Treatment of Outliers:

Firm Heterogeneity in Managerial Incentives and Corporate Innovation Relationship*

Wai-Tung Ho*, Keunkwan Ryu†, and Kam-Ming Wan‡

Abstract

A combination of a single-equation model and ordinary least squares (OLS) is predominately used to quantify the effects of managerial incentives on corporate innovation. The combination is inadequate for two reasons. First, firms in certain industries seldom engage in corporate innovation activities while others actively, resulting in corporate innovation measures having unusual data points in terms of spikes at zero and existence of extreme values. Second, OLS estimates do not capture differential effects of managerial incentives on corporate innovation and are not robust to outliers. Therefore, we propose an alternative combination of a mixture-distribution model and quantile regression. The mixture-distribution model is used to distinguish innovative firms from non-innovative ones. The quantile regression is applied to only innovative firms to estimate the differential effects of managerial incentives on corporate innovation. It is also used to mitigate the outliers' influence. Using only Fama-French industry dummies, our logistic regression well separates innovative firms from the non-innovative ones. Our quantile regression results indicate that managerial incentives play heterogeneous roles on corporate innovation depending on firms' inclination towards innovation. Between the two well-known managerial incentives for corporate innovation, we find that the vega incentive survives our scrutiny while the delta incentive does not.

Current Version: July 5, 2017

Keywords: outlier, vega, delta, corporate innovation, OLS, winsorization, mixture-distribution model, quantile regression

JEL Classifications: G34, G38, J31, J33

---

* We are grateful to Pierre Azoulay, Po-Hsuan Hsu, Kevin J. Murphy, Jim Ohlson, and seminar participants at Seoul National University, and the Third HKU Symposium on Innovation Strategy and Entrepreneurial Finance for their useful comments. We are thankful to excellent research support from Wing-Yung (Hazel) Ho and Chi-Yun (Jessica) Lee. This research project obtained financial support from the Central Research Grants and Departmental General Research Grant of the Hong Kong Polytechnic University (project numbers G-YBL7, G-YBEY, and G-UB34).

* Hong Kong Community College, Hong Kong Polytechnic University, 8 Hung Lok Road, Hung Hom, Hong Kong, Phone: +852-3746-0307; e-mail: ccwtho@hkcc-polyu.edu.hk.

† Department of Economics, Seoul National University, Seoul 08826, Korea, Phone: +82 10 5302-6397; e-mail: ryu@snu.ac.kr.

‡ School of Accounting and Finance, Hong Kong Polytechnic University, Li Ka Shing Tower, Hung Hom, Hong Kong, Phone: +852 2766-7053; e-mail: kmwan@polyu.edu.hk.

I. INTRODUCTION

To estimate the effects of managerial incentives on corporate innovation, researchers predominately use all sample observations ("single-equation model") and ordinary least squares (OLS) regression method. However, this combination is inadequate for two reasons. First, firms in certain industries seldom engage in corporate innovation activities while others actively, resulting in corporate innovation measures having unusual data points in terms of spikes at zero and existence of extreme values. Second, OLS estimates do not capture differential effects of managerial incentives on corporate innovation and are notoriously sensitive to extreme values or outliers.

To mitigate the influence of potential outliers, extant studies usually perform subjective data treatment before applying the least squares regression. The leading data treatment is data winsorization which arbitrarily replaces extreme data values of a variable (or univariate outliers) by specified percentile values. This data treatment presents conceptual challenges because extreme data values are not necessarily outliers once properly conditioned on a model and covariates. Outliers are better treated as a conditional concept in light of a reasonable choice of a model and covariates. This implies that data winsorization distorts informative data if univariate outliers can be properly explained by a reasonable model and covariates. In practice, the extent of data winsorization is subjective and susceptible to confirmation bias. Some winsorize one variable whereas others all variables. Furthermore, the depth of winsorization also varies widely from 0.5% to 5% in the executive compensation literature (Wan, 2014).

To demonstrate the appropriate manner to handle outliers, we investigate the outliers' influence on least squares estimates of managerial incentives on corporate innovation. This empirical relationship is chosen for three reasons. First, corporate innovation is a major decision in many firms and vital to a country's long-term economic growth (Kogan et al. 2016). Second, nearly all key variables of interest—corporate innovation and managerial incentives—have univariate outliers (Hall et al., 2001 & 2005; Coles et al., 2006; and Hirshleifer et al., 2012; Kini and Williams, 2012; Dong et al., 2016). For example, some firms invest heavily in research and development but many others allocate only limited resources to such activities. A case in point is that Microsoft spent over $11 billion on R&D in 2014, compared with just $30 million for the average company in the S&P 1500 index. In terms of the types of managerial incentives, some founder-CEOs rely exclusively on stock incentives but many professional CEOs primarily on stock-options incentives. For instance, William Gates, the co-founder of Microsoft, owned over ten percent of Microsoft stocks in the 1990s, compared with just 0.3 percent of stock ownership for the average CEO in the S&P 1500 index. These examples suggest that the outliers' influence is potentially serious.

Lastly, the structure of managerial incentives, in the form of stocks and stock-options, is instrumental for value-enhancing corporate innovation (Coles et al., 2006; Hirshleifer and Suh, 1992). However, the explosive use of equity-based compensation has many negative unintended consequences. For instance, managerial incentives provided by stock options have been heavily criticized since the unfolding of the option backdating scandal and the drastic growth in CEO compensation in the 1990s (Bebchuk et al., 2010; Lie and Heron, 2007; Lie, 2005; Hall and Murphy, 2003; Bebchuk et al., 2002; Yermack, 1997). Similarly, a growing number of studies show that managerial incentives provided by only stocks (or standard pay-for-performance compensation contracts) is suboptimal for corporate innovation (Ederer and Manso, 2013; Manso, 2011). These reasons compel us to examine the outliers' influence on this research question.

To identify conventional treatments of outliers in the literature, we have first searched the JSTOR database for articles published from 1989 to 2016 in eight leading journals in Accounting and Finance using keywords "delta" and "vega". The list of journals includes the *Journal of Finance*, *Journal of Financial Economics*, *Review of Financial studies*, *Journal of Accounting and Economics*, *Journal of Accounting Research*, *Accounting Review*, *Review of Accounting Studies*, and *Contemporary Accounting Review*. Delta refers to the sensitivity of CEO wealth to stock price ("pay-for-performance sensitivity") and vega the sensitivity of CEO wealth to stock return volatility. Delta and vega are chosen because they are common and complementary measures of managerial incentives. Next, we include only empirical studies and require them to use either delta or vega as an explanatory variable in the empirical analysis.

Twenty six articles satisfy these requirements. Of these, eighteen (69%) use data winsorization to treat outliers while the remaining do not mention any treatments of outliers. Of the eighteen studies that apply data winsorization, thirteen (72%) winsorize nearly all variables while the rest apply winsorization selectively. Data are predominantly winsorized at the 1% level. In sixteen (89%) of these studies, data are winsorized at the first and 99th percentiles.

Logarithmic transformation is another remedy to handle outliers. We are surprised that even though data on managerial incentives are highly right-skewed, seventeen (65%) studies use raw (unscaled) dollar values to measure delta and vega.[1] Nine studies (35%) use log-transformed data on managerial incentives, six by taking logarithm of one plus delta (vega) and three by taking logarithm of delta (vega). In the case of taking logarithm of delta (vega), data

[1] Some studies transform managerial incentives by deflating them by total (cash) compensation. This transformation worsens the outliers' influence because the skewness and kurtosis are greater for the transformed variables than for the original ones. This is because some CEOs receive exceptionally large or unusually small compensation, e.g., one dollar in salary pay (Hamm et al., 2015; Loureiro et al., 2014; and Guthrie et al. 2012). In unreported results, if vega is deflated by total pay, the adjusted skewness and kurtosis of the deflated vega are 4.79 and 10.48, respectively, much higher than 2.34 and 4.32 for the raw vega. Our results are similar if we deflate delta by total (cash) compensation.

truncation occurs because observations with a zero value of delta (vega) are excluded from the sample.

