Data Valuation in the Shadow of Agency Conflicts

Zexin Zheng *

May 28, 2025

Abstract

This paper proposes a novel offline deep reinforcement learning (DRL) framework to quantify agency conflicts at the firm-year level. The framework models management actions under compensation- versus market-cap-maximizing objectives. Actual actions relatively align more with DRL-predicted compensation-maximizing actions indicate higher agency conflicts. Firms with high agency conflicts tend to reduce investment and leverage while increase cash holdings, acquisitions, and net payouts. Among these actions, changes in cash holdings emerge as the strongest signal of agency conflicts. The DRL-derived measure of agency conflicts is significantly associated with corporate misconduct. When applied to a noisy rational expectations model, the framework reveals that investors disvalue management earnings guidance from high-agency-conflict firms, shifting reliance toward analyst forecasts.

Keywords: Deep reinforcement learning, agency conflicts, data valuation.
JEL Codes: G14, G3, D8, C52, C61.

*Hong Kong University of Science and Technology. Email: zzhengaq@connect.ust.hk.

> "If we are to assume that the desire for personal profit is the prime force motivating control, we must conclude that the interests of control are different from and often radically opposed to those of ownership; that the owners most emphatically will not be served by a profit seeking controlling group."
>
> — Adolf Berle and Gardiner Means, *The Modern Corporation and Private Property* (1932)

1 Introduction

Corporate agency conflicts (Jensen and Meckling, 1976), stemming from misaligned incentives between managers and outside shareholders driven by the pursuit of private benefits and the avoidance of costly effort, play a pivotal role in shaping corporate governance frameworks and financial decision-making processes. However, quantifying these conflicts remains challenging, since managers have strong incentives to conceal misbehavior and the observed governance mechanisms are endogenously determined (Nikolov and Whited, 2014).

This paper introduces a novel method to infer firm-year level agency conflicts using offline deep reinforcement learning (DRL). DRL aims to learn an optimal policy function mapping states to actions, maximizing predefined objectives. By setting the objective functions to maximize either personal compensation or market cap, DRL models output the compensation-maximizing and market-cap-maximizing management actions, respectively. Self-interested managers tend to make decisions aligning more closely with compensation-maximizing actions rather than market-cap-maximizing actions. Therefore, the degree of agency conflicts can be captured by comparing whether the actual observed management actions are closer to compensation-maximizing or market-cap-maximizing actions.

This paper examines five key management actions closely related to agency conflicts, including investment decision, cash holdings, acquisition activity, net payout policy, and leverage decision. Action variables represent direct managerial choices, while state variables capture environmental

characteristics not solely determined by management actions. This paper includes five firm-level characteristics and two macroeconomic variables to capture state features from various dimensions, such as profitability, external monitoring, and market-level economic and financial conditions. Managers, modeled as agents, learn to make actions to maximize discounted cumulative rewards based on current state variables. To quantify agency conflicts, this paper considers two types of cumulative rewards: the growth rate of market cap or personal compensation over a two-year period. Management actions affect not only immediate rewards but also the dynamics of state variables for subsequent periods, impacting long-term rewards.

For compensation-maximizing actions, DRL models suggest lowering investment and leverage while increasing cash holdings, acquisitions, and net payouts—aligning with prior research on agency conflicts, such as underinvestment due to unobserved costly effort (DeMarzo et al., 2012), cash hoarding for private benefits (Nikolov and Whited, 2014), and debt reduction to limit discipline (Jensen, 1986). Market-cap-maximizing actions differ significantly, recommending lower cash and payouts, with only leverage advice overlapping. Performance comparisons reveal that DRL-predicted compensation growth rates and market cap growth rates substantially outperform actual outcomes, with lower volatility in both cases.

The degree of agency conflict is quantified as the relative Euclidean distance between actual decisions and the two DRL-recommended action sets. A greater Euclidean distance from the market-cap-maximizing actions, relative to the compensation-maximizing actions, indicates higher agency conflict. This DRL-derived measure of agency conflicts is significantly associated with several common corporate governance proxies, including corporate fraud, earnings management, manager option backdating, opportunistic manager stock trading, and ownership structure. These results remain robust after excluding potential outliers and controlling for agency conflict estimates from previous studies.

Motivated by the fact that agency conflicts deter voluntary disclosure and undermine disclosure quality, this paper leverages the DRL-derived measure of agency conflicts to study whether the market discounts information disclosed by high-agency-conflict firms. Specifically, this paper

estimates the valuation of management earnings guidance, an important form of voluntary disclosure, across firms with varying levels of agency conflicts. The estimation is based on the data valuation model developed by Farboodi et al. (2025), which operates within a noisy rational expectations framework. In this model, mean-variance investors use public and private signals to make portfolio decisions. The model features rich heterogeneity in investors' wealth endowment, price impact effects, and investment styles. The dollar value of data is defined as the amount of risk-free return an investor would require to be indifferent between having the data or not, but receiving additional riskless wealth. In equilibrium, this dollar value of data can be pinned down by a few sufficient statistics, including average conditional and unconditional returns, variances, and forecast errors.

Across various levels of wealth endowment and price impact, the dollar value of management earnings guidance is significantly lower for firms with high agency conflicts compared to those with low conflicts. For instance, with perfect competition and an endowment wealth of $250 million, management earnings guidance is valued at approximately $1,765,800 for the low-agency-conflict portfolio, compared to $1,139,470 for the high-agency-conflict portfolio. These results hold true across different public information sources and investment restrictions, indicating that the market effectively accounts for the negative impact of agency conflicts in processing management disclosures. Additionally, investors tend to value analyst forecasts more when management guidance is less reliable.

This paper contributes to the literature on quantifying agency conflicts to the firm-year granular level by introducing a novel method based on offline DRL. Traditionally, agency conflicts have been quantified using structural models (e.g., Morellec et al., 2012, 2018; Nikolov and Whited, 2014), which require explicit parametric specification of state transition dynamics and reward functions. These models can become computationally intractable when handling multiple interacting management actions due to the curse of dimensionality in traditional estimation methods (e.g., Markov chain Monte Carlo sampling in high-dimensional parameter spaces). This forces existing studies to use simplified models with 1-2 key actions despite the multi-dimensional nature

of management actions. The DRL-derived approach offers several advantages. On one hand, offline DRL provides a scalable framework for estimating agency conflicts in high-dimensional spaces. By utilizing neural networks, DRL implicitly models state transitions through temporal difference learning on historical data, eliminating the need for explicit transition specification. Additionally, the use of adaptive, momentum-based stochastic optimization methods—such as the Adam optimizer—enables efficient training in large parameter spaces, particularly when combined with GPU acceleration. This enables the joint modeling of multiple management actions and their interactions within a unified framework. On the other hand, offline DRL realistically models the dynamic interactions between state and action variables. Dynamic structural models typically focus on how current actions directly affect future rewards, potentially overlooking the intermediary role of future state variables. Offline DRL, instead, explicitly incorporates the feedback effect (Bond et al., 2012; Edmans et al., 2012), which considers how current actions influence future state variables. These future states, along with subsequent actions, then affect future rewards. This comprehensive approach enables offline DRL to better capture the complexity of dynamic agency conflicts, offering a more realistic representation of decision-making processes.

Second, this paper contributes to the growing literature on reinforcement learning (RL) in finance and economics. Existing work in this field can be categorized into two paradigms: online RL, which relies on simulation-based experiments, and offline RL, which is data-driven. For online RL, prior studies demonstrate that algorithmic pricing agents systematically learn collusive strategies in simulated environments. This phenomenon has been documented in product markets (Calvano et al., 2020, 2021; Johnson et al., 2023) and financial markets (Dou et al., 2024). In offline RL applications, recent advances focus on data-driven decision-making. For instance, Cong et al. (2021) employs transformer-based DRL to construct stock portfolios, achieving exceptional risk-adjusted returns. The most closely related study is Campello et al. (2023), which develops a model-based offline DRL framework to optimize management actions for market cap growth. This study extends the literature by modeling incentive misalignment between managers and shareholders, a critical real-world friction absent in prior DRL-derived decision-making frameworks.

Finally, this paper is closely related to the literature on informational efficiency, building on the noisy rational expectations framework. Prior studies, such as those by Bai et al. (2016), Dávila and Parlatore (2025), and Kacperczyk et al. (2025), focus on the *quantity* of information by analyzing the precision of signals about future payoffs embedded in asset prices. In contrast, Farboodi et al. (2025) shifts the focus to the *valuation* of information, demonstrating that a few sufficient statistics derived from the same framework can fully characterize the dollar value of data (or information). This paper extends their model by incorporating heterogeneity in agency conflicts, providing a more nuanced understanding of how divergent manager-shareholder incentive misalignment influence the market value of disclosed information.

The rest of the paper is organized as follows. Section 2 explains applying the DRL framework to management decision-making to estimate firm-year agency conflicts. Section 3 reports the DRL estimation results, validates the DRL-derived measure of agency conflicts, and conducts policy experiments. Section 4 leverages the DRL-derived measure to study heterogeneity in data valuation across varying levels of agency conflicts. Section 5 concludes.

2 Application of Deep Reinforcement Learning (DRL)

2.1 DRL Preliminaries

Reinforcement learning (RL) addresses the problem of learning to control a dynamical system, which is represented as a Markov Decision Process (MDP). An MDP is a tuple $\mathcal{M} := (\mathcal{S}, \mathcal{A}, T, r, \gamma)$, where $\mathcal{S}$ is a set of states $s \in \mathcal{S}$, $\mathcal{A}$ is a set of actions $a \in \mathcal{A}$, T defines the transition dynamics of the subsequent state s' conditional on the current state and action: $T(s_{t+1} \mid s_t, a_t)$, $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ defines a reward function, and $\gamma \in (0, 1]$ is a scalar discount factor.

The objective in RL is to learn a policy π, which maps states to actions. The action-value function Q^π represents the expected cumulative reward from a state-action tuple (s_t, a_t), following the policy π : $Q^\pi(s, a) = \mathcal{B}^\pi Q^\pi(s, a)$, where $\mathcal{B}^\pi$ is a Bellman operator($\mathcal{B}^\pi Q^\pi(s, a) := r(s, a) + \gamma \mathbb{E}[Q^\pi(s', a')]$). An optimal policy π^* maximizes $Q^{\pi^*}(s, a)$.

In continuous state-action spaces, actor-critic methods are a widely used approach for learning near-optimal policies. These methods consist of two components: a parametrized policy π_θ (actor network) and a parametrized action-value function Q_ω (critic network). The critic network updates Q_ω to minimize temporal difference error:

$$\min_{Q_\omega} \mathbb{E}\left[Q_\omega^{\pi_\theta}(s,a) - \mathcal{B}^{\pi_\theta} Q_\omega^{\pi_\theta}(s,a)\right]^2. \tag{1}$$

The actor network updates π_θ to maximize the action-value function:

$$\max_{\pi_\theta} \mathbb{E}\left[Q_\omega^{\pi_\theta}(s,a)\right]. \tag{2}$$

The parameters θ and ω are typically estimated within deep neural networks, enabling the approximation of high-dimensional policies and value functions. The term "deep" in DRL reflects the use of multilayer neural architectures.

In traditional online DRL, transitions (s, a, s', r) are sampled through interactions with the environment, enabling updates to π_θ and Q_ω as more transitions are gathered. However, online interaction can be impractical due to high costs (e.g., firm decision making as explored in this paper, healthcare) or risks (e.g., autonomous driving, healthcare). Offline DRL uses a static dataset of N pre-collected transitions: $\mathcal{D} = \left\{\left(s_i, a_i, r_i, s_i'\right)\right\}_{1:N}$, denoted as $(s, a, r, s') \sim \mathcal{D}$ for simplicity.

Offline DRL based on classical actor-critic methods, as shown in Equations (1) and (2), can suffer from distributional shift—discrepancies between the training data distribution $\mathcal{D}$ and the distribution during testing. This shift can lead to inaccurately high Q-values for out-of-distribution (OOD) actions.[1] The CQL algorithm, developed by Kumar et al. (2020), addresses this by regularizing the action-value function to prevent overestimation for OOD actions. The Cal-QL algorithm by Nakamoto et al. (2023) enhances CQL by modifying the regularizer. Specifically, the Cal-QL

[1] Online DRL can correct errors by executing actions in the environment and observing their true values. However, offline DRL lacks the ability to interact with the environment.

critic network is given by:

$$\min_{Q_\omega} \alpha \underbrace{\left(\underbrace{\mathbb{E}_{s\sim\mathcal{D},a\sim\pi_\theta}\left[\max\left(Q^{\pi_\theta}_\omega(s,a), V^\mu(s)\right)\right]}_{\text{minimize Q-value}} - \underbrace{\mathbb{E}_{s,a\sim\mathcal{D}}\left[Q^{\pi_\theta}_\omega(s,a)\right]}_{\text{maximize Q-value under data}} \right)}_{\text{Cal-QL regularizer}} + \frac{1}{2} \underbrace{\mathbb{E}_{s,a,s'\sim\mathcal{D}}\left[\left(Q^{\pi_\theta}_\omega(s,a) - \mathcal{B}^\pi_\theta \bar{Q}^{\pi_\theta}_\omega(s,a)\right)^2\right]}_{\text{Standard Bellman TD error}}, \tag{3}$$

where α is the penalty weight, and $V^\mu(s)$ is the Q-value of a reference policy μ.[2] The first component of the Cal-QL regularizer aims to minimize Q-values across states, lower-bounded by the Q-value of a sub-optimal reference policy. The second component maximizes the Q-values for state-action tuples that are within the data distribution. Together, the Cal-QL regularizer pushes down on Q-values for OOD actions, addressing potential overestimation errors. The second term is the standard Bellman temporal difference error.

2.2 Empirical Design

This paper applies the Cal-QL algorithm to analyze management decision-making processes. Each year, managers take actions (a) based on current state variables (s). These actions influence not only the immediate reward ($r(s,a)$), but also the dynamics of state variables for the subsequent period (s'), thereby affecting the long-term reward ($\mathbb{E}\left[Q^\pi(s',a')\right]$). Implementation details of the Cal-QL algorithm, the hyperparameter settings, and the network architectures are provided in Appendix A.

Action Variables ($\mathcal{A}$) This paper includes five key management action variables: Firstly, $\Delta Invest$ represents the change in the investment ratio, calculated as the sum of R&D expenses,

[2]Various reference policies can be selected. This study adopts the behavior policy, which is the policy used for data collection, as the reference policy, following Nakamoto et al. (2023).

30% of SG&A expenses,[3] and capital expenditures, all scaled by total assets. Secondly, $\Delta CashHold$ denotes the change in the cash holding ratio, calculated as cash holdings scaled by total assets. Thirdly, $\Delta Acquisition$ measures the change in acquisition expenditure, calculated as acquisition-related costs scaled by total assets. Fourthly, $\Delta NetPayout$ reflects the change in the net payout ratio, calculated as the sum of dividends and net repurchases, scaled by total assets. Finally, $\Delta Leverage$ indicates the change in the leverage ratio, calculated as liabilities divided by total assets.

