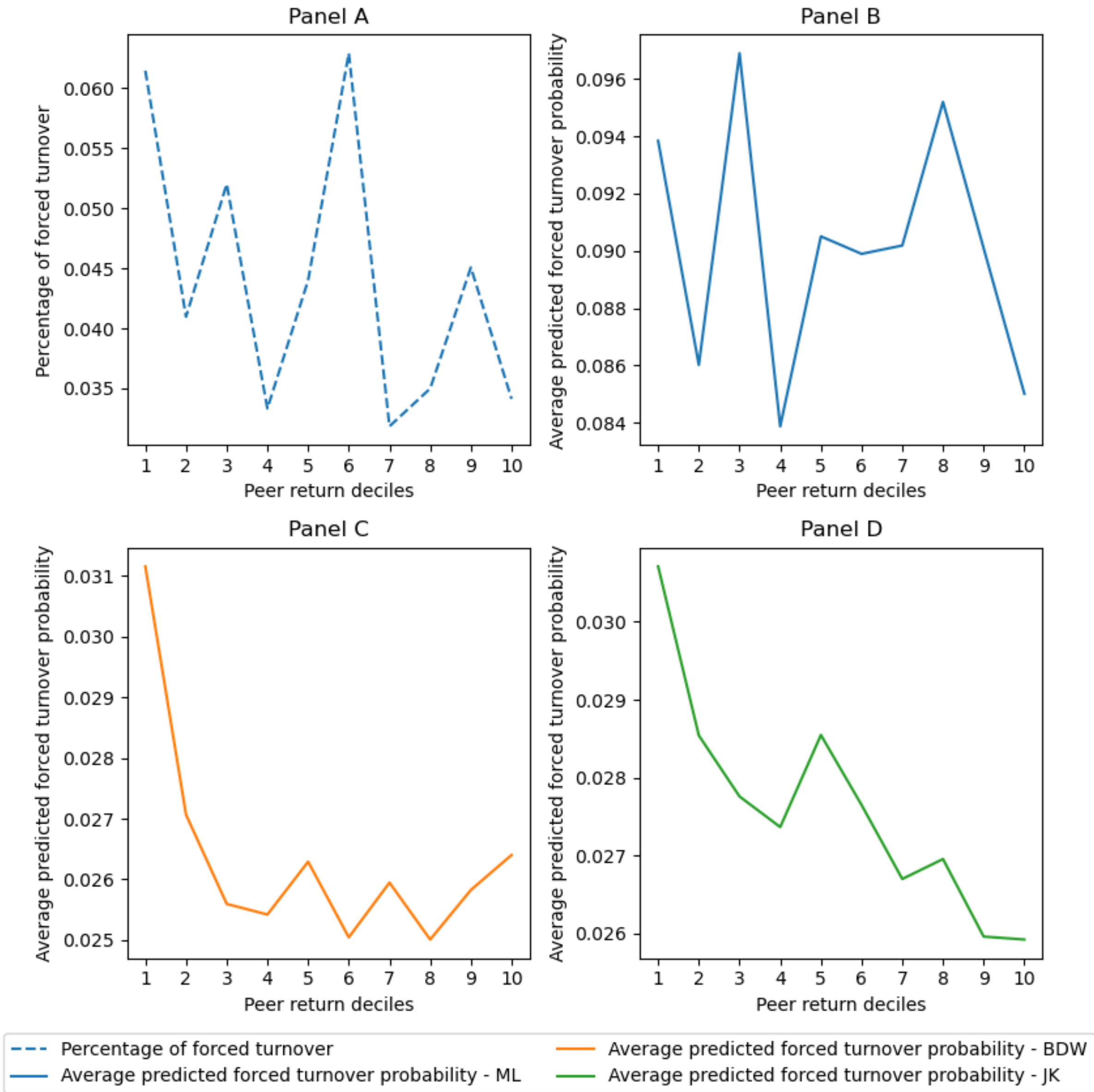


Next, this study investigates the relation between forced CEO turnover and industry-induced performance to identify whether RPE theory in the context of forced CEO turnover follows strong form or weak form. We plot forced CEO turnover rate and average predicted forced CEO probability across 10 deciles of industry-induced performance (i.e., ret_peer_l1), ranging from low (decile 1) to high (decile 10). Under the strong-form assumption, there would be no discernible relation between industry-induced performance and forced CEO turnover. Conversely, according to the weak-form assumption (JK, 2015), a negative relationship would be expected between industry-induced performance and forced CEO turnover.

Panel A of Figure 4 shows that there is no noticeable relation between forced CEO turnover rate and industry-induced performance. The CEO turnover rate varies from 3.2% to 6.3% across different deciles of industry-induced performance. Nevertheless, the absence of a significant relation does not fully imply the absence of any connection between the probability of forced turnover and industry-induced performance. This is because the probability of forced turnover pertains to the risk faced by a CEO of being forced out, rather than the actual occurrence of forced turnover.

Figure 4: Industry-induced performance and forced CEO turnover

This figure shows forced CEO turnover rate and average predicted forced turnover probability across 10 deciles of industry-induced performance (ret_peer_l1). Industry-induced performance are divided into 10 deciles in each year of the test period, ranging from low (decile 1) to high (decile 10). Within each decile, the percentage of forced CEO turnovers is calculated in Panel A, and the average of predicted forced CEO turnover probability is calculated in Panel B, Panel C and Panel D.



Panel C and Panel D show an overall declining pattern of average predicted probability across industry performance deciles, which is consistent with their finding of a negative coefficient on industry-induced performance. In particular, JK's predictions in Panel D show a more distinct decreasing trend compared with BDW's predictions in Panel C do. Because Panel C reveals a lack of relation between the average predicted probability and industry-induced performance after decile 4. While both BDW and JK's predictions demonstrate an overall declining pattern in the average predicted probability, it is important to note that the magnitude of the decline is relatively modest. Specifically, the average predicted probability from BDW (JK)'s model decreases from a high of 3.1% (3.1%) to a low of 2.5% (2.6%), with only a difference of 0.6% (0.5%). However, this declining pattern is not as pronounced compared to the findings in their papers[21]. Furthermore, ML model's predictions in Panel B do not exhibit a significant association with industry-performance deciles. This finding demonstrates that industry-induced performance does not negatively affect forced CEO turnover, which challenges the weak-form RPE theory that industry performance is not filter out in CEO dismissal decisions.

The evidence from ML model's predictions so far cannot fully reject weak-form RPE theory. JK find that "the peer performance effect on CEO turnovers is driven by boards removing many more underperforming (but not outperforming) CEOs in bad times than in good times". If this is true, it is plausible that the inability of industry-induced performance in predicting forced CEO turnover with ML model is driven by CEO-induced performance. In line with JK's approach, we divide CEOs into outperformers and underperformers based on idiosyncratic return. If weak-form RPE holds, we would expect to observe a declining trend of ML's predicted probability across industry-induced performance deciles for underperformers (i.e., *ret_idio_l1* <= 0), and no significant relation across industry-induced performance deciles for outperformers (i.e., *ret_idio_l1* > 0).

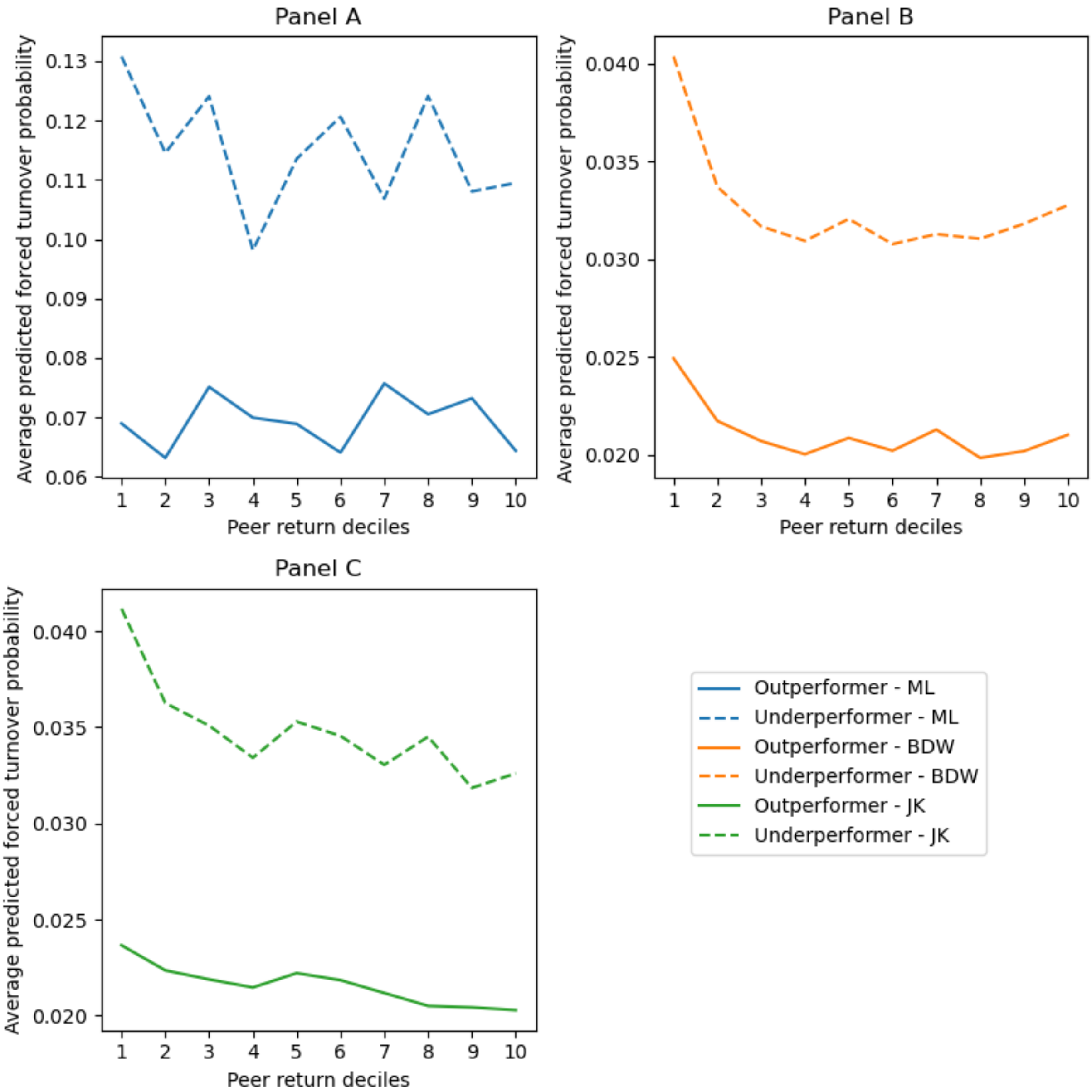
In Panel B and C of Figure 5, while the traditional models' predictions show a relatively stronger (weaker) decreasing trend of forced turnover probability across industry-performance deciles for outperformers (underperformers), the magnitude of this decrease is also modest. For example, Panel C show a decrease of average predicted probability from 4.11% (2.37%) to 3.26% (2.03%) for underperformers (outperformers) across peer return deciles, with only a difference of 0.85% (0.34%). Moreover, in Panel A of Figure 5, the ML predictions clearly show that there is no

[21] In JK's findings, they observe a decrease of 1.93% in the forced turnover probability from the 10th percentile of industry-induced performance (value-weighted) to the 90th percentile of industry-induced performance. In BDW's findings, they observe a one standard deviation increase in industry-induced performance associated with a greater than 0.9% decrease in the probability of forced turnover. In our test sample, the standard deviation of industry-induced performance is 0.199, and the difference between decile 10 and decile 1 is 0.431, which exceeds more than two times the standard deviation. However, our results show that the decrease of forced turnover probability is much smaller than their findings, even the industry-induced performance increases by two standard deviations. Therefore, the decreasing pattern by JK and BDW's predictions is not pronounced in this study.

relation between predicted probability and industry-induced performance for both outperformers and underperformers, contradicting weak-form RPE theory.