Robust estimation methods are rarely used to deal with outliers.[2] Even if they are used, they are used as supplementary evidences to demonstrate the robustness of the least squares estimates. Of the twenty six studies, only one (4%) uses the least absolute deviations method (i.e., median regression) for estimation.[3]

In this paper, we propose an alternative combination of a mixture-distribution model and quantile regression to address the unusual data points of corporate innovation measures and the fragility of OLS estimates to outliers. To handle the spike at zero of corporate innovation, we use a mixture-distribution model to objectively classify firms into innovative firms and non-innovative ones based on their industries. This classification is necessary because it is pointless to estimate the relationships of interest for non-innovative firms as these firms never intend to engage in corporate innovation. Our research methodology involves a two-stage regression procedure. In the first stage, we use logistic regression to separate industries into innovative and non-innovative. In the second stage, after applying logarithmic transformation to our key variables of interests, we use regressions to estimate the effect of managerial incentives on corporate innovation only for those firms in the innovative industries. Quantile regression is also used first to mitigate the outliers' concern, and second to examine the heterogeneity in the relationships between managerial incentives and corporate innovation. The heterogeneity

[2] Occasionally, robust estimation methods such as quantile regression, MM-estimator, and Theil-Sen regression are used in the accounting and finance literature (e.g., Hallock et al., 2010; Adams et al., 2015; and Ohlson and Kim, 2015).

[3] In contrast, median regression is commonly used for estimation in the executive compensation literature (Guthrie et al., 2012; Garvey and Milbourn, 2006; and Aggrawal and Samwick, 1999). In unreported results, we find that approximately 37 percent of empirical studies use median regression as a supplementary test to model the level of executive compensation.

consideration is relevant because even within an innovative industry some firms rarely engage in corporate innovation while others actively.

Just by using Fama-French industry dummies, our logistic regression well separates innovative firms from the non-innovative ones. Our quantile regression results indicate that managerial incentives play heterogeneous roles on corporate innovation depending on firms' inclination towards corporate innovation. Additionally, our quantile regression estimates of vega and delta are robust and stable across different quantiles of the conditional distribution. Congruent with the literature, our quantile regression results indicate that higher sensitivity of CEO wealth to stock return volatility (vega) induces only firms in innovative industries to increase corporate innovation including R&D investments, number of patent counts and citations. In contrast, higher CEO pay-for-performance sensitivity (delta) has no material effect on corporate innovation.

In contrast, our least squares estimates of the sensitivity of CEO wealth to stock return volatility (vega) are highly sensitive to model specifications. The least squares estimates of vega are significantly different from zero in the mixture-distribution model whereas those are statistically indistinguishable from zero in the single-equation model. As the mixture-distribution model uses only firms in the innovative industries whereas the single-equation model uses all the sample firms regardless of whether they belong to the innovative industries, the contrasting results on vega imply that higher vega induces only firms in innovative industries to increase corporate innovation.

Our findings also show that least squares estimates of the CEO pay-for-performance sensitivity (delta) are highly sensitive to outliers. Dropping only one firm from a sample of 635 firms in the innovative industries reduces the least squares estimates of delta by over 250%, rendering them statistically indistinguishable from zero. We also find that the least squares estimates of delta are fragile and change appreciably to different outlier remedies.

To the best of our knowledge, this paper is the first to address the heterogeneity in the managerial incentive effects on corporate innovation and outlier's influence on the corporate innovation. Our paper is novel because conventional treatments of outliers are inadequate to handle spikes at zero. To handle the spike at zero, we use a mixture-distribution model to identify firms in non-innovative industries and to exclude them in the second stage estimation. This approach is less susceptible to sample selection bias than dropping observations with "zero corporate innovation activity" for two reasons. First, the mixture-distribution model uses data objectively to classify innovative and non-innovative industries based on whether constituent firms in an industry ever engage in corporate innovation activities. Second, the second stage of the mixture-distribution model uses all the firm-year observations regardless of whether the corporative innovation activity is carried out or not (i.e., zero) in that year so long as the firm belongs to the innovative industry. The mixture-distribution model coupled with quantile regression in the second stage, also allows for more heterogeneity in the relationship between managerial incentives and corporate innovation. We carry out robustness checks to see whether the effects of delta and vega are sensitive to different treatments of outliers in terms of data winsorization and logarithmic transformation.

II. SAMPLE AND DATA

Our sample is constructed from four data sources. We obtain data on CEO compensation and stock ownership from the ExecuComp database. The ExecuComp database covers firms in the S&P 500, S&P Midcap 400, and S&P Smallcap 600. Next, we obtain accounting data from the

Compustat database. Data on CEO equity incentives are derived from Professor Lalitha Naveen's website and data on patent counts and patent citations from Professor Noah Stoffman's website. We exclude financial service firms (firms with one-digit SIC of 6) and utility firms (firms with two-digit SIC of 49). Our final sample is an unbalanced panel containing a total of 12,379 firm-year observations for 1,948 firms over the period between 1993 and 2004.

We begin our sample in fiscal year 1993 as this is the first year when ExecuComp database includes complete data on stock options. Our sample ends in fiscal year 2004 because this is the last year when companies are not required to record employee stock options as an expense. After June 15, 2005, the Financial Accounting Standards Board implements the FAS 123R which requires companies to record fair value of employee stock options as an expense (see Hayes et al., 2012). Thus, the incentive to use stock options to compensate corporate executives has changed considerably after 2005.

III. METHODOLOGY

To model the relationship between managerial incentives and corporate innovation for chief executive officers, we use the baseline model in Coles et al. (2006):[4]

$$(1) \qquad y_{it} = x_{it}'\beta + u_{it}$$

where y_{it} represents corporate innovation of firm i in year t and x_{it} is a vector of lagged CEO equity incentives joined by a vector of control variables, and u_{it} is a random error. For y_{it}, we take logarithmic transformation of one plus y_{it}, or $\log(1+y_{it})$, because corporate innovation variables have unusual data points in terms of extremely large values (justifying the use of

[4] This model is chosen because it allows us to compare our results directly to those in Coles et al. (2006) which is widely cited with Google Scholar citations of over 1,500 as of June 2017.

logarithmic transformation) and spikes at zero (demanding addition of a positive constant before the transformation).[5]

Measures of Corporate Innovation

We use one input and two output measures for corporate innovation. The input measure is research and development expenditure scaled by book assets (R&D)[6]. This measure quantifies the amount of resources allocated to corporate innovation activities. Following extant studies, we assign a value of zero for observations with missing values in R&D (Coles et al., 2006; and Hirshleifer et al., 2012).[7]

The output measures are the number of a firm's patent counts (#patents) and the number of a firm's patent citations (#patent_cites). Data on the firm's patent counts and citations are obtained from the April 2013 edition of the patent database in Kogan, Papanikolaou, Seru, and Stoffman (see Kogan et al. 2016). This database covers U.S. patent grants and citations from 1926 to 2010. Patents are included in the database only if they are eventually granted.

---

[5] To avoid sample selection bias in the second stage estimation, we also include firm-year observations with zero corporate innovation intensity in that year so long as the firm belongs to the innovative industry. To avoid excluding observations with zero corporate innovation, we add a positive constant (c) to each corporate innovation variable (y_{it}) before applying the logarithmic transformation, or $\log(c+y_{it})$. We follow the literature on corporate innovation and use one for the constant (see Hirshleifer et al., 2012), or $\log(1+y_{it})$. Different choices of the constant result in different distributions of the transformed variable. In unreported results, we confirm that our results are similar in terms of the resulting elasticities whether we add 0.0001, 0.001, or 0.1 instead of 1.

[6] To make our results directly comparable to those in Coles et al, we deflate research and development expenditure by book assets. In unreported results, the adjusted skewness and kurtosis turn out to be greater for the scaled R&D variable than for the raw R&D variable. For example, the adjusted skewness and kurtosis of the scaled R&D are 3.46 and 7.44 compared with 2.16 and 3.41 for the unscaled R&D. Our (unreported) results on least squares and quantile regression estimates are similar in terms of the resulting elasticities whether we use the scaled R&D or the raw R&D in the regressions.

[7] Companies can report a missing value in R&D expenditure despite they allocate real resources to such activities because they can make a conscious choice of not separating R&D expenses from other reported expenses (McVay, 2006). This implies that our treatment of the missing value in R&D underestimates the actual R&D investments. Nevertheless, this underestimation is a minor issue in our study because firms that report a missing value in R&D usually have no innovation outputs. For example, Koh and Reeb (2015) find that only 10.5 percent of firms reporting a missing value in R&D expenditure receive corporate patents. This implies that nearly 45 percent of our sample observations have no R&D investments and patent grants [= 50% $\times$ (100% $-$10.5%)].

