Striking Oil in the Boardroom:
Overpaying Executives through Manipulating Actual Performance Metrics

Vladimir Atanasov
William & Mary
vladimir.atanasov@mason.wm.edu

Dirk Black
University of Nebraska – Lincoln
dirkblack@unl.edu

Maria Boutchkova
University of Edinburgh
maria.boutchkova@ed.ac.uk

November 2024

Running Head: Striking Oil in the Boardroom

We would like to thank Mark Agerton, Ana Albuquerque, Arthur Allen, Seth Armitage, Cecilia Bosman, Jim Bradley, Brian Cadman, Mary Ellen Carter, Yuzhou Chen, Jeff Coles, Claudia Custodio, Peter Cziraki, Jo Danbolt, Shane Dikolli, Alex Edmans, Ellen Engel, Bo Gao, Scott Gibson, Brandon Gipper, Hugh Hetherington, Margot Howard, Denise Jones, Brian Main, Hugh Marble, Mary Marchica, Melissa Martin, Michal Matejka, Chris McCoy, Angus McPhail, John Merrick, Paige Patrick, Andrea Pawliczek, Xiaoxia Peng, Marc Picconi, Robert Pozen, Edward Prysor-Jones, Talis Putnins, Shawn Shi, Bill Skimmyhorn, Kim Smith, Pauline Weetman, Kyle Welch, Ben Whipple, Nick Wilson, Biyu Wu, Moqi Xu, an anonymous director at one of the Big Four auditing firms, anonymous former employees at Chesapeake Energy and Schlumberger Limited, the attendees of the UBC Summer Corporate Finance Conference in Vancouver, seminar participants at the University of Surrey, the University of Sterling, William and Mary, University of Edinburgh, University of Nebraska - Lincoln, the Corporate Governance and Executive Compensation Research Series, and the 2019/2020 student cohort of Investment and Securities Markets at the University of Edinburgh.

Corresponding author: Dirk Black, University of Nebraska - Lincoln, HLH 445K, 730 N. 14th St., Lincoln, NE 68588; 1-402-472-2337; dirkblack@unl.edu

Striking Oil in the Boardroom:

Overpaying Executives through Manipulating Actual Performance Metrics

ABSTRACT

Using hand-collected data, we examine the distribution of performance metrics used to calculate executive compensation in a sample of US oil and gas firms. We find that the distribution of actual-minus-target differences is significantly discontinuous at zero over the 13-year period from 2006 to 2018. Executives are nearly three times more likely to just beat than to just miss performance targets. When we split metrics into transparent and non-transparent measures, we find significant discontinuities only in the non-transparent group, even after 2011 when the US Securities and Exchange Commission (SEC) updated its guidance for non-GAAP performance metrics in proxy statements. Significant discontinuities disappear when firms are financially distressed and have stronger monitoring. Our findings suggest that managers can routinely manipulate realized performance metrics to increase performance-based compensation. Our framework can help investors in their firm governance assessments and the SEC in its redesign and enforcement of disclosure rules for performance metrics.

Keywords: executive compensation, manipulation, transparency, performance measures, discontinuity

JEL Classifications: G34, J33, M12, M52, Q40

Data Availability: Data are available upon request. Wharton Research Data Services (WRDS) was used in preparing this research. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

I. INTRODUCTION

In August 2022, the U.S. Securities and Exchange Commission (SEC) issued its final rules (effective for fiscal years ending on or after December 16, 2022) on Pay vs Performance in the implementation of section 14(i) of the Dodd-Frank Act. Although the final rules make significant progress towards increasing transparency in the way performance-based executive compensation is disclosed, they stop short of addressing the concerns of institutional investors and analysts, echoed in the Wall Street Journal by former SEC Commissioner Jackson and former president of Fidelity Investments Robert Pozen, that non-GAAP adjustments to performance metrics can be "used to justify windfalls to underperforming managers" (Jackson and Pozen 2019). The Council of Institutional Investors (CII) and the Chartered Financial Analyst (CFA) Institute (Bertsch and Mahoney 2019, Schacht and Peters 2019) requested amendments to the SEC rules that "all non-GAAP financial measures presented in the proxy statement Compensation Discussion & Analysis (CD&A) be subject to all disclosure requirements and reconciled with their GAAP counterparts." These amendments supported by the International Corporate Governance Network, Teamsters Union and Americans for Financial Reform Education Fund among others (Waring 2022, Hall 2022, AFREF 2022).

The final SEC rules still help investors analyse the link between executive compensation and firm performance over time and across firms by requiring machine-readable disclosure of the last five years' actual pay and three performance measures: (1) annual total shareholder return (TSR) next to the TSR of the company's peer group, (2) net income, and (3) the "most important" company-selected financial performance measure. However, the lack of reconciliation requirement for the latter and all other performance measures used in determining executive pay perpetuates opacity in the determination of performance-based executive compensation. In this paper we show that such opacity affords managers opportunities to inflate their compensation.

We document that under the current rules, the proxy statement definitions of performance metrics used to determine executive compensation in the oil and gas industry exhibit high variability across firms and time. When we measure how the actual results disclosed in proxy statements correlate with well-defined database metrics, more than half of them do not correlate sufficiently to allow comparability across firms and time and are, therefore, non-transparent.

Looking at the distribution of actual-minus-target differences, we find that the executives in our final sample of 62 public US oil and gas firms over thirteen years (2006-2018) meet non-transparent (transparent) targets 74 (68) percent of the time – statistically significantly higher than 50 percent per a strict definition of randomness and the upper bound of 60 percent as per industry guidelines (Saliba 2016). Moreover, executives are five times more likely to just meet or beat non-transparent performance targets than just miss them. These conclusions are based on a highly statistically significant discontinuity detected by a new test (Bugni and Canay 2021), which builds upon the classic test by McCrary (2008), but relies on weaker conditions. We confirm the findings from the Bugni and Canay (2021) test with results from a test by Byzalov and Basu (2019), which was specifically designed to address known methodological issues in existing earnings discontinuity research.

Having established evidence of discontinuity in actual-minus-target differences in the full sample (1,787 firm-year-metric observations), we further examine which observations exhibit the strongest degree of discontinuity. We develop a new approach for classifying metrics as transparent and non-transparent following the intention behind GAAP – decision usefulness deriving from comparability across firms and consistency over time. To do this, we regress the firm-metric time series of actual performance metric values collected from proxy statements on corresponding database variables with the same definition and classify metrics as non-transparent if no standard database variable explains 80% or more of their variation. These database variables are not necessarily GAAP metrics. However, because database variables are computed identically for all firms and years and clearly defined, we use them as benchmarks for transparency. The classification is only possible for firm-year-metrics that have sufficiently long time series (1,268 observations), of which more than half (656) belong to non-transparent metrics.

We test for significant discontinuity at zero of actual-minus-target differences on sample splits along multiple criteria that have been previously associated with the costs and benefits of manipulation. These sample split tests show that the discontinuity disappears for: 1) transparent metrics; 2) metrics covered by analysts; 3) firms in financial distress whereby creditors add scrutiny; 4) firms with higher governance scores; and 5) firms with a higher quality whistleblower program in place.

We form a simple firm-year measure of executive ability to achieve their targets – the proportion of performance metrics met or exceeded out of all metrics used in a given year – and find that this proportion is greater or equal to ½ 70.5% (76.3%) of the time for transparent (non-transparent) metrics and 70% (77.8%) of the time for non-manipulable (manipulable) metrics. We then regress bonus compensation, equity based compensation, and total compensation on the firm-year measure of executive ability, and find that it is only significant in explaining compensation for non-transparent or manipulable metrics.

We reject several alternative explanations for our findings. We perform robustness tests where we subtract the target of firm j from the actual result of firm i, and show that the discontinuity remains even if part of it may be driven by self-selection in setting easy targets. Moreover, the discontinuity remains only for non-transparent metrics following updated SEC guidance for non-GAAP disclosures in proxy statements in 2011. Untabulated tests show that negative Say-on-pay votes are more likely to be observed following years when the CEO compensation contract is more manipulable. Overall, we conclude that the discontinuity is more likely driven by rent extraction than by efficient determination of executive pay.

Our setting differs from the large literature on meeting or beating earnings goals (most recently reviewed by Burgstahler and Chuk (2017), which is based on reported numbers in annual financial statements that are available in most widely used financial databases (e.g., S&P Capital IQ or Compustat). The widely reported metrics differ from the ones used for performance-based compensation, which are only available in proxy statements. There is one commercially available database that compiles proxy statement compensation data (ISS Incentive Lab), but it records only performance metric targets and does not capture actual results. The difficulty in collecting data on actual results from proxies is one reason for the lack of evidence on the discontinuity of actual-minus-target values until now. This opaque setting, however, offers an opportunity for sharper conclusions about the abuse of proxy statement performance metrics to boost executive pay in contrast to the earnings manipulation literature, where evidence of poor earnings quality can be related to innate firm performance (Dechow, Ge, and Schrand 2010). Specifically, a large part of our data consists of direct metrics of production, costs, reserves, debt, etc., in addition to earnings.

The different information content in the two sets of numbers (those in annual reports versus those in proxies) is also underpinned by the stark contrast in the legal ramifications of their misuse. Not only are adjusted performance metrics reported in form 10-K subject to mandatory reconciliation with their GAAP equivalents, but irregularities found are subject to legal recourse under the Securities Act of 1933. By contrast, proxy statement performance targets are excluded from reconciliation with GAAP (disclosed *actual* performance metrics are required to be reconciled to GAAP – SEC 2018), may be non-earnings-based, and are more legally defensible.

Our focus on a single industry improves the internal validity of the analysis because it offers a greater number of firm-year-metric observations due to the relative homogeneity of the business model in the oil and gas sector. This homogeneity yields a similar mix of performance metrics being used by many firms, such as production volume/growth, reserves, and costs, in addition to earnings-based metrics. The single-industry focus provides greater comparability of the various performance metrics while keeping the hand-collection task manageable.

At the same time, our findings for the oil and gas industry are indicative of similar practices across all US listed firms for several reasons. First, the use of performance metrics for contingent bonus and stock-based compensation is pervasive (as of 2018, around 58 percent of total pay for S&P 500 firms; Roe and Papadopoulos 2019). Second, recent evidence suggests that for subsamples of firms from multiple industries, but only based on earnings-related measures, the use of non-GAAP metrics may be associated with overpaying some executives (Guest, Kothari, and Pozen 2022; Kim and Shin 2019). Third, the oil and gas sector does not have poorer governance relative to US industries along the different dimensions of governance quality compiled by Sustainalytics and Institutional Shareholder Services Inc. (ISS), as well as average executive compensation levels compiled by BoardEx, (industry comparison graphs available upon request).

Our study makes the following contributions. First, we offer discontinuity-based evidence of manipulation of a comprehensive set of performance metrics that is consistent with the indirect rigging evidence provided by Morse, Nanda, and Seru (2011), who show a strong relation of CEO pay to the higher of stock return or accounting return. Second, our sample split results are consistent with opportunism, informing the debate over motives for using adjusted performance metrics (Black, Christensen, Ciesielski, and Whipple 2018, 2021; Bhattacharya, Black, Christensen, and Larson 2003; Isidro and Marques 2015; Young 2014). Third, our findings raise concerns that executive pay practices often fall short of the theoretical optimal prediction that linking pay to performance necessarily addresses agency problems and information asymmetry (Edmans, Gabaix, and Jenter 2017; Murphy and Jensen 2018).

Furthermore, our results have implications for removing the exception in Instruction 5, Item 402(b) of Regulation S-K for performance targets (and enforcing SEC guidance for performance actuals contained in proxy statements). Lastly, our discontinuity tests can be used by the SEC to pre-filter company filings before performing closer examinations and provide grounds to pursue legal action. Such added efficiency is important following the Sarbanes–Oxley Act of 2002 and its Section 304, which does not allow private right of action (Soondar, Major, and Hines 2010) and empowers only the SEC to recapture executive pay in cases of misconduct. Given the SEC's limited resources to pursue cases other than outright fraud, automating detection plus machine readable data on performance metrics starting in proxy season 2023, can improve the SEC's ability to encourage transparency, accountability, consistency, and comparability.

II. INDUSTRY BACKGROUND

We start our analysis by addressing the concern that our choice of the oil and gas industry may make our findings not generalizable, because this industry is very different than most other industries. Based on industry averages of several governance and compensation characteristics over the last decade, the oil and gas industry ranks roughly in the middle relative to other sectors. Among nine governance quality characteristics compiled by Sustainalytics since 2009, the oil and gas sector is different only in board gender diversity, while among six governance characteristics compiled by ISS since 2007, oil and gas firms rank favorably among all sectors by lack of unequal voting rights and dual common class stock. Lastly, by dollar amounts of executive compensation and its components, as well as the ratio of equity-linked pay based on BoardEx data, oil and gas firms rank consistently in the middle among other sectors (all industry comparisons available upon request).

The oil and gas sector may look very similar to other industries on many governance measures, but it provides unique benefits for empirical analysis. The homogeneity of the business model in this sector, and especially throughout the shale technological era (starting around the mid-2000s and coinciding with our sample period), means that firms rely on similar performance metrics, many of which are unique to the oil and gas industry. These metrics include oil and gas production volume/growth, reserves, development costs, production costs, leverage, safety, as well as the typical earnings-related metrics. Table 1 reports 16 different categories of performance measures used by the oil and gas firms in our sample.

Common performance trends in the US oil and gas sector – over-borrowing, over-investment, negative operating cash flows, and low shareholder returns – have been highlighted by practitioners and academics alike (AAPG 2018; Kleinberg, Paltsev, Ebinger, Hobbs, and Boersma 2018; Weijermars and Watson 2011). The similarity of firms within the oil and gas sector enhances our ability to aggregate and reconcile otherwise disparate performance metrics with varying definitions and use between our sample firms and across time during our sample period.

The US oil and gas sector has been criticized for growing production without consideration for long-term sustainability (McLean 2018; Denning 2018). Industry analyst reports often attribute the inefficient excess growth of production (and associated downward pressure on prices) to the ubiquitous tying of oil and gas firm executive compensation to production volume and growth targets. We scrutinize these assertions by examining whether discontinuities are present separately in several groups of metrics including production.

Oil and gas firms appear largely homogeneous in their overall compensation philosophy, in part because they often use the same compensation consultants. Our sample firms most often employ Meridian Compensation Partners, Hewitt Associates (who spun off Meridian after the 2009 disclosure rule change on consulting fees[1]), and Towers Perrin and predecessors, consistent with the compensation consulting analysis in Chacon, Gordon, and Yore (2021). Interestingly, we observe that many of the firms in our sample switch from a multiservice consultant to the related newly spun-off specialist consultant after the 2009 rule disclosure change. This observation is consistent with Chu, Faasse, and Rau (2018), who find that the firms most likely to switch after the 2009 reform have higher CEO pay and boards that are under stronger CEO influence.

In compliance with the 2006 SEC compensation disclosure rules, many firms began reporting performance benchmark targets and their achieved (or actual) levels in their 2007 proxy statements (form DEF14A), reflecting executive compensation for fiscal year 2006.

III. HYPOTHESES

Discontinuity as Evidence for Manipulation

Many previous studies identify suspicious low-probability patterns of discontinuities in financial data. The seminal study of Christie and Schultz (1994) shows clustering of quoted bid and ask prices at round values. Bollen and Pool (2009) find hedge fund return discontinuities. Bernhardt and Davies (2005) document artificially boosted performance at the end of reporting periods by mutual funds. Lie (2005) uncovers extremely favorable timing of executive option grants. Burgstahler and Dichev (1997) and Bartov, Givoly, and Hayn (2002) report sharp discontinuities in the distribution of reported earnings relative to zero or forecasted earnings. The

[1] https://www.sec.gov/rules/final/2009/33-9089.pdf.

identified patterns in these studies have caused regulatory action, private litigation, and other responses that have addressed underlying opportunistic behavior.