Figure 5: Idiosyncratic performance, industry-induced performance and forced CEO turnover

This figure shows average predicted forced turnover probability across 10 deciles of industry-induced performance (ret_peer_l1) for outperformers and underperformers. Outperformers are defined as CEOs with idiosyncratic returns higher than 0. Underperformers are defined as CEOs with idiosyncratic returns lower than 0. Industry-induced performance are divided into 10 deciles in each year of the test period, ranging from low (decile 1) to high (decile 10). Within each decile, the average of predicted forced CEO turnover probability for outperformers and underperformers is calculated in Panel B, Panel C and Panel D.



Fee et al. (2018) highlight that the relation between turnover and industry-induced performance is weak and fragile when considering different timing conventions and sample selections. Their study reveals that weak or no relation exists between forced turnover and industry-induced performance when industry-induced performance is constructed based on yearly timing or when the sample size is limited. The findings from traditional models in Figure 4 and Figure 5 support this perspective, as they indicate a much smaller decrease in the average predicted probability in test set (i.e., a smaller sample size) compared to the reported results in BDW and JK's studies. This implies that the impact of industry-induced performance on forced CEO turnover is not robust. In addition, one common finding in Figure 5 is that the predicted probability for underperformers is consistently higher than outperformers across industry-induced performance deciles. This finding provides additional evidence that idiosyncratic performance plays a dominant role in CEO dismissal decisions, irrespective of industry circumstances. It further supports the strong-form relative performance evaluation (RPE) theory[22].

In summary, the predictions of our ML model provide evidence that higher idiosyncratic performance is linked to a reduced likelihood of dismissal, while industry-induced performance is disregarded in CEO dismissal decisions. In addition, idiosyncratic performance consistently dominates CEO dismissal decisions under different industry circumstances. These findings provide support for the strong-form relative performance evaluation (RPE) theory in the context of forced CEO turnover, thereby rejecting the weak-form RPE theory.

4.3 Other plausible relations

The reliable and accurate predictions from ML model enable to examine additional factors that have caught the attention of researchers concerning forced CEO turnover. In particular, this subsection will briefly discuss the impact of CEO power, CEO tenure, and compensation on forced CEO turnover.

4.3.1 CEO power and forced CEO turnover

Previous studies also examine the relation between CEO power and forced CEO turnover (e.g. Bushman et al., 2010; Hazarika et al., 2012; Jenter & Kanaan, 2015), especially the impact of CEO power on turnover-performance sensitivity (e.g., Denis et al., 1997; Goyal & Park, 2002).

[22] Following another way proposed by JK, we divide industry-induced performance into three groups (terciles) in each year: low, medium, and high. According to JK, turnover-performance relation is more sensitive in bad times than in good times. However, Figure E.1 in Appendix E shows that the turnover-performance sensitivities in the three groups are not significantly different, suggesting that weak-form RPE is not valid.

They find that powerful CEOs are less likely to be dismissed due to poor performance. This phenomenon can be attributed to the entrenchment effects resulting from the CEO power, which enables CEOs to exert influence over the board's decision-making process regarding dismissals.

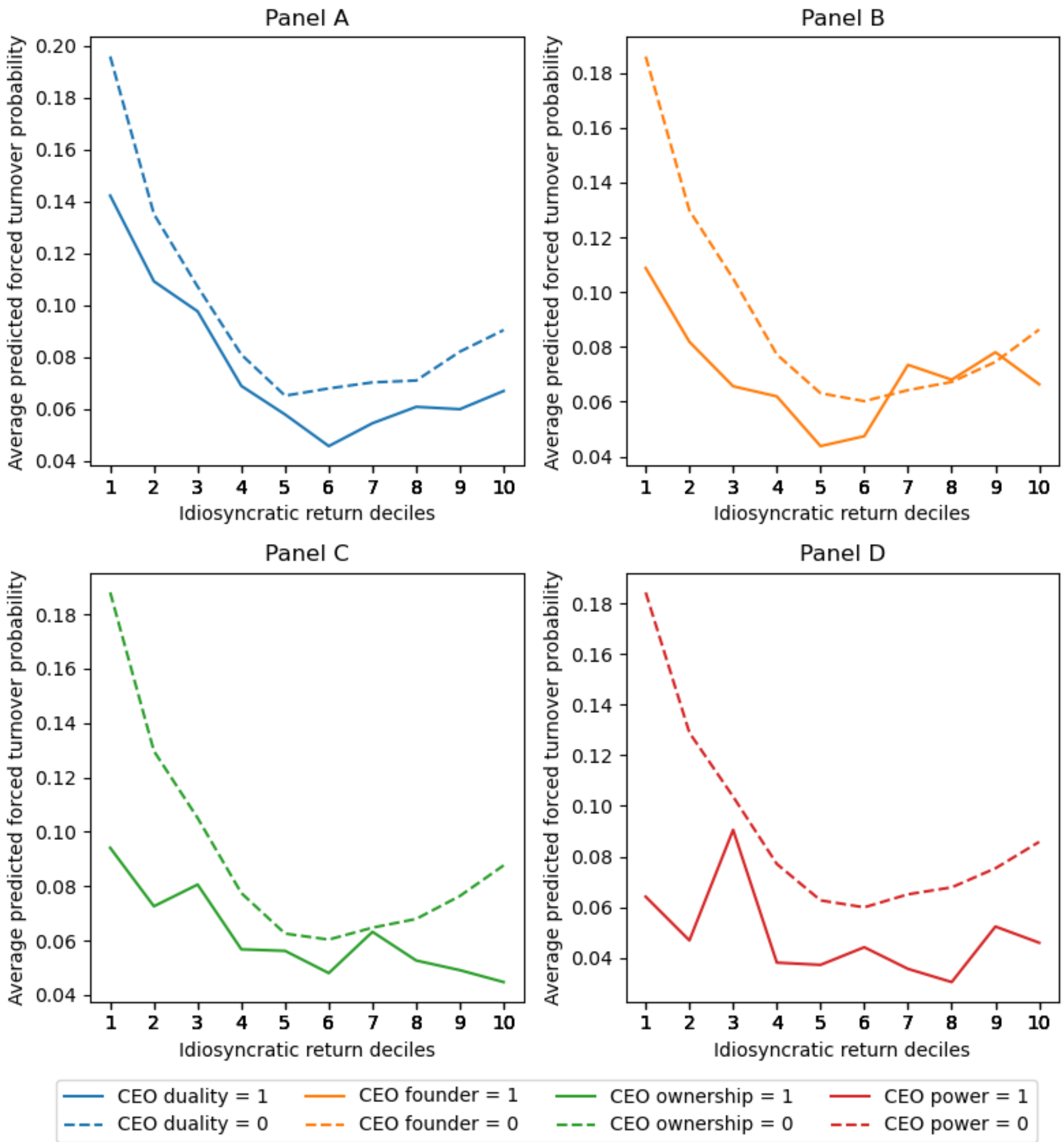
To examine this “CEO power hypothesis”, we follow previous studies to use three proxies to measure CEO power: CEO duality, CEO-founder status and a dummy value indicating whether CEO ownership is larger than 5% or not[23]. In addition, we construct a power indicator as another proxy that takes a value of 1 when a CEO holds both the positions of chairman and founder with a high ownership, otherwise 0. Figure 6 show the plots of ML’s predictions based on different CEO power proxies across idiosyncratic performance deciles.

All panels in Figure 6 illustrate that the forced turnover probability of a powerful CEO displays lower sensitivity to performance, characterized by a slower decrease in the average predicted turnover probability across idiosyncratic return deciles. Moreover, when strictly classify CEO power based on the power indicator in Panel D, idiosyncratic performance show an inability in affecting CEO dismissal decisions for powerful CEOs. Therefore, ML model’s predictions corroborate the hypothesis that powerful CEOs are less sensitive to individual performance in dismissal decisions.

[23] Following JK (2015) and Denis et al. (1997), we select a cut-off of 5% to classify CEOs into high-ownership group and low-ownership group.

Figure 6: CEO power and forced CEO turnover

This figure shows ML's average predicted forced turnover probability across 10 deciles of CEO-induced performance (ret_idio_l1) for powerful CEOs and normal CEOs. Powerful CEOs are defined as CEOs with high ownership, or chairman title, or founder title or power indicator equals 1. CEO-induced performance are divided into 10 deciles in each year of the test period, ranging from low (decile 1) to high (decile 10). Within each decile, the average of predicted forced CEO turnover probability for powerful CEOs and normal CEOs is calculated.



4.3.2 Tenure and forced CEO turnover

The negative and significant relation between tenure and forced CEO turnover has been reported in the results of previous studies (e.g., BDW, 2010; JK, 2015; Dikolli et al., 2014; Jenter &

Lewellen, 2021). According to Hermalin and Weisbach (1998), CEO entrenchment may increase with tenure as the CEO strategically forms more alliances with directors or accumulates greater bargaining power. Specifically, CEOs with a longer tenure are more entrenched and less likely to be dismissed. This study utilizes ML model's predictions to plot the relation between average predicted forced turnover probability and CEO tenure. CEO tenure is divided into seven bins: years 0-2, years 2-4, years 4-6, years 6-8, years 8-11, years 11-14, and over 14 years. Within each tenure bins, the average predicted probability is calculated. Panel A of Figure 7 shows that the predicted probability decreases as CEO tenure increases. This observed decreasing trend suggests that a longer tenure potentially enhances CEOs' entrenchment, leading to a lower likelihood of dismissal.

There are two potential concerns regarding this decreasing pattern. First, Jenter and Lewellen (2021) indicates that the negative relation between tenure and forced CEO turnover probability is partially due to turnover classification scheme. Many turnover classification schemes categorize turnovers above a certain age cut-off (e.g., 60 years old) as "voluntary" (i.e., retirement). However, it is important to note that CEO tenure and age exhibit a strong positive correlation in the context of CEO turnover. This correlation leads to a mechanical reduction in forced turnovers as the tenure increases and more CEOs approach the retirement age. In this study, the turnover classification scheme does not rely on an age cut-off. In addition, the last column in Table 5 provides evidence that the percentage of forced CEO turnover does not decrease as the tenure increases. Therefore, the evidence suggests that the observed decreasing pattern in the predicted probability-tenure relation is unlikely to be a result of misclassifying CEO turnovers.

The second concern is that the observed decreasing pattern in Panel A of figure 7 is possibly due to the entrenchment resulting from CEO power. CEO power and tenure are likely to be highly correlated, that is, powerful CEOs (e.g., founder CEOs) tend to have a longer tenure or CEOs become powerful as the tenure increases (e.g., become a chairman or seize more ownership). In Table 5, we find that the percentage of powerful CEOs (i.e., chairman CEOs and CEOs with a high ownership) increases as tenure increases, and founder CEOs tends to have a longer tenure.