Measures of CEO Equity Incentive

Our measures of managerial incentives include lagged delta (DELTA) and lagged vega (VEGA). Delta is the sensitivity of CEO wealth to stock price. It is defined as the change in the dollar value of the CEO's wealth for a one percentage point change in stock price in the previous year (Jensen and Murphy, 1990). Vega is the sensitivity of CEO wealth to stock return volatility. It is defined as the change in the dollar value of the CEO's wealth for a 0.01 change in the annualized standard deviation of stock returns in the previous year (Coles et al., 2006). Data on delta and vega are derived from Professor Lalitha Naveen's website. As option vega is significantly larger than stock vega, she measures the total vega of the stock and option portfolio by the vega of the option portfolio only. The vega and delta are calculated based on Core and Guay (2002). Detailed computations of delta and vega are available in Coles et al. (2006).

Control Variables

We follow the literature and include a set of control variables capturing firm and CEO characteristics (Coles et al, 2006; Hirshelifer et al., 2012). They are the natural logarithm of sales (log(SALE)) as a proxy for firm size; market-to-book ratio (M/B) for investment opportunity; surplus cash scaled by book assets (SURCASH) for the amount of cash available for corporate innovation; sales growth (SALEGRW) for growth opportunity; stock returns (RET) for firm performance; cash compensation (CASH) for CEO's degree of risk aversion; book leverage (LEVERAGE) for capital structure; and CEO tenure (CEOTenure) for the CEO's experience in her current position. These control variables are measured in the current year.

More specifically, M/B is the market to book ratio of asset values; SALEGRW is the logarithm of the ratio of sales in the current year to the sales in the previous year; RET is the return on common equity over the current year; CASH is the sum of salary and bonus for the CEO; LEVERAGE is the ratio of total book value of debt to book value of total assets; and CEOTenure is the length of time (in year) since the executive takes the CEO position in the firm. Our baseline model also includes industry fixed effects and year fixed effects. The industry fixed effects take into account 48 industries based on the Fama and French classification. To precisely quantify the outliers' influence, all data used in this study are untreated. Appendix 1 contains a detailed description of these variables.

IV. EMPIRICAL RESULTS

A. Descriptive Statistics

Table 2 provides summary statistics of all the variables used in this study in untreated data. Our sample medians are slightly larger but qualitatively similar to those reported in the literature.[8] For example, the median vega is $32,336 in our sample, compared to $34,000 in Coles et al. (2006) and $34,860 in Chava and Purnanandam (2010). The median delta is $203,394 in our sample, compared to $206,000 in Coles et al. and $173,790 in Chava and Purnanandam (2010). Similarly, the median sales is $1,096 million in our sample, compared to $887 million in Coles et al. and $965 million in Chava and Purnanandam (2010).

[INSERT TABLE 2 HERE]

It is clear that every variable has univariate outliers, particularly for our key variables of interest. Univariate outliers are many standard deviations away from their respective means. For example, the maximum value of vega has a *z*-score of 43 and that of delta has a *z*-score of 61. Similarly, the maximum values of all proxies for corporate innovation have a *z*-score of more

[8] The summary statistics reported in this study should be different from those reported in the literature because data are untreated in our study whereas those are winsorized in previous studies (e.g., Coles et al., 2006; and Chava and Punanandam, 2010).

than 25. These correspond to a probability of less than 0.0001% of such extreme values occurring assuming these variables are normally distributed. The large kurtosis of our key variables of interest is another indication of the presence of univariate outliers. These results are expected because extant studies show that means of these variables are many times larger than their respective medians.

It is equally clear that our corporate innovation variables have a discrete spike at zero. This indicates that some firms never engage in corporate innovation. Figures 1A−1C present the histograms of R&D expenditure, the number of patent counts, and the number of patent citations, respectively. Approximately one-half of our sample observations have no R&D investments and production in corporate patents.

Simple data treatment moderates the influence of univariate outliers. If vega is winsorized at the 99th percentile, the maximum value of vega will be set at 1,021.41 (z-score of 3.7) rather than at its original value of 10,840.44 (z-score of 43.25). The winsorized data value is still extreme and corresponds to a probability of less than 0.02% of such an extreme value occurring assuming vega is normally distributed. Alternatively, applying logarithmic transformation on our data moderates the influence of univariate outliers because they are highly right-skewed. However, in unreported results log-transformed data remain extreme in values but less right-skewed.

B. Mixture-distribution Model

As data on corporate innovation exhibit a discrete spike at zero, this indicates that some companies never intend to engage in corporate innovation. It is futile to estimate the relationship between managerial incentives and corporate innovation for such non-innovative firms because managerial incentives are used for purposes other than corporate innovation, e.g., mitigating agency problem (Alchian and Demsetz, 1972; Jensen and Meckling, 1976). We are surprised that nearly all previous studies in the corporate innovation literature include firms in non-innovative industries in the estimation. This implies that their results potentially underestimate the effects of managerial incentives on corporate innovation for firms that truly engage in corporate innovation.

In this paper, we first use a mixture-distribution model to separate industries into innovative and non-innovative. Next, we run several regressions of corporative innovation on managerial incentives by using only those firms belonging to the innovative industries. Specifically, our research methodology involves a two-step regression procedure. In the first stage, we use logistic regression to distinguish innovative industries from non-innovative ones. In the second stage, we use a regression method to estimate the effects of managerial incentives on corporate innovation for firms in the innovative industries.

Equation (2) is the first stage logistic regression:

(2) $c_i^* = z_i'\gamma + v_i$

where c_i^* is a latent variable which measures the propensity that a firm participates in research and development. Depending on whether the propensity exceeds a threshold value or not, the firm is classified as innovative or not: $c_i = 1$ if $c_i^* > 0$ and $c_i = 0$ if otherwise. The dependent variable of the logistic regression, c_i, is a binary variable and takes the value of zero if a firm's research and development expenditure is consistently zero throughout the sample period, and one otherwise. Note that c_i does not change its value over time, which implies that once a firm is counted as innovative, it stays as such whether the firm engages in innovation or not in a particular year. We use R&D expenditure to capture a firm's proclivity to engage in corporate innovation because it captures real resources committed to corporate innovation. Additionally, extant studies show that R&D expenditure is positively correlated with innovation outputs, e.g., patent applications and counts (Hall et al., 2005; Kortum, 1993).

The z_i is a vector of explanatory variables and v_i is a random error following a standard logistic distribution. Our explanatory variables only include a constant and 47 industry dummies to reflect the Fama-French 48-industry classification with the baseline industry being agriculture (Fama and French, 1997). This is because first we think that the nature and intensity of corporate innovation differ meaningfully across industries and second we do not want to select only those firms actively engaging in innovation within an innovative industry. We would like to highlight that our model is a mixture-distribution model, not a sample selection model. There are two types of industries, innovative and non-innovative. Any firm belonging to an innovative industry is counted as innovative while any firm belonging to a non-innovative industry as non-innovative. This methodology does not have sample selection problems. We simply estimate the effect of managerial incentives on corporative innovation using only firms in the innovative industries. We estimate equation (2) using the logistic maximum likelihood estimation method.[9]

To avoid statistical over-fitting, we use out-of-sample prediction to evaluate the performance of the mixture-distribution model. Thus, we first split our sample observations randomly into three sub-samples: 60% in estimation sample, 20% in validation sample, and 20%

[9] The logistic regression fails to identify some industries as either innovative or non-innovative. The failure occurs primarily because the corporate innovation activity is identical (or nearly identical) for all the firms within the same industry, i.e. either one or zero. In such circumstances, we classify all the firms in an industry as innovative if most firms in that industry report R&D activities and as non-innovative if most firms in that industry seldom report R&D activities. According to this classification rule, tobacco, medical equipment, aerospace, and defense industries are classified as innovative whereas precious metal, coal, measuring and control equipment, and transportation industries are classified as non-innovative.

in testing sample.[10] In the first-stage, we use the estimation sample to estimate equation (2). Next, for each firm in the validation sample we use the estimated coefficients obtained in the first-stage logistic regression to compute the probability ($\hat{p}_i$) that a firm is classified as innovative. We use a grid search algorithm to determine the optimal threshold (p_c) for classifying a firm as innovative such that the number of misclassified firms is minimized in the validation sample.[11] Last, we choose only firms in innovative industries in the estimation sample (henceforth "innovative-firm sample") for the second-stage regression using the model specified in equation (1).

In the testing sample, we use the estimated coefficients in equations (1) and (2) and the classification probabilities ($\hat{p}_i$) to estimate the intensity of corporate innovation for each firm-year observation as follows:[12]

(3) $\hat{y}_{it} = 0 \times (1 - \hat{p}_i) + \boldsymbol{x_{it}}\hat{\beta} \times \hat{p}_i$

where $\hat{p}_i$ is the estimated probability that a firm is classified as innovative and $\hat{\beta}$ is the coefficient estimates obtained in the second-stage regression using the estimation sample.