State Variables (S) While action variables represent the direct choices made by managers, state variables capture environmental characteristics that are at least not solely determined by management actions. This paper examines five firm-level characteristics alongside two macroeconomic variables. The firm-level characteristics are defined as follows: Operating Return on Assets ($OROA$) is calculated by dividing operating income before depreciation by total assets, reflecting a firm's operational profitability. Return on Assets (ROA) is calculated as net income over total assets, indicating overall profitability. Gross Profit Margin (GPM) is calculated by subtracting the cost of goods sold from net sales and then dividing by net sales, serving as a proxy for a firm's pricing power. Institutional Ownership ($InstOwn$) is the annual average of quarterly total institutional holdings divided by the quarter-end market cap, capturing the effectiveness of external monitoring by institutional investors. Lastly, the Amihud Illiquidity Measure ($Amihud$) is computed as the logarithm of the annual average of the daily ratio of absolute stock return to its dollar volume, reflecting the cost of trading the firm's stock. In addition, this paper gauges market-level economic and financial conditions using real GDP growth ($GGDP$) and the National Financial Conditions Index ($NFCI$), as per Adrian et al. (2019). $GGDP$, calculated by the Bureau of Economic Analysis, indicates economic expansion or contraction. $NFCI$ offers a weekly measure of U.S. financial conditions across money markets, debt and equity markets, and both traditional and shadow banking systems. A positive $NFCI$ value indicates tighter-than-average financial

[3]Only a small proportion of SG&A is related to investment in intangible organization capital, while the rest of SG&A is related to operating costs that support the current period's profits. The 30% is a rule of thumb used in prior studies (e.g., Eisfeldt and Papanikolaou, 2014).

conditions. The weekly $NFCI$ values are averaged to the annual level in this paper. Furthermore, this paper incorporates the corresponding variables lagged by one year to capture time-dependent patterns. Together, the state space comprises seven concurrent state variables and seven lagged state variables.

Rewards (r) Agency conflicts often arise when managers prioritize their own compensation over the maximization of shareholder value. This misalignment can lead managers to make decisions that benefit themselves at the expense of shareholders. This paper introduces two objectives, inspired by the definition of agency conflicts, to explore this issue. The first objective is to maximize the growth rate of market cap ($\Delta MCap$) over a two-year period. This is represented by the expected value $\mathbb{E}\left[r(s,a) + \gamma r(s',a')\right]$, where $r(s,a)$ and $r(s',a')$ denote the $\Delta MCap$ for the current and subsequent year, respectively. The scalar discount factor, γ, is set at 0.95.[4] A two-year optimization window is chosen to balance the trade-off between management myopia associated with shorter windows and the increased noise from longer windows. The outcome of this maximization is a policy for maximizing market cap, $\pi^{MCap} : \mathcal{S} \rightarrow \mathcal{A}^{MCap}$, which provides a set of optimal values for five action variables ($a^{MCap} \in \mathcal{A}^{MCap}$) based on fourteen state variables for each firm-year. The second objective is to maximize the manager's personal compensation growth rates ($\Delta Compens$) over the same two-year period.[5] Compensation includes all potential personal income sources such as salary, bonus, long-term incentive plan payouts, the value of restricted stock grants, options granted during the year, and any other annual pay received by the manager. The outcome of this maximization is another policy, $\pi^{Compens} : \mathcal{S} \rightarrow \mathcal{A}^{Compens}$, which yields a set of optimal actions ($a^{Compens} \in \mathcal{A}^{Compens}$) to maximize manager's personal compensation for each firm-year. To ensure robustness and mitigate the impact of random initializations in the reinforcement learning process, both π^{MCap} and $\pi^{Compens}$ are estimated multiple times using different random seeds. Specifically, three distinct random seeds are employed to generate three separate estimations for each policy. The final policy is obtained by averaging these estimations,

[4]Results remain robust across alternative values.

[5]In the event that a manager is newly appointed in year t, $\Delta Compens_t$ is assigned a value of 0. Conversely, if a manager departs in year $t+1$, $\Delta Compens_{t+1}$ is assigned a value of -1.

thereby reducing the likelihood that the learned policy is merely a result of favorable initial conditions. This approach enhances the reliability and generalizability of the learned policies.

Agency Conflicts and RL Policy (π) In firms where agency conflicts are severe, management actions tend to align more closely with actions that maximize personal compensation ($\mathcal{A}^{Compens}$) rather than those that maximize market cap ($\mathcal{A}^{Compens}$). This paper quantitatively measures the degree of agency conflicts using the Euclidean distance of observed management actions ($\mathcal{A}^{Actual}$) to these two sets of DRL-recommended actions. Specifically, the agency conflict for firm i in year t is defined as:

$$Agency\ Conflict_{i,t} = \text{Euclidean}\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{MCap}\right) - \text{Euclidean}\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{Compens}\right), \quad (4)$$

where $\text{Euclidean}\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{MCap}\right) = \sqrt{\sum_{j=1}^{5}\left(a_{j,i,t}^{Actual} - a_{j,i,t}^{MCap}\right)^2}$ and $\text{Euclidean}\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{Compens}\right) = \sqrt{\sum_{j=1}^{5}\left(a_{j,i,t}^{Actual} - a_{j,i,t}^{Compens}\right)^2}$, with j indicating the index of the five action variables. A greater Euclidean distance from the market-cap-maximizing actions, relative to the compensation-maximizing actions, indicates higher agency conflicts.

2.3 Sample Construction

This study combines several data sources to construct a sample for estimating agency conflict under the DRL framework. Financial statement data is obtained from Compustat, while stock price information is sourced from the Center for Research in Security Prices (CRSP). Institutional ownership data is derived from the LSEG (formerly Thomson Reuters) 13F database. Macroeconomic data is sourced from the Federal Reserve Economic Data (FRED) for real GDP growth ($GGDP$) and from the Federal Reserve Bank of Chicago for the National Financial Conditions Index ($NFCI$).[6] Managerial compensation data is initially gathered from the ExecuComp database, which provides detailed information on the compensation of top executives within firms comprising the S&P 500,

[6]The $GGDP$ index is available at https://fred.stlouisfed.org/series/A191RL1Q225SBEA, and the $NFCI$ index is available at https://www.chicagofed.org/publications/nfci/index.

MidCap 400, and SmallCap 600 indices. This dataset is further complemented by the Capital IQ People Intelligence database, which extends the coverage to include a broader range of U.S. public and private firms.

The combined sample includes U.S. listed companies with common stocks traded on the NYSE, NASDAQ, and AMEX, covering an annual frequency from 1993 to 2023.[7] The 1993-2018 subsample is used for training the DRL model, while the 2019-2023 subsample is used for testing. Table 1 reports the summary statistics on the mean, standard deviation, and distribution of the key variables. To conserve space, lagged state variables are not included. A comprehensive list of variable definitions is provided in Table B.1. All continuous variables are winsorized at the 2.5% and 97.5% levels to mitigate the influence of outliers.

3 DRL Estimation Result

3.1 Baseline Results

Figure 1 compares the actual actions taken by managers with those recommended by offline DRL models. Panel (a) illustrates the differences between the median values of actual management actions and the compensation-maximizing actions, along with the 95% confidence interval. Five significant patterns emerge:

1. Investment: Managers seeking to maximize personal compensation are advised to lower the investment ratio. This finding aligns with prior research (Panousi and Papanikolaou, 2012; Biais et al., 2010; Lambrecht and Myers, 2012). For instance, DeMarzo et al. (2012) present a dynamic model showing that the investment ratio falls below the first-best level due to unobserved costly effort, with the degree of underinvestment depending on the incentive contract design.

2. Cash Holdings: Managers are encouraged to increase cash holdings to maximize personal

[7]The sample begins in 1993 due to the availability of managerial compensation data, and concludes in 2023 since an additional year of data (2024) is needed to construct the two-year optimization objectives.

compensation, consistent with previous studies (e.g., Kalcheva and Lins, 2007; Dittmar et al., 2003). Nikolov and Whited (2014) demonstrate that high cash holdings can be explained by managers' private benefits from resource diversion, alongside precautionary motives. Gao et al. (2013) provide empirical evidence that high cash holdings in U.S. public firms compared to the private counterpart are due to high agency costs.

3. Acquisitions: Managers are advised to pursue more acquisitions to maximize compensation, in line with past research (Harford, 1999; Harford et al., 2008). Gormley and Matsa (2016) provide causal evidence that managers undertake more value-destroying acquisitions when insulated by antitakeover laws.

4. Net Payout Ratio: Increasing the net payout ratio is recommended for maximizing compensation. This supports Lambrecht and Myers (2012)'s theoretical prediction that rents and net payouts move in tandem.

5. Leverage: Managers are advised to reduce leverage to maximize personal compensation. This aligns with studies on the disciplining effect of debt (Jensen, 1986; Zwiebel, 1996; Morellec, 2004), which suggest that debt limits managerial discretion by reducing free cash flow. Morellec et al. (2012) find that agency costs of 1.5% of equity value can resolve the low-leverage puzzle. Morellec et al. (2018) extend this model to back out agency conflict based on observed leverage decisions in the international setting.

Together, the compensation-maximizing actions estimated from the DRL model resonate with previous theoretical, empirical, and structural studies.[8]

Panel (b) of Figure 1 compares the actual management actions and those aimed at maximizing market cap. A preliminary review reveals that its pattern differs from that of maximizing compensation. Specifically, while managers are encouraged to increase cash holdings and net

[8]It's worth mentioning that some studies offer differing perspectives. For instance, Albuquerue and Wang (2008) suggest that agency conflicts can lead to empire building, rather than underinvestment, due to agents' limited liability; Bates et al. (2009) propose that the rising cash holdings are due to a riskier operating environment rather than agency conflicts.

payout to maximize personal compensation, they are suggested to lower these to maximize market cap. Although the market-cap-maximizing policy also suggests a lower investment ratio and higher acquisition activities, the magnitude is smaller compared to the compensation-maximizing policy. The leverage decision is the only action that receives similar recommendations from both policies. The distinct patterns between Panels (a) and (b) alleviate concerns that actions to maximize compensation might align with those to maximize market cap, potentially biasing the measurement of agency conflicts. Figures B.1 and B.2 provide a more detailed time-series comparison between actual actions and DRL-recommended actions.

Figure 2 illustrates a comparison between actual corporate performance and expected performance based on actions recommended by offline DRL models. Panel (a) presents the median values of actual cumulative growth rates of compensation ($\Delta Compens_t + 0.95\Delta Compens_{t+1}$) against their expected counterparts ($\mathbb{E}[\Delta Compens_t + 0.95\Delta Compens_{t+1}]$) when taking compensation-maximizing actions. The actual median growth rates vary from -0.38% in 2008 to 62.82% in 1996, whereas the expected growth rates predicted by the offline DRL framework is significantly higher, ranging from 71.01% to 94.47%. On average, the expected growth rates outperform the actual growth rates by $65.32\%(= 81.48\% - 16.15\%)$. Additionally, the expected growth rates enjoy lower volatility (0.055) compared to the actual growth rates (0.075). Panel (b) compares the median values of actual cumulative growth rates of market cap ($\Delta MCap_t + 0.95\Delta MCap_{t+1}$) with their expected counterparts ($\mathbb{E}[\Delta MCap_t + 0.95\Delta MCap_{t+1}]$) when taking market-cap-maximizing actions. The actual median growth rates fluctuate between -38.56% in 2008 and 62.82% in 2003, while the expected growth rates from the offline DRL framework ranged from 45.39% to 121.04%. On average, the expected growth rates outperform the actual growth rates by $50.41\%(= 69.70\% - 19.29\%)$, and enjoy lower volatility (0.159 VS 0.205). Although the market cap outperformance is slightly less pronounced than the compensation case, it remains substantial, with the 25th percentile surpassing the actual median growth rate in all periods except 2020. This effect is comparable to findings by Campello et al. (2023), which utilizes another DRL algorithm and demonstrates an average quarterly outperformance of 8.73%, translating to an annual outperformance of 34.92%.

3.2 Validation Tests and Robustness Checks

Building on the DRL-recommended actions and actual management actions outlined in the preceding subsection, this study develops the $AgencyConflict$ measure using Equation (4). Table B.2 presents summary statistics for $AgencyConflict$, including its mean, standard deviation, and distribution. The measure exhibits substantial cross-sectional and time-series variation. Cross-sectionally, the standard deviation ranges from 0.013 to 0.021 across periods, an order of magnitude greater than the mean values. This pronounced heterogeneity aligns with prior research documenting divergent agency conflict patterns across firms (Morellec et al., 2012). Temporally, the mean $AgencyConflict$ fluctuates from −0.010 in 1993 to 0.001 in 2020, underscoring its sensitivity to evolving firm- and macroeconomic-level conditions. These patterns highlight the advantage of granular firm-year estimation over traditional firm-level or market-level approaches, as the finer estimation captures richer informational content about agency frictions. To establish the validity and reliability of the DRL-derived measure, a series of validation tests and robustness checks are conducted, as outlined below.

3.2.1 Corporate Fraud

Corporate fraud includes "false representations, dishonesty and deceit."[9] It is well documented that managers with higher agency conflicts are more likely to engage in accounting fraud (see, e.g., Kumar and Langberg, 2009; Efendi et al., 2007).

This study employs two widely recognized measures to capture corporate fraud. The first measure is financial restatements, collected from the Audit Analytics Database, which records restatements from 1996 onward. Consistent with prior research, two filters are applied to distinguish material financial restatements from unintentional errors: (1) restatements arising from clerical errors are excluded (identified by the Audit Analytics "RES_CLER_ERR" flag) (Wang, 2022); (2) only restatements disclosed via Form 8-K Item 4.02 are included (Cahan et al., 2024), as these are deemed more material due to SEC regulatory requirements (2004, Item 4.02). Following Ali and

[9] See United States v. Grainger, 701 F.2d 308, 311 (4th Cir. 1983), cert. denied, 461 U.S. 947 (1983).

Hirshleifer (2017), the constructed restatement dummy variable ($Restate$) is assigned a value of 1 for fiscal periods with incorrect filings, rather than the period when the restatement is issued.

The second measure is securities class action lawsuit filings, compiled by the Stanford Securities Class Action Clearinghouse (SCAC).[10] The SCAC database includes nearly all securities class action lawsuits filed in U.S. federal courts since the enactment of the Private Securities Litigation Reform Act in 1996. Following previous studies (e.g., Kempf and Spalt, 2023), cases related to initial public offering underwriter allocation, analyst coverage, and mutual funds, rather than firm management, are excluded. The remaining cases predominantly involve financial misrepresentation, accounting manipulations, and insider trading. $AllSuit$ is a dummy variable set to one if a firm is involved in securities class action lawsuits in a given year and zero otherwise. In constructing a more stringent and material non-dismissal lawsuit measure ($NDSuit$), cases that are subsequently dismissed are further excluded, as per Dyck et al. (2010).

In the sample analyzed, 3,950 material restatements, 1,146 lawsuits, and 617 non-dismissed lawsuits are identified, accounting for 5.12%, 1.54%, and 0.83% of firm-year observations, respectively. Table 2 illustrates a positive relationship between agency conflict and the likelihood of corporate fraud. Column (1) uses financial restatement ($Restate$) as a proxy for corporate fraud. The coefficient is significantly positive at 0.108 ($t = 2.47$), suggesting that a one standard deviation increase in agency conflict is associated with a 3.54% increase in the probability of financial restatement. This effect is comparable to other governance-related characteristics, such as insider trading intensity, which is associated with a 2.7% increase in restatement probability, as documented by Ali and Hirshleifer (2017). Columns (2) and (3) employ securities class action lawsuits ($AllSuit$ and $NDSuit$) as proxies for corporate fraud. The coefficients on $AgencyConflict$ remain statistically and economically significant. The coefficient of 0.069 implies that a one standard deviation increase in $AgencyConflict$ is associated with a 13.97% increase in the probability of a non-dismissed lawsuit.

[10]The SCAC lawsuit filings are available at https://securities.stanford.edu/filings.html.

3.2.2 Earnings Management

Earnings management involves managers using judgment in financial reporting and structuring transactions to alter financial reports. This practice can mislead stakeholders regarding the company's true economic performance (Ronen, 2008). Managers with high agency conflicts may resort to earnings management to meet targets, secure bonuses, or manipulate stock prices for personal gain, prioritizing their interests over those of shareholders.