The power indicator this time is different. It is set to 1 if a CEO holds the position of chairman, or is the founder of the company, or possesses a high ownership. Otherwise, the indicator is set to 0. This setting is to fully distinguish powerful CEOs and normal CEOs. In Table 5, the percentage of powerful CEOs based on power indicator also increases with tenure. It is plausible that the observed decreasing pattern in Panel A is potentially driven by powerful CEOs that affects the board's dismissal decisions, and CEO entrenchment may not be attributed to tenure.

We plot ML model's predictions based on different CEO power proxies across tenure bins to examine the relation between tenure and forced turnover probability. From Panel B to Panel D in Figure 7, the average predicted forced turnover probability shows a decreasing pattern as CEO tenure increases for normal CEOs. When we strictly exclude all powerful CEOs from the test sample, Panel E also shows a decreasing pattern in average predicted turnover probability for normal CEOs, ranging from a high of around 11% in the early tenure to a low of around 7% in the late tenure. For powerful CEOs, we do not find any specific pattern across tenue bins. Therefore, the findings in this subsection indicate that a longer CEO tenure can strengthen their position and reduce the likelihood of being dismissed, demonstrating an enhancement of CEO entrenchment.

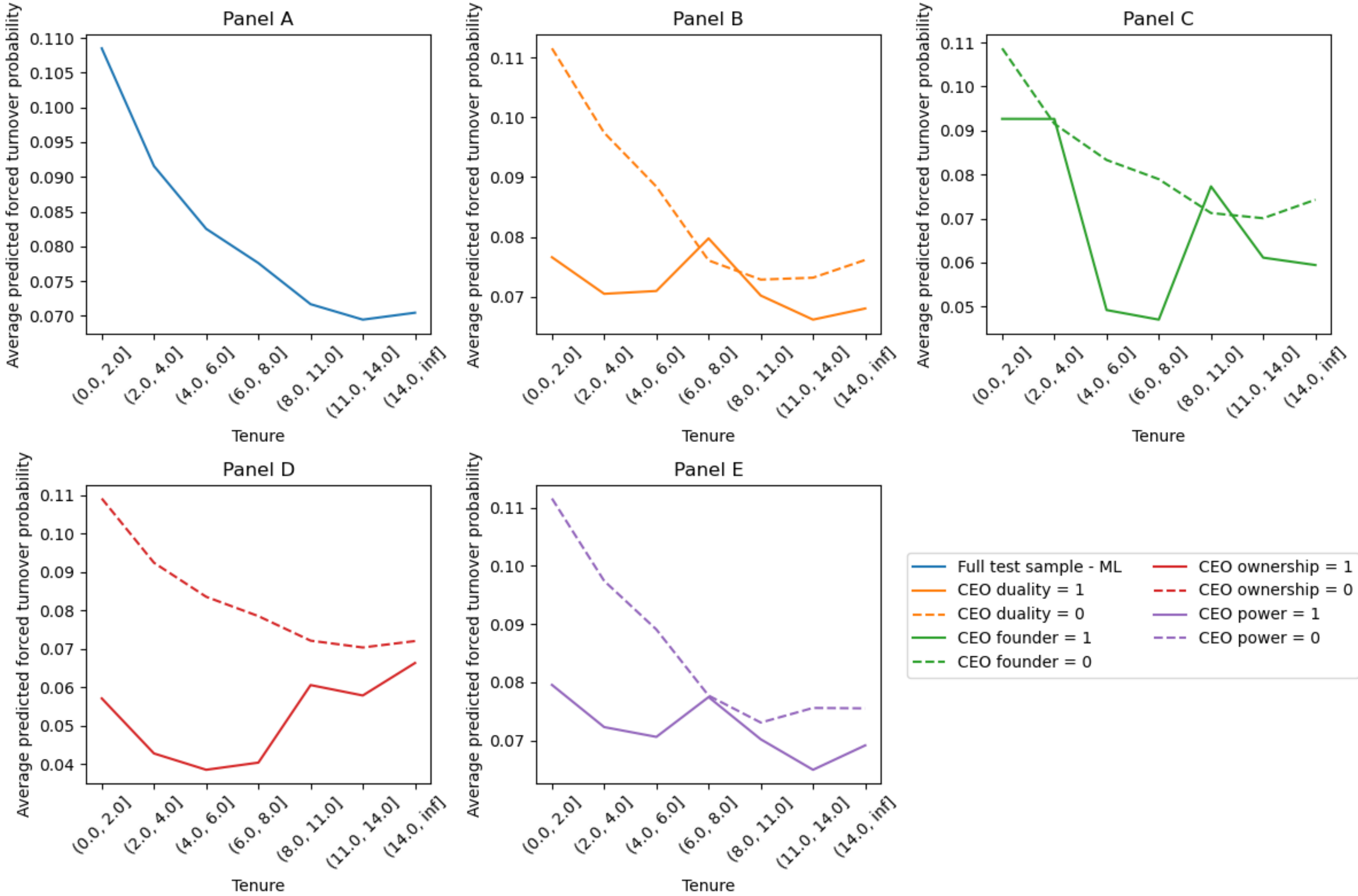
Table 5: Distribution of CEO power across tenure bins

The table shows the percentage of powerful CEOs in each tenure bin based on four different CEO power proxies: CEO duality, CEO-founder status, high (low) ownership and power indicator. The last column shows the percentage of forced turnover in each tenure bin.

Tenure bins	CEO duality	CEO founder	CEO ownership	Power indicator	% Forced turnover
(0.0, 2.0]	8.6%	1.1%	1.1%	0.2%	4.2%
(2.0, 4.0]	21.8%	1.8%	1.7%	0.3%	4.5%
(4.0, 6.0]	33.8%	2.3%	2.2%	0.4%	5.0%
(6.0, 8.0]	41.0%	4.3%	2.4%	1.1%	4.8%
(8.0, 11.0]	45.3%	7.0%	3.8%	1.8%	3.3%
(11.0, 14.0]	53.6%	6.9%	7.1%	2.6%	4.3%
(14.0, inf]	70.6%	25.8%	27.4%	10.6%	4.1%

Figure 7: Tenure and forced CEO turnover

This figure shows ML's average predicted forced turnover probability across 7 tenure bins. Powerful CEOs are defined as CEOs with high ownership, or chairman title, or founder title, or power indicator equals 1. The power indicator in this figure is set to 1 if a CEO holds the position of chairman, or is the founder of the company, or possesses a high ownership. Otherwise, the indicator is set to 0. Within each tenure bin, the average of predicted forced CEO turnover probability for powerful CEOs and normal CEOs is calculated in Panel B, C, D, and E.



4.4 Feature contribution

The contribution of a feature (i.e., a variable) can be assessed based on the magnitude and statistical significance of its coefficient estimates. However, for tree-based ML models the coefficient estimates are not available. To measure the contribution of the features, this study adopts SHAP values (SHapley Additive exPlanations) introduced by Lundberg and Lee (2017).

A SHAP value of a specific feature for a specific observation (x_i) is a measure of the average difference between the prediction made when including the feature (i.e., a full set of features) and the prediction made when excluding the feature (i.e., a set of features without x_i), considering all possible combinations of other features (i.e., different orders of non-x_i that input into a ML model). A positive SHAP value indicates that the feature positively contributes to the prediction, while a negative value suggests a negative contribution. The magnitude of the SHAP value indicates the strength of the feature's influence.

The key advantage of SHAP values is their ability to provide both local and global interpretability. Locally, SHAP values explain the prediction of a specific observation by adding the sum of SHAP values of all features to a base value[24]. Globally, SHAP values summarize the overall impact of features across the entire dataset by averaging the absolute SHAP values[25] for each feature across all instances, enabling insights into feature importance and model behavior.

For tree-based classification models, the original SHAP values for each feature in each observation are calculated in the form of logarithm of odds ratio (log odds ratio). However, log odds ratios are not intuitive and not convenient to explain the contribution. To make the SHAP values more interpretable, this study transforms the original SHAP values into percentage form in two aspects. First, the original base value for predicting testing set is actually the log odds ratio of each prediction in training set, then taking average of these predictions. By transforming, the base value now is the average of predicted forced turnover probability from training set. The

[24] The base value is a starting point for a prediction without any features input. In other words, if an observation does not have any features available, the default prediction of that observation is the base value. The based value is the average of the training sample's predictions.

[25] When assessing the global contribution of each feature, it is not appropriate to directly average the SHAP values for a specific feature across all observations. Consider a scenario with only two observations, ob1 and ob2. Suppose a particular feature has a positive 10% SHAP value for ob1 and a negative 10% SHAP value for ob2. This feature significantly influences the predictions for both ob1 and ob2. However, Simply averaging the SHAP values would result in a global contribution of 0, thereby obscuring the actual impact of this feature. Therefore, while global contribution can indicate the importance or magnitude of a specific feature, it cannot provide a definitive direction of its impact.

SHAP values in the test set now are also transformed to percentage form instead of log odds ratios. The local contribution becomes easier to interpret: for each observation in the test set, the difference between the base value and predicted probability equals the sum of the transformed SHAP values of all features (e.g., if *ret_idio_l1* has a SHAP value of positive 10% for a specific observation, this means this feature increases the forced turnover probability by 10%).

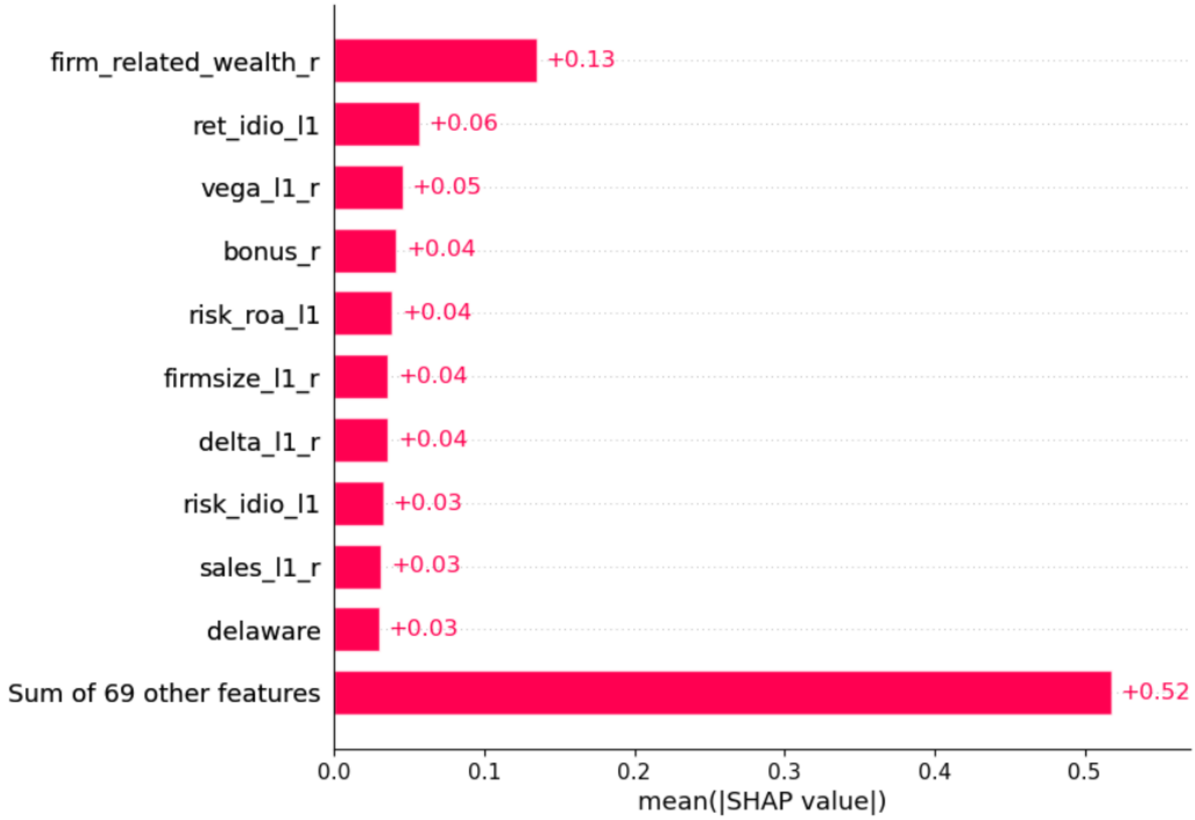
Second, to interpret the global contribution, we take absolute value of those transformed SHAP values and normalize the transformed SHAP values across features in each observation so that the SHAP values for all features of each observation can be aggregated to 100%. Then we take the mean value of each feature across the observations of testing sample and rank mean SHAP values from high to low. In this way, the mean value of absolute SHAP values across features can be added up to 100% and can be easily interpreted (e.g., if *ret_idio_l1* has a mean SHAP value of 10%, this can be explained as this feature has a 10% contribution among all features in predicting forced turnover).