To evaluate the performance of the mixture-distribution model, we use a single-equation model as a benchmark. The single-equation model uses all the observations in the estimation sample regardless of whether it belongs to the innovative industries or not, to estimate the effect of managerial incentives on corporate innovation.[13] For each observation in the testing sample,

[10] The estimation sample has 1,163 firms (or 7,370 firm-year observations). The validation sample and the testing sample have 391 firms (or 2,475 firm-year observations) and 394 firms (or 2,534 firm-year observations), respectively.

[11] Less desirably, we could have used a naïve classification scheme, i.e., classify an industry as innovative if the estimated probability in the first-stage logistic regression is greater than 0.5 [or $P(c_i^*>0) > 0.5$] and non-innovative otherwise.

[12] Alternatively, we could have estimated the R&D intensity as 0 if $\hat{p}_i$ is smaller than p_c, and as $\boldsymbol{x_{it}}\hat{\beta}$ otherwise.

[13] The single-equation model uses all the observations in the estimation sample to estimate the relation between managerial incentives and corporate innovation. In contrast, the mixture-distribution model uses a subset of the observations in the estimation sample, namely firms belonging to those industries which are classified as innovative in the first stage regression.

we compare the mean absolute prediction error and goodness of fit between the mixture-distribution model and the single-equation model.

Table 3A reports the results of the first-stage logistic regression using the estimation sample. Expectedly, our results indicate that traditional R&D intensive industries are more likely to be classified as innovative, e.g., electronic equipment, pharmaceutical products, automobiles and trucks, computers, and aircrafts. In contrast, conventional non-R&D intensive industries are more likely to be classified as non-innovative, e.g., retail, restaurants, and wholesale. The goodness of fit of the first stage logistic regression is good with a pseudo R-square value of 0.44.[14]

[INSERT TABLE 3A HERE]

Table 3B reports the classification accuracy of the first stage regression using the testing sample. Although our explanatory variables only include industry dummies, they are useful to classify firms into innovative and non-innovative. Out of a total of 356 firms in the testing sample, 81.2% of them are correctly classified.[15] Similarly, alternative statistics to measure the classification performance of the first stage regression also point to the same conclusion.

[INSERT TABLE 3B HERE]

Table 4 reports the least squares estimates of the model specified in equation (1) using the estimation sample. The results using the single-equation model are reported in columns (1), (3), and (5) and those using the mixture-distribution model in columns (2), (4), and (6). In terms of

[14] We have decided not to include any firm-specific characteristics in the first stage regression for two reasons. First, the model fitness does not improve appreciably. In unreported results, we observe that the pseudo R-square value increases only slightly to 0.47 when we include firm size and its square as additional explanatory variables in the first stage regression. Second, we do not want to incur sample selection by basing our classification of innovative vs. non-innovative firms only on their industry characteristics. As a result, if a firm in an industry is classified as innovative, all the other firms in that industry are also classified as innovative and vice versa.

[15] The classification accuracy is computed as follows: the ratio of the sum of the number of firms correctly classified as innovative plus the number of firms correctly classified as non-innovative to the total number of firms. The classification accuracy in our testing sample is (171+118)/356, or 0.812.

model performance, our results indicate that the mixture-distribution model is superior to the single-equation model. For each corporate innovation proxy, the out-of-sample mean absolute residual (MAR) is smaller in the mixture-distribution model than in the single-equation model.

[INSERT TABLE 4 HERE]

Similarly, the relative goodness of fit statistic also indicates superior performance of the mixture-distribution model.[16] The relative goodness of fit statistics are consistently greater than 0.5 in favor of the mixture-distribution model.[17] For example, in the R&D regression the relative goodness of fit statistic of the mixture-distribution model is 0.558. This indicates that the absolute prediction error is smaller in the mixture-distribution model than in the single-equation model for 55.8 percent of the testing sample observations.

Our results indicate that least squares estimates of vega are highly sensitive to model specifications. Specifically, they are statistically indistinguishable from zero in the single-equation model whereas those are statistically significant at conventional levels in the mixture-distribution. This implies that higher vega induces only those firms belonging to the innovative industries to increase corporate innovation. If higher managerial incentives are consciously structured to enhance corporate innovation for only firms in innovative industries, we expect that the estimated effects of vega should be stronger in the mixture-distribution model than in the singe-equation model. Expectedly, the least squares estimates of the elasticity of corporate innovation with respect to vega are significantly larger in the mixture-distribution model than those in the single-equation model. For example, the estimated elasticity of corporate patents

[16] Suppose $\hat{y}_i^{(1)}$ and $\hat{y}_i^{(2)}$ are two different predictors for y_i using two competing models (1) & (2) respectively. The goodness of fit of model (2) relative to model (1) is defined as follows: $\sum_{i=1}^{n} I\left[\left|\hat{y}_i^{(2)} - y_i\right| < \left|\hat{y}_i^{(1)} - y_i\right|\right]/n$, where n is the size of the testing sample and y_i is the observed value of the dependent variable. The relative goodness of fit evaluates the proportion of observations that are better predicted by model (2), relative to that predicted by model (1).

[17] Model (2) is considered to be better than model (1) when the goodness of fit of model (2) relative to model (1) is greater than 0.5, that is $\sum_{i=1}^{n} I\left[\left|\hat{y}_i^{(2)} - y_i\right| < \left|\hat{y}_i^{(1)} - y_i\right|\right]/n > 0.5$.

with respect to vega increases from 0.0307 in the single-equation model to 0.062 in the mixture-distribution model, or by 103%.[18]

Our results also show that all the least squares estimates of delta are positive and statistically significant at conventional levels in both models. This implies that higher CEO pay-for-performance sensitivity increases corporate innovation. This conclusion, though, should not be taken seriously because our estimated coefficients of delta change appreciably to different remedies of outliers (to be discussed in the subsequent sections). This implies that the least squares estimates of delta are sensitive to the outliers' influence.[19]

C. Influential Observations

To assess the effect of influential observations on least squares estimates, in the second stage we perform two sets of regression for each corporate innovation proxy. In the first set of regression, we use only firms in the innovative industries of the estimation sample. In the second set, of all the firms in the innovative industries of the estimation sample, we exclude one firm which has the largest combined influence on the estimated coefficients of delta and vega.[20] To quantify the

[18] The elasticity of corporate patents with respect to vega (ε) is computed as follows: $\varepsilon = \beta \times x \times (1+y)/y$, where β is the least square estimate of vega, x is vega, and y is corporate innovation. We evaluate x and y at their respective means using the full sample comprising 12,379 observations. By arranging terms, we can express β as follows: $\beta = \partial \log(1+y)/\partial x = \partial y)/\partial x \times 1/(1+y)$. As elasticity ($\varepsilon$) is defined as $\partial y/\partial x \times x/y$, it is easy to show that $\varepsilon = \beta \times x \times (1+y)/y$.

[19] Additionally, our least squares estimates of delta are unanticipated and contradictory to those reported in the literature. This is because the least squares estimates of delta are significantly smaller in the mixture-distribution model than in the single-equation model. This implies that higher CEO pay-for-performance sensitivity has a weaker effect on corporate innovation for firms in innovative industries than for firms in non-innovative industries. Our results in subsequent sections show that higher CEO pay-for-performance sensitivity has no material effect on corporate innovation after the influence of outliers is mitigated.

[20] As delta and vega are our main variables of interest, we compute the outlier's influence by using a multivariate version of $DFBETA_i$ as follows: $DFBETA_i = \sqrt{\left(\hat{\beta} - \hat{\beta}_{(-i)}\right)^T \hat{V}_{(-i)}^{-1}\left(\hat{\beta} - \hat{\beta}_{(-i)}\right)}$, where $\hat{\beta}$ is a vector of the estimated coefficients of delta and vega using all the firms in innovative industries in the estimation sample and $\hat{\beta}_{(-i)}$ is a vector of the comparable estimates after firm i is excluded from this sample and $\hat{V}_{(-i)}^{-1}$ is defined analogously as the inverse of the estimated variance covariance matrix after firm i is excluded from this sample. The excluded firm i which maximizes $DFBETA_i$ is referred to the most influential firm because dropping it produces the largest

impact of the influential firm, we compare the estimated coefficients of delta and vega from the two sets of regression.

Columns (1), (3), and (5) of Table 5 report the least squares estimates of each corporate innovation proxy in the first set of regressions and columns (2), (4), and (6) of the same Table report those in the second set of regressions. The most influential firm turns out to be Microsoft in all the regressions.