Bennett, Bettis, Gopalan, and Milbourn (2017) apply the idea that discontinuities in the distribution of reported financials are a red flag for manipulation to the study of performance-based compensation. Using a large dataset of companies that employ earnings-per-share (EPS) and other earnings-related metrics as performance targets, they show that the distribution of reported-minus-target performance is discontinuous at zero – companies are much more likely to report accounting numbers that just exceed compensation targets than ones that are just below target. The authors argue that this discontinuity implies that executives take actions to manipulate earnings and other reported financials in order to meet performance goals. Bennett et al. (2017) necessarily restrict their analysis to a small number of widely used metrics. In their sample, while 88 percent of firms tie their grants to a performance metric, only 42 percent have a common metric (EPS) that can be analyzed across multiple firms. In addition, Bennett et al. (2017) do not observe the actual realized values that companies use to determine whether executives have met their performance targets. Instead, they substitute reported earnings, which we show can differ significantly from the actual earnings metrics relevant for compensation.

Other large-sample studies of manipulation of performance metrics relevant for executive compensation use aggregate analyses that often mask important differences in specific industries or firms. For example, Guest et al. (2022) find some evidence that S&P 500 firms that report non-GAAP earnings pay their CEOs excessively. Similarly, using a single year of data for a sample of firms in the S&P 1500 index, Curtis, Li, and Patrick (2021) find that CEOs are less likely to miss compensation minimum earnings thresholds by excluding non-transitory losses and expenses from earnings. Consistent with our findings, Kim and Shin (2019) find that the likelihood to meet or beat performance targets goes hand in hand with boards adjusting actual earnings. Focusing on the informativeness of non-GAAP EPS, Black, Black, Christensen, and Gee (2021) show that non-GAAP EPS are more useful for assessing firm value when disclosed in both earnings announcements and proxy statements (11% of cases in the period 2009-2015).

Discontinuity and Bunching

Optimal incentive contract design is a well-established sub-field of labor economics. Holmstrom and Milgrom (1991) prove that unbalanced multitask incentives can produce distortion, and nonlinear incentive contracts entice manipulation (Baker 2002; Jensen 2003; Murphy 2013). In the context of executive compensation, building on Edmans, Gabaix, Sadzik, and Sannikov (2012), the model in Marinovic and Varas (2019) endogenizes manipulation and shows that non-coopted boards can anticipate manipulation and design optimal contracts with this in mind. In the context of earnings manipulation, Guttman, Kadan, and Kandel (2006) present a model where discontinuities emerge endogenously when an informed manager with stock-based incentives trades off the benefits and costs of manipulation. The only model we are aware of that explicitly allows for a co-opted board is that in Morse et al. (2011) showing how boards can link the CEO contract to a performance metric that is easier to meet or beat. However, these papers rely on data derived from required financial statements for their modelling assumptions.

When we apply existing modelling approaches of incentive compensation to the relatively opaque contracting environment in the oil and gas sector, we note that we are dealing with multiple performance metrics in a nonlinear incentive contract, many of which are likely non-transparent and difficult to understand for the investment community. This lack of transparency also means that any legal claim of measure manipulation is harder to prove using only publicly available data.

Consider a setting in which boards are not co-opted and managers are responding to incentives by increasing effort when they are very close to the target level of a given metric. In the absence of uncontrollable factors or noise, we expect to observe "bunching" of mass just above zero in the distribution of actual-minus-target differences. Importantly, researchers have shown that in the presence of noise or uncontrollable factors, the incentive for exerting extra effort when close to the kink is weakened and the bunching disappears (Brehm, Imberman, and Lovenheim 2017). Therefore, the existence of bunching/discontinuity in the distribution of metrics affected by noise or factors outside the control of managers is most likely caused by managerial manipulation relative to managerial effort.

Several common performance metrics employed by the oil and gas industry are less prone to intentional adjustments (e.g., production volume and reserves). These metrics depend only on physical measures of volume and the price of oil or gas, which is determined by global commodity markets. Furthermore, for the oil and gas industry our period of analysis has been characterized by high technological uncertainty. Shocks to extraction technology, discoveries of previously infeasible reserves and regulatory interventions have resulted in greater unpredictable variation in production, costs and reserves. Therefore, in this industry even the least adjusted metrics are strongly dependent on uncontrollable factors, and any evidence of discontinuity is unlikely to reflect effort as opposed to manipulation.

Adjustments to Executive Compensation Contracts

Lie (2005) and follow-up work (Heron and Lie 2007, 2009) exposed option backdating affecting at least 30 percent of listed firms between 1996 and 2005. These findings led to years of litigation which presumably eliminated the practice (Choi, Wiechman, and Pritchard 2013). The removal of the intrinsic value method of reporting stock option compensation expense by FAS

123R beginning in 2005 further decreased their use. In their yearly report on trends in pay among executives of US listed firms, ISS (Institutional Shareholder Services) report that stock options represented 30 percent of total CEO pay in 2009, but only 17 percent as of 2018 among S&P 500 firms (Roe and Papadopoulos 2019). Instead, firms are shifting towards performance-based stock and bonus plans (58 percent of total pay in 2018).

There are many possible ways in which compensation committees influence performance-based executive pay. For example, executive pay can be inflated by strategically manipulating the composition of peer groups (Albuquerque, De Franco, and Verdi 2013; Faulkender and Yang 2013; Shin 2016). In addition, there are at least four ways that compensation committees may adjust performance targets and/or actuals to benefit executives: 1) strategically choose the set of metrics relevant for compensation; 2) adjust the weight assigned to each performance metric ex post; 3) set target levels at or below projected levels ex ante; and 4) manipulate actual performance levels during the performance period or ex post to meet or beat previously set targets, consistent with the bonus plan hypothesis of Watts and Zimmerman (1990).[2]

Little evidence exists of the prevalence of the first method -- strategic target metric choice, and in the oil and gas industry the set of metrics used is generally stable over time (Figure 1). The second method (switching weights on performance metrics ex post depending on which one is more likely to be achieved) is termed "rigging" by Morse et al. (2011). They provide some indirect panel regression evidence for its use. A recent lawsuit against Netflix (City of Birmingham Relief and Retirement System v. Reed Hastings, Case No. 3:18-cv-02107) alleges that Netflix's

---

[2] One more method highlighted in Kalb, Phillips, and Silberglied (2016) is for firms to choose not to disclose target levels at all, especially following industry or overall economic downturn, thus hiding potential adjustments that benefit executives. Taking this possibility into account means that the proportion of observations meeting targets, and the discontinuities influenced thereby, may be understated relative to the true underlying proportion of firms whose executives meet or beat easier-to-reach, but undisclosed, targets.

compensation committee utilized the third method of ex ante setting target levels just below the predictions of future actual values to ensure that key executives received the maximum possible bonuses.[3]

We focus our main efforts on documenting the usage of the fourth manipulation method – adjusting actual performance measure values during the performance period or ex post such that they meet or beat performance targets. This method has not been explored by much of the existing literature because of the difficulty of obtaining data on actual values of performance metrics from compensation contracts.[4] The actual results given in proxy statements may differ from the closest comparable GAAP metrics (or from other observable metrics in large-sample financial data repositories) if: 1) firms erroneously assume that the exception for reconciling performance targets in compensation contracts to the closest comparable GAAP metric in Instruction 5, Item 402(b) of Regulation S-K also applies to actual results; 2) firms are reporting executive performance prior to updated SEC guidance in 2011 on non-GAAP disclosure that requires reconciliation of actual performance measures to the closest comparable GAAP metric in proxy statements; 3) firms are not aware of updated SEC guidance from 2011 that requires reconciliation of actual performance to the closest comparable GAAP metric in proxy statements; 4) firms purposely choose regulatory non-compliance (Robinson, Xue, and Yu 2011); 5) firms deem non-standard performance metrics as more indicative of manager effort (regardless of their reconciliation choice); or 6) firms choose performance metrics with no GAAP counterpart (such as production). Bennett et al. (2017) focus exclusively on performance targets and reported earnings results deriving from financial statement filings to infer actual-minus-target differences. With our hand-collected data, we perform a more

[3] The defendants (Netflix and its directors) have agreed to cover up to $800,000 of the plaintiff's legal fees and up to $10,000 in penalties – a settlement approved in July 2020.
[4] Armstrong, Chau, Ittner, and Xiao (2022) is a notable and recent exception.

refined analysis by directly observing and examining differences between performance targets and actual performance disclosed directly in the CD&A sections of firms' proxy statements.

We note that both the third and fourth method of manipulation (setting target levels at or below projected levels ex ante and manipulating actual performance levels during the performance period or ex post to meet or beat previously set targets) will result in a discontinuity around zero in the distribution of actual-minus-target differences. In our first hypothesis, we do not distinguish between the two methods and formulate it as follows:

Hypothesis 1: The distribution of actual–minus–target differences of performance metrics will exhibit a discontinuity at zero.

Hypothesis 1 is not necessarily obvious as may appear in light of Bennett et al. (2017) and the broader earnings discontinuity literature. In our sample of oil and gas firms a substantial number of observations represent physical measures of production or reserve volume or verifiable dollar amounts reflecting oil and gas prices (revenue and dollar value of reserves), which may make manipulation more difficult. In addition, an analysis on the full sample is important because sub-sample splits by type of metric may suffer from low statistical power (sample split results are presented in the online appendix).

Prior research indicates that some managers manipulate earnings (Dechow, Richardson, and Tuna 2003; Bergstresser and Philippon 2006; Cheng, Harford, and Zhang 2015), with some evidence that non-GAAP metrics can be prone to manipulation in ways that benefit executives (Guest et al. 2022; Jackson and Pozen 2019). Moreover, studying differences between non-GAAP metrics and standardized or GAAP-based metrics based on audited financial statements from large databases like Capital IQ allows us to distinguish whether compensation committees or managers choose to manipulate targets (i.e., third manipulation method) or actual values (i.e., fourth manipulation method). Manipulating targets requires predictable values of performance metrics. In fact, the Netflix lawsuit alleges that Netflix used a GAAP measure of revenue that was very predictable to set target levels. In contrast, manipulating actual values to exceed preset targets requires flexibility to add or exclude various items in the calculation of actual performance. These modifications necessarily result in actual performance metrics that diverge from standardized or GAAP-based metrics (assuming targets cannot be achieved using unmanipulated performance).

Anecdotal examples (Morgenson 2011, 2014), supported by survey evidence (Wilkins, Hermanson, and Cohen 2016) confirm that compensation committees adjust target performance metrics, often excluding the impact of extraordinary items.

We test whether types of metrics that are either easier to manipulate ex ante, or harder to detect if manipulation had happened ex post, are more likely to exhibit discontinuity by constructing two new measures: manipulability and transparency. Manipulability is a binary indicator equal to 0 if a given metric is objectively verifiable and simple to compute and 1 otherwise. We next develop a novel method to classify proxy statement actual performance metrics as transparent and non-transparent and study their distributions separately. Our measure of transparency is designed to capture the ease of detecting irregularities in actual performance metrics ex post across firms or time by analysists or observers. In addition to the splits by manipulability and transparency, we analyze the clarification made by the SEC in 2011 that reported actual performance metrics are also subject to the GAAP reconciliation requirement and not exempt as per Instruction 5, Item 402(b) of Regulation S-K applicable to the CD&A section of the proxy statement. The reconciliation requirement should reduce information asymmetry and make it easier for analysts to detect irregularities and thereby reduce the incentives of executives to engage in manipulation.

We formulate our second hypothesis as:

Hypothesis 2: The distributions of actual–minus–target levels of manipulable (non-transparent) performance metrics will exhibit stronger discontinuities than will the distributions of non-manipulable (transparent) metrics (and especially so before the 2011 SEC clarification).

Factors Affecting the Costs and Benefits of Actual Performance Metric Adjustments

We examine a series of internal and external factors that increase the scrutiny executives face. Financial distress and potential bankruptcy proceedings add a level of legal scrutiny over executives that they are normally immune to. Private litigation in the US is not very effective in recovering presumed excessive executive compensation. The main legal grounds for raising such claims are the Securities Exchange Act of 1934 (§10(b) and §14(a)) and state law claims for breach of fiduciary duty and waste of corporate assets. These actions rarely survive motions to dismiss (Soondar et al. 2010). SOX §304, the Troubled Asset Relief Program (TARP), and some state laws empower the SEC and other federal authorities to pursue legal action and recoup compensation in case of misconduct. However, the resources and capacity of these institutions are limited; thus, these institutions pursue only the most egregious cases of fraud. Even one of the most notorious cases of an overpaid executive, the class action suit against Cheniere Energy seeking recovery of $2 billion worth of stock awards, was settled with the defendant covering legal fees and agreeing to limit future stock awards but without a clawback of pay.[5] The additional provisions on disclosure

[5] Cheniere Energy's CEO Charif Souki made national headlines as the highest-paid executive in the U.S. when the vast majority of his total compensation came from stock awards — $132.93 million of $141.95 million in 2013. Cheniere lost $507.9 million that year and had not made a profit until then. The class action not only disputed the excessive pay but also had a basis in the fact that the abstentions in the shareholder vote for the incentive plan were not counted correctly under Delaware law, and the board, chaired by Souki, who announced the outcome of the vote to be "in favor" of the plan stood to benefit greatly by the stock awards (Eaton 2014).

of executive to median pay ratio, voluntary claw-backs, and say on pay, mandated by the Dodd-Frank Act of 2010, do not appear to have altered the very low success rate of motions to recoup executive compensation (Crane 2013).

When firms become financially distressed, they face increased scrutiny by debtholders who have additional rights to recoup compensation under the Bankruptcy Code. At the extreme, when a firm becomes insolvent, the definition of fiduciary duty of managers changes to include all remaining claimholders, not just shareholders (Becker and Strömberg 2012). The threat that excessive executive compensation can be deemed a "fraudulent conveyance" (Atanasov, Black, and Ciccotello 2011) imposes greater risk and costs when attempting to overpay executives or alter performance metrics.

Finally, we examine other factors expected to affect the relative costs and benefits of actual performance measure adjustment. Specifically, we test how the following factors influence actual-minus-target discontinuities around zero: 1) analyst following and analyst metric coverage; 2) short interest; 3) effective tax rate; 4) making a profit or a loss in a given year; 5) corporate governance score; 6) bribery and corruption; 7) whistleblower programs; 8) CEO-Chair duality; and 9) board independence.

Burgstahler and Dichev (1997) first document that discontinuities in earnings are weaker among firms with lower recent profitability, consistent with the hypothesis that these firms have lower net benefits of managing earnings. This line of analysis is related to the increased probability of CEO dismissal following poor performance which may also increase the incentives for achieved results adjustment. Burgstahler and Chuk (2017) show that the discontinuity in earnings is less prominent for firms with greater analyst following. We extend this idea at the metric level and classify metrics based on whether analysts provide forecasts close to the metric's target.

Bhattacharya, Christensen, Liao, and Ouyang (2021) find that the threat of increased short selling significantly curbs aggressive non-GAAP disclosures. The effect is stronger when pre-disclosure information asymmetry is high and corporate governance is weak.

Firm sensitivity to taxation depending on its ability to shield revenues with interest payments or other deductible expenses will affect the benefits of performance metric management (Dyreng, Hanlon, and Maydew 2008). Finally, firms with weaker corporate governance have been shown to exclude more persistent income statement items from non-GAAP earnings (Frankel, McVay, and Soliman 2011). Summarizing all the evidence on disappearing discontinuities under greater regulatory, market, and corporate governance scrutiny, we formulate our last hypothesis:

Hypothesis 3: A discontinuity in the distribution of actual–minus–target differences of performance metrics is more likely when the net benefit of meeting or beating targets is higher.