Earnings management in this study is measured by the absolute value of discretionary accruals (DA), calculated using the Jones (1991), Dechow et al. (1995), and Kothari et al. (2005) models. For the Jones (1991) model, discretionary accruals are estimated using cross-sectional regressions for each industry, classified by two-digit SIC codes with at least 10 observations, as follows:

$$\frac{TA_{i,t}}{Asset_{i,t-1}} = \alpha_1 \frac{1}{Asset_{i,t-1}} + \alpha_2 \frac{\Delta Sale_{i,t}}{Asset_{i,t-1}} + \alpha_3 \frac{PPE_{i,t}}{Asset_{i,t-1}} + \epsilon_{i,t} \tag{5}$$

$Assets_{i,t-1}$ represents total assets of firm i at time $t-1$, $\Delta Sale_{i,t}$ is the change in revenues from the previous year, and $PPE_{i,t}$ is the gross value of property, plant, and equipment of firm i at time t. $TA_{i,t}$ denotes the total accruals of firm i at time t, calculated as $TA_{i,t} = IB_{i,t} - OCF_{i,t}$, where $IB_{i,t}$ is income before extraordinary items and $OCF_{i,t}$ is operating cash flows. The estimated coefficients from Equation (5) are used to derive the firm's normal accruals ($NA_{i,t}$):

$$NA_{i,t}^{Jones} = \hat{\alpha}_1 \frac{1}{Asset_{i,t-1}} + \hat{\alpha}_2 \frac{\Delta Sale_{i,t}}{Asset_{i,t-1}} + \hat{\alpha}_3 \frac{PPE_{i,t}}{Asset_{i,t-1}}$$

Discretionary accruals are then calculated as the absolute difference between total accruals and fitted normal accruals: $DA_{i,t}^{Jones} = \left| \frac{TA_{i,t}}{Asset_{i-1}} - NA_{i,t}^{Jones} \right|$.[11] The Dechow et al. (1995) model modifies the Jones model by adjusting the change in revenues for changes in receivables:

$$NA_{i,t}^{Dechow} = \hat{\alpha}_1 \frac{1}{Asset_{i,t-1}} + \hat{\alpha}_2 \frac{\Delta Sale_{i,t} - \Delta Rec_{i,t}}{Asset_{i,t-1}} + \hat{\alpha}_3 \frac{PPE_{i,t}}{Asset_{i,t-1}}$$

[11] Unsigned accruals are commonly used in the earnings management literature (e.g., Hazarika et al., 2012) since earnings cannot be consistently managed in a single direction.

where $\Delta Rec_{i,t}$ is the change in net receivables. Discretionary accruals are adjusted accordingly: $DA_{i,t}^{Dechow} = \left| \frac{TA_{i,t}}{Asset_{i-1}} - NA_{i,t}^{Dechow} \right|$. This adjustment is based on the premise that managing earnings through credit sales is easier than through cash sales. The Kothari et al. (2005) model controls for performance effects by matching firms on ROA and adjusting discretionary accruals: $DA_{i,t}^{Kothari} = \left| DA_{i,t}^{Jones} - DA_{k,t}^{Jones} \right|$, where $DA_{k,t}^{Jones}$ is the discretionary accrual of a performance-matched counterpart with the closest ROA in the same industry.

This paper uses firm-level financial data from Compustat to calculate discretionary accruals. Higher discretionary accruals indicate a greater likelihood of earnings manipulation. The final regression sample consists of 63,955 firm-years during 1993-2023. Table 3 presents the regression of earnings management on $AgencyConflict$, controlling for firm and year fixed effects. The coefficients of interest are significantly positive across different models. Specifically, in Column (3), the coefficient is 0.293 ($t = 6.43$), with discretionary accruals estimated using the Kothari et al. (2005) model. Economically, a one standard deviation increase in $AgencyConflict$ is associated with a 4% increase in earnings management.

3.2.3 Manager Option Backdating

Option grant backdating involves retroactively setting the grant date of at-the-money stock options to a period when the stock price is lower. This practice allows managers to purchase stock at a reduced price, potentially selling it at a higher market value, thereby increasing the profitability of these options. According to Bebchuk et al. (2010), such opportunistic timing practices are indicative of governance breakdowns and raise significant governance concerns. Managers experiencing higher agency conflicts are more inclined to manipulate stock option grant dates to enhance their personal compensation, often to the detriment of shareholder interests.

To investigate this issue, this paper constructs a dataset of option grants awarded to managers from 1993 to 2023, using the LSEG Table 2 Derivative Transactions database (formerly Thomson Reuters). Records with cleanse codes R ("data verified through the cleansing process") or H ("cleansed with a very high level of confidence") are included. The price data is sourced from

the CRSP database. Data processing procedures are akin to those utilized by Bebchuk et al. (2010). Specifically, grants that are considered scheduled are excluded, as they are less likely to be opportunistically timed. A grant is deemed scheduled if another grant is awarded on the same date within a ±2-day window in the previous year. Additionally, only grants where the strike price is within 1% of the closing price on the grant date or one day before or after are included. Managers are identified as individuals listed as CEO or president in the LSEG database. A backdating event is defined as occurring when the strike price is no more than 101% of the lowest daily price within the grant month. To capture backdating activities, two metrics are employed: an insider backdating dummy ($BDDummy$), which equals one if at least one backdating event occurred in a firm in a given year, and the backdating fraction ($BDFraction$), calculated as the number of backdating events divided by the total manager option grant events within a firm-year. The final sample spans 14,423 firm-year observations from 1993 to 2023, with 16.14% of these observations flagged for backdating.

Table 4 reveals a positive relationship between $AgencyConflict$ and the likelihood of manager option backdating. In Column (1), the coefficient of 0.462 ($t = 2.26$) indicates that a one standard deviation increase in agency conflict is associated with a 4.8% increase in the probability of backdating, measured at the mean backdating of 16.14%. This effect is comparable to the impact of other governance factors, such as board independence, which prior research by Collins et al. (2009) found to increase backdating probability by 1.5% for a one standard deviation increase in the measure.

3.2.4 Opportunistic Manager Trading in Stock Markets

Opportunistic insider trading on non-public information undermines the integrity of financial markets and is prohibited by SEC Rule 10b5-1. Managers with greater agency conflicts are more likely to exploit non-public information for personal gain.

To identify manager trades likely based on non-public information, this paper uses the price pattern ratio developed by Rozanov (2008). This ratio is calculated as the proportion of post-

transaction buy-and-hold returns to pre-transaction returns:

$$PricePattern = \frac{\prod_{k=-\tau}^{-1}(1+R_{i,k})}{\prod_{k=1}^{\tau}(1+R_{i,k})},$$

where τ is the estimation window. Two windows are considered: $[-10,+10]$ and $[-20,+20]$, with $k = 0$ as the manager trading date. $R_{i,k}$ is either the raw or market-adjusted daily return of firm i. For managers with multiple trades in a year, $PricePattern$ is averaged for each firm-year observation, following Liu (2016).

This paper obtains insider stock transaction data from the LSEG Table 1 Stock Transactions database (formerly Thomson Reuters), focusing on records with cleanse codes R ("data verified through the cleansing process") or H ("cleansed with a very high level of confidence"). An insider is defined as a manager if he or she is identified in the LSEG database either as a CEO or the president of the company (role code CEO or P). This paper focuses on manager purchases, as research (e.g., Ravina and Sapienza, 2010) shows executives achieve positive abnormal returns on purchases but not on sales. The final sample consists of 9,037 firm-year observations in 1993-2023.

Table 5 presents the regression of price pattern ratios against $AgencyConflict$, controlling for firm and year fixed effects. The coefficients on $AgencyConflict$ are significantly positive across various estimation windows and return patterns. For instance, the coefficient is 1.093 ($t = 3.95$) when using raw returns to calculate the price pattern in a $[-20,+20]$ window, as shown in Column (2). This suggests that a one standard deviation increase in a firm's agency conflict corresponds to a 1.6% increase in the price pattern measure. The effect is economically significant compared to the impact of other governance-related factors, such as corporate corruption culture. According to estimates by Liu (2016), a one standard deviation increase in corruption culture is linked to a 0.86% increase in the likelihood of opportunistic insider trading.

3.2.5 Ownership Structure

The ownership structure, a traditional proxy for corporate governance, is intricately linked to agency conflicts. For instance, Aghion et al. (2013) demonstrate that high institutional ownership mitigates agency conflicts by shielding managers from career-related concerns. Alternatively, Hartzell and Starks (2003) argue that institutional investors play a critical monitoring role in alleviating agency issues. Besides institutional ownership, blockholder ownership is also pivotal in curbing agency problems, either through direct intervention (Shleifer and Vishny, 1986) or the threat of exit (Edmans, 2009).

Building on these studies, this paper develops two governance measures based on ownership structure. The first measure, institutional ownership ($InstOwn$), is calculated as the annual average of quarterly total institutional holdings divided by the quarter-end market cap. Data on institutional holdings is sourced from the LSEG 13F database. The second measure is outside blockholder ownership ($BlockOwn$), defined as the proportion of stock held by outside blockholders. Blockholder data is compiled by Schwartz-Ziv and Volkova (2025) and accessible via WRDS, covering the period from 1998 to 2023 on an annual basis. To avoid the confounding effects from inside blockholdings, this paper focuses on the blockholdings by financial institutions (identified by the "block_type" flag).[12] A higher level of institutional or outside blockholder ownership suggests enhanced governance and consequently, reduced agency conflicts. Table 6 presents the regression results. The coefficients are significantly negative across various measures, supporting the hypothesis that an improved ownership structure mitigates agency conflicts.

3.2.6 Robustness Checks

The first robustness check addresses the potential impact of outliers. The offline DRL framework assumes that management objectives are either maximizing market cap or personal compensation.

[12] Schwartz-Ziv and Volkova (2025) categorizes blockholders into three types: institutional, individual, and other. Individual blockholders include both outsiders and insiders, which cannot be differentiated in the dataset. The "other" category primarily consists of product market suppliers/customers, which are often established to address inefficiencies from incomplete contracting and financial market frictions (Fee et al., 2006), and less related to corporate governance issues.

While these objectives are common and well-documented in the literature, there may be outliers where managers prioritize alternative objectives, such as stakeholder interests. To ensure the significant association between the DRL-derived measure of agency conflicts and corporate governance proxies is not driven by such outliers, this paper excludes observations where actual management actions deviate substantially from DRL-recommended actions. Specifically, firm-year observations with both Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{MCap}\right)$ and Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{Compens}\right)$ outside two standard deviations of the sample mean are excluded. This process eliminates 5.38% of observations (4,156 out of 77,186). As shown in Table 7, the five validation tests remain consistent after excluding outliers, with the relationship between $Agency\ Conflict$ and corporate governance proxies retaining statistically and economically significant.

The second robustness check compares $Agency\ Conflict$ with the structural agency conflict index ($\hat{\phi}$) proposed by Morellec et al. (2018, hereafter MNS). MNS derive agency conflicts from leverage choice distortions in a dynamic capital structure model, developing a firm-level index that varies across firms and countries. This part re-examines the association between $AgencyConflict$ and different corporate governance proxies while controlling for $\hat{\phi}$, leveraging MNS's publicly available data.[13] Table 8 presents results from this combined sample, which is smaller in sample size due to differences in sample periods and filters between this study and MNS. Year fixed effects are excluded because $\hat{\phi}$ is time-invariant. For regressions using opportunistic price patterns as a governance proxy, firm fixed effects are omitted to avoid subsuming $\hat{\phi}$ in small-sample estimations. The significant relationship between $AgencyConflict$ and governance proxies persists even after accounting for $\hat{\phi}$, reinforcing the robustness of the findings.

3.3 Policy Experiments

The preceding subsection demonstrates that $AgencyConflict$, constructed using all five management actions, significantly associated with corporate governance proxies, affirming its efficacy in capturing agency conflict severity. To identify which individual management action most strongly

[13] I sincerely thank the authors for providing these estimates at https://www.epfl.ch/labs/sfi-em/.

signals agency conflicts, this subsection conducts counterfactual experiments. Specifically, each of the five management actions is sequentially excluded from the $AgencyConflict$ calculation (Equation 4). For instance, when muting the investment decision ($\Delta Invest$), its compensation- and market-cap-maximizing values are set equal to its realized action. This adjustment renders the counterfactual Euclidean distance independent of the investment dimension, relying solely on the remaining four actions. If the muted action strongly signals agency conflicts, the validation results using the counterfactual measure should exhibit reduced statistical significance.

Table 9 presents the findings. The first row replicates the baseline validation results from Tables 2 to 6 for comparison, where coefficients of $AgencyConflict$ are statistically significant (at least 5% level) with expected signs across all 14 specifications. Subsequent rows detail counterfactual results. Muting $\Delta CashHold$ renders 4 of 14 validation variables statistically insignificant, whereas muting $\Delta Leverage$ preserves significance across all specifications. These results suggest $\Delta CashHold$ serves as the most salient indicator of agency conflicts among the five actions.

4 Data Valuation

4.1 Motivating Facts

The relationship between management disclosure policies and agency conflicts has been extensively documented in firm disclosure literature. The notion that persistent conflicts of interest between managers and shareholders deter information disclosure originates from seminal works by Dye (1983) and Penno (1984), who posit that maintaining information asymmetries enables managers to make decisions consistent with their own self-interests. Beyond merely reducing disclosure frequency, agency problems undermine disclosure quality as well. Empirical evidence from Feng et al. (2009) demonstrates that weak internal controls correlate with decreased accuracy in management earnings forecasts.

Table 10 presents evidence analyzing the agency conflict-disclosure nexus. $GuideFrequency$ measures the frequency of management guidance, calculated as the number of annual and quarterly

management guidance records at the firm-year level. Column (1) incorporates all types of management forecasts, such as earnings, sales, and capital expenditure. Column (2) focuses on earnings guidance, which is the most common and notable guidance. Consistent across specifications, the negative coefficients on $AgencyConflict$ reveal a negative effect of agency conflicts on guidance frequency. This pattern persists when substituting frequency with a binary disclosure indicator ($GuideDummy$), as shown in Columns (3) and (4). Extending to the disclosure quality dimension, Column (5) employs guidance error ($GuideError$) — measured as the absolute deviation between actual earnings and management forecasts, divided by pre-announcement stock price — as a proxy for disclosure quality. The positive coefficient confirms that guidance tends to be less accurate in firms with higher agency conflicts.

These findings collectively raise a informational efficiency question: To what extent do investors take this negative effect into consideration when processing management guidance? Particularly, does the market assign less value to guidance from high-agency-conflict firms? To answer this question, this paper utilizes the structural valuation framework developed by Farboodi et al. (2025), whose theoretical and methodological details are elaborated in the subsequent subsections.

4.2 Overview of the Data Valuation Model

The model is built on the noisy rational expectations framework, where investors use public and private signals to select high excess return portfolios, controlling for risk.

In each period t, n overlapping generations investors with mean-variance preferences, $i \in [0, 1]$, are born. The number of investors can be finite in the case of imperfectly competitive market. The economy contains N risky assets and a risk-free asset with return r. Each risky asset is a claim to a stream of dividends $\{d_t\}_{t=0}^{\infty}$, where the dividend vector d_t follows the autoregressive process:

$$d_{t+1} = \mu + G(d_t - \mu) + y_{t+1},$$

with $y_{t+1} \sim \mathcal{N}(0, \Sigma_d)$ being an i.i.d. exogenous dividend innovation shock across time. The per capita supply of risky assets is $\bar{x} + x$, where $\bar{x}$ is a positive constant vector and $x \sim N(0, \Sigma_x)$ represents shocks from liquidity, hedging, noise trading, or other trading activities unrelated to dividend innovation. The risky asset supply contributes to price noise and creates an incentive to learn.