[INSERT TABLE 5 HERE]

Our results indicate that the outliers' influence on the least squares estimates of delta is pronounced. For example, after excluding Microsoft from the sample of 635 firms in the innovative industries, the least square estimate of the elasticity of corporate patents with respect to delta drops from 0.0089 (p-value of 0.001%) in column (3) to –0.0247 (p-value of 27.87%) in column (4), or by 377%. Similarly, removing Microsoft from the sample lowers the estimated elasticity of R&D with respect to delta by 583% and that of the elasticity of patent citations with respect to delta by 279%. In all cases, removing Microsoft renders the least square estimates of delta statistically indistinguishable from zero, compared with statistical significance at conventional levels when Microsoft is included in the sample.

In contrast, the outliers' influence on the least squares estimates of vega is much smaller. After excluding Microsoft from the sample, the least squares estimates of vega increase slightly with a range from 10.43% in the patent citations regression to 14.17% in the corporate patents regression. Besides, all the least squares estimates of vega are statistically distinguishable from zero, regardless of whether Microsoft is included or excluded from the sample. Our results indicate that the vega effect is robust to the outliers' influence whereas the delta effect is not.

combined influence on the estimated coefficients of vega and delta. Besides, the most influential firm can be viewed as an (conditional) outlier.

This compels us to investigate in the next section whether our results vary meaningfully to different remedies of outliers.

D. Remedies of Outliers

To investigate the influence of different remedies of outliers on our least squares estimates, we perform three conventional remedies of outliers including (i) logarithmic transformation of managerial incentive proxies, (ii) data winsorization, and (iii) median regression.[21] Table 6 reports least squares estimates under different remedies of outliers for firms in innovative industries in the estimation sample: log transformation of one plus each managerial incentive proxy ("log-transformed") in column (2); winsorization of dependent variable—R&D, #patents, and #patent_cites—at the first and the 99th percentiles ("partially-winsorized") in column (3); winsorization of all variables at the first and the 99th percentiles ("fully-winsorized") in column (4); and median regression in column (5).[22]

[INSERT TABLE 6 HERE]

Our results in Table 6 show that the estimated effects of delta and vega change appreciably to different remedies of outliers. The statistical significance of the least squares estimates of delta alters significantly depending on whether data are partially- or fully-winsorized. When only dependent variables are winsorized, the least squares estimates of delta are economically and statistically identical to those using untreated data. In contrast, when data are fully-winsorized, the least squares estimates of delta are meaningfully different from those

[21] Data trimming—discard observations with extreme data points—is occasionally used to treat outliers. Nevertheless, data trimming has several shortcomings. First, least squares estimates are biased if data are selectively trimmed (Kothari et al., 2005). Second, data trimming is subjective and susceptible to confirmation bias. Third, it removes many observations if data are mechanically trimmed at a customary level, e.g., the first and 99th percentiles. In untabulated results, when data are fully-trimmed our least squares estimates are qualitatively similar to those when data are fully-winsorized.

[22] In untabulated results, our conclusion is qualitatively similar and robust to other winsorization cutoff levels.

using untreated data.[23] For instance, the least squares estimate of the elasticity of corporate patents with respect to delta drops from 0.00892 (*p*-value of 0.001%) using untreated data in column (1) to –0.0457 (*p*-value of 56.91%) using fully-winsorized data in column (4), or by 612%.

Expectedly, the outliers' influence is not moderated when only dependent variables are winsorized. This is because this remedy fails to modify observations with extreme data points in independent variables (i.e., high leverage points) which are usually the source of outliers.[24] Furthermore, least squares estimates are biased when only dependent variables are winsorized (Goldberger, 1981). Although the outliers' influence is moderated when data are fully-winsorized, this remedy has many shortcomings. First, excessive data winsorization distorts informative data. Second, it reduces efficiencies of least squares estimators under most common circumstances (Lien and Balakrishnan, 2005). Third, this remedy is subjective and susceptible to confirmation bias.

Our results also indicate that the statistical significance of the least squares estimates of vega is robust and distinguishable from zero, regardless of whether data are treated or not. However, the magnitude of these least squares estimates changes significantly under different winsorization treatments. For example, the estimated elasticity of corporate patents with respect to vega increases from 0.062 (*p*-value of 0.06%) using untreated data in column (1) to 0.149 (*p*-value of 0.02%) using fully-winsorized data in column (4), or by 141%. This implies that the estimated effects of vega vary under different data treatments.

Our least squares estimates also differ meaningfully when we apply logarithmic

[23] In unreported tables, when only explanatory variables are winsorized at the first and 99th percentiles the least squares estimates of delta and vega are qualitatively identical to those using fully-winsorized data.

[24] In statistical jargon, high leverage points refer to observations with an independent variable which has extreme values in its distribution.

transformation on the managerial incentive variables, compared with those using untreated data. In particular, when we use log-transformed data the least squares estimates of vega are economically and statistically distinguishable from zero whereas those of delta are not. This implies that logarithmic transformation is useful to moderate the outliers' influence in our model. In general, logarithmic transformation is commonly used to deal with extreme data points because it compresses highly right-skewed data without distorting raw data. However, log-transformed data may exacerbate the outliers' influence because logarithmic transformation may introduce more skewness than untreated data if the underlying distribution is not log-normal (Feng et al., 2012).

As indicated by the results in Table 5, the outliers' influence on the least squares estimates of vega are small whereas those of delta are much larger. Thus, we anticipate that the estimated effects of vega should be qualitatively similar between median regression and least squares estimates. We are surprised that the median regression estimates of vega differ meaningfully from the least squares estimates using untreated data. In particular, our median regression estimates of vega are statistically indistinguishable from zero whereas those using least squares regressions in untreated data are not.[25] Quite the contrary, the estimated effects of delta are economically and statistically similar between median regressions and least squares regressions. These discrepancies compel us to use quantile regression method in the following section. This is because it allows us to investigate the extent of heterogeneity in the relationship at different points of the conditional distribution rather than at a single quantile point (or median).

[25] Overall, our least squares estimates of vega are qualitatively similar to those reported in extant studies (e.g., Coles et al., 2006; Hirshleifer et al., 2012) in that the least squares estimates of vega are positive and statistically significant at conventional levels.

E. Quantile Regression

In the second-stage procedure, we use quantile regression method for estimation because their estimates are robust to outliers (e.g., Koenker and Basset, 1978; Koenker and Hallock, 2001). Quantile regression also allows for different estimates to be computed at different quantiles of the conditional distribution. The latter consideration is relevant if the relationship differs meaningfully across different quantiles of the conditional distribution. In practice, many economic relationships differ significantly across individuals. For example, the pay-for-performance sensitivity varies considerably across different CEOs, firm types, and firm sizes (e.g., Hallock et al., 2010; Conyon and Schwalbach, 2000; Baker and Hall, 2004; and Schaefer, 1998). The relationship between managerial incentives and corporate innovations differs meaningfully not only across the innovative industries and the non-innovative industries, but also across the firms within the innovative industries.

We use equation (1) to demonstrate the latter point. The least squares regression minimizes the sum of squared residuals (e_{it}^2) with respect to β:

(4) $$\Sigma e_{it}^2 = \Sigma(y_{it} - x_{it}'\beta)^2$$

where y_{it} is the dependent variable, x_{it} is a vector of explanatory variables, β is a vector of regression coefficients, and e_{it} is a residual. The least squares estimates are vulnerable to the outlier's influence because they are estimated based on a sum of squared residuals. In contrast, quantile regression minimizes the sum of absolute residuals (asymmetrically weighted):

(5) $$\Sigma| y_{it} - \mathbf{x}_{it}'\beta(\tau) | \times [\tau\times I(y_{it} > \mathbf{x}_{it}'\beta(\tau)) + (1-\tau)\times I(y_{it} \leq \mathbf{x}_{it}'\beta(\tau))]$$

where $\beta(\tau)$ is a vector of regression coefficients that depend on τ, the quantile being estimated (where $0 < \tau < 1$), and I is the usual indicator function that takes the value of one if the condition in the parentheses is true, and zero otherwise. The formula in equation (5) gives a weight of τ to any observations that are greater than their respective predicted values and a weight of $(1 - \tau)$ to any observations that are smaller than their respective predicted values.

When $\tau = 0.5$, the quantile regression reduces to median regression and equation (5) can be simplified as follows:

(6) $\Sigma|y_{it} - \mathbf{x}_{it}'\beta|/2$.

Quantile regression allows the regression estimates to differ by quantiles, that is, $\beta(\tau)$ can be different across different quantile points being estimated (τ).