4.4.1 Global contribution

Figure 8 illustrates the top 10 features that have the greatest impact on predicting forced CEO turnover. Among these features, firm-related wealth, which represents the total value of the CEO's equity portfolio, stands out as the most influential feature. It contributes significantly more to the prediction of forced turnovers, with a SHAP value of 13%, which is more than twice the SHAP value of idiosyncratic return. On the one hand, increasing compensation or granting more deferred equity pay can be an effective method to retain talented executives (Jochem et al., 2018; Mehran & Yermack, 1997). One the other hand, Peters and Wagner (2014) suggest that CEO turnover risk is priced in compensation. In addition, Gao et al. (2012) find that a pay cut is used an substitute for forced turnover, helping to explain the rarity of forced CEO turnover. In either case, it is likely that the level of compensation reflects the satisfaction of the board on CEO’s performance[26]. Idiosyncratic return is ranked in the second place following risk-taking incentive (i.e., vega) and bonus. In the top 10 feature importance, incentive related features, risk-taking related features, performance-related features play an important role in predicting forced turnover.

[26] Figure E.2 in Appendix E shows a decreasing pattern of ML model’s predictions across firm-related wealth deciles, suggesting that CEOs having higher firm-related wealth are less likely to be dismissed.

Figure 8: Global contribution



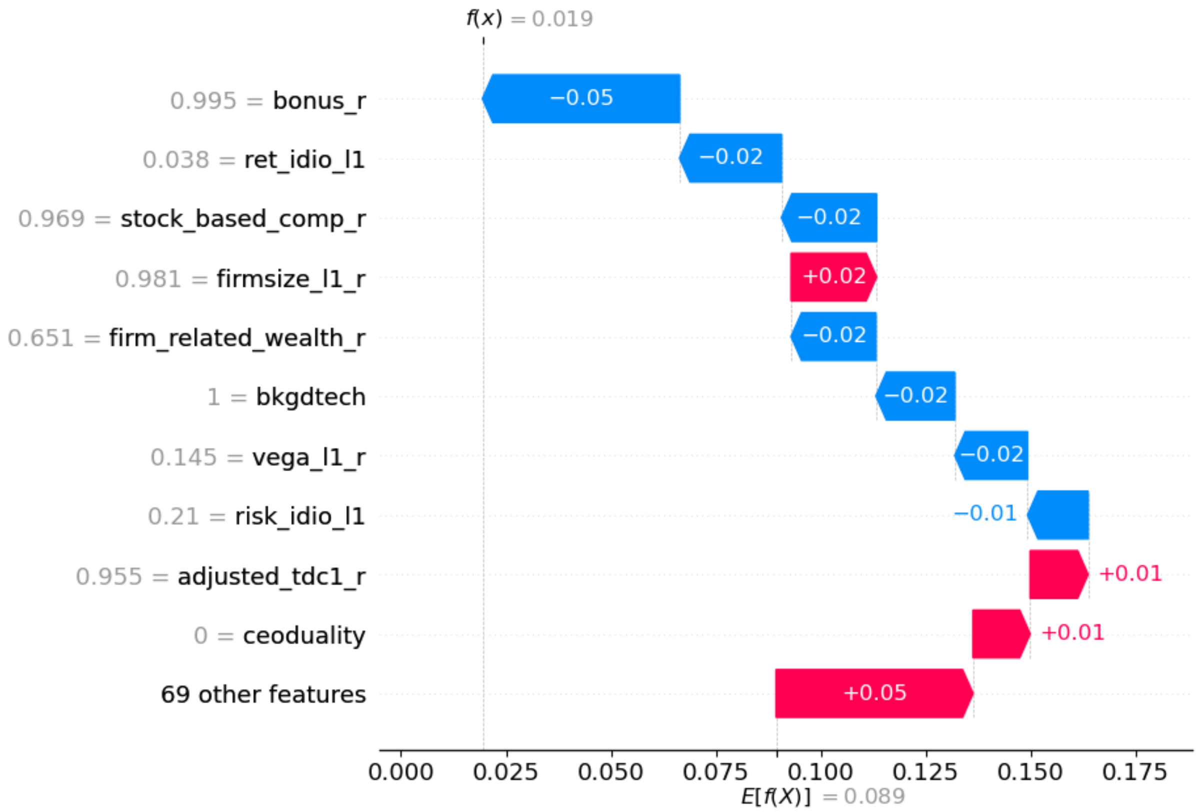
4.4.2 Local contribution

We show two examples about how ML interprets feature importance in predicting forced turnover. The base value for predicting forced turnover in test set is 8.9%, which is the average of predicted probabilities of the training set.

Figure 9.A shows an example of non-turnover in Microsoft in 2016. The waterfall plot exhibits 10 most important features in predicting forced turnover probability. For the CEO who received high bonus and incentive pay, the forced turnover probability is reduced by 7% (0.05 + 0.02). a high idiosyncratic return also helps reduce forced turnover probability by 2%. CEO duality that equals 0 means the CEO is not chairman on the board. This increases the probability of being forced out by 1% as the CEO may not be powerful to affect the board's decision of dismissal.

Interestingly, ML model also detects the effect of background. Forced turnover probability is reduced by 2% for the CEO with technology background. This finding implies that skill-matching between CEOs and firms affect forced turnover probability. Overall, the forced turnover probability for the CEO in Microsoft is reduced by 7% from the base value to 1.9%, which indicates that the CEO is highly unlikely to be dismissed.

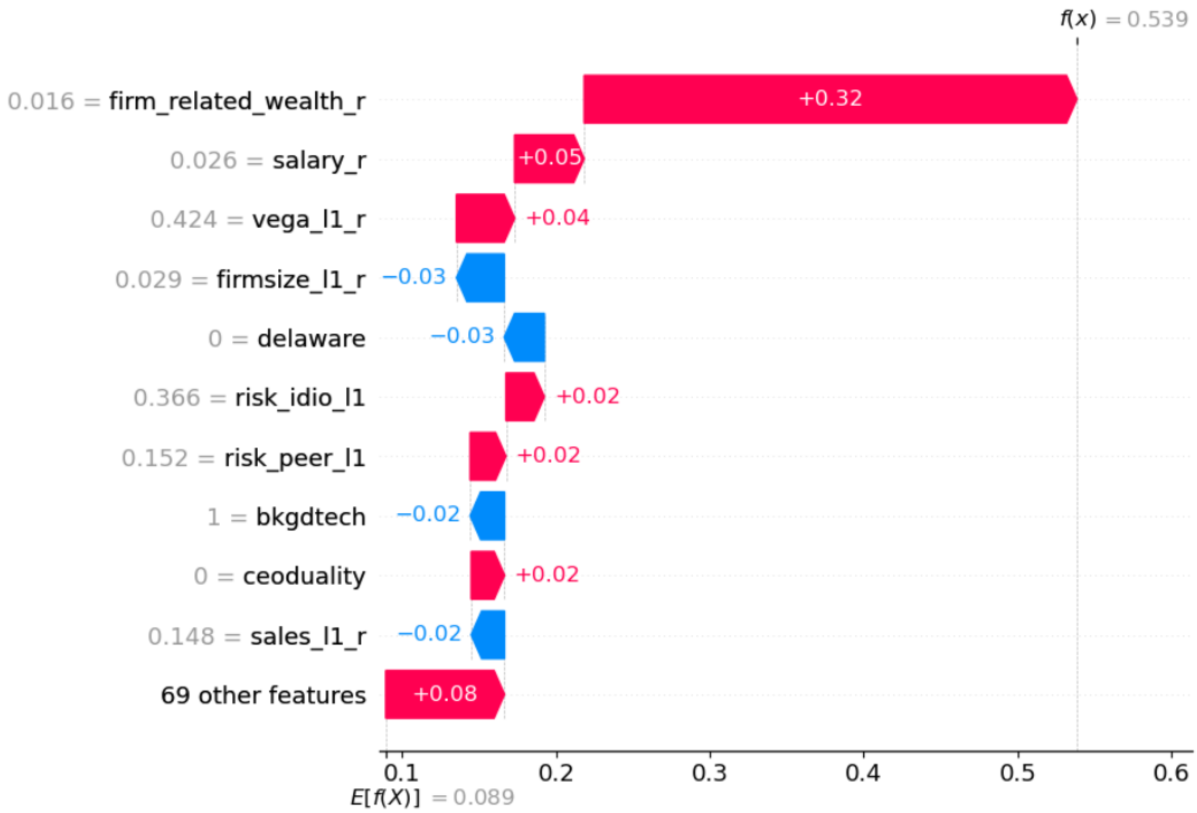
Figure 9.A: Non-turnover example - Microsoft Corp. (GVKEY-12141, FF49-36)



To demonstrate how ML can predict forced turnover and detect skill-matching, we provide another example of Computer Task Group in 2016. Both Computer Task Group and Microsoft operate in the Computer Software industry (the FF49 code is 36). Figure 9.B shows that the probability of forced turnover increases by 37% when there is low firm related wealth and a low salary. On the other hand, small firm size and non-Delaware incorporation[27] reduce the likelihood of forced turnover. Additionally, the figure also reveals that having a technology background reduces the probability of forced turnover, which indicates that ML can detect skill-matching between CEOs and firms. In summary, the forced turnover probability for the CEO in Computer Task Group increased by 45% from the base value to 53.9%. Notably, this predicted probability of forced turnover was subsequently realized in the reality.

[27] Small firms are easier for CEOs to capture (e.g., Taylor, 2010). Delaware firms have a higher tendency to dismiss CEOs, irrespective of firm performance (Jagannathan & Pritchard, 2017).

Figure 9.B: Forced turnover example - Computer Task Group (GVKEY-3342, FF49-36)



4.6 Machine predicted turnovers versus actual turnovers

In this subsection, this study utilizes ML model's predictions to partially reconcile two phenomena: why CEOs who are predicted to be dismissed are still in position (i.e., false positives) and why CEOs who are not predicted to be dismissed but are fired (i.e., false negatives).

The pattern that CEOs who are predicted to be dismissed are still in position could be explained by two reasons: 1) Following competition theory, DeFond & Park (1999) find that the frequency of CEO turnover is greater in highly competitive industries than in less competitive industries. It is likely that those false positives are in low competitive environment, so that they have less likelihood to be replaced; 2) Based on previous corporate governance literature, these firms with false positives are likely to have problematic corporate governance, i.e., CEOs are entrenched. Consequently, these firms tend to undervalue CEOs' performance and overvalue CEO power in dismissal decisions.