According to Bayes' law, the information investors learn from their private data, public prices, as well as other public data can be linearly combined into one composite signal. Specifically, investors receive an unbiased signal s_{it} about the dividend innovation y_{t+1}:

$$s_{it} = y_{t+1} + \zeta_{it} z_{t+1} + \xi_{it},$$

where $z_{t+1} \sim \mathcal{N}(0, \Sigma_z)$ is public signal noise from learning prices or other data, observed at the end of period t; $\xi_{it} \sim \mathcal{N}\left(0, K_{it}^{-1}\right)$ is the private signal noise. Both z_{t+1} and ξ_{it} are i.i.d across time.

The model unfolds over two periods. At time t, investors choose their portfolio holdings of risky assets subject to a budget constraint:

$$\begin{aligned} \max_{q_{it}} \rho_i \mathbb{E}\left[c_{it+1} \mid \mathcal{I}_{it}\right] - \frac{\rho_i^2}{2} \mathbb{V}\left[c_{it+1} \mid \mathcal{I}_{it}\right] \\ c_{it+1} = r\left(\bar{w}_{it} - q_{it}' p_t\right) + q_{it}'\left(p_{t+1} + d_{t+1}\right). \end{aligned} \tag{6}$$

Here, $\bar{w}_{it}$ denotes initial endowment and p_t denotes price of risky assets. Investor i's information set is $\mathcal{I}_{it} = \{\mathcal{I}_{t-1}, y_t, d_t, x_t, z_t, p_t, s_{it}\}$, where $\mathcal{I}_{t-1}$ is the information set of the average investor at time $t-1$.

In the following period $t+1$, investors sell their assets, consume dividends and proceeds from asset sales, and exit the model. The assumption of two-period-lived investors simplifies the equilibrium price to a Markov structure, dependent only on current innovations. The price schedule is assumed to be a linear function:

$$p_t = A_t + B(d_t - \mu) + C_t y_{t+1} + D_t x_{t+1} + F_t z_{t+1}.$$

In equilibrium, the expected utility of investors can be expressed as:

$$\tilde{U}\left(\boldsymbol{\mathcal{I}}_{it}\right) \approx \mathbb{E}\left[R_{it}\right]' \hat{\tilde{V}}_{it}^{-1} \mathbb{E}\left[R_{it}\right] + \operatorname{Tr}\left[\left(\mathbb{V}\left[R_{it}\right] - \mathbb{V}\left[R_{it} \mid \boldsymbol{\mathcal{I}}_{it}\right]\right) \hat{\tilde{V}}_{it}^{-1}\right] + r\rho_i \bar{w}_{it}, \text{[14]} \tag{7}$$

where

$$\hat{\tilde{V}}_{it} := \tilde{V}_{it}\left(I - \frac{1}{2}\mathbb{V}\left[R_{it} \mid \boldsymbol{\mathcal{I}}_{it}\right] \tilde{V}_{it}^{-1}\right)^{-1} \quad \text{and} \quad \tilde{V}_{it} := \left(\mathbb{V}\left[R_{it} \mid \boldsymbol{\mathcal{I}}_{it}\right] + \frac{1}{\rho_i}\frac{dp}{dq_i} \oslash p_t p_t'\right).$$

Here, $\oslash$ represents element-by-element division of matrices, and Tr is the matrix trace. $R_{it} := (p_{t+1} + d_{t+1}) \oslash p_t - r$ denotes the excess return on risky assets. The dollar value of data is the amount of risk-free return an investor would require to be indifferent between having the data or not, but receiving additional riskless wealth. Dividing the utility difference by the absolute risk aversion coefficient yields a certainty equivalent amount:

$$\$ \text{ Value of Data}_i = \frac{1}{\rho_i}\left(\tilde{U}\left(\boldsymbol{\mathcal{I}}_{it} \cup Data\right) - \tilde{U}\left(\boldsymbol{\mathcal{I}}_{it}\right)\right) \tag{8}$$

Equation (8) demonstrates that valuing a stream of data requires only a few key statistics: average conditional and unconditional returns, variances, and forecast errors.

4.3 Procedure of Structural Estimation

This subsection outlines the method used to estimate the dollar value of management earnings guidance.

Portfolio Construction To investigate whether investors' valuation of management earnings guidance varies with the degree of a firm's agency conflicts, this study constructs two value-weighted portfolios using the DRL-derived measure of agency conflicts ($AgencyConflict$). The weights are based on market cap at the end of the previous fiscal quarter. Management earnings

[14] The approximate sign is used because this expression assumes that the ex ante variation in p_t is relatively small, allowing p_t to be treated as a constant from an ex ante perspective.

guidance is provided quarterly, while $AgencyConflict$ is calculated annually. Therefore, firms are classified into high-agency-conflict and low-agency-conflict portfolios based on whether their annual $AgencyConflict$ is above or below the sample median for the year corresponding to a given quarter.

Unconditional Return and Variance The estimates for the unconditional expected return $\mathbb{E}[R_t]$ and variance $\mathbb{V}[R_t]$ are derived from time series moments. Specifically, $\widehat{\mathbb{E}[R_t]} = \frac{1}{T}\sum_{t=1}^{T} R_t$ and $\widehat{\mathbb{V}[R_t]}$ is calculated using the Newey-West estimator for the spectrum at frequency zero of $R_t - \widehat{\mathbb{E}[R_t]}$, with the Andrews rule for HAR inference bandwidth $0.75T^{\frac{1}{3}}$.

Conditional Variance Let X_t represent the management earnings guidance data, which is the primary focus of this study. Z_t denotes publicly available data already included in the investor's information set. This includes the value-weighted portfolio price at the end of the last fiscal year (p_t). Additionally, it is assumed that investors have access to market-level economic and financial conditions data, specifically $GGDP$ and $NFCI$. These variables are incorporated into Z_t and have been utilized in the DRL framework. $\mathbb{V}[R_t \mid \boldsymbol{\mathcal{I}}]$ is estimated as $\widehat{\mathrm{Cov}}\left(\varepsilon_t^Z\right)$ by projecting R_t onto the existing data, Z_t:

$$R_t = \gamma Z_t + \varepsilon_t^Z. \tag{9}$$

Similarly, $\mathbb{V}[R_t \mid \boldsymbol{\mathcal{I}} \cup Data]$ is estimated as $\widehat{\mathrm{Cov}}\left(\varepsilon_t^{XZ}\right)$ by projecting R_t onto Z_t and X_t:

$$R_t = \alpha X_t + \beta Z_t + \varepsilon_t^{XZ} \tag{10}$$

Other Parameters The price impact term, $\frac{dp}{dq_i} \oslash p_t p_t'$, is set at 1.5×10^{-8}, following Farboodi et al. (2025). The coefficient of absolute risk aversion, ρ_i, varies with wealth. One way to impute such a value is to assume the investor has constant relative-risk aversion (CRRA, denoted as σ_{CRRA}) and back out ρ_i from $-\exp(-\rho_i c_{it}) = c_{it}^{1-\sigma_{CRRA}} / (1 - \sigma_{CRRA})$. For relative-risk aversion of 2, ρ_i can be

expressed as a function of investor wealth c_{it} as $\rho_i = \ln(c_{it})/c_{it}$. Investors' endowment, $\bar{w}_{it}$, is set at \$500,000 and \$250 million, reflecting the wealth levels of the average U.S. household (Badarinza et al., 2016) and hedge fund (Yin, 2016), respectively.

Substituting these parameters into Equation (8) yields the estimated dollar value of management earnings guidance data. To determine the standard errors for this estimate, a wild bootstrap method is applied with 500 samples, using Equation (9) as the data generating process and the two-point distribution from Mammen (1993).

4.4 Sample and Result of Structural Data Valuation

Management earnings guidance data is obtained from the LSEG IBES database, which collects information from public disclosures, including press releases and conference calls. Only guidance announced within the prediction quarter is included. For firms issuing multiple guidance within a quarter, only the most recent guidance is used. When management provides a range instead of a single EPS value, the midpoint of the range is taken as the guidance value. The guidance surprise for firm i in quarter q is calculated as $GuideSurprise = \left(Guide_{i,q} - \overline{Actual}_{i,q}\right)/Price_{i,q}$, where $Guide$ represents the latest management earnings guidance, $\overline{Actual}_{i,q}$ denotes the average earnings realized over the previous four quarters, and $Price$ refers to the stock price on the trading day prior to the guidance announcement date. This firm-quarter-level guidance surprise measure is then aggregated to the portfolio-quarter level using the last-quarter-end market capitalization as the weight. Due to the limited availability of management guidance data and DRL-derived measure of agency conflicts before 1998, the final sample spans from 1998Q1 to 2023Q4.

Table 11 Column (1) shows the dollar value of management earnings guidance ($GuideSurprise$) for the low-agency-conflict portfolio, considering investors with varying initial wealth levels and price impact. The results highlight several notable patterns observed by Farboodi et al. (2025). Firstly, wealthier investors assign a higher dollar value to management earnings guidance. For example, an investor with \$250 million in wealth in a competitive setting values this data 340 times more than an investor with \$0.5 million in wealth. This is due to wealth effects influencing

their local risk aversion ρ_i. Second, price impact reduces the dollar value of management earnings guidance. For a \$250 million investor, accounting for price impact decreases the data's value by 80.18%. This occurs because price impact hinders effective information incorporation into trading strategies, as prices move against the investor, reducing data valuation. Thirdly, the price impact effect grows with wealth. While it significantly lowers data valuation for \$250 million investors, it has little effect among \$0.5 million investors. This is because wealthier investors have a lower absolute risk aversion ρ_i, and price impact dp/dq_i scales with ρ_i, as shown in Equation (7).

Column (2) of Table 11 shows results for the high-agency-conflict portfolio. Across various levels of initial wealth and price impact, the dollar value of management earnings guidance in this portfolio is notably lower than in the low-agency-conflict portfolio. For instance, with perfect competition and an endowment wealth of \$250 million, management earnings guidance is valued at approximately \$1,765,800 for investors holding the low-agency-conflict portfolio, compared to \$1,139,470 for investors holding the high-agency-conflict portfolio. This supports the hypothesis that investors consider agency conflicts when evaluating management guidance information.

Tables 12 and Table B.3 present a series of robustness checks. To conserve space, only the dollar values for investors with \$250 million in wealth are reported. The first robustness check assesses the impact of potential outliers in $AgencyConflict$. Firms whose actions deviate by more than two standard deviations from both market-cap-maximizing and compensation-maximizing actions are excluded. Despite these exclusions, the patterns of data valuation across varying degrees of agency conflicts remain consistent, as shown in Panel A of Table 12, although the dollar values are slightly lower than those in Table 11.

The second robustness check examines the sensitivity of management earnings guidance dollar values to other public information sources. In the baseline analysis, investors are assumed to access market-level economic and financial conditions indexes ($GGDP$ and $NFCI$), along with the value-weighted portfolio price. Panel B of Table 12 investigates alternative market-level public data sources. The first row reports the dollar value for investors with access to the S&P 500 D/P ratio, sourced from Shiller (2015), averaged from monthly to quarterly. The second

row uses VIX from FRED,[15] reflecting market expectations of near-term volatility conveyed by stock index option prices, averaged from daily to quarterly. The third row considers inflation and unemployment rates, with inflation calculated as the quarterly CPI growth rate from Shiller (2015) and unemployment sourced from FRED,[16] both averaged from monthly to quarterly. The fourth row examines macro and financial uncertainty indexes from Jurado et al. (2015), derived from hundreds of macroeconomic and financial indicators using Markov chain Monte Carlo methods and also averaged from monthly to quarterly. Although management guidance dollar values fluctuate with different public data sources, values in the low-agency-conflict portfolio consistently exceed those in the high-agency-conflict portfolio.

The third robustness check relaxes investment restrictions. In the baseline analysis in Table 11, dollar values are estimated for investors holding either high-agency-conflict or low-agency-conflict portfolios. Now, investors are allowed to hold both. In this scenario, multi-asset investors can leverage management guidance from the high-agency-conflict portfolio not only to decide whether to buy more or less within the high-agency-conflict portfolio, but also to adjust asset allocations across portfolios. Panel C of Table 12 presents the results, showing that dollar values in the low-agency-conflict portfolio remain higher than those in the high-agency-conflict portfolio, even after considering asset allocation effects.

The fourth robustness check uses equal weighting to construct portfolios. The results reported in Panel A of Table B.3 mirror the value-weighted results. The fifth robustness check adjusts the price impact value from 1.5×10^{-8} to 5×10^{-9}, in accordance with Gabaix and Koijen (2021). As reported in Panel B of Table B.3, this reduced price impact results in higher dollar values compared to the baseline, particularly for wealthy investors. Additionally, the divergent pattern of dollar values persists.

[15] Available at https://fred.stlouisfed.org/series/VIXCLS.
[16] Available at https://fred.stlouisfed.org/series/UNRATE.

4.5 Further Study: Valuation of Analyst Forecast Data

The results in the previous subsection indicate that management earnings guidance tends to be low-quality for firms with high agency conflicts, leading investors to devalue this information. This raises a question: Do investors rely more on alternative information sources when management guidance is unreliable? Specifically, do they place greater value on analyst forecasts for firms with high agency conflicts?

To explore this question, this paper obtains analyst earnings forecasts data from the LSEG IBES database, and goes through the following steps to construct the analysts forecast surprise. First, only forecasts announced within the prediction quarter are included. Second, for analysts who issue multiple forecasts for the same firm within a quarter, only the most recent forecast is considered. Finally, the analyst forecast surprise is calculated as $AnalystSurprise = \left(Consensus_{i,q} - Consensus_{i,q-1}\right) / Price_{i,q}$, where $Consensus$ is the median of analyst forecasts for the firm-quarter, and $Price_{i,q}$ is the average stock price at the end of the current and previous fiscal quarters. This firm-quarter-level measure is then aggregated to the portfolio-quarter level. Table 13 shows the dollar values of analyst earnings forecasts ($AnalystSurprise$), which are smaller than those of management earnings guidance reported in Table 11. This aligns with Kim and Song (2015), which shows that stock prices respond more strongly to management earnings guidance than to analyst forecast revisions, as the latter are generally perceived to be less informative. Crucially, the dollar value of analyst forecasts in the high-agency-conflict portfolio is significantly higher than in the low-agency-conflict portfolio, across various levels of initial wealth and price impact effects. For example, with perfect competition and an endowment wealth of \$250 million, the dollar value of analyst earnings forecasts for investors holding the high-agency-conflict portfolio is twice that of those holding the low-agency-conflict portfolio (\$997,500 VS \$470,370).

5 Conclusion

This paper introduces a novel approach to quantifying corporate agency conflicts at the firm-year level using offline DRL. By training offline DRL models to simulate compensation-maximizing and market-cap-maximizing management policies, this paper derives actionable benchmarks to evaluate real-world decisions across five key domains: investment, cash holdings, acquisitions, net payouts, and leverage. The relative Euclidean distance between observed decisions and these benchmarks provides a granular and dynamic measure of agency conflicts, validated through robust associations with established corporate governance proxies such as fraud, earnings management, and insider trading. The analysis reveals that compensation-maximizing actions—characterized by lower investment, more cash holdings, and reduced leverage—deviate from market-cap-maximizing actions, underscoring the prevalence of hidden agency conflicts. On top of the DRL-derived agency conflicts measure, this paper demonstrates that the market internalizes these conflicts: earnings guidance from high-agency-conflict firms is discounted in valuation, with investors instead relying more on analyst forecasts. This finding, derived from a noisy rational expectations framework, highlights the economic consequences of manager-shareholder incentive misalignment for informational efficiency.