We use quantile regressions to estimate equation (1). We also perform logarithmic transformation of our managerial incentive variables as follows: log(1+vega) and log(1+delta). This is because in unreported results we find that the quantile regression estimates of vega and delta are significantly more stable across different conditional quantiles after the transformation. This indicates that logarithmic transformation is effective to mitigate the outliers' influence. Next, we run nine quantile regressions (where τ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9) only using firms in innovative industries in the estimation sample.

Table 7 reports the results of the nine quantile regressions of each measure of corporate innovation using only firms in innovative industries of the estimation sample. The estimates of nine quantile regressions are reported in columns (2)–(10). For better comparison, we also report the least squares estimates in column (1).

[INSERT TABLE 7 HERE]

Our results show that nearly all quantile regression estimates of vega are positive. This indicates that higher vega induces firms in innovative industries to increase corporate innovation. The positive relationship is most relevant for a firm at the middle quantiles of the conditional distribution (where $0.5 \leq \tau \leq 0.8$) because only those estimates are statistically distinguishable from zero. In contrast, the quantile regression estimates at the lowest quantiles of the conditional distribution (where $\tau = 0.1$ or 0.2) are close to zero and statistically indistinguishable from zero. For example, the results in Panel B of Table 7 show that the estimated elasticity of corporate patents with respect to vega is 0.164 for a firm in the eightieth percentile of the conditional distribution ($\tau = 0.8$), compared with zero for a firm in the tenth percentile of the conditional distribution ($\tau = 0.1$). The difference in these two estimates is 0.164. This implies that there is a considerable heterogeneity in the estimated effects of vega on corporate innovation. The heterogeneity in the estimated effects of vega can be clearly seen in Figures 2A, 3A, and 4A which plot the nine quantile regression estimates of vega for the R&D, corporate patents, and patent citations regressions, respectively.

For the estimated effects of vega the least squares and median regression estimates are remarkably similar. The results in panel A of Table 7 indicate that the least squares estimate of the elasticity of R&D with respect to vega is 0.124 whereas that of the median regression ($\tau = 0.5$) is 0.126. Besides, the quantile regression estimates of vega are similar for the quantiles greater than 0.2 ($\tau > 0.2$). The results in panel A of Table 7 show that the estimated elasticity of R&D with respect of vega ranges from 0.0837 ($\tau = 0.8$) to 0.126 ($\tau = 0.5$) for the quantiles greater than 0.2. Overall, our quantile regression results are in congruent with findings in extant studies in that higher sensitivity of CEO wealth to stock volatility induces firms to increase corporate innovation (Coles et al., 2006; Hirshelifer et al., 2016).

In contrast, none of the quantile regression estimates of delta are statistically distinguishable from zero. This implies that higher CEO pay-for-performance sensitivity (delta) has no material effect on corporate innovation, regardless of the expected intensity of corporate innovation of the firm. Similarly, the heterogeneity in the estimated effects of delta can be seen in Figures 2B, 3B, and 4B which plot the nine quantile regression estimates of delta for the R&D, corporate patents, and patent citations regressions, respectively.

As indicated clearly in Figure 2B, the least squares estimate of the elasticity of R&D with respect of delta is slightly positive whereas those of the quantile regression estimates are negative except for those at the twentieth and twentieth-fifth quantiles. This contradictory result is attributable to an outlier, namely Microsoft. In statistical parlance, Microsoft is a vertical outlier because it invests heavily in R&D.[26] Microsoft is one of the biggest spenders on R&D investments in the U.S. during the recent decades. Microsoft is also a high leverage point because the company's CEOs have unusually large equity incentives (i.e., exceptionally large delta) but no stock-options (i.e., zero vega).[27] William Gates and Steve Ballmer, the former CEOs of the company never received a single stock-option grant as they were given substantial stock incentives prior to the company went public. Note that this unique managerial incentive structure is not uncommon in the U.S. In practice, many founder-CEOs rely primarily on stock (rather than stock-option) incentives to implement risky policies including corporate innovation.[28]

F. Robustness Tests

To further understand the merit of the mixture-distribution model, we re-run quantile regressions using all the observations in the estimation sample. Table 8 reports the results of the nine quantile regressions of each measure of corporate innovation after firms in non-innovative

---

[26] Vertical outliers refer to observations with extremely large error terms in a given model.
[27] The sensitivity of CEO wealth to stock price (delta) is exceptionally large and economically meaningful for William Gates and Steve Ballmer. For example, in 1998 the wealth of William Gates changed by approximately $709.8 million for a one percentage point change in Microsoft's stock price.
[28] For example, the compensation is primarily cash-based for Jeff Bezos and Steve Jobs, the founders of Amazon and Pixar, respectively. Additionally, they never received a single stock option grant due to their substantial stock ownership in the company.

industries are also included in the sample. The estimates of nine quantile regressions are reported in columns (2)–(10) and that of the least squares estimate in column (1).

[INSERT TABLE 8 HERE]

Our quantile regression estimates in Table 8 differ meaningfully from those in Table 7 which are obtained by using only firms in innovative industries. Specifically, the results in Table 8 show that our quantile regression estimates of vega at the lower quantiles ($0.1 \le \tau \le 0.4$) are close to zero and statistically indistinguishable from zero. In contrast, those at the higher quantiles ($\tau \ge 0.5$) is positive. However, of twenty-seven quantile regression estimates of vega, only two are statistically significant at conventional levels. Besides, none of these estimates are statistically distinguishable from zero for the corporate patents regressions. Including firms in non-innovative industries in the sample meaningfully weakens the effects of vega except for (conditionally) very innovative firms ($\tau = 0.9$). Expectedly, none of the quantile regression estimates of delta are statistically significant at conventional levels.

The heterogeneity in the estimated effects of vega is noticeably weaker when we include firms in non-innovative industries in the quantile regressions. Figures 5A–7A plot the nine quantile regression estimates of vega for the R&D, corporate patents, and patent citations regressions. For the estimated effects of vega, the least squares and median regression estimates are remarkably similar when we use only firms in innovative industries in the estimation. They are quite different, though, when firms in non-innovative industries are included in the estimation. For instance, the results in Figure 5A indicate that the least squares estimates of the elasticity of R&D with respect to vega increase as vega increases. In contrast, those using median regressions ($\tau = 0.5$) are constant and do not change when vega increases.

As data on corporate innovation has a discrete spike at zero, some studies remove those observations from the least squares regressions, e.g., Hirshleifer et al. (2012). However, this approach has two shortcomings. First, it discards observations with useful information because a firm's corporate innovation output could be irregular over time. For instance, a company may receive patent grants in some years but none in other years. In unreported results, we find that over 25 percent of our sample firms have zero corporate patents in some years but non-zero corporate patents in other years.

Second, selectively removing such firm-year observations with zero corporate innovation, first incurs sample selection bias and second weakens the heterogeneity in the relationship between managerial incentives and corporate innovation within an industry. In unreported results, when we exclude data with zero corporate patents, our sample size decreases from 7,370 to 3,127, or by 58%. Besides, the ranges of our quantile regression estimates are meaningfully smaller than those reported in Table 7. For example, the difference in the quantile regression estimates of vega between the lowest quantile ($\tau = 0.1$) and the highest quantile ($\tau = 1.0$) is significantly smaller when we exclude firm-year observations with zero corporate patents than that reported in Panel A of Table 7.

Conclusion

The mixture-distribution model is a preferred specification to study issues on corporate innovation. This is because many firms never commit resources on corporate innovation and should be excluded from the estimation. The mixture-distribution model allows us to better pin down the impact of managerial incentives on corporate innovation objectively for firms that truly care about corporate innovation.

Second, OLS regression method is inappropriate to estimate the relationship between managerial incentives and corporate innovation. This is because least squares estimates are susceptible to the outliers' influence given our data are highly right-skewed with obvious outliers. Conventional remedies of outliers (e.g., data winsorization) are inadequate because outliers are a conditional concept and should be treated based on a reasonable choice of a model and covariates. Furthermore, least squares estimates are fragile and vary appreciably to different data treatments. Last, data treatment is subjective and vulnerable to the confirmation bias.