IV. DATA AND METHODS

We collect all proxy statements for the period 2006 – 2018 for 86 oil and gas firms that have stock price data available and at least seven proxy statements during 2006 – 2018.[6] Our decision to require at least seven proxies focuses the hand collection effort on firms that are more likely to allow classification of metrics into transparent and non-transparent types as well as cover the established firms that are more likely to have active ongoing business operations. Of these 86 firms, we retain 69 firms that mention, use, or report performance metrics. We further limit our final sample to 62 firms when we require data to compute actual-minus-target differences of performance metrics. We collect financial, executive compensation, and stock market data from

[6] There are 204 (149) listed energy firms with at least one (three) proxy statement(s) that have stock price data before 2010 to meet the minimum time-series of four annual observations.

Capital IQ for the period 2006 – 2018. We obtain information on board composition for the period 2006 – 2018 from BoardEx. Finally, we collect Environmental, Sustainability, and Governance (ESG) ratings from Sustainalytics.

The typical firm in our sample employs a portfolio of metrics, for which achieving their desired target levels will trigger the award of a cash bonus and/or stock-based compensation for its top executives. In some cases, different metrics are applied for bonus and stock awards. Some firms report the weights applied for each metric and details on how they trigger different types of compensation, but disclosure practices are not consistent. Therefore, we opt for collecting all available data regardless of the level of detail.

We collect performance targets and actual results for fiscal years 2006 through 2018 and compute an actual-minus-target difference measure.[7] This measure is defined as the difference between the actual result and the target divided by the target to arrive at a relative rate of beating or falling short of the target regardless of units of measurement. Compensation committees will often set two targets for a given metric – one more optimistic (stretch) and another easier to reach (target).[8] We only use the latter that triggers the payment of bonus or incentive compensation. For positive performance benchmarks such as return on capital, production volume, or reserve replacement, the actual-minus-target difference is positive (negative) when the desired outcome is above (below) the set target. For negative performance benchmarks such as net debt, general and administrative (G&A) costs, or finding costs, the actual-minus-target difference is positive (negative) when the desired outcome is below (above) the set target (i.e., we multiply these actual-

[7] See Appendix A for variable descriptions. A workflow description of collecting all performance-based compensation data in our study is available upon request.
[8] Occasionally, there are three goalposts in setting target levels: threshold, target and stretch. For consistency with compensation contracts with two or only one goalpost, we use target across the board.

minus-target differences by negative one).

In coding metrics as positive or negative, we note that the desired direction is typically stated in the proxy statements themselves. In the case of leverage, the desired level could be approached from above or from below given our theoretical understanding of optimal leverage. In our hand-collected data, the desired direction for the level of debt was indicated by compensation committees to be downwards because stretch goals were always lower than target goals.

The first characteristic on which we perform a sample split discontinuity analysis is manipulability. The oil and gas industry is attractive because it uses a relatively large number of objectively verifiable metrics, which would be harder to manipulate. We apply the following criteria in designating a metric as non-manipulable ex ante: a physical quantity tracked by an independent agency (e.g., U.S. Energy Information Administration (EIA) or the Railroad Commission of Texas (RRC)), a dollar amount resulting from multiplying a quantity by the observable market price of oil/gas, or an amount verifiable by a counterparty in a capital market transaction, e.g., bonds outstanding or stock market capitalization or the sale price of a fixed asset. The types of metrics in the non-manipulable group include: production volume, revenues, reserves, debt outstanding, total shareholder return and capital expenditure. All other internally reported metrics, e.g., costs, earnings, return ratios and their adjusted variants are classified as manipulable or equal to 1. We concede that this classification approach may be debatable and construct a second metric whereby we refrain from subjective judgement and let the data determine the classification.

In particular, we develop a statistical procedure to classify which actual performance metrics are transparent and which are not. This procedure directly establishes the difference between a metric used for compensation and the most similar database metric reported by the same firm. In most cases, proxy statements do not disclose whether a target performance compensation metric follows GAAP (as allowed by the exception in Instruction 5, Item 402(b) of Regulation S-K).

In developing the transparent/non-transparent classification, we attempt to rely on the intuition espoused by the Financial Accounting Standards Board (FASB) to encourage firms to provide investors with decision-useful information that is consistent across time and comparable across firms (Black, Christensen, Ciesielski, and Whipple 2021). Database metrics available in widely used vendor data products like Capital IQ and Compustat from Standard and Poor's, Bloomberg, and Thomson Reuters have clear definitions that generally provide consistency and comparability similar to GAAP metrics found in the audited financial reports upon which these data repositories rely. In our case, we assume that metrics disclosed in proxy statements for executive compensation purposes are increasingly transparent as they are increasingly explained by variation in database variables. We use S&P Capital IQ as our benchmark database but verify that the definitions of the metrics are similar in Bloomberg and Thomson Reuters. In Appendix B, we report the words used to define database metrics (on the left) and proxy metrics (on the right). We use boldface type for words that are present in the left side list to highlight the variability in language used for proxy definitions.

In Table 1, we group all performance metrics into thematic groups (Debt, Earnings, Accounting Return, Total Shareholder Return, Production, Reserves, Capital Expenditure, Cash Flow, Costs, Enterprise Value Added, Safety, ESG, and Operating Efficiency). We use a data-driven approach to map the actual results in the proxies (proxy metrics) to the corresponding numbers in S&P Capital IQ (database metrics). For each metric group, we select all closely related candidate variables in Capital IQ and regress the numbers from the proxies on the Capital IQ numbers of each candidate variable by firm-metric. We require at least four annual observations of the same firm-metric to estimate any regressions.[9] If the R-squared of a regression is greater than 0.8, we classify that metric as transparent and assign the database variable with the highest R-squared to be the one mapped to that proxy metric. Metrics with the highest R-squared below 0.8 are classified as non-transparent.[10]

The guiding principles in designing this data-driven, R-squared-based approach also follow the spirit of disclosure improvement revealed in SEC regulations, comment letters, and reviews. For example, in its 2007 review of the newly adopted 2006 Executive Compensation Disclosure regulations, the Corporate Finance division of the SEC provides examples of comment letters it sent to companies regarding performance targets: "Where a company presented a non-GAAP financial figure as a performance target and … did not disclose how it would calculate that figure, consistent with Instruction 5 to Item 402(b)(2), we asked it to disclose how it would do so. For example, where a company disclosed total shareholder return as a performance target, we asked the company to disclose how it would calculate total shareholder return."[11] The R-squared-based approach allows us to identify performance metrics from proxy statements that are very close to publicly available, machine-readable variables that are calculated consistently across firms. Proxy statement performance metrics that are heavily modified will have lower R-squared levels.

The R-squared measure is a simple measure of similarity between two series of values. It is not by itself an indicator of firm transparency. For example, one of our firms (EOG Resources) uses several non-GAAP metrics and reconciles them with their GAAP equivalents, thereby demonstrating a high degree of transparency. Among the metrics EOG uses and clearly designates

[9] The discontinuity results are unchanged if we require five observations for the proxy-on-database metric regressions or when we only include the firm-metrics which have between five and eight observations to maintain similarity in the number of observations producing the R-squared measure.
[10] We show results for other cut-off points in Panel C of Table 2.
[11] https://www.sec.gov/divisions/corpfin/guidance/execcompdisclosure.htm.

as non-GAAP, our R-squared measure identifies some as very close to, and others as very far from, their GAAP/public dataset equivalents. Thus, the R-squared measure is simply one descriptor of a firm-metric that is not necessarily related to how transparent the firm reporting is in general.

Table 1, Panels A-D list all performance metrics reported by sample firms in their proxy statements by sub-period.[12] We also report the number of firms that use a given metric. The metrics employed by the highest number of firms are: Production (31 firms), Reserve Replacement (31 firms), and Development Costs (27 firms). The average R-squared values from the procedure used to classify a metric as transparent or non-transparent indicate wide variation across metrics.

In Table 1, Panel E, we report the firm counts and average time series lengths by metric type. One potential concern about the design of the R-squared classification procedure is that metrics with shorter time series lengths may mechanically yield lower R-squared values and be classified as non-transparent not because of genuine low similarity but due to low power. We partially address this by requiring five observations in robustness tests. In addition, columns (4) and (5) in Table 1, Panel E indicate that the time series lengths of the metric groups with R-squared < .8 (non-transparent) are often longer than the time series lengths of the metric groups with R-squared > .8 (transparent).

Figure 1 shows firm usage of the most common metrics over time. The choice of performance metrics by each firm is relatively stable over time, with some firms switching from transparent to non-transparent versions of metrics. The most consistently used metrics are Production, Reserves, and Earnings. Different types of cost-related metrics are also common, with greater intensity in the latter part of the period. The last group of metrics shown in Panel D of Figure 1 (Enterprise Value Added (EVA), Safety, ESG, and Operating Efficiency) have few (if

[12] Determined as before and after SEC disclosure guidance in 2011 on non-GAAP metrics in firm proxy statements.

any) database equivalents. Notably, ESG metrics are used by very few firms in the sample. Metrics that appear variably used over time include: Capital Expenditure, Debt, Accounting Return and Total Shareholder Return (TSR). For example, companies that use adjusted debt-related metrics (namely Debt, Debt/Book, and Debt/EBITDAX) throughout the period are: ConocoPhillips, Pioneer, and Lilis Energy. EOG uses Interest Expense only in the period 2012-2017.

The distributional characteristics of the performance measures within our sample firms and across time necessitate the use of multiple tests for distributional discontinuity. The discontinuity test we employ in our baseline results is the Bugni-Canay approximate sign test for continuity of a density (Bugni and Canay 2021). This is an alternative test for the presence of a discontinuity of the density of the running variable at a given cut-off, introduced first by McCrary (2008), but appropriate for small samples and free of parametric distributional assumptions. The test does not require that the analyzed random variables are independent and identically distributed. We confirm our findings with the Byzalov and Basu (2019) semi-parametric discontinuity test. This test potentially has higher power but relies on stronger assumptions than does the Bugni-Canay test.

In our last set of tests we perform panel regressions of bonus, equity and total compensation on the proportion of achieved performance targets in a given firm-year plus controls. We also compute the proportion of achieved performance targets separately for non-manipulable/manipulable and transparent/non-transparent metrics. In Figure 2 we show heatmaps of this proportion for all firm-years. The firms (along the y-axis) are ranked by their average achievement record with the best achievers at the top. For each heatmap we compute the percentage of firm-year observations with proportion of achieved targets greater than 0.5. Executives have a better achievement record when metrics are non-transparent: 76.32% vs 70.45% when metrics are transparent. Similarly, their success rate is 77.84% when metrics are manipulable vs. 70.05% when metrics are non-manipulable.

V. TESTS FOR ACTUAL -MINUS- TARGET DIFFERENCES

In this section, we first test Hypothesis 1 by assessing discontinuities in distributions of performance metrics reported by all sample firms. We then test Hypotheses 2 and 3 by examining discontinuities in the distributions of various sub-samples of metrics.

Discontinuity in the Full Sample

In Figure 3, we show a distribution of actual-minus-target differences. The discontinuity at zero is clear and strongly statistically significant. The Bugni-Canay discontinuity test uses a data driven approach to select the bin width around the threshold. This approach typically results in smaller bin width when there are more observations. In Table 2, Panel A, we show the data-driven bin width for the entire sample of 1,787 observations is 0.0012. For comparability with all tests in the paper, where the number of observations varies for different sample splits, we also report the number of observations at and above zero and below zero within a common bin width of 0.01. For the full sample, just meeting or beating the target by up to one percent is 2.8 times more likely than just missing it by up to one percent.

The proportion of observations at or above target is 71 percent – significantly higher (p-value < 0.0001 from a one-sample test of the proportion of observations being > 0.6; untabulated) than the benchmark that targets should be achieved 50-60 percent of the time (Saliba 2016).

We also implement alternative statistical tests and research designs that can address some endogeneity concerns. First, we confirm that actual–target differences are indeed discontinuous at zero with a different test, recently developed by Byzalov and Basu (2019). This discontinuity test combines the Burgstahler and Dichev (1997) meet-or-just-beat intuition with a semi-parametric approach of fitting a polynomial function of up to pre-selected power *n* to the running variable. In their Figure 7, Byzalov and Basu (2019) illustrate the performance of their approach on the Compustat universe under different modelling choices and bin widths. We use the same specifications as in that figure and present the results for the distribution of actual-minus-target differences in Figure 4. The two sets of options of the Byzalov-Basu discontinuity testing procedure as implemented in their original paper require pre-selected bin widths (0.0025 and 0.001) and corresponding minimum numbers of estimation bins (16 and 40) and small profit/loss bins (4 and 10). The Byzalov-Basu procedure cuts off the estimation sample to the number of estimation bins times the selected bin width above and below the discontinuity, which results in a subset of our original dataset (429 observations from 46 firms). The p-values reported beneath each fitted model correspond to the coefficients of the four polynomial terms (Smooth0 – Smooth3) and the Kinky indicator variable. The specified models (i) or (ii) in the estimation procedure do not affect the estimated polynomial coeffcients; they affect only the coefficient on the kinky term. Under all specifications, the Byzalov-Basu discontinuity test (i.e., the kinky term) is significant, documenting the existence of an actual-minus-target discontinuity around zero.

Second, we consider different potential mechanisms that may cause discontinuities. Suppose that boards and compensation committees choose the optimal amount to pay their CEOs (i.e., the amount that maximizes shareholder value while retaining talent and motivating effort). However, regulatory frictions make it impossible for the board to award that optimal amount unless they engage in some form of performance metric management. For example, Marinovic and Varas (2019) model optimal compensation with endogenous manipulation, while Albuquerque, Bennett, Custodio, and Cvijanovic (2022) show empirically that pay-for-luck, which some authors interpret as evidence of overpaying executives, can be explained with efficient rewarding of executives' actions in anticipation of lucky events. The way to address this possibility analytically is to design a theoretical model predicting optimal compensation across firms, simulate the actual–minus–target differences, and measure the discontinuity produced by the simulated distribution relative to the distribution we observe in our sample. We leave this for another study.

In the present analysis, we point to theoretical research demonstrating that incentives to manipulate performance are embedded in compensation plans with discrete jumps and nonlinearities in payoffs (summarized by Murphy 2013). When outcomes are predictable and do not contain much noise, models show increased incentive to exert extra effort when very close to the target without necessarily engaging in misinformation or bad faith adjustments (Gibbs 2013); however, the strength of this incentive to exert extra effort disappears in the presence of noise or uncontrollable factors (Brehm et al. 2017). The oil & gas sector is an especially good example of unpredictability and lack of control by managers, as Bertrand and Mullainathan (2001) first argue in their seminal paper on pay for luck. Therefore, the discontinuities we find are likely to be the result of some combination of the following board actions: 1) strategically choose the set of metrics relevant for compensation; 2) adjust the weight assigned to each performance metric ex post; 3) set target levels at or below projected levels ex ante; and 4) manipulate actual performance levels during the performance period or ex post to meet or beat previously set targets. The first mechanism cannot solely explain the discontinuities because there are few patterns of many firms switching in and out of a particular metric at the same time in Figure 1.

To isolate the effect of the manipulation of actual results from the manipulation of targets, we perform an i-vs-j analysis. Instead of subtracting the target from the actual level of a metric for the same firm, we use the target of firm j and the actual result of firm i. To match every firm-metric-year observation i to a comparable paired observation j, we sort the data by year, metric, and target. We then select as pairs every two adjacent observations within the same metric-year as long as they satisfy the following condition: $\left|\frac{target_k - target_{k-1}}{target_{k-1}}\right| < 1$, where k denotes the sorting order. We then compute our actual-minus-target measure for each $i - j$ pair as $\frac{actual_i - target_j}{target_j}$. Figure 5 shows that the discontinuity remains significant, which suggests that even after removing the firm-specific selection in setting a target level, the actual results from the proxy statement continue to indicate that the probability of just beating the target is disproportionately high. This robustness test has a limitation in that the $i - j$ differences may me mechanically very close to our main analysis simply because of proximity. In unreported results due to very low number of observations we employ a matching approach in selecting the metric-year achieved result of firm j that is closest to the metric-year target of firm i.