The findings carry practical implications for governance design, disclosure regulation, and investment analysis. The DRL-derived measure offers policymakers a tool to identify firms prone to agency concerns, while investors can leverage it to adjust for disclosure reliability. By bridging machine learning with corporate finance theory, this study underscores the transformative potential of computational methods in unraveling complex financial frictions.

References

Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4):1263–1289.

Aghion, P., Van Reenen, J., and Zingales, L. (2013). Innovation and institutional ownership. *American economic review*, 103(1):277–304.

Albuquerue, R. and Wang, N. (2008). Agency conflicts, investment, and asset pricing. *the Journal of Finance*, 63(1):1–40.

Ali, U. and Hirshleifer, D. (2017). Opportunism as a firm and managerial trait: Predicting insider trading profits and misconduct. *Journal of financial economics*, 126(3):490–515.

Badarinza, C., Campbell, J. Y., and Ramadorai, T. (2016). International comparative household finance. *Annual Review of Economics*, 8(1):111–144.

Bai, J., Philippon, T., and Savov, A. (2016). Have financial markets become more informative? *Journal of Financial Economics*, 122(3):625–654.

Bates, T. W., Kahle, K. M., and Stulz, R. M. (2009). Why do us firms hold so much more cash than they used to? *The journal of finance*, 64(5):1985–2021.

Bebchuk, L. A., Grinstein, Y., and Peyer, U. (2010). Lucky ceos and lucky directors. *The Journal of Finance*, 65(6):2363–2401.

Biais, B., Mariotti, T., Rochet, J.-C., and Villeneuve, S. (2010). Large risks, limited liability, and dynamic moral hazard. *Econometrica*, 78(1):73–118.

Bond, P., Edmans, A., and Goldstein, I. (2012). The real effects of financial markets. *Annu. Rev. Financ. Econ.*, 4(1):339–360.

Cahan, S. F., Chen, C., and Chen, L. (2024). In financial statements we trust: institutional investors' stockholdings after restatements. *The Accounting Review*, 99(2):143–168.

Calvano, E., Calzolari, G., Denicolo, V., and Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297.

Calvano, E., Calzolari, G., Denicoló, V., and Pastorello, S. (2021). Algorithmic collusion with imperfect monitoring. *International journal of industrial organization*, 79:102712.

Campello, M., Cong, L. W., and Zhou, L. (2023). Alphamanager: A data-driven-robust-control approach to corporate finance. *Available at SSRN 4590323*.

Collins, D. W., Gong, G., and Li, H. (2009). Corporate governance and backdating of executive stock options. *Contemporary accounting research*, 26(2):403–445.

Cong, L. W., Tang, K., Wang, J., and Zhang, Y. (2021). Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai.

Dávila, E. and Parlatore, C. (2025). Identifying price informativeness. *Review of Financial Studies (Forthcoming)*.

Dechow, P. M., Sloan, R. G., and Sweeney, A. P. (1995). Detecting earnings management. *Accounting review*, pages 193–225.

DeMarzo, P. M., Fishman, M. J., He, Z., and Wang, N. (2012). Dynamic agency and the q theory of investment. *The journal of Finance*, 67(6):2295–2340.

Dittmar, A., Mahrt-Smith, J., and Servaes, H. (2003). International corporate governance and corporate cash holdings. *Journal of Financial and Quantitative analysis*, 38(1):111–133.

Dou, W. W., Goldstein, I., and Ji, Y. (2024). Ai-powered trading, algorithmic collusion, and price efficiency. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Dyck, A., Morse, A., and Zingales, L. (2010). Who blows the whistle on corporate fraud? *The journal of finance*, 65(6):2213–2253.

Dye, R. A. (1983). Communication and post-decision information. *Journal of Accounting Research*, pages 514–533.

Edmans, A. (2009). Blockholder trading, market efficiency, and managerial myopia. *The journal of finance*, 64(6):2481–2513.

Edmans, A., Goldstein, I., and Jiang, W. (2012). The real effects of financial markets: The impact of prices on takeovers. *The Journal of Finance*, 67(3):933–971.

Efendi, J., Srivastava, A., and Swanson, E. P. (2007). Why do corporate managers misstate financial statements? the role of option compensation and other factors. *Journal of financial economics*, 85(3):667–708.

Eisfeldt, A. L. and Papanikolaou, D. (AER2014). The value and ownership of intangible capital. *American Economic Review*, 104(5):189–194.

Farboodi, M., Singal, D., Veldkamp, L., and Venkateswaran, V. (2025). Valuing financial data. *The Review of Financial Studies*, 38(3):938–980.

Fee, C. E., Hadlock, C. J., and Thomas, S. (2006). Corporate equity ownership and the governance of product market relationships. *The Journal of Finance*, 61(3):1217–1251.

Feng, M., Li, C., and McVay, S. (2009). Internal control and management guidance. *Journal of accounting and economics*, 48(2-3):190–209.

Gabaix, X. and Koijen, R. S. (2021). In search of the origins of financial fluctuations: The inelastic markets hypothesis. Technical report, National Bureau of Economic Research.

Gao, H., Harford, J., and Li, K. (2013). Determinants of corporate cash policy: Insights from private firms. *Journal of Financial Economics*, 109(3):623–639.

Gormley, T. A. and Matsa, D. A. (2016). Playing it safe? managerial preferences, risk, and agency conflicts. *Journal of financial economics*, 122(3):431–455.

Harford, J. (1999). Corporate cash reserves and acquisitions. *The journal of finance*, 54(6):1969–1997.

Harford, J., Mansi, S. A., and Maxwell, W. F. (2008). Corporate governance and firm cash holdings in the us. *Journal of financial economics*, 87(3):535–555.

Hartzell, J. C. and Starks, L. T. (2003). Institutional investors and executive compensation. *The journal of finance*, 58(6):2351–2374.

Hazarika, S., Karpoff, J. M., and Nahata, R. (2012). Internal corporate governance, ceo turnover, and earnings management. *Journal of financial Economics*, 104(1):44–69.

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *The American economic review*, 76(2):323–329.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360.

Johnson, J. P., Rhodes, A., and Wildenbeest, M. (2023). Platform design when sellers use pricing algorithms. *Econometrica*, 91(5):1841–1879.

Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of accounting research*, 29(2):193–228.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.

Kacperczyk, M., Nosal, J., and Sundaresan, S. (2025). Market power and price informativeness. *Review of Economic Studies*, 92(3):1955–1986.

Kalcheva, I. and Lins, K. V. (2007). International evidence on cash holdings and expected managerial agency problems. *The review of financial studies*, 20(4):1087–1112.

Kempf, E. and Spalt, O. (2023). Attracting the sharks: Corporate innovation and securities class action lawsuits. *Management Science*, 69(3):1805–1834.

Kim, Y. and Song, M. (2015). Management earnings forecasts and value of analyst forecast revisions. *Management Science*, 61(7):1663–1683.

Kothari, S. P., Leone, A. J., and Wasley, C. E. (2005). Performance matched discretionary accrual measures. *Journal of accounting and economics*, 39(1):163–197.

Kumar, A., Zhou, A., Tucker, G., and Levine, S. (2020). Conservative q-learning for offline reinforcement learning. *Advances in neural information processing systems*, 33:1179–1191.

Kumar, P. and Langberg, N. (2009). Corporate fraud and investment distortions in efficient capital markets. *The RAND Journal of Economics*, 40(1):144–172.

Lambrecht, B. M. and Myers, S. C. (2012). A lintner model of payout and managerial rents. *The journal of finance*, 67(5):1761–1810.

Liu, X. (2016). Corruption culture and corporate misconduct. *Journal of Financial Economics*, 122(2):307–327.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The annals of statistics*, 21(1):255–285.

Morellec, E. (2004). Can managerial discretion explain observed leverage ratios? *The Review of Financial Studies*, 17(1):257–294.

Morellec, E., Nikolov, B., and Schürhoff, N. (2012). Corporate governance and capital structure dynamics. *The journal of finance*, 67(3):803–848.

Morellec, E., Nikolov, B., and Schürhoff, N. (2018). Agency conflicts around the world. *The Review of Financial Studies*, 31(11):4232–4287.

Nakamoto, M., Zhai, S., Singh, A., Sobol Mark, M., Ma, Y., Finn, C., Kumar, A., and Levine, S. (2023). Cal-ql: Calibrated offline rl pre-training for efficient online fine-tuning. *Advances in Neural Information Processing Systems*, 36:62244–62269.

Nikolov, B. and Whited, T. M. (2014). Agency conflicts and cash: Estimates from a dynamic model. *The Journal of Finance*, 69(5):1883–1921.

Panousi, V. and Papanikolaou, D. (2012). Investment, idiosyncratic risk, and ownership. *The Journal of finance*, 67(3):1113–1148.

Penno, M. (1984). Asymmetry of pre-decision information and managerial accounting. *Journal of Accounting Research*, pages 177–191.

Ravina, E. and Sapienza, P. (2010). What do independent directors know? evidence from their trading. *The Review of Financial Studies*, 23(3):962–1003.

Ronen, J. (2008). *Earnings management*. Springer.

Rozanov, K. A. (2008). *Corporate governance and insider trading*. PhD thesis, Massachusetts Institute of Technology.

Schwartz-Ziv, M. and Volkova, E. (2025). Is blockholder diversity detrimental? *Management Science*, 71(2):1356–1390.

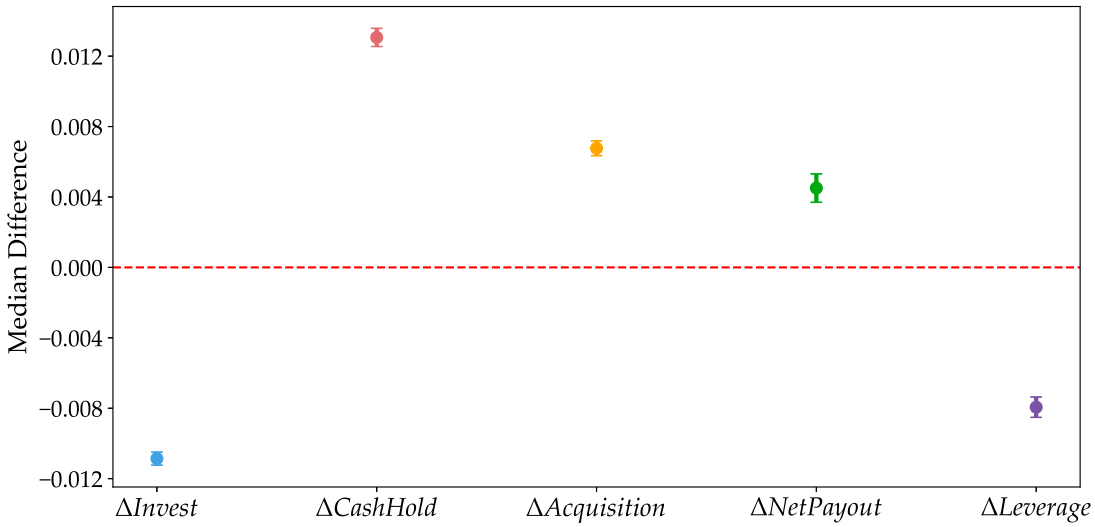
Shiller, R. J. (2015). Irrational exuberance: Revised and expanded third edition.

Shleifer, A. and Vishny, R. W. (1986). Large shareholders and corporate control. *Journal of political economy*, 94(3, Part 1):461–488.

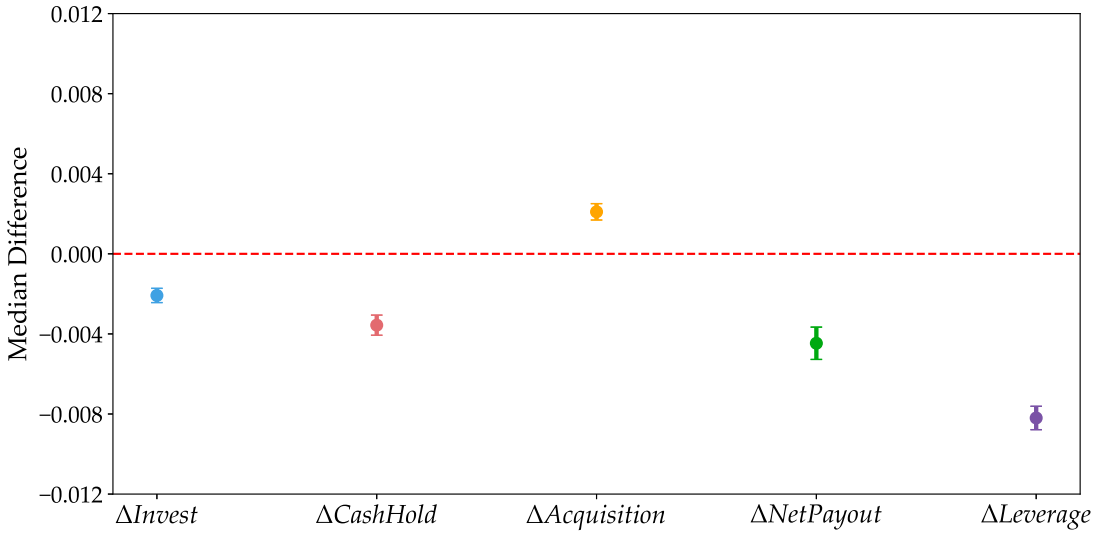
Wang, W. (2022). Accounting restatements and bank liquidity creation. *The Accounting Review*, 97(6):445–473.

Yin, C. (2016). The optimal size of hedge funds: Conflict between investors and fund managers. *The Journal of Finance*, 71(4):1857–1894.

Zwiebel, J. (1996). Dynamic capital structure under managerial entrenchment. *The American economic review*, pages 1197–1215.