To explore the reason, we construct a sample consisting of the observations with ML's predicted probabilities in the top deciles (i.e., decile 10 in Figure 2)[28]. This is to ensure that CEOs are at the same risk level of being dismissed, i.e., CEOs in decile 10 are highly likely to be dismissed and are all predicted to be dismissed now. Within the sample, we divide the CEOs into a treatment group that contains CEOs who are not dismissed (i.e., false positives) and a control group that contains CEOs who are dismissed (i.e., true positives). We compare treatment group to control group to explore which features affect dismissal decisions to deviate from the recommendation made by the ML model.

Table 6 reports the characteristics for firms with true positives and firms with false positives. Compared to true positives, false positives have a higher ownership and are more likely to be a chairman, i.e., those CEOs are more powerful and entrenched. Additionally, we find that the mean ROE for treatment group is negative and significantly lower than the mean ROE for control group. Furthermore, despite the treatment group having a higher mean idiosyncratic return (industry adjusted ROA) compared to the control group, it is important to note that the idiosyncratic return (industry adjusted ROA) still exhibits a negative value. This suggests a poor CEO performance in the treatment group is undervalued in dismissal decisions. Conversely, we find that short-term liquidity measures (i.e., *current_ratio_l1* and *cash_ratio_l1*) are likely to be overvalued in dismissal decisions, whereby they are significantly higher for treatment group than control group. However, we do not find any significant difference between treatment group and control group in the mean of competition measures (i.e., *competition_l1*, *hhi_sale_l1*, and *hhi_mktval_l1*). Taken together, these findings imply that the occurrence of false positives is mainly due to CEO entrenchment, where the interests' misalignment exists between CEOs and shareholders in treatment group. Those CEOs are powerful and are likely to increase the liquidity to expropriate the benefits at the expenses of shareholders' interests[29], thereby affecting the dismissal decisions.

[28] We do not follow a random guessing threshold (i.e., 50%, which is used by ML models to label predictions) to classify forced CEO turnovers. The reason for this is that the threshold for dismissal can vary across firms, years, and industries. For example, technology firms, which often operate in highly dynamic and competitive environments, may have lower tolerance for underperformance or strategic missteps, leading to a lower dismissal threshold. On the other hand, utility firms, which operate in regulated and stable markets, may have different criteria for CEO dismissals, placing greater emphasis on long-term stability and regulatory compliance.

[29] For example, Elyasiani and Zhang (2015) indicate that entrenched CEOs "hold more liquidity because it helps reduce their firm's risks, provides them with job and wealth security, and gives them discretion in pursuing personal objectives".

The pattern that CEOs who are not predicted to be dismissed but are fired is under-explored in previous studies. Therefore, this study intends to explore it empirically using ML's predictions. Similarly, we also construct a sample consisting of the observations with ML's predicted probabilities in the bottom deciles (i.e., decile 1 in Figure 2). This time, all CEOs in decile 1 are predicted to be retained in their positions. Within decile 1, CEOs who are dismissed are labeled as false negatives and assigned to treatment group; CEOs who are not dismissed are labeled as true negatives and assigned to control group.

Table 7 finds that the firms in treatment group face higher competition (i.e., lower market concentration) compared to firms in control group. For performance measures, both accounting-based and market-based, we do not find any significant difference between treatment firms and control firms and those performance measures are positive. One exception is that the firms in treatment group are likely to be in financial distress, with significantly lower Altman Z-score (*z_score_l1*). For the CEOs in treatment group, they have a significantly lower ownership than the CEOs in control group, indicating that they are less powerful than the CEOs in control group. Although the firm size for treatment group is not significantly higher than for control group, the firm size is measured by percentile ranking and indicates that the average firm size of treatment group is relatively large (i.e., above 50% of firms). According to the findings in Table 7, the less powerful CEOs are likely to be misattributed for pushing their firms into financial distress. One plausible explanation could be that directors of poorly performing large firms may choose to dismiss CEOs to protect themselves from accountability and preserve their positions and reputations (Taylor, 2010).

Table 6: A comparison between true positives and false positives

This table reports the mean of performance, entrenchment, and firm-related features for firms in the test sample with CEOs whom our ML algorithm predicted would be dismissed but are retained (treatment group with false positives) and compares it to the mean for firms with CEOs whom our ML algorithm predicted would be dismissed and are dismissed in reality (control group with true positives).

varname	obs (control)	mean (control)	obs (treatment)	mean (treatment)	mean-diff	p
firmsize_l1_r	126	0.370	617	0.405	-0.035	0.223
sales_l1_r	126	0.445	615	0.436	0.010	0.747
tenure	126	5.230	617	5.464	-0.234	0.606
ceoownership	126	0.007	617	0.012	-0.005***	0.000
ceoage	126	57.082	617	56.767	0.315	0.605
ceofounder	126	0.032	617	0.037	-0.006	0.751
ceoduality	126	0.183	617	0.250	-0.067*	0.085
roe_l1	126	0.087	617	-0.111	0.198*	0.068
current_ratio_l1	124	2.233	592	2.656	-0.422**	0.021
cash_ratio_l1	124	0.852	592	1.204	-0.352**	0.014
profit_cl_l1	124	0.376	592	0.483	-0.107	0.426
z_score_l1	125	3.242	616	2.429	0.814*	0.077
roa_ind_adj_l1	126	-0.045	617	-0.014	-0.032*	0.086
competition_l1	126	256.992	617	295.511	-38.518	0.388
risk_roa_l1	121	3.661	575	3.257	0.403	0.304
ret_peer_l1	121	0.102	581	0.118	-0.016	0.407
ret_idio_l1	121	-0.211	581	-0.116	-0.095**	0.019
risk_peer_l1	121	0.155	581	0.164	-0.009**	0.037
risk_idio_l1	121	0.380	581	0.363	0.017	0.368
boardsize_l1	126	9.889	616	9.843	0.046	0.859
inddir_pct_l1	126	0.853	616	0.857	-0.005	0.471
hhi_sale_l1	126	0.090	617	0.081	0.009	0.207
hhi_mktval_l1	126	0.103	617	0.092	0.011	0.230

Table 7: A comparison between true negatives and false negatives

This table reports the mean of performance, entrenchment, and firm-related features for firms in the test sample with CEOs whom our ML algorithm predicted would not be dismissed but are dismissed (treatment group with false negatives) and compares it to the mean for firms with CEOs whom our ML algorithm predicted would not be dismissed and are still in the position (control group with true negatives).

varname	obs (control)	mean (control)	obs (treatment)	mean (treatment)	mean-diff	p
firmsize_l1_r	738	0.452	10	0.520	-0.068	0.409
sales_l1_r	738	0.409	10	0.372	0.037	0.696
tenure	738	8.997	10	8.583	0.414	0.796
ceoownership	738	0.030	10	0.013	0.017***	0.007
ceoage	738	59.499	10	56.810	2.689	0.206
ceofounder	738	0.110	10	0.100	0.010	0.925
ceoduality	738	0.466	10	0.400	0.066	0.697
roe_l1	738	0.141	10	0.090	0.051	0.373
current_ratio_l1	655	2.879	7	3.128	-0.250	0.856
cash_ratio_l1	655	1.157	7	1.210	-0.052	0.962
profit_cl_l1	655	1.108	7	1.494	-0.386	0.596
z_score_l1	738	5.213	10	1.375	3.838***	0.000
roa_ind_adj_l1	738	0.052	10	0.082	-0.030	0.626
competition_l1	738	487.978	10	531.300	-43.322	0.835
risk_roa_l1	592	1.083	6	0.679	0.403	0.142
ret_peer_l1	597	0.113	7	0.023	0.091	0.380
ret_idio_l1	597	0.122	7	0.118	0.003	0.947
risk_peer_l1	597	0.154	7	0.172	-0.018	0.554
risk_idio_l1	597	0.210	7	0.257	-0.047	0.397
boardsize_l1	728	9.710	10	9.900	-0.190	0.824
inddir_pct_l1	728	0.847	10	0.854	-0.007	0.825
hhi_sale_l1	738	0.076	10	0.044	0.032***	0.004
hhi_mktval_l1	738	0.083	10	0.044	0.039***	0.009

5.0 Conclusion

This study uses a machine learning approach, i.e., LightGBM model, to directly predict forced CEO turnover. LightGBM model provides more reliable and accurate predictions, as it outperforms the traditional models as well as LASSO under multiple performance metrics. By utilizing machine learning's predictions, this study provides new insights into the determinants of forced CEO turnover. First, this study validates the RPE theory and finds that CEO-induced performance is a robust determinant of forced CEO turnover. Second, this study rejects the weak-form RPE theory raised by JK (2015). Instead, machine learning's predictions support strong-form RPE, where forced CEO turnover is irrelevant to industry-induced performance. Third, this study validates the effects of CEO power on turnover-performance sensitivity. In addition, this study demonstrates that entrenchment is enhanced as tenure increases and the entrenchment effects of CEO tenure is not resulted from turnover misclassification or CEO power.

To interpret the machine learning model, this study utilizes SHAP value to quantify the contribution of each feature. Globally, we find that incentive related features, performance related features, and risk-taking related features play an import role in predicting forced CEO turnover. Locally, we find that machine learning can somewhat detect skill-matching between CEOs and firms, which provides additional evidence that machine learning can deal with high-dimensional interactions and nonlinearities to improve the quality of a dismissal decision.

Finally, this study explores two phenomena in CEO dismissal decisions: why CEOs who are predicted to be dismissed are still in position (i.e., false positives) and why CEOs who are not predicted to be dismissed but are fired (i.e., false negatives). In terms of false positives, we find that powerful and entrenched CEOs are likely to be retained, regardless of their poor performance. In terms of false negatives, we find that CEOs are likely to be dismissed in relatively larger firms that are facing financial distress, suggesting that the directors in large firms misattribute the poor performance to CEOs to preserve their reputations and positions. These findings suggest that some CEO dismissal decisions are suboptimal for shareholders but are value-maximizing for managers and directors.

In conclusion, the study provides a comprehensive picture about how CEOs are dismissed. For researchers, this study helps to identify the most robust factors in predicting forced CEO turnovers. For practitioners, this study provides a novel machine learning method to complement and improve the quality of CEO dismissal decisions.

Appendix A: CEO classifications and definitions

Code	Title	Brief Description
1	Involuntary - CEO death	The CEO died while in office and did not have an opportunity to resign before health failed.
2	Involuntary - CEO illness	Required announcement that the CEO was leaving for health concerns rather than removed during a health crisis.
3	Involuntary – CEO dismissed for job performance	The CEO stepped down for reasons related to job performance. This included situations where the CEO was immediately terminated as well as when the CEO was given some transition period, but the media coverage was negative. Often the media cited financial performance or some other failing of CEO job performance (e.g., leadership deficiencies, innovation weaknesses, etc.).
4	Involuntary - CEO dismissed for legal violations or concerns	The CEO was terminated for behavioral or policy-related problems. The CEO's departure was almost always immediate, and the announcement cited an instance where the CEO violated company HR policy, expense account cheating, etc.
5	Voluntary - CEO retired	Voluntary retirement based on how the turnover was reported in the media. Here the departure did not sound forced, and the CEO often had a voice or comment in the succession announcement. Media coverage of voluntary turnover was more valedictory than critical. Firms use different mandatory retirement ages, so we could not use 65 or older and facing mandatory retirement as a cut off. We examined coverage around the event and subsequent coverage of the CEO's career when it sounded unclear.
6	Voluntary - new opportunity (new career driven succession)	The CEO left to pursue a new venture or to work at another company. This frequently occurred in startup firms and for founders.
7	Other	Interim CEOs, CEO departure following a merger or acquisition, company ceased to exist, company changed key identifiers so it is not an actual turnover, and CEO may or may not have taken over the new company.
8	Missing	Despite attempts to collect information, there was not sufficient data to assign a code to the turnover event. These will remain the subject of further investigation and expansion.