Comparing Distribution Discontinuities of Manipulable vs. Non-manipulable and Transparent vs. Non-transparent Metrics

Our next analysis examines distributional discontinuities of actual-minus-target differences of manipulable/non-manipulable and transparent/non-transparent performance metrics (Hypothesis 2). We designate each metric for manipulablity as described in the previous section and transparency using the regression-based R-squared criterion, and we proceed with estimating the discontinuity test separately for the two sets of sub-samples of metrics. Figure 6 shows the overlapped histograms of the two groups and reports the p-values of the discontinuity tests at zero for each group. The distribution of manipulable (non-transparent) metrics is strongly discontinuous with a p-value < 0.01. In contrast, the test cannot reject the hypothesis that the distribution transparent metrics is continuous at zero (Table 2, Panel B, column 3, p-value > 0.15) and is marginally insignificant for non-manipulable metrics (Table 2, Panel B, column 1, p-value > 0.09).

In Table 2, Panel C, we further split the performance metrics into three groups based on cutoff values from the R-squared distribution. We report two sample splits. The first one uses cutoffs of 0.62 (the 33rd percentile of the distribution) and 0.80 to classify metrics into low, medium, and high R-squared groups. The second split uses 0.77 (the 50th percentile) and 0.90 as the cutoffs. In both sample splits, the only metrics with a significant discontinuity in the distribution of actual-minus-target values are in the lowest R-squared group. These results confirm the analysis in Figure 5 and suggest that performance metrics that are most dissimilar to the transparent metrics derived from Capital IQ have the highest likelihood of being manipulated.[13]

We also examine whether the discontinuities for these metrics changed surrounding updated SEC disclosure guidance for non-GAAP metrics in proxy statements from July 8, 2011. This guidance clarified that the reconciliation exemption provided for non-GAAP compensation targets does not apply to non-GAAP compensation actuals.[14] We split the full sample, and the manipulable/non-manipulable and transparent/non-transparent subsamples, into 2006-2011 (i.e., the pre-period) and 2012-2018 (i.e., the post-period) groups and test for the actual-minus-target discontinuity. We present the results in Figure 7. We find that the discontinuity exists in both the pre- and post-periods for the full sample. However, we find that the discontinuity, while present in the pre-period, is absent in the post-period for non-manipulable (transparent) metrics. The discontinuity exists in both the pre- and post-periods for manipulable (non-transparent) metrics but appears to be attenuated in the post-period. Thus, the updated SEC guidance appears to be

---

[13] Our inferences from Figure 5 and Table 2, Panels B and C remain unchanged when we recalculate the R-squared measure whereby we require five observations for the proxy-on-database metric regressions or when we only include firm-metrics that have between five and eight observations to maintain similarity in the number of observations producing the R-squared measure.

[14] https://www.sec.gov/divisions/corpfin/guidance/nongaapinterp.htm, Section 108.

associated with an attenuated discontinuity for both non-manipulable (transparent) and manipulable (non-transparent) metrics but is associated with a complete removal of the discontinuity for non-manipulable (transparent) metrics in the post-period with p-values > 0.7 and = 1 (a p-value of one under the Bugni-Canay test is possible by design since it is defined as a step function which takes round values of zero and one for boundary conditions).

Our manipulability and R-squared analysis or any other measure of concordance between transparent financial metrics and compensation-relevant performance metrics reported in proxy statements can be used by rating agencies, regulators, and investors to determine which performance metrics are likely to be heavily modified. Our analysis shows that such metrics are at higher risk of manipulation.

Distribution Discontinuity and Financial Distress

In Table 3, we test Hypothesis 3 with respect to legal scrutiny introduced by financial distress. The sample splits are done based on two proxies for financial distress. In Panel A, we create an indicator that takes the value of one for firm-year observations when the firm was classified as in default (i.e., S&P credit rating D); the indicator is also equal to one for the two years prior to the firm receiving S&P credit rating D. In Panel B, we also include firm-year observations where the firm has any rating below CC (i.e., D – default, SD – selective default, or R – firm is under regulatory supervision owing to its financial condition). The indicator is also equal to one for the two years prior to the firm receiving any rating below CC. Thus, Panel B includes all distressed firms in Panel A plus those with SD or R ratings.

Consistent with our predictions that manipulation is harder under intense creditor scrutiny, the discontinuity is not significant for distressed firms in either Panel A or Panel B of Table 3. We also note that healthy firms manifest the discontinuity in both panels, perhaps consistent with firms managing performance to avoid financial distress in the first place (Watts and Zimmerman 1990).

Distribution Discontinuity and Factors Influencing the Net Benefit of Manipulation

Table 4 reports split sample results testing the remaining factors under Hypothesis 3. The sample splits by analyst coverage and short interest are defined as the 25th and 75th percentile groups (defining them as below and above the median does not change the results). The sample split along analyst coverage of metrics is defined using a similar R-squared procedure as the one defining metric transparency. Specifically, we download all analyst forecasts from Bloomberg and Capital IQ and run regressions of metric target values on the analyst consensus (i.e., mean) forecasts. We run regressions of each target on each analyst consensus forecast in the same group. We classify a metric as having external analyst coverage if it has at least one regression on analyst forecasts with an R-squared of at least 80%.

The sample splits along the effective tax rate (ETR) are defined as highest ETR group > 40%, lower ETR group (0%, 40%] and non-payers of income tax. The profitability split is based on an indicator variable equal to one if a firm makes a profit in a given year and zero if it makes a loss. In four of the first five tests in Table 4, the discontinuity remains significant, suggesting that better informational environment, market discipline, taxation exposure, and firm profitability do not affect the tendency to just beat performance targets. The only test in which the discontinuity disappears in one of the two subsamples is that in which we separate metrics with external analyst coverage from metrics without coverage (i.e., "Internal"). As expected, only metrics without analyst coverage exhibit significant discontinuity.

We next perform sample splits along several governance characteristics. The discontinuity persists if we split the sample by CEO-Chair duality or board independence. However, the discontinuity disappears for observations where the firm has an above-median governance score, a high-quality bribery and corruption policy, or whistleblower program. These findings are consistent with the conclusion that the presence of discontinuity is evidence of manipulation. Moreover, the results are consistent with Stubben and Welch (2020) who find that whistleblower report volume is associated with fewer and less material government fines and lawsuits.

The governance-related findings in Table 4 make it highly improbable that the discontinuities are the result of good faith effort on behalf of managers or optimal compensation setting on behalf of compensation committees because there is no reason for these two mechanisms to obtain only under bad governance or lack of a whistleblower program.

Executive compensation regressions on proportion of achieved targets

In Table 5 we show regressions of bonus compensation (column 1), equity based compensation (column 2) and total compensation (column 3) on the proportion of performance metrics met in a given firm-year, controlling for size, metric and year fixed effects. We compute the proportion of achieved targets separately for non-manipulable (Panel A), manipulable (Panel B), non-transparent (Panel C), and transparent (Panel D) subsamples. Executive compensation is only significantly affected by the success rate of the executives only when metrics are manipulable (bonus and total compensation) or non-transparent (total compensation). The lack of significance in Panels A and D sends an equally strong message that compensation contracts may not be doing what boards intend if they do not respond to achieving targets in non-manipulable and transparent metrics.

Distribution Discontinuity by Metric Type

In untabulated results (available upon request), we examine distribution discontinuities by metric type. We find that earnings, ESG and other, safety, debt, production per share, and reserves are discontinuous ($p < 0.05$), while stock price, cash, costs, production volume, and CAPEX are not ($p > 0.05$). Not surprisingly, stock-price-based measures are less prone to manipulation as stock prices are less prone to direct management by firms, while earning-based measures are strongly discontinuous. The different results for production volume and production per share may be explained by firm management of the number of shares outstanding in the denominator of the latter (e.g., Cheng et al. 2015).

Monitoring Effects

In this subsection, we report results from untabulated tests of whether firms adjust CEO pay monitoring in anticipation of, or in response to, the degree of potential performance metric manipulation. First, we examine whether say-on-pay voting is associated with the degree of performance metric manipulability. We classify metrics as difficult to manipulate if they represent objective production volume, are based on items at the top of the income statement (e.g., revenues and costs), or are based on the company stock price or stock return. We classify metrics as easy to manipulate if they are defined on a per share basis, are measured in terms of a growth rate, or if they are adjusted or based on entries at the bottom of the income statement (e.g., EPS, net income). We then calculate the average annual manipulability across all performance metrics by firm-year and regress this measure on two indicators (one for years t and t+1; one for years t-1 and t-2) for a say-on-pay vote lower than 80%. We find only modest evidence that current and future negative say-on-pay votes are more likely when metrics are classified as more manipulable, and no evidence that past negative say-on-pay votes alter the overall manipulability of compensation contracts.

VI. CONCLUSION

Our detailed analysis of 13 years of performance-based compensation disclosures from the proxy statements of 62 firms in the oil and gas industry finds several patterns that have low probability of occurring naturally in the absence of strategic behavior by firm managers and compensation committees. Over the period, managers have been five (1.6) times more likely to barely meet than barely miss non-transparent (transparent) compensation targets. Statistical tests of the likelihood of such discontinuities reject the null of no discontinuity. The discontinuities are only significant for the subgroup of performance metrics that differ substantially from publicly available database numbers or analyst forecasts and for firm-year observations when the firm is not in financial distress. These results confirm our hypothesis that manipulation is more likely when it is harder to detect and not subject to intense analyst or creditor scrutiny. The discontinuities disappear (remain) in the latter half of the sample period for non-manipulable/transparent (manipulable/non-transparent) measures following SEC guidance for non-GAAP measures in proxy statements in 2011, and are more likely to occur in earnings, production per share, safety, and leverage-related metrics.

Regression evidence reveals that the proportion of metrics met in a given firm-year only significantly increases executive compensation for manipulable and non-transparent metrics and is unrelated to this proportion for non-manipulable and transparent metrics. These findings imply inefficiencies in the way executives are incentivized by boards and support our conclusions based on the discontinuity evidence.

Our analysis provides a benchmark for future work that can benefit from the machine readable data mandated by the new disclosure rules and execute manipulability and transparency analysis to measure their impact. One limitation of our manual data collection is that we could not perform the typical just-under vs over-target logit analysis in a panel framework due to extremely small number of observations. This is a type of analysis that will be feasible after the proxy statements in 2023 become available that will provide a 3-year history for all publicly listed firms.

Our results provide support for the petition by former SEC officials, the Council of

Institutional Investors, and the Chartered Financial Analyst Institute requesting that the SEC change the disclosure and reconciliation rules for non-GAAP performance targets used to determine executive compensation. We hypothesize that these requested improvements in disclosure and metric transparency can reduce the likelihood of opportunistic metric manipulation designed to increase executive pay. Overall, we document significant costs of the original exclusion of performance targets from the standard disclosure and reconciliation rules for reported non-GAAP metrics, which we believe exceed the potential benefits associated with this exclusion.

REFERENCES

American Association of Petroleum Geologists (AAPG). 2018. Unconventional energy resources: 2017 review. *Natural Resources Research* 28 (4): 1661–1751. doi: 10.1007/s11053-018-9432-1.

Albuquerque, A. M., B. Bennett, C. Custodio, and D. Cvijanovic. 2022. CEO compensation and Real Estate Prices: Pay for Luck or Pay for Action? *Review of Accounting Studies* (forthcoming). doi: 10.2139/ssrn.3212867.

Albuquerque, A. M., G. De Franco, and R. Verdi. 2013. Peer choice in CEO compensation. *Journal of Financial Economics* 108 (1): 160–181. doi: 10.1016/j.jfineco.2012.10.002.

Armstrong, C., J. Chau, C. D. Ittner, and J. J. Xiao, J. J. 2022. *Earnings per share goals and CEO incentives*. Working paper, University of Pennsylvania, The Chinese University of Hong Kong, and the University of Rochester. doi: 10.2139/ssrn.2919478.

Atanasov, V. A., B. Black, and C. Ciccotello. 2011. Law and tunneling. *Journal of Corporation Law* 37 (1): 1–49. doi:

Baker, G. 2002. Distortion and risk in optimal incentive contracts. *Journal of Human Resources* 37 (4): 728–751. doi: 10.2307/3069615.

Bartov, E., D. Givoly, and C. Hayn. 2002. The rewards to meeting or beating earnings expectations. *Journal of Accounting and Economics* 33 (2): 173–204. doi: 10.1016/S0165-4101(02)00045-9.

Becker, B., and P. Strömberg. 2012. Fiduciary duties and equity-debtholder conflicts. *Review of Financial Studies* 25:(6): 1931–1969. doi: 10.1093/rfs/hhs006.

Bennett, B., J. C. Bettis, R. Gopalan, and T. Milbourn. 2017. Compensation goals and firm performance. *Journal of Financial Economics* 124 (2): 307–330. doi: 10.1016/j.jfineco.2017.01.010.

Bergstresser, D., and T. Philippon. 2006. CEO incentives and earnings management. *Journal of Financial Economics* 80 (3): 511–529. doi: 10.1016/j.jfineco.2004.10.011.

Bernhardt, D., and R. J. Davies. 2005. Painting the tape: Aggregate evidence. *Economics Letters* 89 (3): 306–311. doi: 10.1016/j.econlet.2005.06.015.

Bertrand, M., and S. Mullainathan. 2001. Are CEOs rewarded for luck? The ones without principals are. *The Quarterly Journal of Economics* 116 (3): 901–932. doi: 10.1162/00335530152466269.

Bertsch, K. A., and J. P. Mahoney. 2019. *Rulemaking petition regarding disclosures on use of non-GAAP financials in proxy statement Compensation Discussion and Analysis*. https://www.sec.gov/rules/petitions/2019/petn4-745.pdf

Bhattacharya, N., E. L. Black, T. E. Christensen, and C. R. Larson. 2003. Assessing the relative informativeness and permanence of pro forma earnings and GAAP operating earnings. *Journal of Accounting and Economics* 36 (1–3): 285–319. doi: 10.1016/j.jacceco.2003.06.001.

Bhattacharya, N., T. E. Christensen, Q. Liao, and B. Ouyang. (2021). Can Short Sellers Constrain Opportunistic Non-GAAP Reporting? *Review of Accounting Studies* (forthcoming). doi: 10.1007/s11142-021-09621-9.

Black, D. E., E. L. Black, T. E. Christensen, and K. H. Gee. 2021. Comparing non-GAAP EPS in earnings announcements and proxy statements. *Management Science* (forthcoming). doi: 10.2139/ssrn.2343140.

Black, D. E., T. E. Christensen, J. T. Ciesielski, and B. C. Whipple. 2021. Non-GAAP earnings: A consistency and comparability crisis? *Contemporary Accounting Research* (forthcoming). doi: 10.2139/ssrn.2759312.

Black, D. E., T. E. Christensen, J. T. Ciesielski, and B. C. Whipple. 2018. Non-GAAP reporting: Evidence from academia and current practice. *Journal of Business Finance & Accounting* 45 (3–4): 259–294. doi: 10.1111/jbfa.12298.

Bollen, N. P. B., and V. K. Pool. 2009. Do hedge fund managers misreport returns? Evidence from the pooled distribution. *The Journal of Finance* 64 (5): 2257–2288. doi: 10.1111/j.1540-6261.2009.01500.x.

Brehm, M., S. A. Imberman, and M. F. Lovenheim. 2017. Achievement effects of individual performance incentives in a teacher merit pay tournament. *Labour Economics* 44: 133–150. doi: 10.1016/j.labeco.2016.12.008.

Bugni, F. A., and I. A. Canay. 2021. Testing continuity of a density via g-order statistics in the regression discontinuity design. *Journal of Econometrics* 221 (1): 138-159. doi: 10.1016/j.jeconom.2020.02.004.

Burgstahler, D., and E. Chuk. 2017. What have we learned about earnings management? Integrating discontinuity evidence. *Contemporary Accounting Research* 34 (2): 726–749. doi: 10.1111/1911-3846.12301.

Burgstahler, D., and I. Dichev. 1997. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics* 24 (1): 99–126. doi: 10.1016/S0165-4101(97)00017-7.