To mitigate the outliers' influence, we first use quantile regressions for estimation and second apply logarithmic transformation on key variables of interest. Our quantile regression results indicate that higher sensitivity of CEO wealth to stock return volatility (vega) induces firms in innovative industries to increase R&D investment, corporate patents, and patent citations. In addition, the effects of vega are on average stronger for conditionally (expected) more innovative firms than for conditionally (expected) less innovative firms. In contrast, higher CEO pay-for-performance sensitivity (delta) has no material effect on corporate innovation. The latter result adds depth to our understanding of the relevance of the standard pay-for-performance compensation contracts on creativity and corporate innovation. Our results are consistent with findings in recent studies showing that standard pay-for-performance incentive schemes are not necessarily optimal to enhance corporate innovation (Ederer and Manso, 2013; Manso, 2011; Gneezy and Rustichini, 2000).

FIGURE 1A: Histogram of Research and Development Expenditure scaled by Total Asset

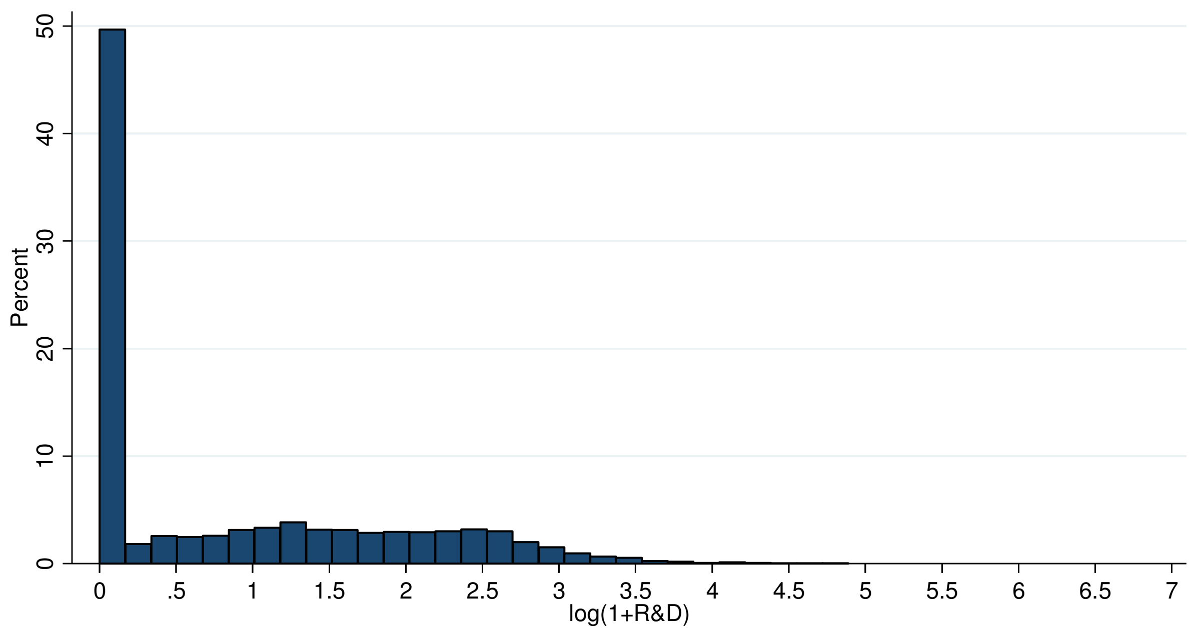


FIGURE 1B: Histogram of the Number of Patent Counts

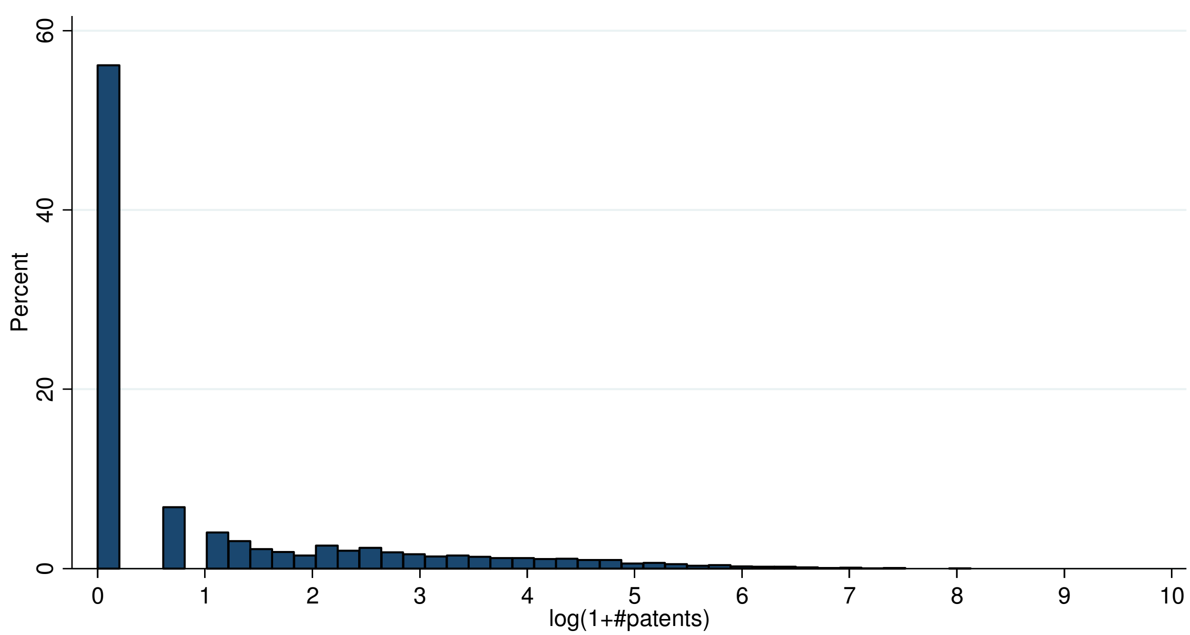


FIGURE 1C: Histogram of the Number of Patent Citations

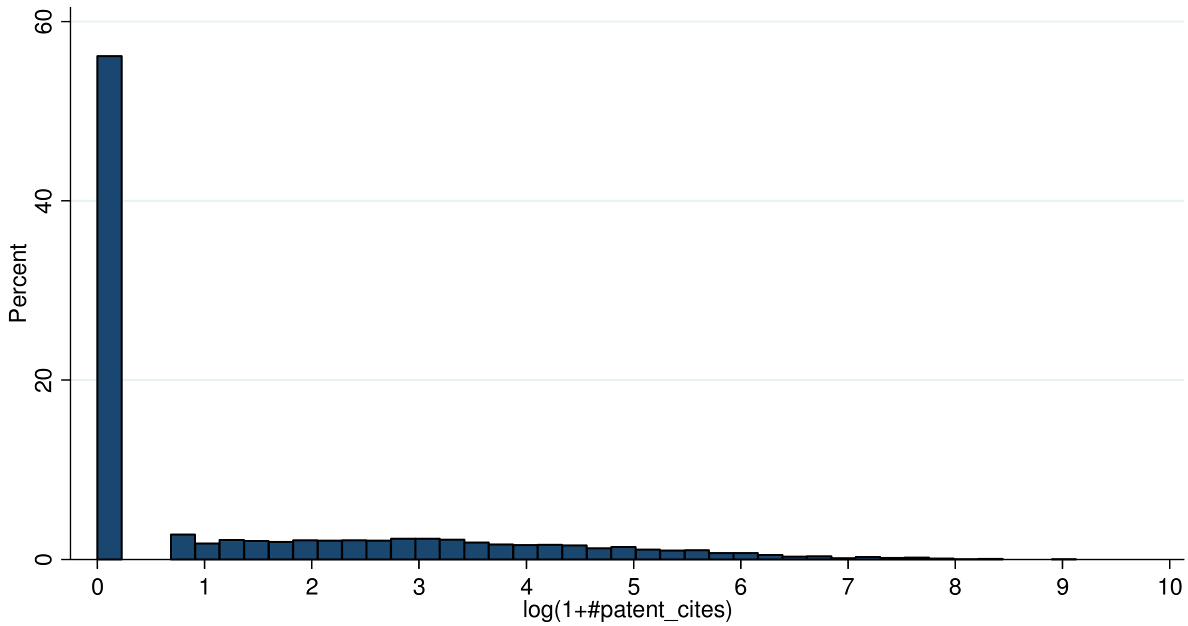


FIGURE 2A: Quantile Regression Estimates of Vega — R&D Expenditure scaled by Total Asset