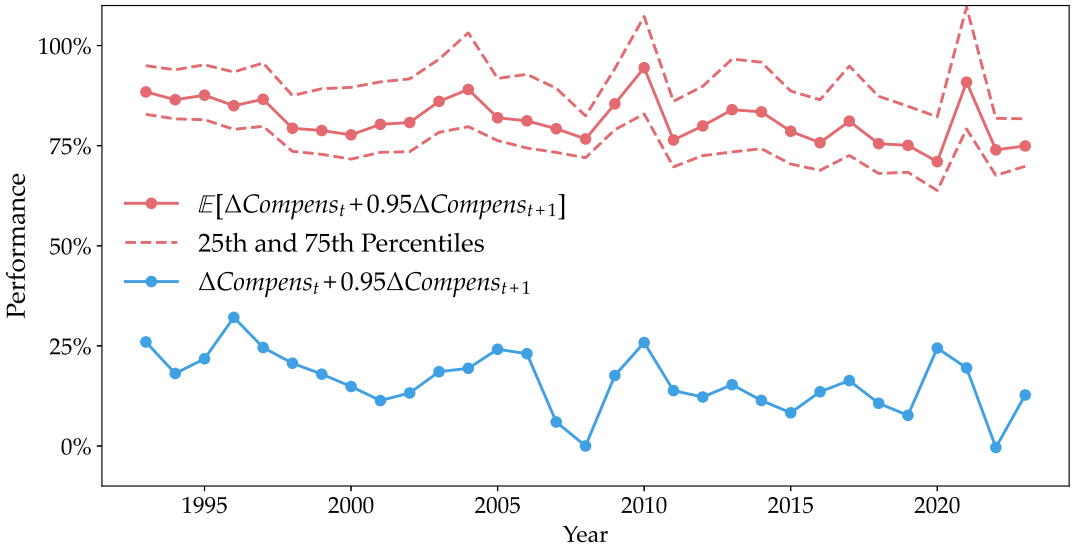
(a) Compensation-Maximizing Actions VS Actual Actions



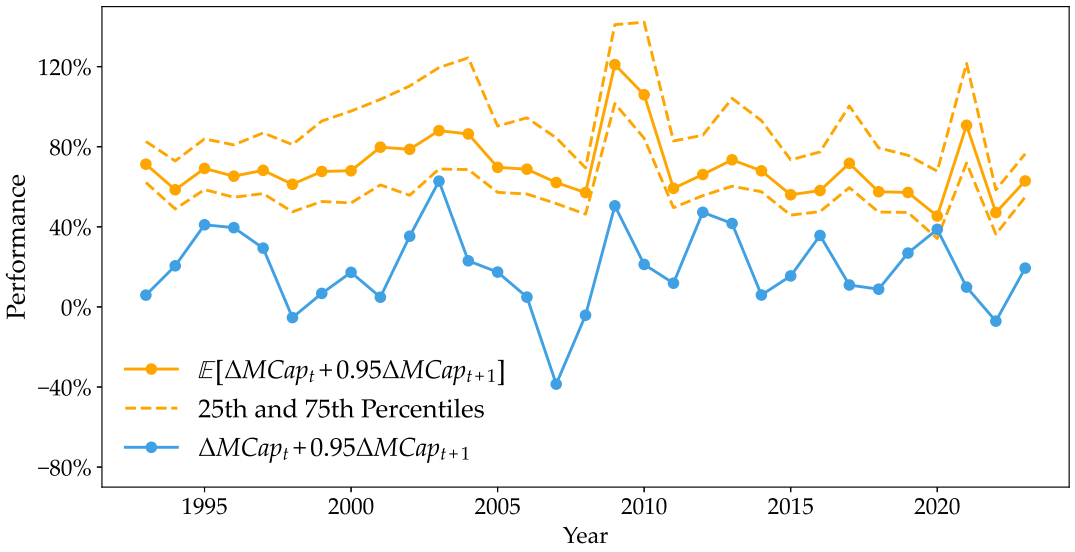
(b) Market-Cap-Maximizing Actions VS Actual Actions

Panel (a) illustrates the differences between the median values of actual management actions ($\mathcal{A}^{Actual}$) and the compensation-maximizing actions ($\mathcal{A}^{Compens}$), along with the 95% confidence interval. Panel (b) instead compares the actual management actions ($\mathcal{A}^{Actual}$) and those aimed at maximizing market cap ($\mathcal{A}^{MCap}$). Five management actions are included: (1) Change in the investment ratio ($\Delta Invest$), calculated as the sum of R&D expenses, 30% of SG&A expenses, and capital expenditures, all scaled by total assets. (2) Change in the cash holding ratio ($\Delta CashHold$), calculated as cash holdings scaled by total assets. (3) Change in acquisition expenditure ($\Delta Acquisition$), calculated as acquisition-related costs scaled by total assets. (4) Change in the net payout ratio ($\Delta NetPayout$), calculated as the sum of dividends and net repurchases, scaled by total assets. (5) Change in the leverage ratio ($\Delta Leverage$), calculated as liabilities divided by total assets.

Figure 1: DRL-Recommended Actions VS Actual Actions



(a) Cumulative Growth Rate of Compensation



(b) Cumulative Growth Rate of Market Cap

This figure illustrates a comparison between actual corporate performance and expected performance based on DRL-recommended actions. Panel (a) presents the median values of actual cumulative growth rates of compensation ($\Delta Compens_t + 0.95\Delta Compens_{t+1}$) against their expected counterparts ($\mathbb{E}[\Delta Compens_t + 0.95\Delta Compens_{t+1}]$) when taking compensation-maximizing actions. Panel (b) compares the median values of actual cumulative growth rates of market cap ($\Delta MCap_t + 0.95\Delta MCap_{t+1}$) with their expected counterparts ($\mathbb{E}[\Delta MCap_t + 0.95\Delta MCap_{t+1}]$) when taking market-cap-maximizing actions.

Figure 2: DRL-Predicted Performance VS Actual Performance

Table 1: Summary Statistics

This table presents the summary statistics for the variables used in the offline DRL framework. The sample has an annual frequency and spans from 1993 to 2023. All continuous variables are winsorized at the top and bottom 2.5% to mitigate the influence of outliers. Variable definitions are provided in Table B.1.

Variable	N	Mean	SD	P10	P25	P50	P75	P90
Panel A: Action Variables								
$\Delta Invest$	77181	-0.001	0.049	-0.048	-0.013	0.000	0.013	0.046
$\Delta CashHold$	77181	-0.004	0.071	-0.086	-0.026	0.000	0.023	0.072
$\Delta Acquisition$	77181	-0.001	0.057	-0.040	0.000	0.000	0.000	0.039
$\Delta NetPayout$	77181	0.013	0.114	-0.061	-0.009	0.000	0.016	0.097
$\Delta Leverage$	77181	0.009	0.082	-0.073	-0.026	0.001	0.034	0.103
Panel B: Firm-level State Variables								
$OROA$	77181	0.064	0.160	-0.086	0.021	0.089	0.150	0.215
ROA	77181	-0.013	0.168	-0.190	-0.009	0.022	0.066	0.116
GPM	77181	0.348	0.465	0.115	0.227	0.379	0.586	0.770
$InstOwn$	77181	0.558	0.312	0.081	0.288	0.610	0.825	0.941
$Amihud$	77181	-4.510	3.407	-8.681	-7.106	-5.053	-2.226	0.653
Panel C: Market-level State Variables								
$GGDP$	31	0.026	0.018	0.010	0.020	0.027	0.038	0.044
$NFCI$	31	-0.411	0.430	-0.730	-0.657	-0.486	-0.319	-0.220
Panel D: Rewards								
$\Delta MCap$	77181	0.016	0.058	-0.044	-0.019	0.006	0.036	0.083
$\Delta Compens$	77181	0.020	0.068	-0.038	-0.009	0.003	0.029	0.088

Table 2: Validation: Corporate Fraud

This table presents OLS estimates examining the relationship between corporate fraud and agency conflicts. $AgencyConflict$ is quantified as the difference in Euclidean distance between actual management actions and market cap-maximizing actions, and actual management actions and compensation-maximizing actions. Both market cap-maximizing and compensation-maximizing actions are derived from offline DRL models. The dependent variable in Column (1) is $Restate$, a dummy variable set to one if a firm has material financial restatements in a given year and zero otherwise. In Columns (2), the dependent variable is $AllSuit$, a dummy variable set to one if a firm is involved in securities class action lawsuits in a given year and zero otherwise. The dependent variable in Column (3) is $NDSuit$, which further excludes cases that are subsequently dismissed, compared to $AllSuit$. Refer to Table B.1 for detailed variable definitions. All specifications include firm and year fixed effects. t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Dependent Variable:	$Restate$	$AllSuit$	$NDSuit$
	(1)	(2)	(3)
$AgencyConflict$	0.108**	0.086***	0.069***
	(2.47)	(2.96)	(3.18)
Observations	75,421	72,655	72,655
R-squared	0.385	0.136	0.119
Firm Fixed Effect	Y	Y	Y
Year Fixed Effect	Y	Y	Y

Table 3: Validation: Earnings Management

This table presents OLS estimates examining the relationship between earnings management and agency conflicts. $AgencyConflict$ is quantified as the difference in Euclidean distance between actual management actions and market cap-maximizing actions, and actual management actions and compensation-maximizing actions. Both market cap-maximizing and compensation-maximizing actions are derived from offline DRL models. Earnings management involves managers using judgment in financial reporting and structuring transactions to alter financial reports. Earnings management is measured by the absolute value of discretionary accruals (DA), calculated using the models by Jones (1991), Dechow et al. (1995), and Kothari et al. (2005). Higher discretionary accruals indicate a greater likelihood of earnings manipulation. Refer to Table B.1 for detailed variable definitions. All specifications include firm and year fixed effects. t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Dependent Variable:	DA^{Jones}	DA^{Dechow}	$DA^{Kothari}$
	(1)	(2)	(3)
$AgencyConflict$	0.196***	0.195***	0.293***
	(7.27)	(7.15)	(6.43)
Observations	63,955	63,955	63,955
R-squared	0.316	0.312	0.269
Firm Fixed Effect	Y	Y	Y
Year Fixed Effect	Y	Y	Y

Table 4: Validation: Manager Option Backdating

This table presents OLS estimates examining the relationship between option backdating and agency conflicts. $AgencyConflict$ is quantified as the difference in Euclidean distance between actual management actions and market cap-maximizing actions, and actual management actions and compensation-maximizing actions. Both market cap-maximizing and compensation-maximizing actions are derived from offline DRL models. Option grant backdating involves retroactively setting the grant date of at-the-money stock options to a period when the stock price is lower. A backdating event is defined as occurring when the strike price is no more than 101% of the lowest daily price within the grant month. To capture backdating activities, two metrics were employed: Column (1) considers the insider backdating dummy, $BDDummy$, which equals one if at least one backdating event occurred in a firm in a given year; Column (2) examines the backdating fraction, $BDFraction$, calculated as the number of backdating events divided by the total manager option grant events within a year. Refer to Table B.1 for detailed variable definitions. All specifications include firm and year fixed effects. t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Dependent Variable:	$BDDummy$	$BDFraction$
	(1)	(2)
$AgencyConflict$	0.462**	0.387**
	(2.26)	(2.07)
Observations	14,423	14,423
R-squared	0.256	0.257
Firm Fixed Effect	Y	Y
Year Fixed Effect	Y	Y

Table 5: Validation: Price Pattern of Opportunistic Manager Trading

This table presents OLS estimates examining the relationship between opportunistic manager trading in stock markets and agency conflicts. $AgencyConflict$ is quantified as the difference in Euclidean distance between actual management actions and market cap-maximizing actions, and actual management actions and compensation-maximizing actions. Both market cap-maximizing and compensation-maximizing actions are derived from offline DRL models. Opportunistic manager trading is identified by the price pattern ratio, $PricePattern$, calculated as $PricePattern = \prod_{k=-\tau}^{-1}(1+R_{i,k})/\prod_{k=1}^{\tau}(1+R_{i,k})$, where τ is the estimation window. Two windows are considered: $[-10,+10]$ in Columns (1) and (3), and $[-20,+20]$ in Columns (2) and (4), with $k=0$ as the manager trading date. $R_{i,k}$ is the raw daily return of firm i in Columns (1)-(2), and market-adjusted daily return of firm i in Columns (3)-(4). For managers with multiple trades in a year, the price pattern ratios are averaged for each firm-year observation. Refer to Table B.1 for detailed variable definitions. All specifications include firm and year fixed effects. t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Dependent Variable:	$PricePattern$			
Estimation Return:	Raw Return		Market-adjusted Return	
Estimation Window:	[-10, 10]	[-20, 20]	[-10, 10]	[-20, 20]
	(1)	(2)	(3)	(4)
$AgencyConflict$	0.613***	1.093***	0.471***	0.731***
	(3.08)	(3.95)	(2.73)	(3.14)
Observations	9,037	9,035	9,037	9,035
R-squared	0.397	0.409	0.405	0.412
Firm Fixed Effect	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y

Table 6: Validation: Ownership Structure

This table presents OLS estimates examining the relationship between ownership structure and agency conflicts. $AgencyConflict$ is quantified as the difference in Euclidean distance between actual management actions and market cap-maximizing actions, and actual management actions and compensation-maximizing actions. Both market cap-maximizing and compensation-maximizing actions are derived from offline DRL models. Column (1) examines institutional ownership ($InstOwn$), calculated as the annual average of quarterly total institutional holdings divided by the quarter-end market cap. Column (2) considers outside blockholder ownership ($BlockOwn$), defined as the proportion of stock held by outside blockholders. Higher levels of institutional or outside blockholder ownership suggest enhanced governance. Refer to Table B.1 for detailed variable definitions. All specifications include firm and year fixed effects. t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Dependent Variable:	$InstOwn$	$BlockOwn$
	(1)	(2)
$AgencyConflict$	-0.131***	-0.056**
	(-4.37)	(-2.37)
Observations	75,421	75,421
R-squared	0.852	0.679
Firm Fixed Effect	Y	Y
Year Fixed Effect	Y	Y

Table 7: Robustness of DRL Model: Impact of Outliers

This table replicates Tables 2-6, but excludes firm-year observations where both Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{MCap}\right)$ and Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{Compens}\right)$ outside two standard deviations of the sample mean. Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{MCap}\right)$ represents the Euclidean distance between actual management decisions $(\mathcal{A}_{i,t}^{Actual})$ and market cap-maximizing actions $(\mathcal{A}_{i,t}^{MCap})$, and Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{Compens}\right)$ denotes the Euclidean distance between actual management decisions $(\mathcal{A}_{i,t}^{Actual})$ and compensation-maximizing actions $(\mathcal{A}_{i,t}^{MCap})$ for firm i in year t. Panel A examines corporate fraud. $Restate$ equals one if a firm has material financial restatements in a given year and zero otherwise. $AllSuit$ equals one if a firm is involved in securities class action lawsuits in a given year and zero otherwise. $NDSuit$ excludes cases that are subsequently dismissed, compared to $AllSuit$. Panel B examines earnings management, measured by the absolute value of discretionary accruals (DA), calculated using the models by Jones (1991), Dechow et al. (1995), and Kothari et al. (2005). Panel C considers manager option backdating. A backdating event is defined as occurring when the strike price of option granted to a manager is no more than 101% of the lowest daily price within the grant month. $BDDummy$ equals one if at least one backdating event occurred in a firm in a given year and zero otherwise, while $BDFraction$ measures the frequency of backdating events. Panel D examines the price pattern ratio of opportunistic manager trading: $PricePattern$, calculated as $PricePattern = \prod_{k=-20}^{-1}(1+R_{i,k})/\prod_{k=1}^{20}(1+R_{i,k})$, where $k=0$ is the manager trading date and $R_{i,k}$ is either the raw or market-adjusted daily return of firm i. Panel E examines ownership structure. institutional ownership ($InstOwn$) is calculated as the annual average of quarterly total institutional holdings divided by the quarter-end market cap. Outside blockholder ownership ($BlockOwn$) is defined as the proportion of stock held by outside blockholders. Refer to Table B.1 for detailed variable definitions. All specifications include firm and year fixed effects. t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Panel A: Corporate Fraud			
	$Restate$	$AllSuit$	$NDSuit$
	(1)	(2)	(3)
$AgencyConflict$	0.121***	0.066**	0.053**
	(2.66)	(2.25)	(2.43)
Observations	71,379	68,664	68,664
R-squared	0.385	0.133	0.119
Firm & Year Fixed Effects	Y	Y	Y
Panel B: Earnings Management			
	DA^{Jones}	DA^{Dechow}	$DA^{Kothari}$
$AgencyConflict$	0.168***	0.169***	0.270***
	(7.27)	(7.22)	(6.44)
Observations	60,084	60,084	60,084
R-squared	0.316	0.309	0.243
Firm & Year Fixed Effects	Y	Y	Y
Panel C: Manager Option Backdating			
	$BDDummy$		$BDFraction$
$AgencyConflict$	0.580***		0.505**
	(2.66)		(2.53)
Observations	13,413		13,413
R-squared	0.256		0.257

Continued on next page

Table 7 – *Continued*

Firm & Year Fixed Effects	Y	Y
Panel D: Price Pattern of Opportunistic Manager Trading		
	PricePattern	
Estimation Return	Raw Return	Mkt-adj Return
AgencyConflict	0.993***	0.625***
	(3.51)	(2.65)
Observations	8,541	8,541
R-squared	0.409	0.409
Firm & Year Fixed Effects	Y	Y
Panel E: Ownership Structure		
Dependent Variable:	*InstOwn*	*BlockOwn*
AgencyConflict	-0.122***	-0.066***
	(-3.95)	(-2.68)
Observations	71,379	71,379
R-squared	0.854	0.682
Firm & Year Fixed Effects	Y	Y