Byzalov, D., and S. Basu. 2019. Modeling the determinants of meet-or-just-beat behavior in distribution discontinuity tests. *Journal of Accounting and Economics* 68 (2–3): 101266. doi: 10.1016/j.jacceco.2019.101266.

Chacon, R., R. Gordon, and A. S. Yore. 2021. *Compensation consultants: Whom do they serve? Evidence from consultant changes*. Working paper, University of Colorado - Colorado Springs, Towson University, and University of Missouri. doi: 10.2139/ssrn.3281133.

Cheng, Y., J. Harford, and T. Zhang. 2015. Bonus-driven repurchases. *Journal of Financial and Quantitative Analysis* 50 (3): 447–475. doi: 10.1017/S0022109015000149.

Choi, S. J., A. C. Wiechman, and A. C. Pritchard. 2013. Scandal enforcement at the SEC: The arc of the option backdating investigations. *American Law and Economics Review* 15 (2): 542–577. doi: 10.1093/aler/aht007.

Christie, W. G., and P. H. Schultz. 1994. Why do NASDAQ market makers avoid odd-eighth quotes? *The Journal of Finance* 49 (5): 1813–1840. doi: 10.1111/j.1540-6261.1994.tb04782.x.

Chu, J., J. Faasse, and P. R. Rau. 2018. Do compensation consultants enable higher CEO pay? A disclosure rule change as a separating device. *Management Science* 64 (10): 4915–4935. doi: 10.1287/mnsc.2017.2845.

City of Birmingham Relief and Retirement System v. Reed Hastings, Case No. 3:18-cv-02107 (US District Court Northern District of California).

Crane, D. 2013. Excessive compensation without a remedy. *Arizona State Law Journal* 45: 281-298.

Curtis, A., V. Li, and P. H. Patrick. 2021. The use of adjusted earnings in performance evaluation. *Review of Accounting Studies* 26: 1290-1322. doi: 10.1007/s11142-021-09580-1.

Dechow, P. M., W. Ge, and C. Schrand. 2010. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics* 50 (2–3): 344–401. doi: 10.1016/j.jacceco.2010.09.001.

Dechow, P. M., S. A. Richardson, and I. Tuna. 2003. Why are earnings kinky? An examination of the earnings management explanation. *Review of Accounting Studies* 8 (2/3): 355–384. doi: 10.1023/A:1024481916719.

Denning, L. 2018. Shale CEOs lack incentive to find your holy grail. *Bloomberg* (June 13): https://www.bloomberg.com/opinion/articles/2018-06-13/oil-e-p-company-ceo-pay-lacking-incentive.

Dyreng, S. D., M. Hanlon, and E. L. Maydew. 2008. Long-run corporate tax avoidance. *The Accounting Review* 83 (1): 61–82. doi: 10.2308/accr.2008.83.1.61.

Eaton, C. 2014. Investors claw back in Cheniere compensation fight. *Houston Chronicle* (June 13): https://www.houstonchronicle.com/business/energy/article/Investors-claw-back-in-compensation-fight-at-5551788.php.

Edmans, A., X. Gabaix, and D. Jenter. 2017. Executive compensation: A survey of theory and evidence. *The Handbook of the Economics of Corporate Governance* 1: 383–539. doi: 10.1016/bs.hecg.2017.11.010.

Edmans, A., X. Gabaix, T. Sadzik, and Y. Sannikov. 2012. Dynamic CEO compensation. *The Journal of Finance* 67 (5): 1603–1647. doi: 10.1111/j.1540-6261.2012.01768.x.

Faulkender, M., and J. Yang. 2013. Is disclosure an effective cleansing mechanism? The dynamics of compensation peer benchmarking. *Review of Financial Studies* 26 (3): 806–839. doi: 10.1093/rfs/hhs115.

Frankel, R., S. McVay, and M. Soliman. 2011. Non-GAAP earnings and board independence. *Review of Accounting Studies* 16 (4): 719–744. doi: 10.1007/s11142-011-9166-3.

Gibbs, M. 2013. Design and implementation of pay for performance. In *The Oxford Handbook of Managerial Economics*, edited by C. R. Thomas and W. F. Shughart, 396–423. Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780199782956.013.0018.

Guest, N. M., S. Kothari, and R. Pozen. 2022. *Why do large positive non-GAAP earnings adjustments predict abnormally high CEO pay?* Working paper, Cornell University and Massachusetts Institute of Technology. doi: 10.2139/ssrn.3030953.

Guttman, I., O. Kadan, and E. Kandel. 2006. A rational expectations theory of kinks in financial reporting. *The Accounting Review* 81 (4): 811–848. doi: 10.2308/accr.2006.81.4.811.

Heron, R. A., and E. Lie. 2007. Does backdating explain the stock price pattern around executive stock option grants? *Journal of Financial Economics* 83 (2): 271–295. doi: 10.1016/j.jfineco.2005.12.003.

Heron, R. A., and E. Lie. 2009. What fraction of stock option grants to top executives have been backdated or manipulated? *Management Science* 55 (4): 513–525. doi: 10.1287/mnsc.1080.0958.

Ho, S. 2020. SEC Commissioner Jackson says rule change on non-GAAP measures for CEO pay should be considered. *Thomson Reuters Tax & Accounting* (January 22): https://tax.thomsonreuters.com/news/sec-commissioner-jackson-says-rule-change-on-non-gaap-measures-for-ceo-pay-should-be-considered/.

Holmstrom, B., and P. Milgrom. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization* 7 (Special): 24–52. doi: 10.1093/jleo/7.special_issue.24.

Isidro, H., and A. Marques. 2015. The role of institutional and economic factors in the strategic use of non-GAAP disclosures to beat earnings benchmarks. *European Accounting Review* 24 (1): 95–128. doi: 10.1080/09638180.2014.894928.

Jackson, R. J., and R. C. Pozen. 2019. Executive Pay Needs a Transparent Scorecard. *The Wall Street Journal* (October 4): https://www.wsj.com/articles/executive-pay-needs-a-transparent-scorecard-11554936540.

Jensen, M. C. 2003. Paying people to lie: The truth about the budgeting process. *European Financial Management* 9 (3): 379–406. doi: 10.1111/1468-036X.00226.

Kalb, R., R. Phillips, and S. Silberglied. 2016. *Commodity Crash: An ISS Analysis Goal Setting, Attainment, and Say-on-Pay Vote Results at U.S. Energy Companies*. ISS. https://www.issgovernance.com/library/goal-setting-attainment-say-pay-vote-results-u-s-energy-companies/.

Kim, S., and J. Y. Shin. 2019. *Subjective adjustments to objective performance measures in executive annual bonus contracts*. Working paper, Monash University and Seoul National University. doi: 10.2139/ssrn.3018199.

Kleinberg, R. L., S. Paltsev, C. K. Ebinger, D. Hobbs, and T. Boersma. 2018. Tight oil market dynamics: Benchmarks, breakeven points, and inelasticities. *Energy Economics* 70: 70-83.

Lie, E. 2005. On the timing of CEO stock option awards. *Management Science* 51 (5): 802–812. doi: 10.1287/mnsc.1050.0365.

Marinovic, I., and F. Varas. 2019. CEO horizon, optimal pay duration, and the escalation of short-termism. *The Journal of Finance* 74 (4): 2011–2053. doi: 10.1111/jofi.12770.

McCrary, J. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142 (2): 698–714. doi: 10.1016/j.jeconom.2007.05.005.

McLean, B. 2018. The next financial crisis lurks underground. *The New York Times* (September 1): https://www.nytimes.com/2018/09/01/opinion/the-next-financial-crisis-lurks-underground.html.

Morgenson, G. 2011. Moving the goal posts on pay. *The New York Times* (July 5): http://www.nytimes.com/2011/05/08/business/ 08gret.html?pagewanted=all.

Morgenson, G. 2014. Making ends meet at Walmart. *The New York Times* (October 5): http://www.nytimes.com/2014/05/11/business/ making-ends-meet-at-walmart.html?_r=0.

Morse, A., V. Nanda, and A. Seru. 2011. Are incentive contracts rigged by powerful CEOs? *The Journal of Finance* 66 (5): 1779–1821. doi: 10.1111/j.1540-6261.2011.01687.x.

Murphy, K. J. 2013. Chapter 4 - Executive compensation: Where we are, and how we got there. *Handbook of the Economics of Finance* 2 (A): 211–356. doi: 10.1016/B978-0-44-453594-8.00004-5.

Murphy, K. J., and M. C. Jensen. 2018. *The politics of pay: The unintended consequences of regulating executive compensation*. Working paper, University of Southern California and Harvard Business School. doi: 10.2139/ssrn.3153147.

Robinson, J. R., Y. Xue, and Y. Yu. 2011. Determinants of disclosure noncompliance and the effect of the SEC review: Evidence from the 2006 mandated compensation disclosure regulations. *The Accounting Review* 86 (4): 1415–1444. doi: 10.2308/accr-10033.

Roe, J., and K. Papadopoulos. 2019. *2019 U.S. Executive Compensation Trends*. ISS. https://www.issgovernance.com/library/2019-us-executive-compensation-trends/.

Saliba, R. 2016. *Latest Trends in Executive Pay Are Performance Awards Paying Out As Intended?* ISS. https://www.issgovernance.com/library/latest-trends-in-executive-pay-are-performance-awards-paying-out-as-intended/.

Schacht, K. N., and S. J. Peters. 2019. *Letter of support for Council of Institutional Investors Petition for Rulemaking Regarding Disclosures on Use of Non-GAAP Financials in Proxy Statement CD&As*. https://www.sec.gov/comments/4-607/4607-14.pdf.

Securities and Exchange Commission (SEC). 2018. *Non-GAAP Financial Measures*. https://www.sec.gov/divisions/corpfin/guidance/nongaapinterp.htm.

Shin, T. 2016. Fair pay or power play? Pay equity, managerial power, and compensation adjustments for CEOs. *Journal of Management* 42 (2): 419–448. doi: 10.1177/0149206313478186.

Soondar, S. L., A. Major, and C. Hines. 2010. Litigation and recoupment of executive compensation. *Hastings Business Law Journal* 6 (2): 397–466.

Stubben, S. R., and K. T. Welch. 2020. Evidence on the use and efficacy of internal whistleblowing systems. *Journal of Accounting Research* 58 (2): 473-518. doi: 10.1111/1475-679X.12303.

Watts, R. L., and J. L. Zimmerman. 1990. Positive accounting theory: A ten year perspective. *Accounting Review* 65 (1): 131–156. https://www.jstor.org/stable/247880.

Weijermars, R., and S. Watson. 2011. Unconventional-natural-gas business: TSR benchmark and recommendations for prudent management of shareholder value. *SPE Economics & Management* 3 (04): 247–261. https://doi.org/10.2118/154056-PA.

Wilkins, A. M., D. R. Hermanson, and J. R. Cohen. 2016. Do compensation committee members perceive changing CEO incentive performance targets mid-cycle to be fair? *Journal of Business Ethics* 137 (3): 623–638. doi: 10.1007/s10551-015-2567-7.

Young, S. 2014. The drivers, consequences and policy implications of non-GAAP earnings reporting. *Accounting and Business Research* 44 (4): 444–465. doi: 10.1080/00014788.2014.900952.

Appendix A: Variable Definitions

Variable/Concept	Definition	Source
Actual-minus-target Difference	The difference between the achieved result and the target divided by the target. For positive performance benchmarks, the result-to-target difference is positive (negative) when the desired outcome is above (below) the set minimum; such as: return on equity, return on capital, stock return, production growth, production volume, reserve replacement ratio, capital expenditure, cash holdings, ESG and management efficiency targets, etc. For negative performance benchmarks, the result-to-target difference is positive (negative) when the desired outcome is below (above) the set maximum; such as: Debt/Book, Net Debt, Debt/EBITDAX, general and administrative (G&A) costs, operating and G&A costs and finding costs, safety targets, expressed as number of incidents or fatalities	Authors calculations based on proxy data
Performance Metric Manipulability	A metric is classified a non-manipulable (manipulability = 0) if it represents: (1) either a physical quantity verifiable by independent agencies, eg. production volume and reserves; (2) or a dollar value equal to a physical quantity times a market-wide price of oil or gas, eg. revenue; (3) or a dollar amount verifiable by a counterparty in a capital market transaction, eg. bonds outstanding or stock market capitalization or the sale price of a fixed asset. In all other cases manipulability = 1.	Assigned as per given rules
Performance Metric Transparency	We consider a firm-metric transparent if a time-series regression of the achieved results reported in firm proxies on equivalent metric values from Capital IQ produces an R–squared over 0.8. (We offer robustness results based on different cutoffs in Table 2)	Authors calculations based on proxy and Capital IQ data
Analyst Following	Number of analysts following each firm and producing performance metric estimates	Capital IQ
Metric Coverage by Analysts	We consider a firm-metric covered by analysts if a time-series regression of the target numbers in the proxy on the consensus estimates by analysts of an equivalent metric values from Capital IQ produces an R–squared over 0.8. The sample split along analyst coverage of metrics is defined using a similar R-squared procedure as the one defining metric transparency. Specifically, we download all analyst forecasts from Bloomberg and Capital IQ and run regressions of metric target values on the analyst consensus (i.e., mean) forecasts. We run regressions of each target on each analyst consensus forecast in the same group. We classify a metric as having external analyst coverage if it has at least one regression on analyst forecasts with an R-squared of at least 80%.	Authors calculations based on proxy, Capital IQ, and Bloomberg data
Short Interest	Percentage of float in short positions at the end of the current year	Capital IQ
ETR	Effective tax rate for the current fiscal year = Income Tax Expense / Net Income	Capital IQ
Profitability	Indicator variable = 1 if Net Income Including Unusual Items > 0 and zero otherwise	Capital IQ
Governance Score	Weighted average of all indicators in the G (governance) category	Sustainalytics
Policy on Bribery and Corruption	G_1_1 – indicator for the presence of a Policy on Bribery and Corruption as part of the Governance Section of a firm's ESG score	"
Whistleblower Program	G_1_2 – indicator for the presence of a Whistleblower Program as part of the Governance Section of a firm's ESG score	"
CEO-Chair Duality	An indicator variable equal to one if in the current year the CEO position and the Chair of the Board were held by separate individuals and zero otherwise	Boardex
Board Independence	Proportion of board members classified as independent out of total board size	"
Proportion performance metrics met	For each firm-year we take the proportion of performance metrics met or exceeded out of all performance metrics used; this variable is computed for all observation and separately for manipulable/non-manipulable and transparent/non-transparent metrics	Authors calculations based on proxy data

Variable/Concept	**Definition**	**Source**
Firm size	Ln(Total Assets)	Capital IQ

APPENDIX B

Database Metrics Mapping to Proxy Metrics and Usage by Sample Firms

We list the most common words in the definitions of database metrics (Capital IQ, and consistent with Bloomberg and Thomson Reuters) and proxy metrics ordered by frequency of use down to at least 2 occurrences per word, reported by metric group. Common words in the descriptions of all metrics as well as company names are not reported, eg. oil, gas, barrels, etc. The words in bold on the right are the same as words on the left. A glossary of abbreviations is available upon request. The analysis covers 69 E&P energy firms, which have stock price data and have filed at least 7 proxies in the period 2007-2019.