9	Execucomp error	If a researcher were to create a dataset of all potential turnovers using execucomp (co_per_rol != l.co_per_rol), several instances will appear of what looks like a turnover when there was no actual event. This code captures those.

Source: The user manual of CEO turnover dataset by Gentry et al. (2021) .

Appendix B: Variable list

Variable name	Definition	Source
CEO level		
tenure	Number of years on CEO position.	Execucomp
salary	CEO annual salary (in thousands).	Execucomp
bonus	CEO bonus (in thousands).	Execucomp
adjusted_tdc1	Total compensation (in thousands). Adjusted to align the pre-2006 with post-2006 compensation. Following Walker (2009).	Execucomp
stock_based_comp	CEO stock awards (in thousands).	Execucomp
option_based_comp	CEO option awards (in thousands).	Execucomp
equity_based_comp	CEO stock awards plus option awards (in thousands).	Execucomp
delta	Pay-performance sensitivity (in thousands). The change in the dollar value of the executive's wealth for a one percentage point change in stock price. Following Coles et al. (2006).	Execucomp
optiondelta	Pay-performance sensitivity (in thousands). The change in the dollar value of the executive's option wealth for a one percentage point change in stock price. Following Coles et al. (2006).	Execucomp
sharedelta	Pay-performance sensitivity (in thousands). The change in the dollar value of the executive's stock wealth for a one percentage point change in stock price. Following Coles et al. (2006).	Execucomp
firm_related_wealth	The sum of the value of the stock and option portfolio held by the executive (in thousands). Following Coles et al. (2006).	Execucomp
vega	Risk-taking incentives (in thousands). The change in the dollar value of the executive's wealth for a 0.01 change in the annualized standard deviation of stock returns. Following Coles et al. (2006).	Execucomp
ceo_turnover	A dummy variable equals 1 if the CEO is dismissed, else 0.	Gentry et al. (2021)
forcedturnover	A dummy variable equals 1 if the CEO is forced out, else 0.	Gentry et al. (2021)
voluntaryturnover	A dummy variable equals 1 if the CEO voluntarily resign his/her job, else 0.	Gentry et al. (2021)
Female	A dummy variable equals 1 if the CEO is Female, else 0.	Execucomp & BoardEx
foreign	A dummy variable equals 1 if the CEO's nationality is not US.	BoardEx
ceoage	CEO age.	Execucomp & BoardEx

ceofounder	A dummy variable equals 1 if the CEO is also the founder of the firm, else 0.	Execucomp & BoardEx
ceoduality	A dummy variable equals 1 if the CEO is also the chairman of the firm in year t, else 0.	Execucomp & BoardEx
mba	Equals 1 if the CEO holds an MBA degree.	BoardEx
ivyleague	Equals 1 if the CEO graduated from an university in Ivy League.	BoardEx
phd	Equals 1 if the CEO holds a PhD degree.	BoardEx
bkgdacademic	Equals 1 if job history includes in title one of the following: "professor", "academic", "lecturer", "teacher", "instructor", "faculty", "fellow", "dean", "teaching".	BoardEx
bkgdfinance	Equals 1 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".	BoardEx
bkgdhr	Equals 1 if job history includes in title one of the following: "hr ", "recruitment", "human resource".	BoardEx
bkgdlaw	Equals 1 if job history includes in title one of the following: "lawyer", "legal", "attorney", "judge", "judicial".	BoardEx
bkgdmanager	Equals 1 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", "Employer", "Associate", "Head of Division".	BoardEx
bkgdmarketing	Equals 1 if job history includes in title one of the following: "Marketing", "Publisher", "MKTG", "Sales", "Brand Manager", "Regional Manager", "Communication", "Merchandising", "Comms", "Distribution", "Media".	BoardEx
bkgdmilitary	Equals 1 if job history includes in title one of the following: "Captain", "Soldier", "Lieutenant", "Admiral", "Military", "Commanding", "Commander", "Commandant", "Infantry", "Veteran", "Sergeant", "Army".	BoardEx

bkgdpolitician	Equals 1 if job history includes in title one of the following: "Politician", "Senator", "Political", "Governor".	BoardEx
bkgdscience	Equals 1 if job history includes in title one of the following: "Researcher", "Medical", "Doctor", "Scientist", "Physician", "Engineer", "Biologist", "Geologist", "Physicist", "Metallurgist", "Science", "Scientific", "Pharmacist".	BoardEx
bkgdtech	Equals 1 if job history includes in title one of the following: "Technology", "Software", "Programmer", " IT ", "Chief Information Officer", "Database", "System Administrator", "Developer".	BoardEx
private	Equals 1 if the CEO has work experience in private firms.	BoardEx
armedforces	Equals 1 if the CEO has work experience in Armed Forces.	BoardEx
charities	Equals 1 if the CEO has work experience in Charities.	BoardEx
clubs	Equals 1 if the CEO has work experience in Club.	BoardEx
government	Equals 1 if the CEO has work experience in Government.	BoardEx
medical	Equals 1 if the CEO has work experience in Medical.	BoardEx
partnership	Equals 1 if the CEO has work experience in partnership firms	BoardEx
quoted	Equals 1 if the CEO has work experience in quoted firms	BoardEx
sporting	Equals 1 if the CEO has work experience in Sporting.	BoardEx
universities	Equals 1 if the CEO has work experience in Universities	BoardEx
job_intl	Equals 1 if the CEO has a job position outside the US.	BoardEx
job_intl_Africa	Equals 1 if the CEO has a job position in Africa.	BoardEx
job_intl_Asia	Equals 1 if the CEO has a job position in Asia.	BoardEx
job_intl_Europe	Equals 1 if the CEO has a job position in Europe.	BoardEx
job_intl_North_America	Equals 1 if the CEO has a job position in North America.	BoardEx
job_intl_South_America	Equals 1 if the CEO has a job position in South America.	BoardEx
job_intl_Oceania	Equals 1 if the CEO has a job position in Oceania.	BoardEx
ceoownership	The percentage of shares owned by CEO.	Execucomp
Board level		
boardsize	Number of direcors on the board.	BoardEx
inddir_pct	The percentage of independent directors on the board. Based on the definition in WRDS, directors labeled as "Supervisory Director" are regarded as independent directors in BoardEx.	BoardEx

E-index	Entrenchment index	ISS governance
Firm level		
roa	Return on assets (Income before extraodinary items divided by total assets).	Compustat
tangibility	The total value of property, plant and eqquipment divided by total assets.	Compustat
roe	Return on equity (Income before extraodinary items divided by value of common equity).	Compustat
firmsize	Natural logarithm of total assets.	Compustat
m2b	Market to book value.	Compustat
capex	Capital expenditure divided by total assets (Missing values are replaced with 0).	Compustat
rdintensity	Research and development expense divided by total assets (Missing values are replaced with 0).	Compustat
booklev	Book leverage. Total debts divided by the sum of total debts and common equity.	Compustat
cashflow	(EBITDA - tax payment - interest and related expense)/total assets.	Compustat
ceoso_e	Equals 1 if the CEO is exempt from filing Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002.	Compustat
ceoso_y	Equals 1 if the CEO has filed Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002.	Compustat
ceoso_n	Equals 1 if the CEO has not filed Certification Documents as required under section 302 of the Sarbanes-Oxley Act of 2002.	Compustat
delaware	Equals 1 if the firm is incorporated in Delaware.	Compustat
ebitda_at	EBITDA/total assets	Compustat
sales	Natural logarithm of net sales.	Compustat
dividends_at	dividends/at	Compustat
divratio	Dividends payment divided by EBITDA.	Compustat
divpayer	Equals 1 if the firm pays dividends, else 0.	Compustat
current_ratio	Current assets/Current liability	Compustat
cash_ratio	Cash and short-term investments/Current liability	Compustat
profit_cl	EBITDA/Current liability	Compustat
z_score	Altman Z-score.	Compustat

hhi_mktval	Herfindahl-Hirschman Index on market value	Compustat
hhi_sale	Herfindahl-Hirschman Index on market value	Compustat
roa_ind_adj	Industry adjusted ROA (Industry classification is based on Fama-French 49 industry classification).	Compustat
risk_roa	The standard deviation of quarterly industry median adjusted earnings growth over the past four years (data should be at least 8 of 16 quarters available).	Compustat
Industry and market level		
ind_roa	Industry median ROA(Industry classification is based on Fama-French 49 industry classification).	Compustat
ret_idio_l1	Regress individual firm daily returns on value-weighted industry returns in the fiscal year t-1. The annual idiosyncratic return is equal to firm annual return minus industry induced annual return.	CRSP & Fama-French 49 industry returns
ret_peer_l1	Regress individual firm daily returns on value-weighted industry returns in the fiscal year t-1. The peer return is equal to annualised expected return.	CRSP & Fama-French 49 industry returns
risk_idio_l1	Regress individual firm daily returns on value-weighted industry returns in the fiscal year t-1. The idiosyncratic risk is equal to the standard deviation of residuals.	CRSP & Fama-French 49 industry returns
risk_peer_l1	Regress individual firm daily returns on value-weighted industry returns in the fiscal year t-1. The industry-related risk is equal to the standard deviation of industry-induced daily returns.	CRSP & Fama-French 49 industry returns
Miscellaneous		
gvkey	Firm identifier from Compustat.	Compustat
cusip	Firm identifier from Compustat.	Compustat
cik	Firm identifier from Compustat. Also SEC firm identifier.	Compustat
companyid	Firms identifier from BoardEx.	BoardEx
directorid	Individual identifier from Boardex.	BoardEx
coperol	Firm-executive unique identifier.	Execucomp
execid	Individual identifier from Execucomp.	Execucomp
permno	In this dataset, each firm is assigned with its main permno.	CRSP
ceo_dismissal	Identifies whether the turnover is forced or voluntary.	Gentry et al. (2021)