Table 8: Robustness of DRL Model: Comparison with Existing Measure

This table replicates Tables 2-6, but controls for the structural agency conflict index ($\hat{\phi}$) proposed by Morellec et al. (2018). Panel A examines corporate fraud. *Restate* equals one if a firm has material financial restatements in a given year and zero otherwise. *AllSuit* equals one if a firm is involved in securities class action lawsuits in a given year and zero otherwise. *NDSuit* excludes cases that are subsequently dismissed, compared to *AllSuit*. Panel B examines earnings management, measured by the absolute value of discretionary accruals (*DA*), calculated using the models by Jones (1991), Dechow et al. (1995), and Kothari et al. (2005). Panel C considers manager option backdating. A backdating event is defined as occurring when the strike price of option granted to a manager is no more than 101% of the lowest daily price within the grant month. *BDDummy* equals one if at least one backdating event occurred in a firm in a given year and zero otherwise, while *BDFraction* measures the frequency of backdating events. Panel D examines the price pattern ratio of opportunistic manager trading: *PricePattern*, calculated as $PricePattern = \prod_{k=-20}^{-1}(1+R_{i,k})/\prod_{k=1}^{20}(1+R_{i,k})$, where $k=0$ is the manager trading date and $R_{i,k}$ is either the raw or market-adjusted daily return of firm *i*. Panel E examines ownership structure. institutional ownership (*InstOwn*) is calculated as the annual average of quarterly total institutional holdings divided by the quarter-end market cap. Outside blockholder ownership (*BlockOwn*) is defined as the proportion of stock held by outside blockholders. Refer to Table B.1 for detailed variable definitions. Year fixed effects are excluded because $\hat{\phi}$ is time-invariant. Firm fixed effects are also omitted in Panel D to avoid subsuming $\hat{\phi}$ in small-sample estimations. *t*-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Panel A: Corporate Fraud

	Restate	*AllSuit*	*NDSuit*
	(1)	(2)	(3)
AgencyConflict	0.146**	0.115***	0.072**
	(2.33)	(2.89)	(2.45)
$\hat{\phi}$	0.022	0.017	0.011
	(0.01)	(0.01)	(0.01)
Observations	40,625	39,003	39,003
R-squared	0.333	0.107	0.088
Firm Fixed Effect	Y	Y	Y
Year Fixed Effect	N	N	N

Panel B: Earnings Management

	DA^{Jones}	DA^{Dechow}	$DA^{Kothari}$
AgencyConflict	0.186***	0.190***	0.321***
	(5.69)	(5.82)	(5.83)
$\hat{\phi}$	-0.650	-0.534	0.072
	(-0.53)	(-0.44)	(0.03)
Observations	40,029	40,029	40,029
R-squared	0.257	0.258	0.205
Firm Fixed Effect	Y	Y	Y
Year Fixed Effect	N	N	N

Panel C: Manager Option Backdating

	BDDummy	*BDFraction*
AgencyConflict	0.450*	0.443*

Continued on next page

Table 8 – *Continued*

	(1.71)	(1.84)
$\hat{\phi}$	41.766	41.761
	(0.81)	(0.88)
Observations	8,516	8,516
R-squared	0.224	0.229
Firm Fixed Effect	Y	Y
Year Fixed Effect	N	N

Panel D: Price Pattern of Opportunistic Manager Trading

	$PricePattern$	
Estimation Return	Raw Return	Mkt-adj Return
$AgencyConflict$	1.626***	1.274***
	(4.56)	(4.26)
$\hat{\phi}$	0.199	0.232*
	(1.24)	(1.72)
Observations	5,243	5,243
R-squared	0.004	0.004
Firm Fixed Effect	N	N
Year Fixed Effect	N	N

Panel E: Ownership Structure

Dependent Variable:	$InstOwn$	$BlockOwn$
$AgencyConflict$	-0.375***	-0.301***
	(-7.96)	(-7.99)
$\hat{\phi}$	3.652**	3.217**
	(2.05)	(2.26)
Observations	40,625	40,625
R-squared	0.764	0.502
Firm Fixed Effect	Y	Y
Year Fixed Effect	N	N

Table 9: Counterfactual Policies

This table reports the validation results of Tables 2 to 6 using counterfactual policies. When a chosen action is muted, its compensation- and market-cap-maximizing values are set equal to its realized action. This adjustment renders the counterfactual Euclidean distance independent of the chosen action, relying solely on the remaining four actions.

	Expected Sign (1)	[Significant at 5%] (2)	(Significant at 10%) (3)
Baseline	14	[14]	(14)
Mute $\Delta Invest$	14	[12]	(14)
Mute $\Delta CashHold$	14	[10]	(10)
Mute $\Delta Acquisition$	14	[11]	(13)
Mute $\Delta NetPayout$	14	[10]	(12)
Mute $\Delta Leverage$	14	[14]	(14)

Table 10: Motivating Fact: Management Guidance

This table presents OLS estimates examining the relationship between management guidance and agency conflicts. *GuideFrequency* denotes the frequency of management guidance, calculated as the number of annual and quarterly management guidance records within a fiscal year for each firm. Column (1) incorporates all types of management guidance, such as earnings per share (EPS), sales, and capital expenditure, while Column (2) focuses on EPS guidance. *GuideDummy* equals one if a firm announces either annual or quarterly management guidance within a fiscal year and zero otherwise. Column (3) incorporates all types of management guidance, while Column (4) focuses on EPS guidance. *GuideError* denotes the earnings guidance accuracy, calculated as the absolute deviation between actual earnings and management forecasts, divided by the pre-announcement stock price. A higher value of *GuideError* implies lower guidance accuracy. All specifications include firm and year fixed effects. *t*-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Dependent Variable:	*GuideFrequency*		*GuideDummy*		*GuideError*
Guidance Type:	All	EPS	All	EPS	EPS
	(1)	(2)	(3)	(4)	(5)
AgencyConflict	-1.720***	-1.434***	-0.222***	-0.229***	0.154***
	(-3.62)	(-3.04)	(-3.00)	(-3.18)	(5.11)
Observations	75,421	75,421	75,421	75,421	24,471
R-squared	0.677	0.632	0.653	0.633	0.606
Firm Fixed Effect	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y

Table 11: Valuation of Management Earnings Guidance

This table reports the dollar values of quarterly management earnings guidance. Firms are classified into high- and low-agency-conflict portfolios based on whether their annual DRL-derived measure of agency conflicts is above or below the sample median for the year corresponding to a given quarter. Management earnings guidance is calculated as $GuideSurprise = \left(Guide_{i,q} - \overline{Actual}_{i,q}\right)/Price_{i,q}$. Here, $Guide$ represents the latest management earnings forecasts, $\overline{Actual}_{i,q}$ denotes the average earnings realized over the previous four quarters, and $Price$ refers to the stock price on the trading day prior to the guidance announcement date. This firm-quarter-level measure is then aggregated to the portfolio-quarter level using the last-quarter-end market capitalization as the weight. Dollar values of $GuideSurprise$ are calculated according to Equation (8). The sample covers 1998Q1 to 2023Q4. Panel A examines the setting of perfect completion, while Panel B considers the setting with price impact. All dollar values are annualized and in thousands of USD. Standard errors are calculated using a wild bootstrap with 500 samples and reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Valuation Data:	$GuideSurprise$	
Agency Conflict:	Low	High
	(1)	(2)
Panel A: Perfect Competition		
Dollar Value (in $000) for Investor	5.2***	3.36***
with $500,000 Wealth	(0.59)	(0.41)
Dollar Value (in $000) for Investor	1765.8***	1139.47***
with $250m Wealth	(199.67)	(139.93)
Panel B: With Price Impact		
Dollar Value (in $000) for Investor	5.2***	3.36***
with $500,000 Wealth	(0.59)	(0.41)
Dollar Value (in $000) for Investor	349.97***	300.94***
with $250m Wealth	(37.54)	(34.91)

Table 12: Robustness of Valuation: Outliers, Alternative Public Data, and Investment Restriction

This table replicates Table 11 with several modifications. Panel A excludes firm-year observations where both Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{MCap}\right)$ and Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{Compens}\right)$ outside two standard deviations of the sample mean when constructing portfolios. Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{MCap}\right)$ represents the Euclidean distance between actual management decisions ($\mathcal{A}_{i,t}^{Actual}$) and market cap-maximizing actions ($\mathcal{A}_{i,t}^{MCap}$), and Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{Compens}\right)$ denotes the Euclidean distance between actual management decisions ($\mathcal{A}_{i,t}^{Actual}$) and compensation-maximizing actions ($\mathcal{A}_{i,t}^{MCap}$) for firm i in year t. Panel B investigates alternative market-level public data sources, including S&P 500 D/P ratio, VIX index, inflation rate, unemployment rate, macro uncertainty, and financial uncertainty. Panel C relaxes investment restrictions so that investors are allowed to hold both high- and low-agency-conflict portfolios. Standard errors are calculated using a wild bootstrap with 500 samples and reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Valuation Data:	*GuideSurprise*	
Agency Conflict:	Low	High
Panel A: Exclusion of Outliers in Agency Conflict		
Without price impact	1654.0***	1089.24***
	(183.68)	(138.48)
With price impact	316.26***	292.47***
	(34.03)	(34.11)
Panel B: Alternative Publicly Available Data (With price impact)		
SP500 D/P	394.63***	320.18***
	(49.54)	(40.73)
VIX	429.54***	285.95***
	(46.97)	(38.36)
Inflation & Unemployment	346.26***	284.7***
	(37.95)	(34.61)
Macro & Financial Uncertainty	308.27***	190.53***
	(41.56)	(28.26)
Panel C: Relaxed Investment Restriction		
Without price impact	1736.08***	1302.3***
	(300.87)	(191.06)
With price impact	575.15***	387.05***
	(62.28)	(39.64)

Table 13: Valuation of Analyst Consensus Earnings Forecasts

This table reports the dollar values of quarterly analyst consensus earnings forecasts. Firms are classified into high- and low-agency-conflict portfolios based on whether their annual DRL-derived measure of agency conflicts is above or below the sample median for the year corresponding to a given quarter. Analyst consensus earnings forecasts are calculated as $AnalystSurprise = \left(Consensus_{i,q} - Consensus_{i,q-1}\right) / Price_{i,q}$, where $Consensus$ is the median of analyst forecasts for the firm-quarter, and $Price_{i,q}$ is the average stock price at the end of the current and previous fiscal quarters. This firm-quarter-level measure is then aggregated to the portfolio-quarter level using the last-quarter-end market capitalization as the weight. Dollar values of $AnalystSurprise$ are calculated according to Equation (8). The sample covers 1998Q1 to 2023Q4. Panel A examines the setting of perfect completion, while Panel B considers the setting with price impact. All dollar values are annualized and in thousands of USD. Standard errors are calculated using a wild bootstrap with 500 samples and reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Valuation Data:	$AnalystSurprise$	
Agency Conflict:	Low	High
	(1)	(2)
Panel A: Perfect Competition		
Dollar Value (in $000) for Investor	1.39*	2.94***
with $500,000 Wealth	(0.77)	(1.06)
Dollar Value (in $000) for Investor	470.37*	997.5***
with $250m Wealth	(260.41)	(360.85)
Panel B: With Price Impact		
Dollar Value (in $000) for Investor	1.38*	2.94***
with $500,000 Wealth	(0.77)	(1.06)
Dollar Value (in $000) for Investor	54.67**	166.99***
with $250m Wealth	(27.61)	(48.8)

Appendix

A More Details about Offline DRL Models

A.1 Implementation Details of Cal-QL

Let $J_Q(\omega)$ denote the Cal-QL critic network:

$$J_Q(\omega) := \alpha \left(\mathbb{E}_{s\sim\mathcal{D}, a\sim\pi_\theta} \left[\max\left(Q_\omega^{\pi_\theta}(s,a), V^\mu(s) \right) \right] - \mathbb{E}_{s,a\sim\mathcal{D}} \left[Q_\omega^{\pi_\theta}(s,a) \right] \right) + \frac{1}{2} \mathbb{E}_{s,a,s'\sim\mathcal{D}} \left[\left(Q_\omega^{\pi_\theta}(s,a) - \mathcal{B}_\theta^\pi \bar{Q}_\omega^{\pi_\theta}(s,a) \right)^2 \right].$$

Algorithm 1 Cal-QL pseudo-code

1: Initialize Q-function Q_ω, and optionally a policy π_θ
2: **for** step t in $\{1, \ldots, N\}$ **do**
3: Train the Q-function using gradient descent steps on $J_Q(\omega)$:

$$\omega_t := \omega_{t-1} - \eta_Q \nabla_\theta J_Q(\omega)$$

4: Improve policy π_θ via gradient descent steps on θ with SAC-style entropy regularization:

$$\theta_t := \theta_{t-1} + \eta_\pi \mathbb{E}_{s\sim\mathcal{D}, a\sim\pi_\theta(\cdot|s)} \left[Q_\omega(s,a) - \log \pi_\theta(a|s) \right]$$

5: **end for**

A.2 Hyperparameters and Network Architectures

This subsection details the hyperparameters and network architectures used in this paper.

A.2.1 General Settings

- Observation space: 14-dimensional vector (state variables)
- Action space: 5 continuous management actions
- Batch size: 1024 (samples per training batch)
- Discount factor (γ): 0.95
- Target network update rate (τ): 0.005 (controls soft updates for stability).

A.2.2 Network Architectures

For the actor network:

- Structure: 2 hidden layers (256 units each) with ReLU activation.
- Output layer: Tanh activation to ensure the output action values are bounded within -1 to 1.
- Regularization:
  - Batch normalization to stabilize training
  - Dropout rate at 0.2 to prevent overfitting

For the critic network:

- Structure: 2 hidden layers (256 units each) with ReLU activation.
- Q-function: Quantile regression with 64 quantiles to model value distribution. Quantile regression models the distribution of returns rather than mean values, improving robustness to uncertainty.
- Ensemble: 2 critics (reduces overestimation bias)

A.2.3 Learning Rates

All learning rates employed a cosine annealing schedule, starting at the following initial values and decaying to zero over 4000 iterations. This schedule smoothly reduces learning rates during training, aiding convergence without manual tuning.