Words used in defining database metrics	Words used in defining proxy metrics
<u>Debt</u> debt, total, long, term, capital, net, equity	**debt,** EBITDAX, **net,** reduction, EBITDA, book, **capital**
<u>Earnings</u> excluding, EBITDA, based, stock, EPS, affiliates, equity, incurred, diluted, basic, items, extra, EBIT, normalized, EBITDAX, net, income	adjusted, **income, EBITDA, EBITDAX, net, EPS,** earnings, share, growth, operating, sales, ESG, peer, downstream, year, compared, consolidated, segment, annual, companies, volume, special, throughput, crude, **items,** tax, competitive, retail, fuel, base, regulated, comparison, pre, pipeline, average, margin, revenue
<u>Acct Return</u> return, capital, invested, assets, common, equity	**return, capital,** employed, drilling, rate, efficiency, ROCE, ROE, ROC, ROR, tax, program, average
<u>TSR</u> total, return, shareholder, 1yr, 4yrs, 3yrs, 5yrs	**total, return,** stock, **shareholder,** TSR, price, annual, year, absolute, performance
<u>Production</u> production, share, growth, average, daily	**production, growth, share,** net, adjusted, debt, total, volumes, operated, upstream, volume, **daily,** liquids, **average,** CFO, domestic, year, rate, annual, SCO, organic, tertiary
<u>Reserves</u> total, reserves, replacement, reserve, undeveloped, developed, percentage, ratio, proved	**reserve, reserves, proved,** growth, **replacement,** additions, share, **developed,** year, organic, added, **ratio,** domestic, convert, revisions, multiple, **total,** net, debt, develop, production, adjusted, upgrade, inventory, **undeveloped,** natural
<u>Capital Expenditure</u> capital, expenditure	**capital,** expenditures, wells, growth, share, CAPEX, development, **expenditure,** net, new, NAV, additions, total, value, appraisal, drilled, resources, resource, asset, installation

APPENDIX B (continued)

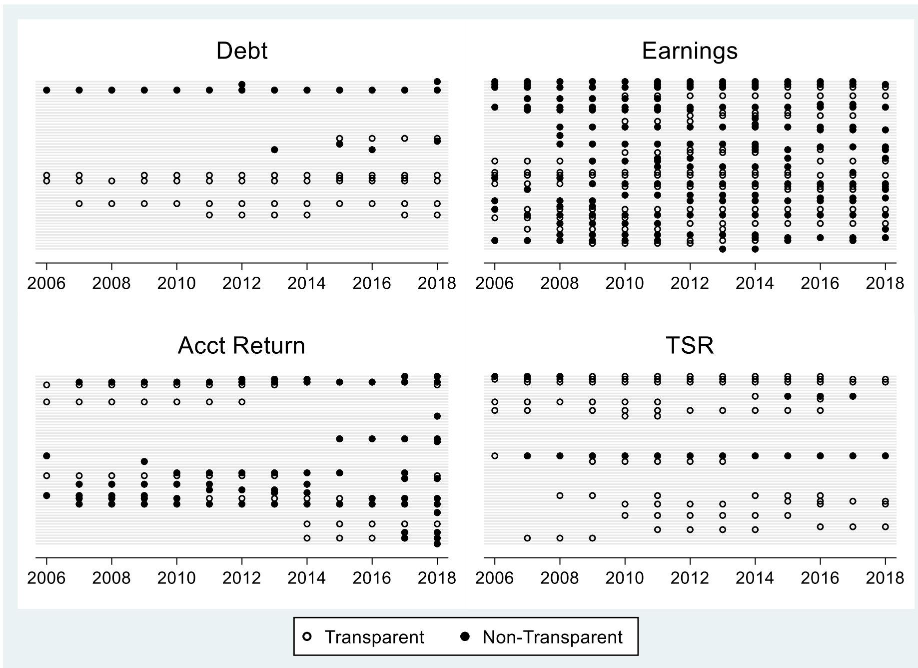
Cash Flow CF, operations, unlevered, free, cash, share, distributable	**cash,** flow, discretionary, operating, **operations, share,** adjusted, **free,** net, **CF,** flows, activities, growth, trigger, CFO, provided
Development Costs costs, production, future, acquisition, cost, average, general, expense, administrative, total, unproven, development, proven	**cost,** finding, **costs, development,** F&D, drilling, control, bit, completion, drill, year, unit, budgeted, expenditure, capital
G&A Costs general, expense, administrative, costs, total	G&A, **costs,** expenses, LOE, **general,** cash, **expense, administrative**
Production Costs production, average, cost, costs, future, total, taxes	operating, LOE, lease, expenses, **costs,** expense, **cost,** cash, unit, O&M, transportation, control, direct, net, base, TOTI, **total,** OPEX, excluding, controllable
Other Costs costs, production, acquisition, proven, average, cost, future, administrative, expense, development, exploration, general	**costs, cost,** cash, G&A, unit, operating, **expense, production,** control, controllable, transportation, gathering, fees, **exploration**
EVA diluted, enterprise, exercisable, total, value	**value,** EVA, economic, added, improvement, creation, year, annual, high, retention
Safety	rate, safety, incident, recordable, environmental, TRIR, lost, time, health, total, spills, DART, downstream, incidents, HSSE, accidents, reduction, injury, case, performance, event, serious, frequency, expressed, reportable, process, significant, upstream
ESG	strategic, visits
Operating Efficiency	availability, mechanical, downstream, refining, days, outstanding

FIGURE 1

Firm Usage of Performance Metrics over the Sample Period

Each row on the figures represents one sample firm. We show all metrics grouped by type and whether they are transparent or non-transparent. The sample consists of all 69 E&P energy firms, which have stock price data, have filed at least 7 proxies in the period 2007-2019, and mention, use, or report performance metrics.

Panel A: Debt, Earnings, Accounting Return, and Total Shareholder Return



Panel B: Production, Reserves, CAPEX, and Cash Flow

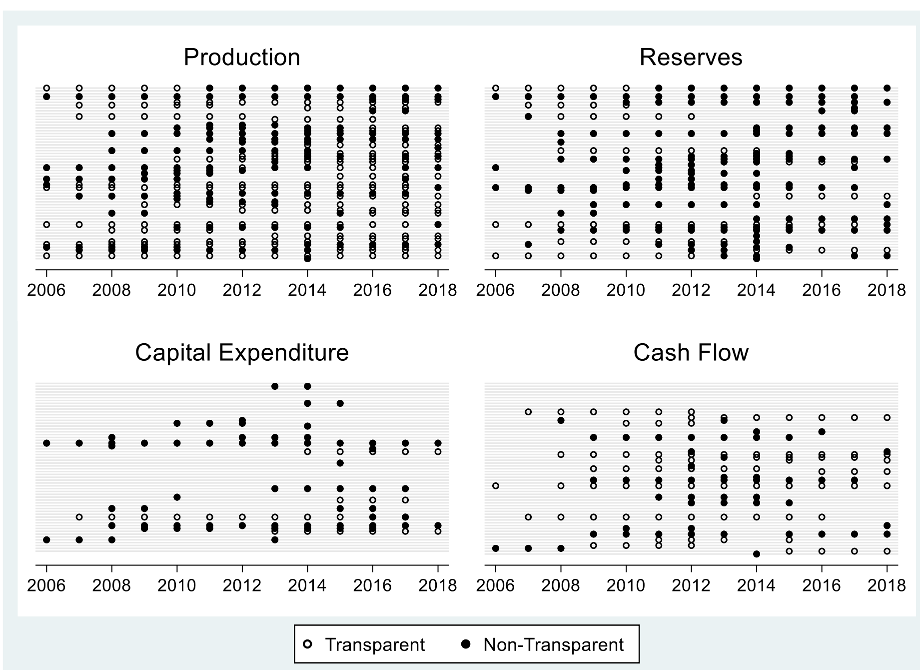
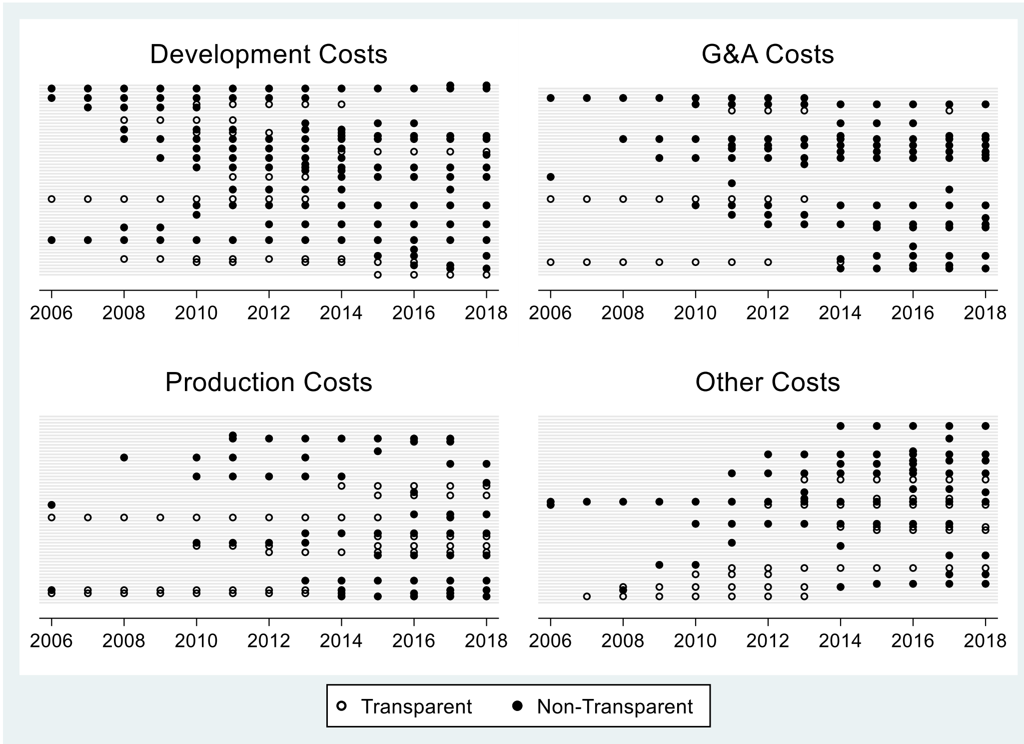


FIGURE 1 (continued)

Panel C: Costs



Panel D: Enterprise Value Added, Safety, ESG, and Operating Efficiency

FIGURE 2

Proportion of achieved performance metrics by firm-year

We show heatmaps of the proportion of achieved targets in each firm-year observation overall and separately by non-manipulable/manipulable and transparent/non-transparent metrics. The firms are ranked by their overall average rate of achieving targets, thereby the lightest (highest achievers) are at the top of the graphs. The percentages at the top of each graph capture the proportion of cells in each heatmap where the proportion of targets achieved is greater than 50%. The analysis covers 62 E&P energy firms, which have stock price data, have filed at least 7 proxies in the period 2007-2019, and report performance targets and actuals.

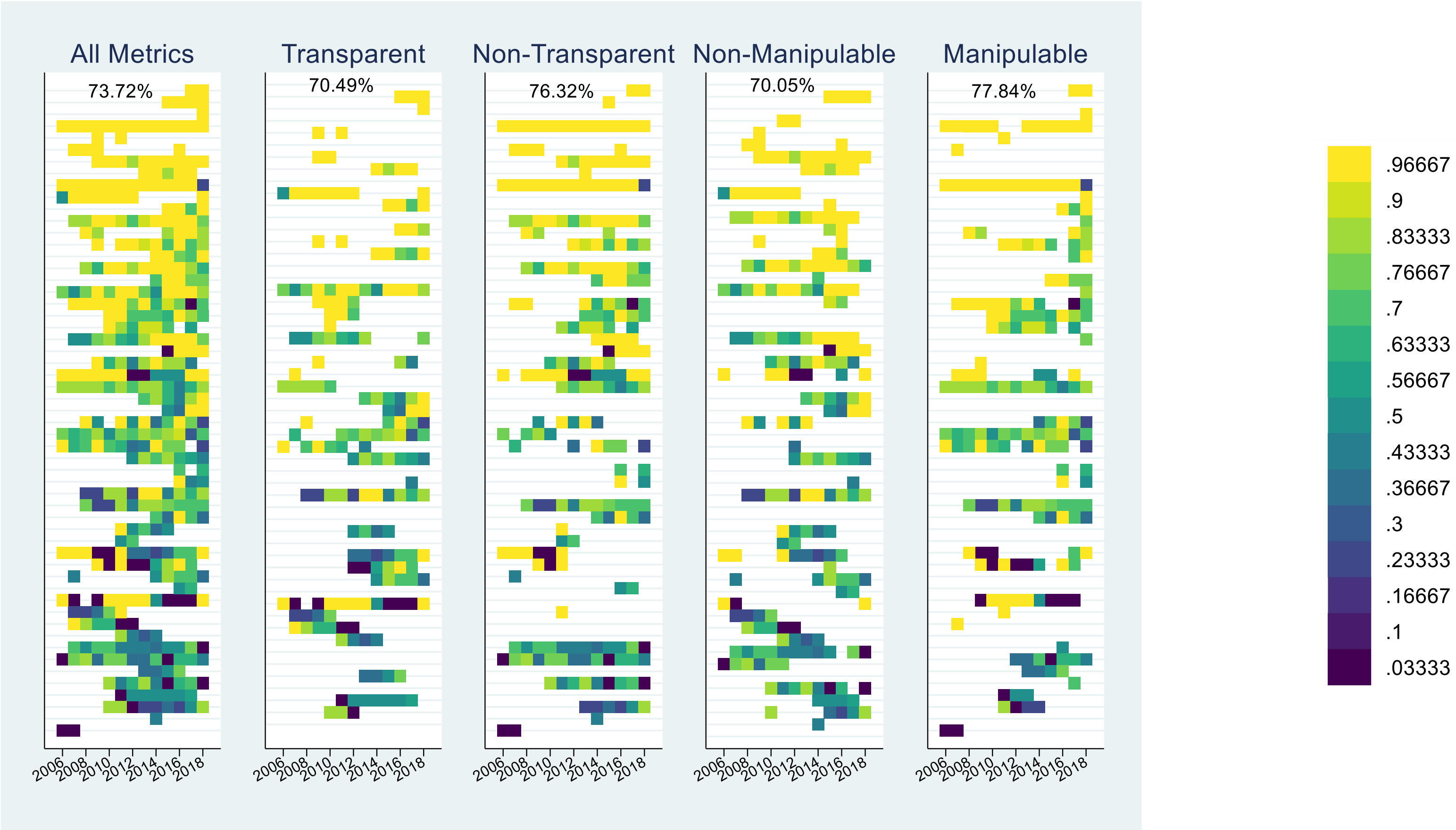


FIGURE 3

Frequency Distribution of Actual-minus-target Differences

This figure shows the frequency distribution of actual-minus-target differences for all performance benchmarks used by our sample firms during the period 2006-2018. For positive performance benchmarks, the actual-minus-target difference is positive (negative) when the desired outcome is above (below) the set minimum, such as return on equity, return on capital, production growth, production volume, and reserve replacement. For negative performance benchmarks, the actual-minus-target difference is positive (negative) when the desired outcome is below (above) the set maximum, such as debt/book, net debt, debt/EBITDAX, general and administrative costs, and development costs. The number of observations is lower than the number in Table 2 because we limit the scale of the histogram to the range [-0.5;0.5] so as to zoom in on the discontinuity at 0 (the result is unchanged if all observations are considered).