Appendix C: Confusion matrix

Figure C.1.A: Confusion matrix for LightGBM model

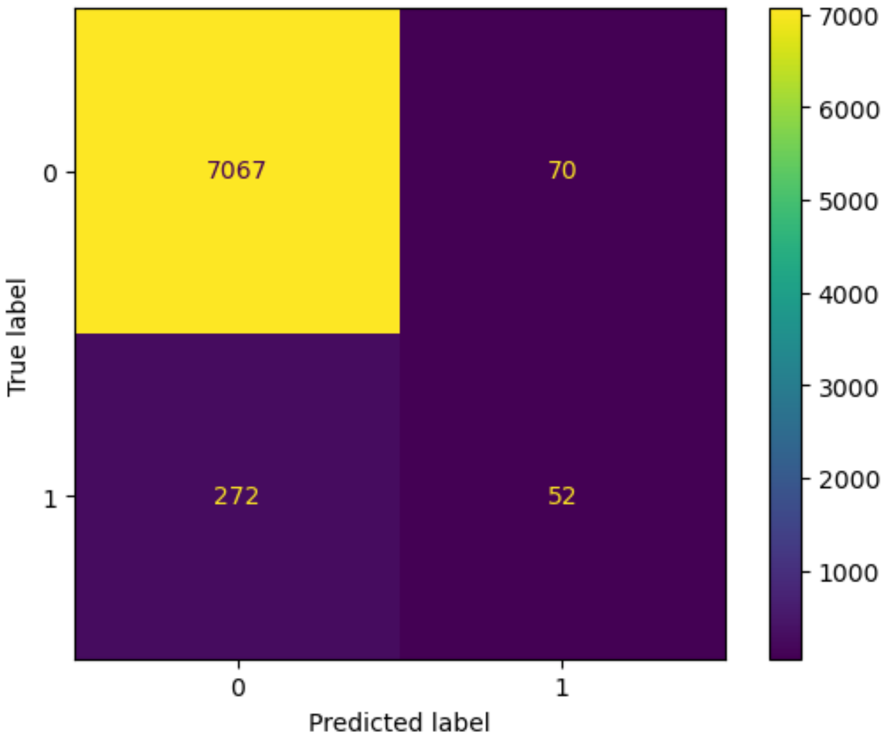


Figure C.1.B: Confusion matrix for LASSO model

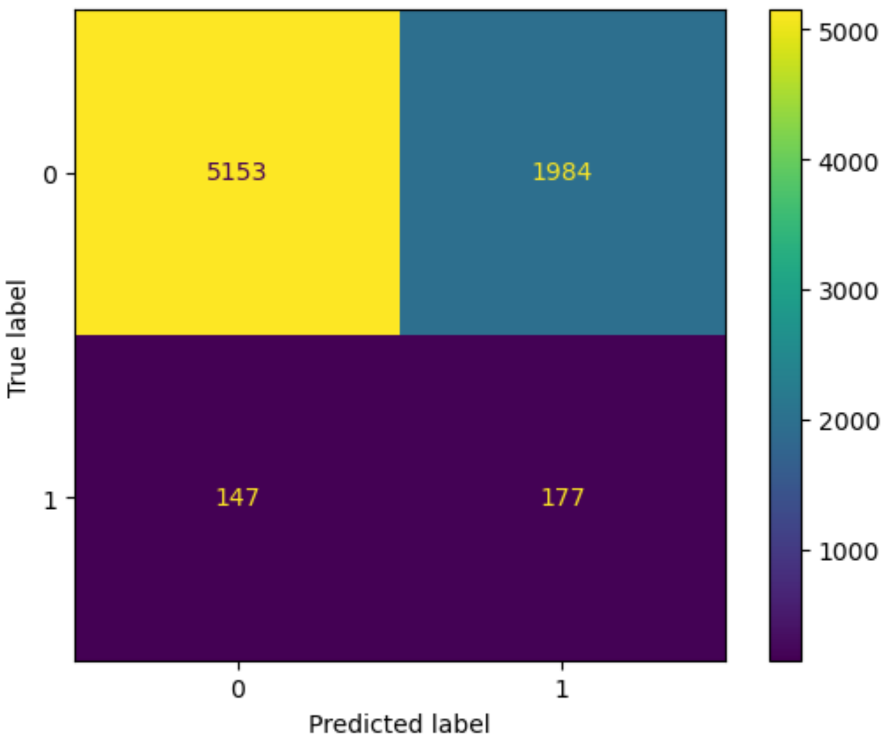


Figure C.1.C: Confusion matrix for BDW model

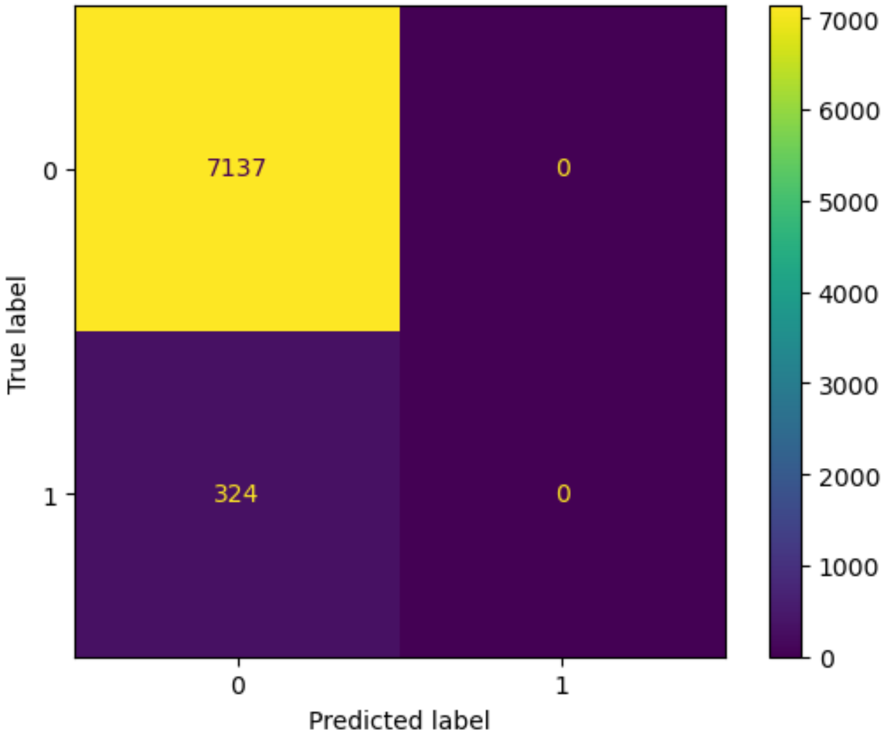
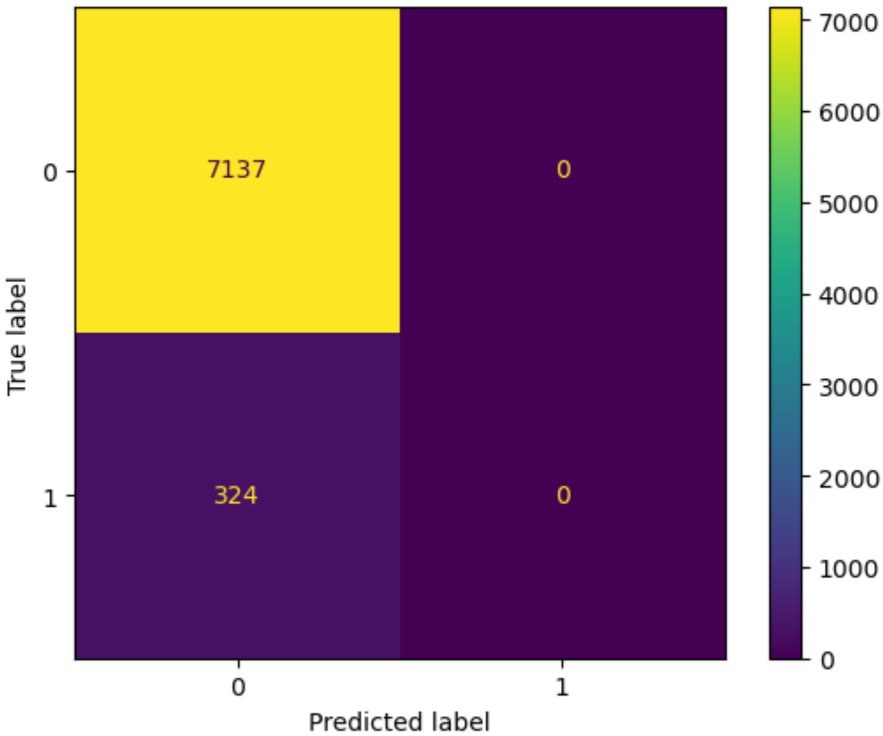


Figure C.1.D: Confusion matrix for JK model



Appendix D: NDCG scores under different cut-offs

Model	NDCG@1.0 %	NDCG@2.0 %	NDCG@4.0 %	NDCG@6.0 %	NDCG@10.0 %
LightGBM	0.58	0.47	0.36	0.37	0.42
LASSO	0.28	0.24	0.20	0.23	0.27
BDW	0.12	0.11	0.10	0.12	0.19
JK	0.11	0.13	0.12	0.15	0.20

Appendix E: Additional tests

Figure E.1: Turnover-performance sensitivity under different industry circumstances

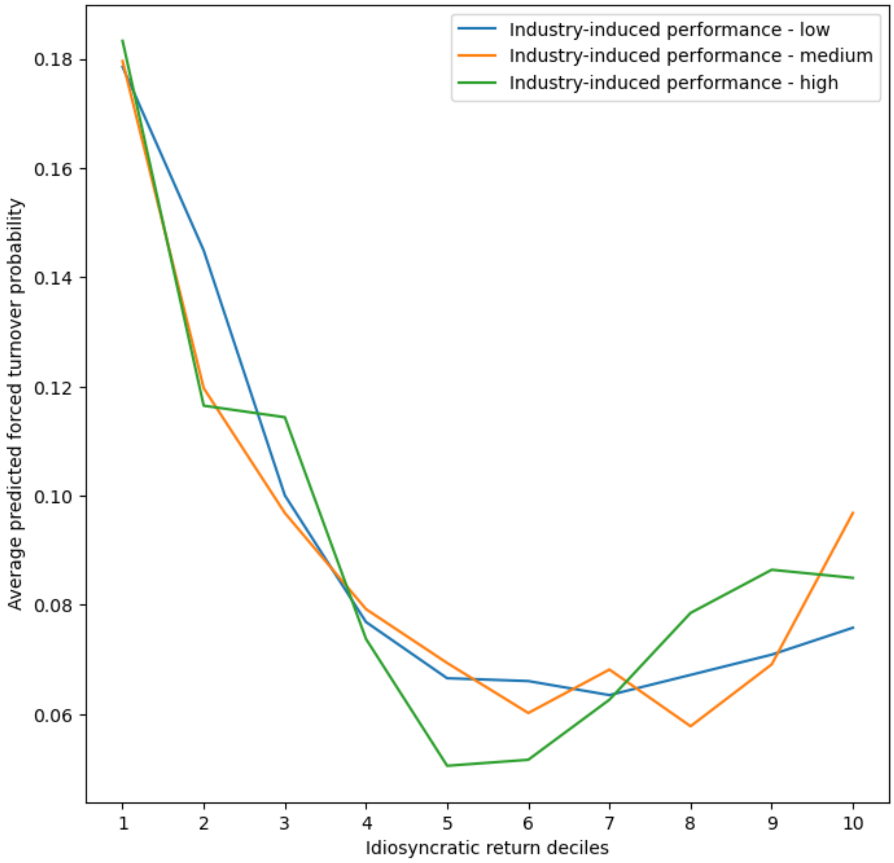
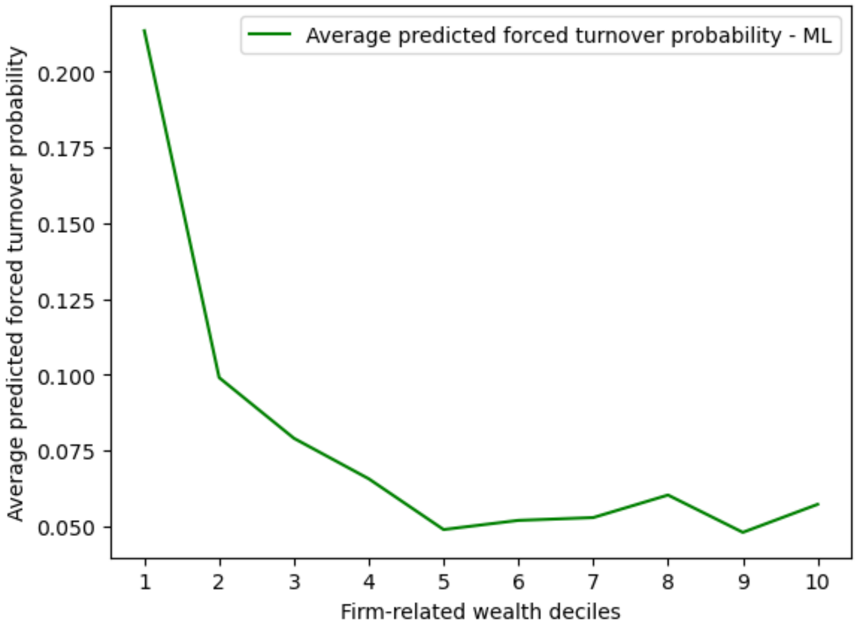