- Actor: 1×10^{-4}
- Critic: 1×10^{-4}
- Temperature: 1×10^{-4} (entropy regularization)
- Alpha: 1×10^{-4} (conservative loss weighting)

A.2.4 Optimizer Settings

The Adam optimizer was used for all networks with:

- Exponential decay rates for moment estimates: $\beta_1 = 0.9$, $\beta_2 = 0.999$.
- Numerical stability term: $\epsilon = 1 \times 10^{-8}$
- L2 regularization: 1×10^{-4}

A.2.5 CalQL-Specific Parameters

- Conservative loss weight: 10.0. Higher values enforce stronger punishment on Q-values for out-of-distribution actions. A weight of 10 is in the upper range compared to prior work (often 1–10), emphasizing robustness in management decision-making.
- Action samples to approximate Q-values: 20

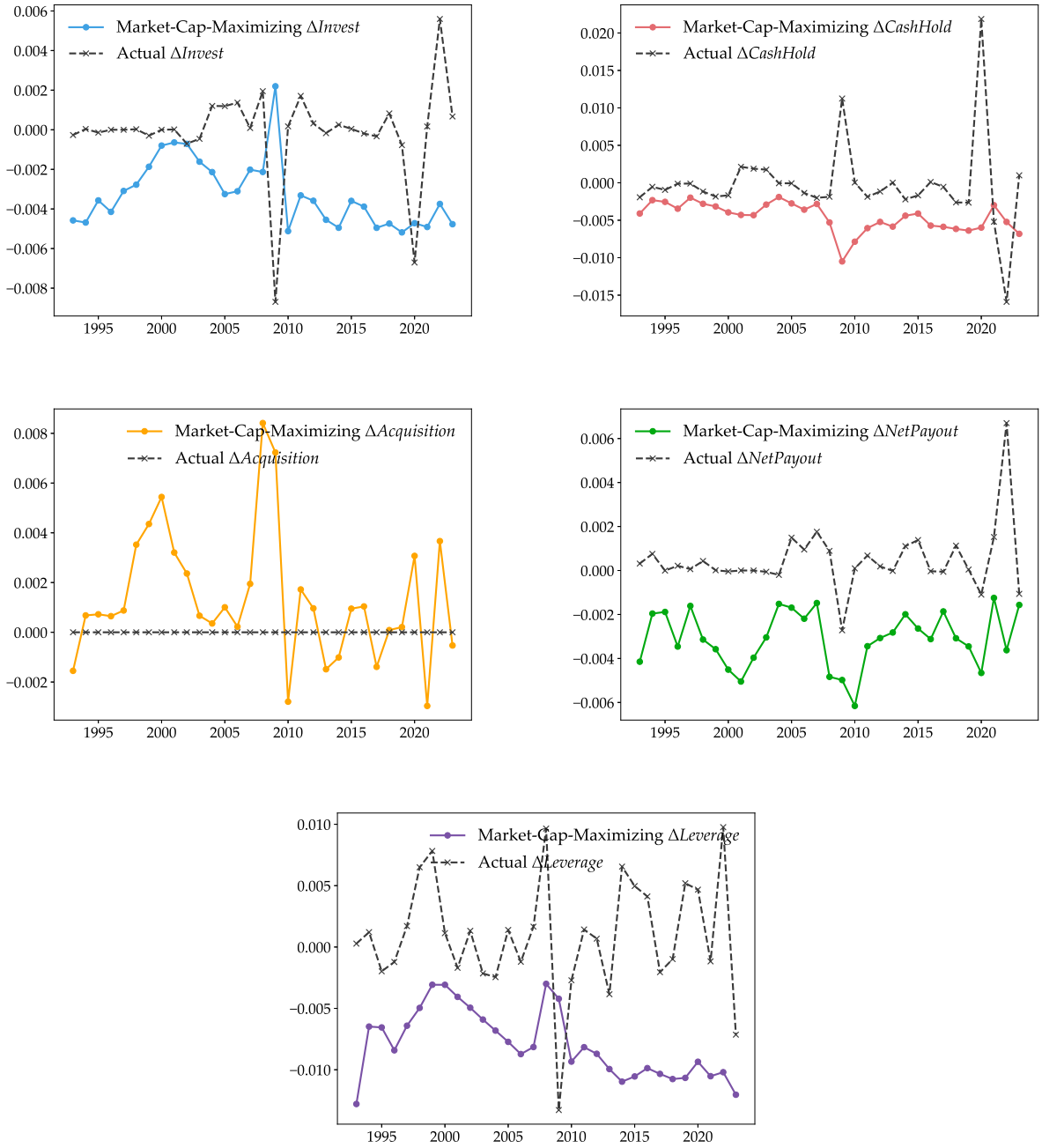
- Initial temperature: 1.0. It determines how much the learned policy is allowed to deviate from the behavior policy (the one that generated the dataset). A low temperature makes the policy stick closely to the behavior policy (more conservative), while a higher initial temperature encourages the policy to be more stochastic. It is automatically adjustable during training by minimizing a specific loss function that encourages the policy to maintain a desired level of entropy. As a result, the temperature is automatically and gradually reduced during training.
- Alpha threshold: 10.0. Alpha threshold sets an upper bound on the Lagrange multiplier α used to enforce conservatism in Q-value updates. α scales the CalQL regularizer, which penalizes overestimation of Q-values for out-of-distribution actions. The threshold prevents α from growing excessively (e.g., due to unstable training), ensuring conservative updates remain bounded. Note that α is a learnable parameter (Lagrange multiplier) that dynamically adjusts the strength of the conservative regularization during training, whereas conservative loss weight is a fixed hyperparameter that directly scales the magnitude of the conservative regularization term.

B Additional Figures and Tables



This figure plots the median values of compensation-maximizing actions and actual actions over the years. Five management actions are included: (1) Change in the investment ratio ($\Delta Invest$) (2) Change in the cash holding ratio ($\Delta CashHold$) (3) Change in acquisition expenditure ($\Delta Acquisition$) (4) Change in the net payout ratio ($\Delta NetPayout$) (5) Change in the leverage ratio ($\Delta Leverage$). Refer to Table B.1 for detailed variable definitions.

Figure B.1: Time-Series Compensation-Maximizing and Actual Actions



This figure plots the median values of market-cap-maximizing actions and actual actions over the years. Five management actions are included: (1) Change in the investment ratio ($\Delta Invest$) (2) Change in the cash holding ratio ($\Delta CashHold$) (3) Change in acquisition expenditure ($\Delta Acquisition$) (4) Change in the net payout ratio ($\Delta NetPayout$) (5) Change in the leverage ratio ($\Delta Leverage$). Refer to Table B.1 for detailed variable definitions.

Figure B.2: Time-Series Market-Cap-Maximizing and Actual Actions

Table B.1: Variable Definitions

Variable	Description
$\Delta Invest$	The change in the investment ratio: $\frac{R\&D_t+0.3\times SG\&A_t+CAPX_t}{Asset_t} - \frac{R\&D_{t-1}+0.3\times SG\&A_{t-1}+CAPX_{t-1}}{Asset_{t-1}}$.
$\Delta CashHold$	The change in the cash holding ratio: $\frac{CashHold_t}{Asset_t} - \frac{CashHold_{t-1}}{Asset_{t-1}}$.
$\Delta Acquisition$	The change in acquisition expenditure: $\frac{Acquisition_t}{Asset_t} - \frac{Acquisition_{t-1}}{Asset_{t-1}}$, where $Acquisition$ represents acquisition-related costs (Compustat item "Aqc").
$\Delta NetPayout$	The change in the net payout ratio: $\frac{Dividend_t+Repurchase_t-Issuance_t}{Asset_t} - \frac{Dividend_{t-1}+Repurchase_{t-1}-Issuance_{t-1}}{Asset_{t-1}}$.
$\Delta Leverage$	The change in the leverage ratio: $\frac{Liability_t}{Asset_t} - \frac{Liability_{t-1}}{Asset_{t-1}}$.
$\Delta MCap$	Growth rate of market cap, calculated as $(MCap_t - MCap_{t-1})/MCap_{t-1}$, where $MCap_{t-1}$ and $MCap_t$ denote the fiscal-year-end market capitalization at fiscal year $t-1$ and t, respectively.
$\Delta Compens$	Growth rate of manager's personal compensation, calculated as $(Compens_t - Compens_{t-1})/Compens_{t-1}$. $Compens_{t-1}$ and $Compens_t$ represent the fiscal-year-end compensation at fiscal year $t-1$ and t, respectively. Compensation includes all potential personal income sources such as salary, bonus, long-term incentive plan payouts, the value of restricted stock grants, options granted during the year, and any other annual pay received by the manager.
$OROA$	Operating return on assets, calculated by dividing operating income before depreciation by total assets.
ROA	Return on assets, calculated as net income scaled by total assets.
GPM	Gross profit margin, calculated as $\frac{NetSale_t-COGS_t}{NetSale_t}$, where $COGS$ is the cost of goods sold.
$InstOwn$	Institutional ownership, calculated as the annual average of quarterly total institutional holdings divided by the quarter-end market cap.
$Amihud$	Natural logarithm of the annual average Amihud illiquidity index: $\log\left(\frac{1}{N}\sum_{d=1}^{N}\frac{\lvert r_d\rvert}{DVolume_d}\right)$, where N is the number of trading days within the year, r_d denotes the daily return at day d, and $DVolume_d$ represents the dollar volume at day d.
$GGDP$	Real GDP growth rate, obtained from FRED.
$NFCI$	National financial conditions index, obtained from the Federal Reserve Bank of Chicago.
$Agency\ Conflict$	The degree of agency conflict, calculated as the Euclidean distance of observed management actions to the RL-outcome actions: Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{MCap}\right)$ − Euclidean $\left(\mathcal{A}_{i,t}^{Actual}, \mathcal{A}_{i,t}^{Compens}\right)$. Five management actions are considered, including $\Delta Invest$, $\Delta CashHold$, $\Delta Acquisition$, $\Delta NetPayout$, and $\Delta Leverage$.

Continued on next page

Table B.1 – *Continued*

Variable	Description
$Restate$	A dummy variable set to one if a firm has material financial restatements in a given year and zero otherwise.
$AllSuit$	A dummy variable set to one if a firm is involved in securities class action lawsuits in a given year and zero otherwise.
$NDSuit$	$NDSuit$ excludes cases that are subsequently dismissed, compared to $AllSuit$.
DA	Discretionary accruals calculated using the models by Jones (1991), Dechow et al. (1995), and Kothari et al. (2005).
$BDDummy$	A dummy variable set to one if at least one backdating event occurred in a firm in a given year. A backdating event is defined as occurring when the strike price is no more than 101% of the lowest daily price within the grant month.
$BDFraction$	The number of backdating events divided by the total manager option grant events within a year.
$PricePattern$	Price pattern ratio, calculated as the proportion of post-transaction buy-and-hold returns to pre-transaction returns: $PricePattern = \prod_{k=-\tau}^{-1}(1+R_{i,k}) / \prod_{k=1}^{\tau}(1+R_{i,k})$, where τ is the estimation window. Two windows are considered: $[-10,+10]$ and $[-20,+20]$, with $k=0$ as the manager trading date. $R_{i,k}$ is either the raw or market-adjusted daily return of firm i.
$BlockOwn$	Proportion of stock held by outside blockholders.
$\hat{\phi}$	Agency conflict index estimated from the structural model developed by Morellec et al. (2018).
$GuideFrequency$	Frequency of management guidance, calculated as the number of annual and quarterly management guidance records within a fiscal year for each firm.
$GuideDummy$	A dummy set to one if a firm announces either annual or quarterly management guidance within a fiscal year and zero otherwise.
$GuideError$	Earnings guidance accuracy, calculated as the absolute deviation between actual earnings and management forecasts, divided by the pre-announcement stock price.
$GuideSurprise$	Management earnings guidance, calculated as $\left(Guide_{i,q} - \overline{Actual}_{i,q}\right) / Price_{i,q}$. $Guide$ represents the latest management earnings forecast, $\overline{Actual}_{i,q}$ denotes the average earnings realized over the previous four quarters, and $Price$ refers to the stock price on the trading day prior to the guidance announcement date.

Continued on next page

Table B.1 – *Continued*

Variable	Description
$AnalystSurprise$	Analyst consensus earnings forecasts, calculated as $\left(Consensus_{i,q} - Consensus_{i,q-1}\right) / Price_{i,q}$, where $Consensus$ is the median of analyst forecasts for the firm-quarter, and $Price_{i,q}$ is the average stock price at the end of the current and previous fiscal quarters.

Table B.2: Summary Statistics: $AgencyConflict$ Estimated from DRL Models

This table reports the summary statistics for $AgencyConflict$ estimated from DRL models, which is calculated according to Equation (4).

Year	N	Mean	SD	P10	P25	P50	P75	P90
All	77186	-0.005	0.017	-0.025	-0.016	-0.005	0.006	0.017
1993	345	-0.010	0.014	-0.026	-0.021	-0.011	0.000	0.009
1994	1043	-0.006	0.013	-0.023	-0.015	-0.005	0.003	0.011
1995	1390	-0.006	0.014	-0.024	-0.016	-0.005	0.004	0.012
1996	1376	-0.005	0.014	-0.023	-0.014	-0.004	0.004	0.012
1997	1511	-0.006	0.014	-0.024	-0.016	-0.007	0.003	0.012
1998	1930	-0.005	0.014	-0.023	-0.016	-0.005	0.004	0.013
1999	2571	-0.004	0.015	-0.022	-0.014	-0.004	0.006	0.015
2000	3200	-0.003	0.016	-0.023	-0.014	-0.003	0.007	0.018
2001	3384	-0.003	0.017	-0.023	-0.015	-0.004	0.007	0.018
2002	3222	-0.003	0.016	-0.022	-0.014	-0.004	0.008	0.019
2003	3075	-0.004	0.015	-0.023	-0.015	-0.005	0.006	0.016
2004	2977	-0.005	0.017	-0.025	-0.016	-0.006	0.006	0.016
2005	2874	-0.005	0.016	-0.025	-0.016	-0.005	0.006	0.016
2006	2818	-0.006	0.016	-0.027	-0.017	-0.007	0.004	0.015
2007	2832	-0.005	0.016	-0.026	-0.017	-0.006	0.005	0.015
2008	3221	-0.005	0.021	-0.030	-0.019	-0.006	0.008	0.020
2009	3108	-0.001	0.020	-0.023	-0.014	-0.002	0.012	0.023
2010	2930	-0.002	0.019	-0.024	-0.013	-0.002	0.009	0.020
2011	2851	-0.005	0.016	-0.024	-0.016	-0.005	0.005	0.014
2012	2811	-0.006	0.016	-0.025	-0.016	-0.005	0.004	0.014
2013	2770	-0.005	0.016	-0.026	-0.016	-0.005	0.006	0.016
2014	2763	-0.006	0.016	-0.027	-0.017	-0.006	0.005	0.014
2015	2736	-0.006	0.017	-0.026	-0.017	-0.007	0.004	0.015
2016	2666	-0.005	0.016	-0.025	-0.016	-0.006	0.005	0.015
2017	2614	-0.006	0.017	-0.026	-0.017	-0.006	0.005	0.016
2018	2618	-0.006	0.017	-0.026	-0.017	-0.007	0.005	0.016
2019	2626	-0.005	0.017	-0.026	-0.017	-0.006	0.006	0.015
2020	2624	0.001	0.017	-0.021	-0.011	0.001	0.014	0.023
2021	2657	-0.006	0.018	-0.028	-0.018	-0.006	0.006	0.017
2022	2837	-0.006	0.019	-0.029	-0.020	-0.007	0.006	0.017
2023	806	-0.006	0.016	-0.026	-0.019	-0.007	0.005	0.015

Table B.3: Robustness of Valuation: Equal Weighting and Alternative Price Impact Value

This table replicates Table 11 with several modifications. Panel A uses equal weighting to construct portfolios. Panel B adjusts the price impact value from 1.5×10^{-8} to 5×10^{-9}. Management earnings guidance is calculated as $GuideSurprise = \left(Guide_{i,q} - \overline{Actual}_{i,q}\right) / Price_{i,q}$. Here, $Guide$ represents the latest management earnings forecast, $\overline{Actual}_{i,q}$ denotes the average earnings realized over the previous four quarters, and $Price$ refers to the stock price on the trading day prior to the guidance announcement date. This firm-quarter-level measure is then aggregated to the portfolio-quarter level using the last-quarter-end market capitalization as the weight. Dollar values of $GuideSurprise$ are calculated according to Equation (8). All dollar values are annualized and in thousands of USD. Standard errors are calculated using a wild bootstrap with 500 samples and reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10%, respectively.

Valuation Data:	*GuideSurprise*	
Agency Conflict:	Low	High
Panel A: Equal-Weighted Portfolio		
0.5m wealth	4.51***	2.91***
without price impact	(0.39)	(0.51)
250m wealth	1531.55***	988.75***
without price impact	(132.95)	(173.28)
0.5m wealth	4.51***	2.91***
with price impact	(0.39)	(0.51)
250m wealth	460.41***	363.8***
with price impact	(36.46)	(54.65)
Panel B: Alternative Price Impact Value		
0.5m wealth	5.2***	3.36***
	(0.59)	(0.41)
250m wealth	803.3***	633.89***
	(86.16)	(73.61)