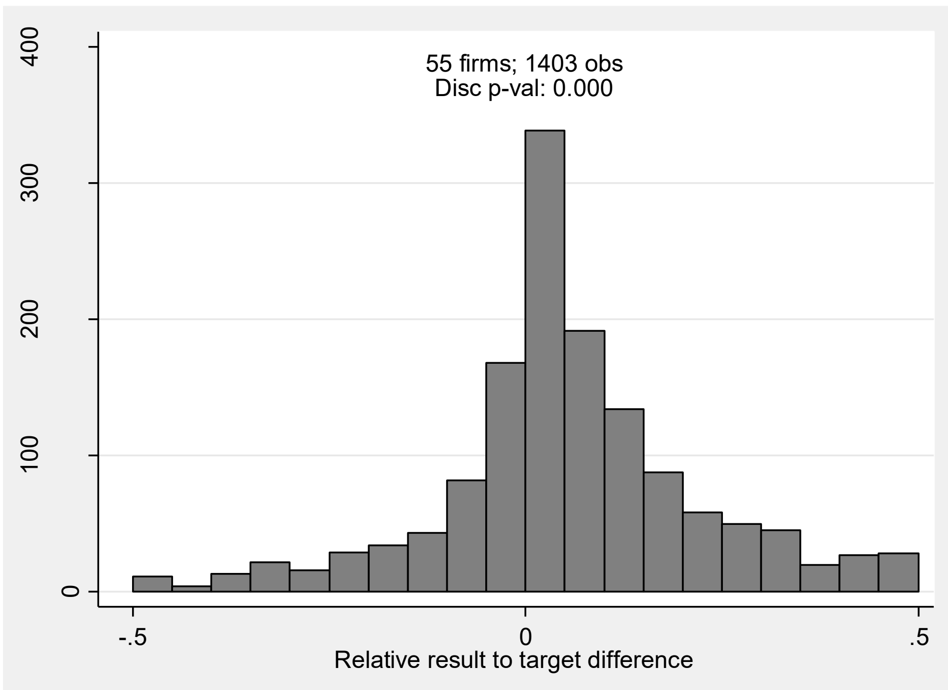
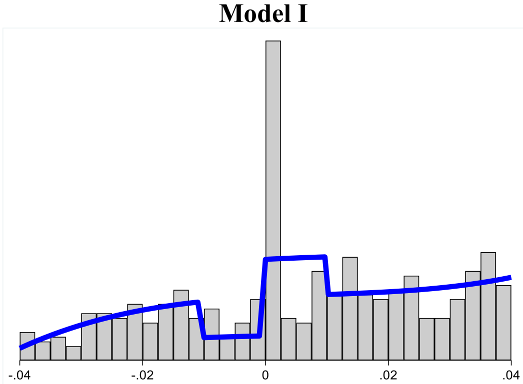


FIGURE 4

Byzalov and Basu Discontinuity Test

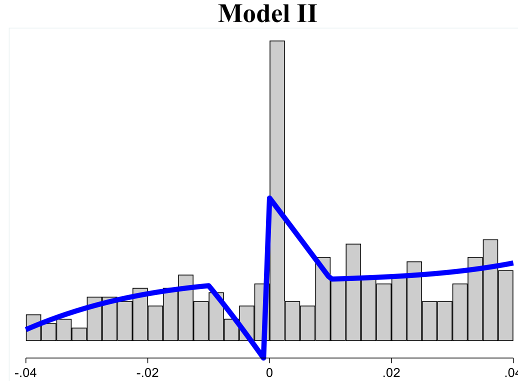
We implement specifications in Figure 7 of Byzalov and Basu (2019). The top two graphs are generated with the following Stata command: `kinkyX y, binwidth(.0025) est_bins(16) em_bins(4) em_type(i or ii) degree(3) cluster(firmid)`. The bottom two graphs are generated with the following Stata command: `kinkyX y, binwidth(.001) est_bins(40) em_bins(10) em_type(i or ii) degree(3) cluster(firmid)`. We report the p-values of the estimated coefficients resulting from the Byzalov and Basu procedure. The Smooth0, 1, 2 and 3 labels correspond to the coefficients on the zeroth (intercept), first, second and third degree of the explanatory variable, while Kinky detects the significance of a discontinuity. All four estimations use 429 observations from 46 firms, which is a subset of our overall data resulting from the required number of observations for a binwidth of 0.0025 (0.001) with 16 (40) estimation bins and 4 (10) small profit bins in the kinkyX command options.

Model I

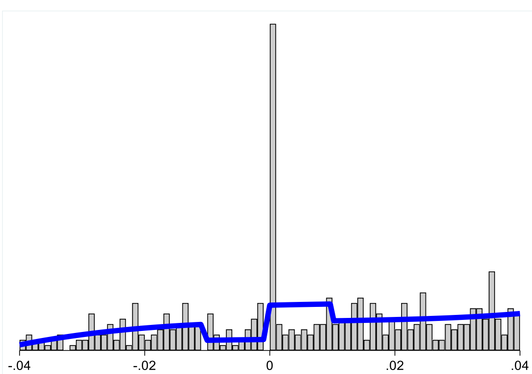


	Coef.	p-val.
Smooth0	0.031	0.000
Smooth1	0.156	0.512
Smooth2	-0.489	0.182
Smooth3	1.775	0.380
Kinky	0.614	0.001

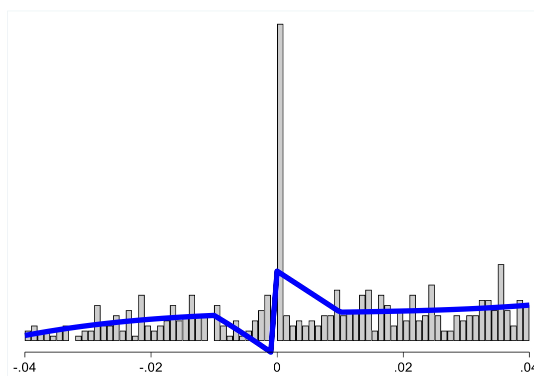
Model II



	Coef.	p-val.
Smooth0	0.031	0.000
Smooth1	0.156	0.512
Smooth2	-0.489	0.182
Smooth3	1.775	0.380
Kinky	0.716	0.000



	Coef.	p-val.
Smooth0	0.013	0.000
Smooth1	0.067	0.474
Smooth2	-0.197	0.173
Smooth3	0.665	0.397
Kinky	0.614	0.001



	Coef.	p-val.
Smooth0	0.013	0.000
Smooth1	0.067	0.474
Smooth2	-0.197	0.173
Smooth3	0.665	0.397
Kinky	0.782	0.000

FIGURE 5

Discontinuity in *i*-vs-*j* Actual-minus-target Differences

Instead of calculating the actual-minus-target difference of a metric for the same firm, we take the target of firm i and the actual result of firm j. To match every firm-metric-year observation i to a comparable paired observation j we sort the data by year, metric, and target. We then select as pairs every two adjacent observations within the same metric year as long as they satisfy the following condition: $\left|\frac{target_k - target_{k-1}}{target_{k-1}}\right| < 1$, where k denotes the sorting order. We then compute our actual-minus-target measure for each $i - j$ pair as $\frac{actual_i - target_j}{target_j}$. To zoom in on the discontinuity at 0, we limit the histogram to actual-minus-target differences [-1;1], which does not affect the discontinuity test and reduces the number of observations and firms to 653 and 49, respectively.

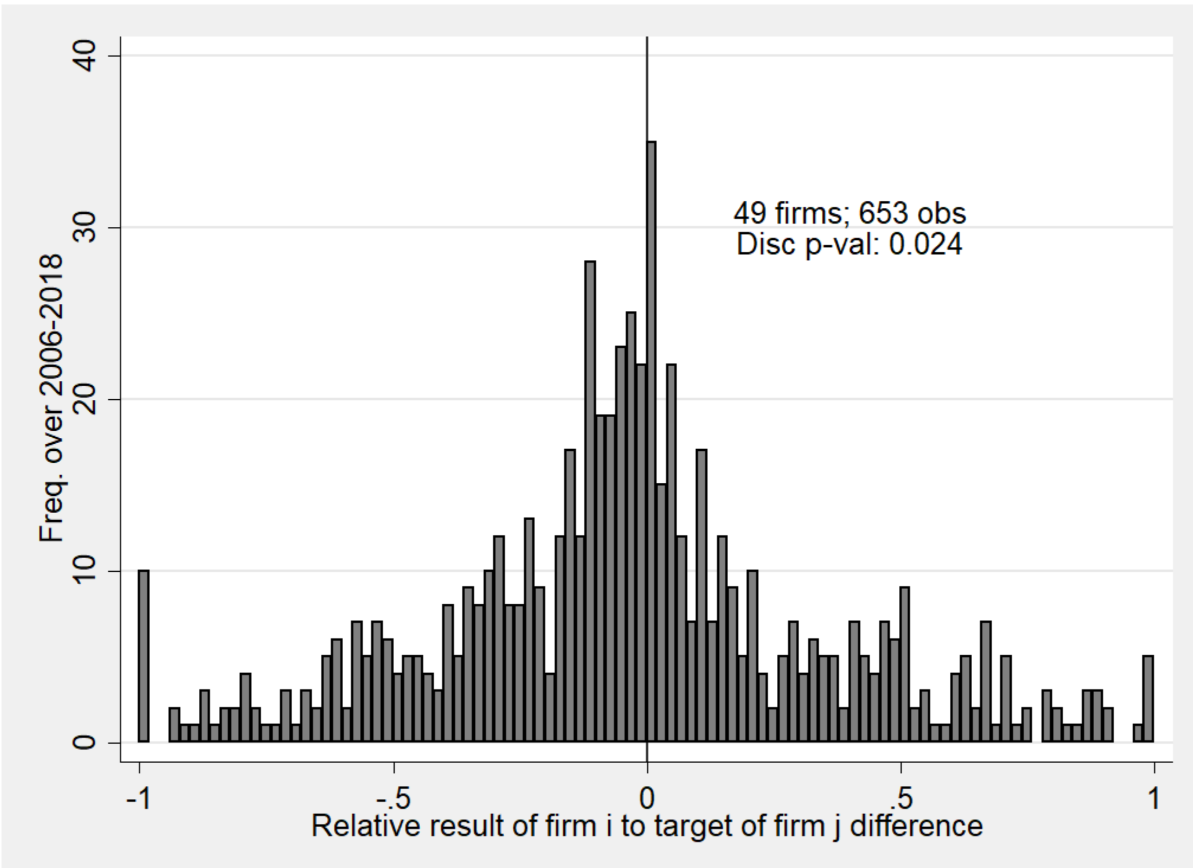
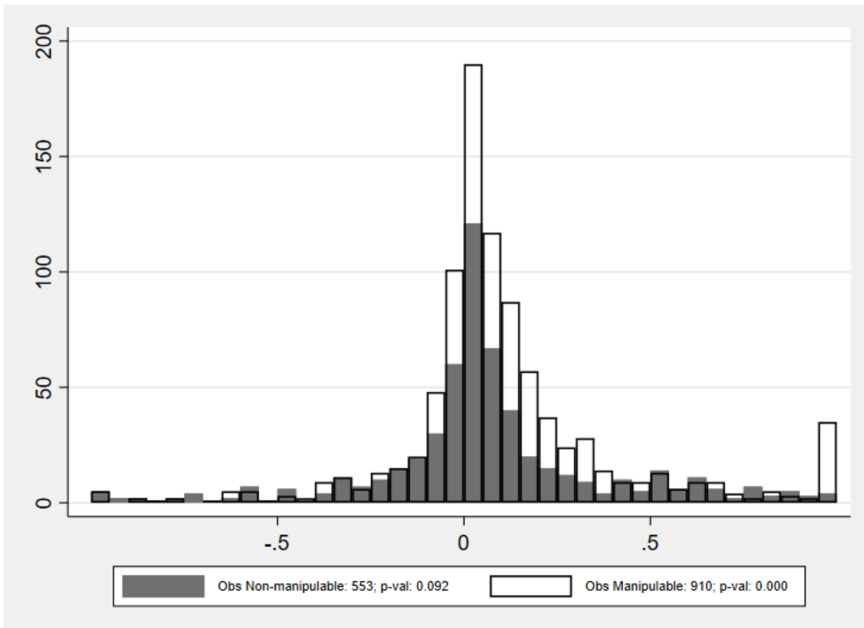


FIGURE 6

Frequency Distribution of Actual-minus-target Differences for Non-manipulable vs. Manipulable and Transparent vs. Non-transparent Metrics

We designate a metric as transparent based on a statistical procedure described in section IV. The p-values are from the Bugni-Canay permutation test of discontinuity at 0. The number of observations is determined by the scale of the histogram to the range [-1;1] and for transparency to be defined the requirement that the metric has a corresponding variable in Capital IQ (the result is unchanged if all observations are considered). The analysis covers 62 E&P energy firms, which have stock price data, have filed at least 7 proxies in the period 2007-2019, and report performance targets and actuals.



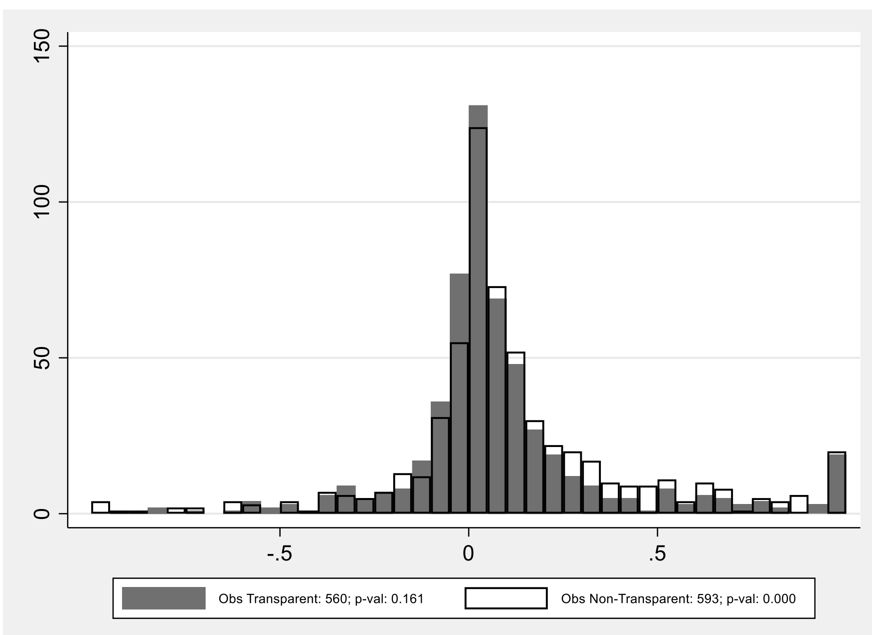
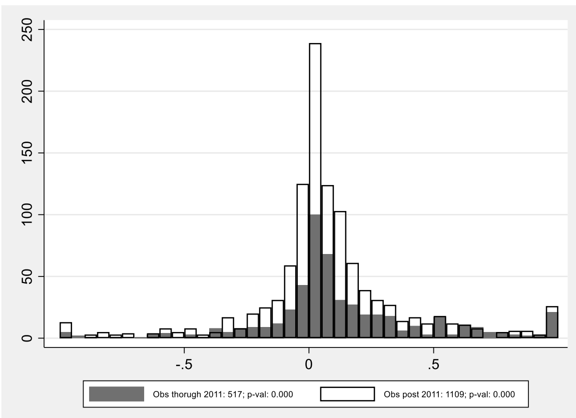
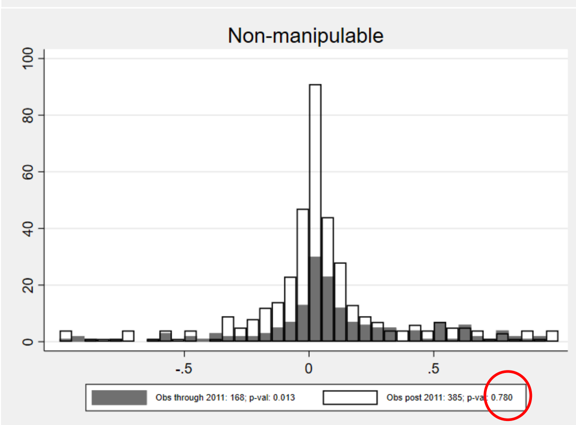


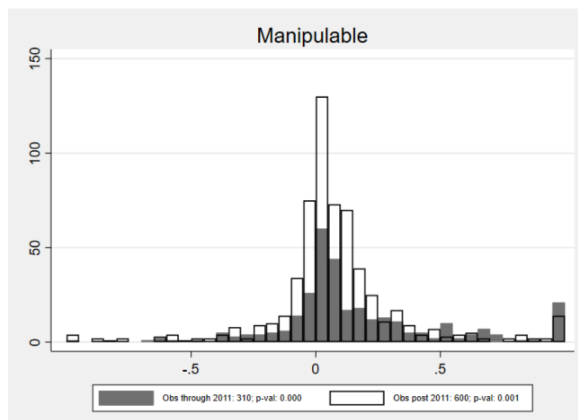
FIGURE 7

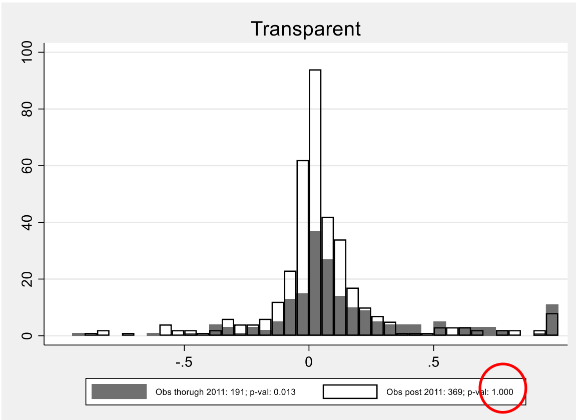
Frequency Distribution of Actual-minus-target Differences for the Full Sample, Non-manipulable vs. Manipulable and Transparent vs. Non-transparent Metrics before and after SEC Guidance in 2011

We designate a metric as transparent based on a statistical procedure described in section IV. The p-values are from the Bugni-Canay permutation test of discontinuity at 0. The number of observations is determined by the scale of the histogram to the range [-1;1] and the requirement that the metric has a corresponding variable in Capital IQ (the result is unchanged if all observations are considered). The indicator variable Post 2011 = 0 for proxies dated before June 2011 and Post 2011 = 1 for proxies dated after June 2011. Note that a p-value of 1 under the Bugni-Canay test is possible by design since it is defined as a step function which takes round values of 0 and 1 for boundary conditions. The analysis covers 62 E&P energy firms, which have stock price data, have filed at least 7 proxies in the period 2007-2019, and report performance targets and actuals.