Figure E.2: CEO wealth and forced turnover



Reference

Allgood, S., & Farrell, K. A. (2000). The Effect of Ceo Tenure on the Relation Between Firm Performance and Turnover. *Journal of Financial Research*, *23*(3), 373–390. https://doi.org/10.1111/j.1475-6803.2000.tb00748.x

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research*, *58*(1), 199–235. https://doi.org/10.1111/1475-679X.12292

Beneish, M. D., Marshall, C. D., & Yang, J. (2017). Explaining CEO retention in misreporting firms. *Journal of Financial Economics*, *123*(3), 512–535. https://doi.org/10.1016/j.jfineco.2016.12.004

Bertomeu, J., Cheynel, E., Floyd, E., & Pan, W. (2021). Using machine learning to detect misstatements. *Review of Accounting Studies*, *26*(2), 468–519. https://doi.org/10.1007/s11142-020-09563-8

Bhagat, S., Bolton, B. J., & Subramanian, A. (2010). *CEO Education, CEO Turnover, and Firm Performance* (SSRN Scholarly Paper No. 1670219). https://doi.org/10.2139/ssrn.1670219

Bushman, R., Dai, Z., & Wang, X. (2010). Risk and CEO turnover. *Journal of Financial Economics*, *96*(3), 381–398. https://doi.org/10.1016/j.jfineco.2010.03.001

Campbell, T. C., Gallmeyer, M., Johnson, S. A., Rutherford, J., & Stanley, B. W. (2011). CEO optimism and forced turnover. *Journal of Financial Economics*, *101*(3), 695–712. https://doi.org/10.1016/j.jfineco.2011.03.004

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, X., Cho, Y. H. (Tony), Dou, Y., & Lev, B. (2022). Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data. *Journal of Accounting Research*, *60*(2), 467–515. https://doi.org/10.1111/1475-679X.12429

Coles, J., Daniel, N., & Naveen, L. (2006). Managerial incentives and risk-taking☆. *Journal of Financial Economics*, *79*(2), 431–468. https://doi.org/10.1016/j.jfineco.2004.09.004

Coles, J. L., Daniel, N. D., & Naveen, L. (2014). Co-opted Boards. *The Review of Financial Studies*, *27*(6), 1751–1796. https://doi.org/10.1093/rfs/hhu011

DeFond, M. L., & Park, C. W. (1999). The effect of competition on CEO turnover1We appreciate helpful comments from Steve Baginski, Daniel Beneish, Walt Blacconiere, John Core (the referee), Pat Hopkins, S.P. Kothari (the editor), Jevons Lee, Eric Noreen, Jerry Salamon, Jim Seida, Bert Steece, Kay Stice, Nathan Stuart, Mark Vargus, Jim Wahlen, T.J. Wong, and an anonymous referee. We also benefited from the comments of workshop participants at Arizona State University, UCLA, USC and Vanderbilt University.1. *Journal of Accounting and Economics*, *27*(1), 35–56. https://doi.org/10.1016/S0165-4101(98)00044-5

Denis, D. J., Denis, D. K., & Sarin, A. (1997). Ownership structure and top executive turnover. *Journal of Financial Economics*, *45*(2), 193–221. https://doi.org/10.1016/S0304-405X(97)00016-0

Dikolli, S. S., Mayew, W. J., & Nanda, D. (2014). CEO tenure and the performance-turnover relation. *Review of Accounting Studies*, *19*(1), 281–327. https://doi.org/10.1007/s11142-013-9247-6

Eisfeldt, A. L., & Kuhnen, C. M. (2013). CEO turnover in a competitive assignment framework. *Journal of Financial Economics*, *109*(2), 351–372. https://doi.org/10.1016/j.jfineco.2013.02.020

Elyasiani, E., & Zhang, L. (2015). CEO entrenchment and corporate liquidity management. *Journal of Banking & Finance*, *54*, 115–128. https://doi.org/10.1016/j.jbankfin.2015.01.014

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, *8*(1), 140. https://doi.org/10.1186/s40537-021-00516-9

Engel, E., Hayes, R. M., & Wang, X. (2003). CEO turnover and properties of accounting information. *Journal of Accounting and Economics*, *36*(1), 197–226. https://doi.org/10.1016/j.jacceco.2003.08.001

Erel, I., Stern, L. H., Tan, C., & Weisbach, M. S. (2021). Selecting Directors Using Machine Learning. *The Review of Financial Studies*, *34*(7), 3226–3264. https://doi.org/10.1093/rfs/hhab050

Ertugrul, M., & Krishnan, K. (2011). Can CEO dismissals be proactive? *Journal of Corporate Finance*, *17*(1), 134–151. https://doi.org/10.1016/j.jcorpfin.2010.09.008

Farrell, K. A., & Whidbee, D. A. (2003). Impact of firm performance expectations on CEO turnover and replacement decisions. *Journal of Accounting and Economics*, *36*(1), 165–196. https://doi.org/10.1016/j.jacceco.2003.09.001

Fee, C. E., Hadlock, C. J., Huang, J., & Pierce, J. R. (2018). Robust Models of CEO Turnover: New Evidence on Relative Performance Evaluation. *The Review of Corporate Finance Studies*, *7*(1), 70–100. https://doi.org/10.1093/rcfs/cfx018

Gao, H., Harford, J., & Li, K. (2012). CEO pay cuts and forced turnover: Their causes and consequences. *Journal of Corporate Finance*, *18*(2), 291–310. https://doi.org/10.1016/j.jcorpfin.2012.01.001

Geertsema, P., & Lu, H. (2023). Relative Valuation with Machine Learning. *Journal of Accounting Research*, *61*(1), 329–376. https://doi.org/10.1111/1475-679X.12464

Gentry, R. J., Harrison, J. S., Quigley, T. J., & Boivie, S. (2021). A database of CEO turnover and dismissal in S&P 1500 firms, 2000–2018. *Strategic Management Journal*, *42*(5), 968–991. https://doi.org/10.1002/smj.3278

Gibbons, R., & Murphy, K. J. (1990). Relative Performance Evaluation for Chief Executive Officers. *Industrial and Labor Relations Review*, *43*(3), 30S-51S. https://doi.org/10.2307/2523570

Goyal, V. K., & Park, C. W. (2002). Board leadership structure and CEO turnover. *Journal of Corporate Finance*, *8*(1), 49–66. https://doi.org/10.1016/S0929-1199(01)00028-1

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Guo, L., & Masulis, R. W. (2015). Board Structure and Monitoring: New Evidence from CEO Turnovers. *The Review of Financial Studies*, *28*(10), 2770–2811. https://doi.org/10.1093/rfs/hhv038

Hazarika, S., Karpoff, J. M., & Nahata, R. (2012). Internal corporate governance, CEO turnover, and earnings management. *Journal of Financial Economics*, *104*(1), 44–69. https://doi.org/10.1016/j.jfineco.2011.10.011

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

Huson, M. R., Parrino, R., & Starks, L. T. (2001). Internal Monitoring Mechanisms and CEO Turnover: A Long-Term Perspective. *The Journal of Finance*, *56*(6), 2265–2297. https://doi.org/10.1111/0022-1082.00405

Jagannathan, M., & Pritchard, A. C. (2017). Do Delaware CEOs get fired? *Journal of Banking & Finance*, *74*, 85–101. https://doi.org/10.1016/j.jbankfin.2016.10.008

Jenter, D., & Kanaan, F. (2015). CEO Turnover and Relative Performance Evaluation: CEO Turnover and Relative Performance Evaluation. *The Journal of Finance*, *70*(5), 2155–2184. https://doi.org/10.1111/jofi.12282

Jenter, D., & Lewellen, K. (2021). Performance-Induced CEO Turnover. *The Review of Financial Studies*, *34*(2), 569–617. https://doi.org/10.1093/rfs/hhaa069

Jochem, T., Ladika, T., & Sautner, Z. (2018). The Retention Effects of Unvested Equity: Evidence from Accelerated Option Vesting. *The Review of Financial Studies*, *31*(11), 4142–4186. https://doi.org/10.1093/rfs/hhy017

Kaplan, S. N., & Minton, B. (2006). *How has CEO Turnover Changed? Increasingly Performance Sensitive Boards and Increasingly Uneasy CEOs* (Working Paper No. 12465). National Bureau of Economic Research. https://doi.org/10.3386/w12465

Koh, P.-S., & Reeb, D. M. (2015). Missing R&D. *Journal of Accounting and Economics*, *60*(1), 73–94. https://doi.org/10.1016/j.jacceco.2015.03.004

Laux, V. (2012). Stock option vesting conditions, CEO turnover, and myopic investment. *Journal of Financial Economics*, *106*(3), 513–526. https://doi.org/10.1016/j.jfineco.2012.06.003

Liu, X.-Y., & Zhou, Z.-H. (2013). Ensemble Methods for Class Imbalance Learning. In *Imbalanced Learning* (pp. 61–82). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118646106.ch4

Liu, Y. (2014). Outside options and CEO turnover: The network effect. *Journal of Corporate Finance*, *28*, 201–217. https://doi.org/10.1016/j.jcorpfin.2014.03.004

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Mehran, H., & Yermack, D. (1997). *Compensation and Top Management Turnover* (SSRN Scholarly Paper No. 1297755). https://papers.ssrn.com/abstract=1297755

Moritz, B., & Zimmermann, T. (2016). *Tree-Based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns* (SSRN Scholarly Paper No. 2740751). https://doi.org/10.2139/ssrn.2740751

Murphy, K. J., & Zimmerman, J. L. (1993). Financial performance surrounding CEO turnover. *Journal of Accounting and Economics*, *16*(1), 273–315. https://doi.org/10.1016/0165-4101(93)90014-7

Parrino, R. (1997). CEO turnover and outside succession A cross-sectional analysis. *Journal of Financial Economics*, *46*(2), 165–197. https://doi.org/10.1016/S0304-405X(97)00028-7

Taylor, L. A. (2010). Why Are CEOs Rarely Fired? Evidence from Structural Estimation. *The Journal of Finance*, *65*(6), 2051–2087. https://doi.org/10.1111/j.1540-6261.2010.01610.x

Walker, D. I. (2009). *Evolving Executive Equity Compensation and the Limits of Optimal Contracting* (SSRN Scholarly Paper No. 1443170). https://doi.org/10.2139/ssrn.1443170

Weisbach, M. S. (1988). Outside directors and CEO turnover. *Journal of Financial Economics*, *20*, 431–460. https://doi.org/10.1016/0304-405X(88)90053-0

Zhou, Z.-H. (2021). Ensemble Learning. In Z.-H. Zhou (Ed.), *Machine Learning* (pp. 181–210). Springer. https://doi.org/10.1007/978-981-15-1967-3_8