Transparent
0
20
40
60
80
100
-.5
0
.5
Obs thorugh 2011: 191; p-val: 0.013
Obs post 2011: 369; p-val: 1.000

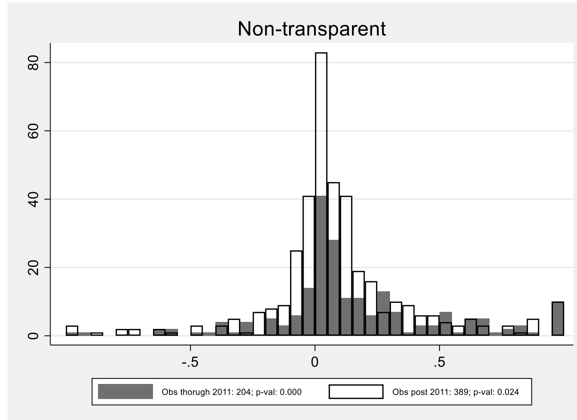

Non-transparent
0
20
40
60
80
-.5
0
.5
Obs thorugh 2011: 204; p-val: 0.000
Obs post 2011: 389; p-val: 0.024

TABLE 1

Performance Metrics Used for Compensation in our Sample

In Panels A-D, we report the target and actual performance metrics reported by the 62 companies in our final sample. The first (last) proxy in our data is submitted in 2007 (2019) and covers the compensation for 2006 (2018). We split the metrics into different groups and report the number of unique firms that use them in different time periods. The period splits are determined based on regime periods before and after SEC disclosure guidance in 2011. We also report the average R-squared for the same metric from the firm-metric time-series procedure described in section IV that we use to designate a given metric as transparent or non-transparent. Metrics with missing R-squared typically do not have any corresponding database metrics or do not have sufficient firm-level observations to estimate the regressions. The last column shows whether a corresponding variable is available in public financial statements and/or electronic databases and if not, the R-squared regressions use the most closely related metrics available in the same category.

Panel A: Debt, Earnings, Accounting Return, and Total Shareholder Return

	pre-2011	post-2011	# unique firms	Avg R-sq	Financials
Debt	2	7	7	0.691	1
Debt/Book	2	2	2	0.796	1
Debt/EDITDA	1	1	1	0.940	1
Interest Exp	0	1	1	0.353	1
Net Debt/Book	1	2	2	0.933	1
EBITDA	7	16	18	0.600	1
EBITDA - Adj	1	9	10	0.903	0
EBITDA Growth	1	3	3	0.579	1
EPS	3	3	3	0.814	1
EPS - Adj	0	1	1	0.993	0
Net Income	5	4	6	0.943	1
Net Income - Adj	1	2	3	0.963	0
Operating Income	8	8	11	0.673	1
Operating Income - Adj	2	2	3	0.839	0
Operating Income Growth	0	2	2		1
ROC	9	13	15	0.649	1
ROE	4	3	4	0.897	1
ROR	0	5	5	0.912	0

Panel B: Production, Reserves, CAPEX, and Cash Flow

	pre-2011	post-2011	# unique firms	Avg R-sq	Financials
Debt-Adj Production/sh Growth	2	6	6	0.614	0
Net Production	2	3	3	0.799	0
Net Production/sh	0	1	1		0
Production	13	29	31	0.929	1
Production Growth	8	11	12	0.874	1
Production/sh	1	1	2	0.831	1
Production/sh Growth	3	4	4	0.588	1
Proved Reserves	4	5	6	0.906	1
Reserve Repl.	17	28	31	0.518	1
Reserve Repl./sh	2	4	4	0.935	1
Reserve Repl./sh - Debt Adj	0	2	2		0
Capital Expenditure	6	13	14	0.437	1
Free/Discretionary CF	6	11	12	0.954	1
Free/Discretionary CF/sh	1	1	1	0.761	1
Operating CF	2	8	8	0.950	1
Operating CF - Adj	1	2	3	0.995	0

TABLE 1 (continued)

Panel C: Costs

	pre-2011	post-2011	# unique firms	Avg R-sq	Financials
Development Costs	16	24	27	0.695	1
Administrative Costs	9	23	25	0.582	1
Operating Costs	5	19	21	0.809	1
Total or Other Costs	8	21	23	0.773	1

Panel D: Enterprise Value Added, Safety, ESG, and Operating Efficiency

	pre-2011	post-2011	# unique firms	Avg R-sq	Financials
EVA	1	1	2	0.002	0
Health & Safety	2	9	10		0
Incident Rate	5	12	13		0
ESG	1	1	2		0
Strategic/Relationship Mgt	1	3	3		0
DPO	0	1	1		0
DSO	0	1	1		0
Equipment Run Times	0	1	1		0
Mechanical Availability	2	0	2		0

Panel E: Firm and Observation Counts

In Panel E, we provide the number of firms in each metric group that are mapped to database metrics with an R-squared greater than our chosen cut-off value of 0.8 in column (1) and the corresponding avg. number of year observations per firm-metric in column (4). We also present the firm count below the R-squared cut-off of 0.8 in column (2) with the corresponding avg. number of year observations in column (5). Finally, we present the number of firms that use a given metric and disclose target and actual results included in our discontinuity tests but have an insufficient number of annual observations to be included in the R-squared computation in column (3) with the corresponding avg. number of year observations in column (6).

Metric Group	Number of firms			Avg number of yr obs per firm-metric		
	R-sq > .8	R-sq < .8	Unused obs	R-sq > .8	R-sq < .8	Unused obs
	(1)	(2)	(3)	(4)	(5)	(6)
Debt	5	2	4	6.0	7.5	1.0
Earnings	17	16	17	4.3	6.3	1.3
Acct Return	5	7	11	6.9	6.9	1.2
TSR	4	2	4	5.3	10.0	1.3
Production	28	11	18	5.6	7.4	2.0
Reserves	10	14	22	5.4	6.5	1.4
Capital Expenditure	4	6	8	5.8	8.0	1.3
Cash Flow	11	2	9	5.1	5.0	1.8
Development Costs	7	11	15	4.4	6.8	1.9
G&A Costs	3	10	15	7.3	5.1	1.5
Production Costs	7	6	13	5.1	4.8	2.4
Other Costs	8	5	14	4.3	6.2	1.8
EVA	0	1	2	.	6.0	2.0
Safety			18			2.9
ESG			5			2.8
Operating Efficiency			5			3.0

TABLE 2

Tests for Discontinuities of Performance Metrics for the Full Sample and Grouped by R-squared Concordance Measure

In Panel A, we present results for the full sample of 62 companies that report targets and achieved results for performance metrics. In Panels B and C, we classify each performance metric reported in the proxy statement based on the highest R-squared of regressing this metric on all database metrics reported in Capital IQ. In Panel B, we present results for a single split at R-squared = 0.8. In Panel C, we split the metrics into a low, medium, and high R-squared based on different cutoff values of the R-squared distribution. We perform two separate splits: 1) At R-squared cutoffs of 0.62 (the 33rd percentile) and 0.80; and, 2) At cutoffs of 0.77 (50th percentile) and 0.90. We report p-values of the Bugni-Canay approximate sign test for continuity of a density.

Panel A: Full Sample

	All obs
Obs	1,787
Ratio Meet or Beat / All	0.71
Obs [0;0.01)	104
Obs [-0.01;0)	37
Ratio 0 and just above / just below	2.8
B&C p-value	0.0000
B&C data driven bin width	0.0012

Panel B: Two Group Split at R-squared of 0.8

	Non-Manipulable	**Manipulable**	**Transparent**	**Non-Transparent**
Obs	627	988	612	656
Ratio Meet or Beat / All	0.67	0.73	0.68	0.74
Obs [0;0.01)	29	57	28	40
Obs [-0.01;0)	16	20	17	8
Ratio 0 and just above / just below	1.8	2.9	1.6	5.0
B&C p-value	0.0919	0.0000	0.1608	0.0000
B&C data driven bin width	0.0115	0.0078	0.0114	0.0107

Panel C: Three Group Split at Different Cutoffs

	Group Transparent	**Group Non-Transparent**	**Group Non-Transparent**
	Highest R-sq	**Medium R-sq**	**Lowest R-sq**
Groups split at 0.619 (33th %ile) and 0.8			
P-value	0.1608	0.1435	0.0001
Obs	560	211	382
Groups split at 0.772 (50th %ile) and 0.9			
P-value	0.4011	0.1435	0.0000
Obs	406	172	575

TABLE 3

Tests for Discontinuities of Performance Metrics Grouped by Financial Distress

We report discontinuity tests for sample splits by financial distress for 62 companies that report targets and achieved results for performance metrics. In Panel A, the observations in the second column reflect firm-year observations, plus the two years before, when the firm was classified as in default (i.e., S&P credit rating D. In Panel B, we also include the firm-year observations where the firm has any rating below CC as distressed or in default (i.e., D – default, SD – selective default, or R – firm is under regulatory supervision owing to its financial condition).

Panel A: S&P Credit Rating D

	Healthy	**In Default**
Obs	1,673	114
Ratio Meet or Beat / All	0.71	0.63
Obs [0;0.01)	94	10
Obs [-0.01;0)	32	5
Ratio 0 and just above / just below	2.9	2.0
B&C p-value	0.0000	0.3323
B&C data driven bin width	0.0014	0.0124

Panel B: S&P Credit Rating below CC

	Healthy	**Distressed or In Default**
Obs	1,647	140
Ratio Meet or Beat / All	0.71	0.65
Obs [0;0.01)	94	10
Obs [-0.01;0)	31	6
Ratio 0 and just above / just below	3.0	1.7
B&C p-value	0.0000	0.6291
B&C data driven bin width	0.0014	0.0085

TABLE 4

Tests for Discontinuities of Performance Metrics Grouped by Factors Affecting the Net Benefit of Manipulation

We report discontinuity tests by sample splits along different factors affecting the net benefit of manipulation, including analyst following, metric coverage by analysts, short interest, effective tax rate (ETR), profit/loss in a given year, overall governance score, policy on bribery & corruption, whistleblower programs, CEO-chair duality, and board independence. The sample consists of 62 companies that report targets and actual results for performance metrics.

	Obs	Ratio Meet or Beat / All	Obs [0;0.01)	Obs [-0.01;0)	Ratio 0 and just above / just below	B&C p-value	B&C data driven bin width
Analyst Following							
75%ile	662	0.77	46	12	3.8	0.0000	0.0083
25%ile	321	0.67	16	5	3.2	0.0037	0.0214
Metric Coverage by Analysts							
Internal	1,459	0.71	83	27	3.1	0.0000	0.0019
External	328	0.69	21	10	2.1	0.1742	0.0133
Short Interest							
75%ile	650	0.71	43	11	3.9	0.0000	0.0096
25%ile	204	0.73	16	3	5.3	0.0127	0.0070
Effective Tax Rate (ETR)							
ETR > 40%	264	0.72	25	8	3.1	0.0003	0.0014
0% < ETR <= 40%	486	0.71	24	11	2.2	0.0241	0.0143
ETR 0 or n/a	876	0.70	55	18	3.1	0.0000	0.0094
Profitability							
Profitable	699	0.70	42	14	3.0	0.0002	0.0094
Unprofitable	867	0.70	58	22	2.6	0.0001	0.0081
Governance Score							
> Median	440	0.72	24	15	1.6	0.2624	0.0133
< Median	389	0.76	25	5	5.0	0.0004	0.0143
Policy on Bribery & Corruption							
> Median	449	0.72	23	12	1.9	0.1608	0.0137
< Median	380	0.76	26	8	3.3	0.0004	0.0155
Whistleblower Programs							
> Median	200	0.69	12	7	1.7	0.1435	0.0076
< Median	629	0.76	37	13	2.8	0.0006	0.0104
CEO & Board Chair Duality							
Different CEO & Board Chair	876	0.68	57	23	2.5	0.0005	0.0081
Same person is CEO & Board Chair	750	0.73	47	14	3.4	0.0000	0.0076
Board independence							
> Median	637	0.72	43	13	3.3	0.0001	0.0094
< Median	989	0.69	61	24	2.5	0.0001	0.0059

TABLE 5

Regressions of executive compensation on proportion of performance metrics met

We report linear probability model regressions of bonus compensation (column 1), equity based compensation (column 2) and total compensation (column 3) on the proportion of performance metrics met in a given firm-year, controlling for size, metric and year fixed effects. The different panels present subsamples by manipulability and transparency of performance metrics, whereby the proportion of performance metrics met is computed out of the total number of manipulable or not manipulable and transparent or not transparent metrics in each firm-year. Standard errors reported below coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Panel A: Non-Manipulable

	(1)	(2)	(3)
Prop perf metrics met	-1.838	-0.414	0.245
	(1.900)	(0.479)	(0.345)
Size	1.543	1.591***	1.402***
	(0.968)	(0.244)	(0.176)
Constant	6.868	2.647	4.268**
	(9.319)	(2.351)	(1.693)
Obs	237	237	237
Adj R-sq	0.420	0.887	0.647

Panel B: Manipulable

	(1)	(2)	(3)
Prop perf metrics met	3.700**	-0.021	0.266***
	(1.634)	(0.575)	(0.097)
Size	-0.008	0.474	0.231***
	(1.207)	(0.425)	(0.072)
Constant	14.215	13.278***	14.058***
	(9.987)	(3.515)	(0.594)
Obs	170	170	170
Adj R-sq	0.485	0.593	0.911

Panel C: Non-Transparent

	(1)	(2)	(3)
Prop perf metrics met	1.238	0.576	0.668**
	(1.543)	(0.648)	(0.323)
Size	2.144**	1.534***	1.235***
	(0.904)	(0.380)	(0.189)
Constant	-1.247	2.481	5.579***
	(8.332)	(3.500)	(1.743)
Obs	224	224	224
Adj R-sq	0.489	0.793	0.638

(TABLE 5 cont.)

Panel D: Transparent

	(1)	(2)	(3)
Prop perf metrics met	0.626	-0.010	0.079
	(1.783)	(0.117)	(0.072)
Size	-0.523	0.213**	0.199***
	(1.355)	(0.089)	(0.055)
Constant	25.425**	13.309***	14.123***
	(11.267)	(0.738)	(0.457)
Obs	183	183	183
Adj R-sq	0.414	0.894	0.923