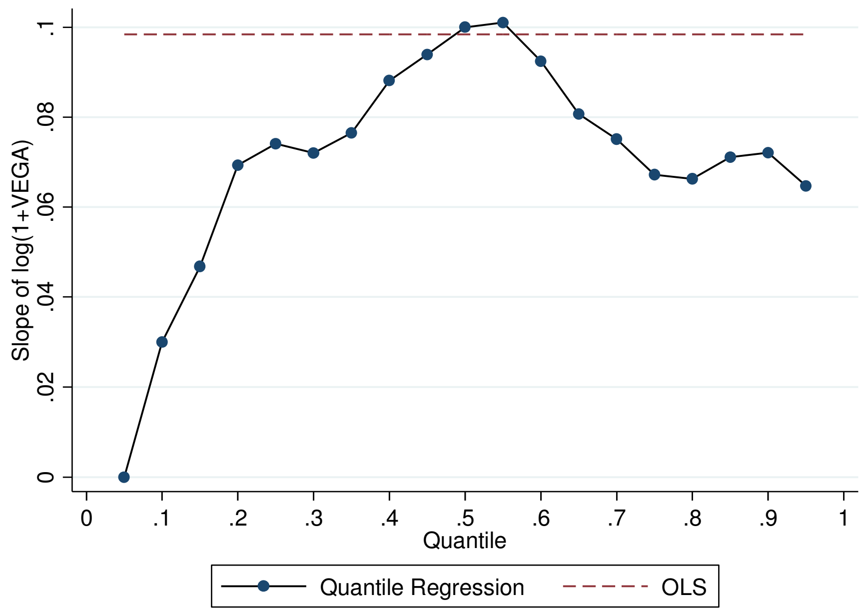


FIGURE 2B: Quantile Regression Estimates of Delta — R&D Expenditure scaled by Total Asset

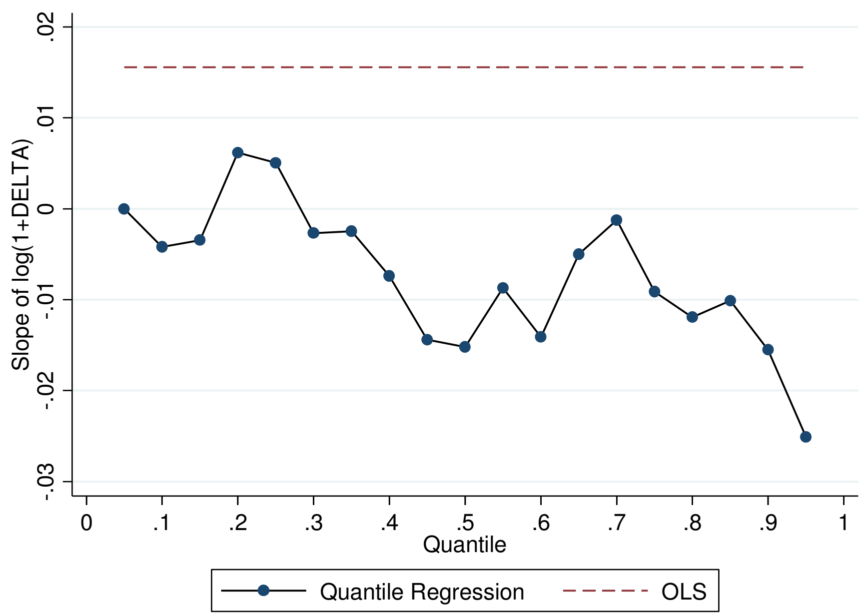


FIGURE 3A: Quantile Regression Estimates of Vega — Number of Patent Counts

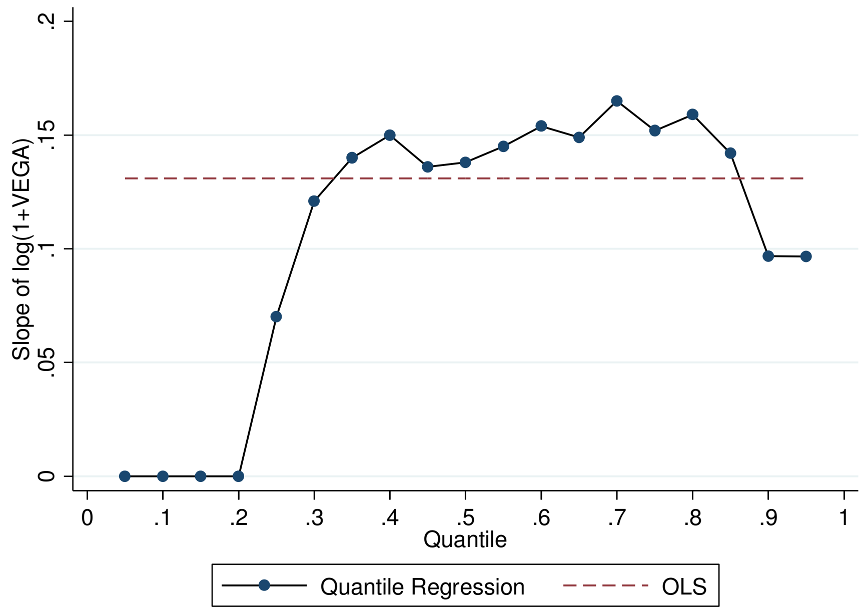


FIGURE 3B: Quantile Regression Estimates of Delta — Number of Patent Counts

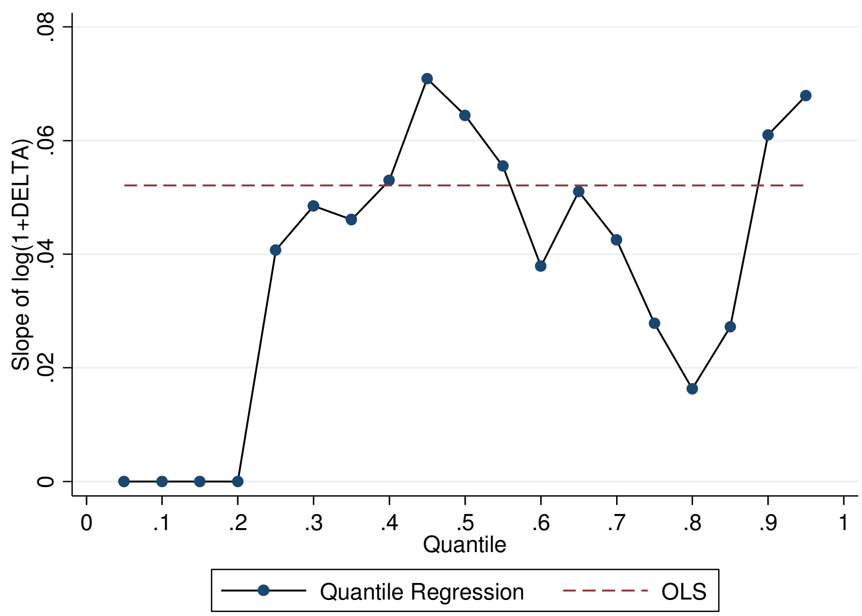


FIGURE 4A: Quantile Regression Estimates of Vega — Number of Patent Citations

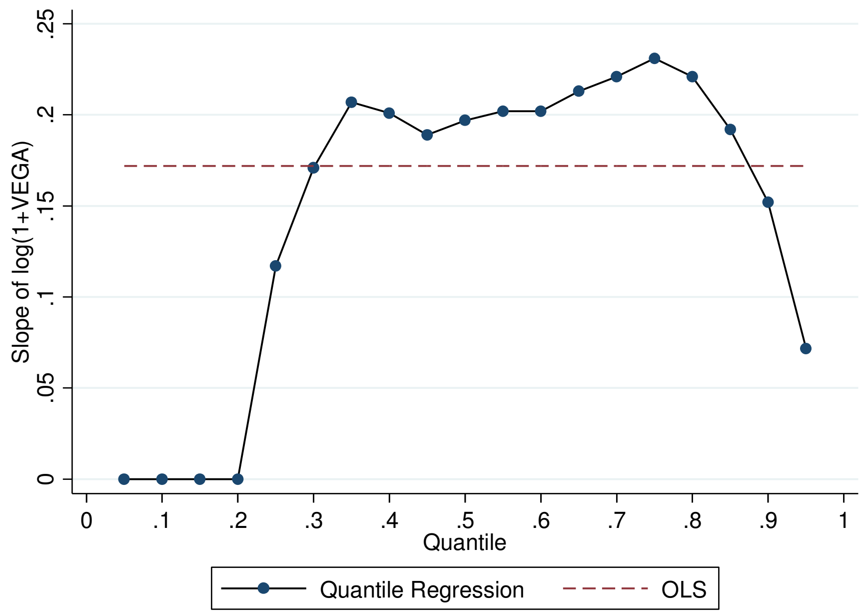


FIGURE 4B: Quantile Regression Estimates of Delta — Number of Patent Citations

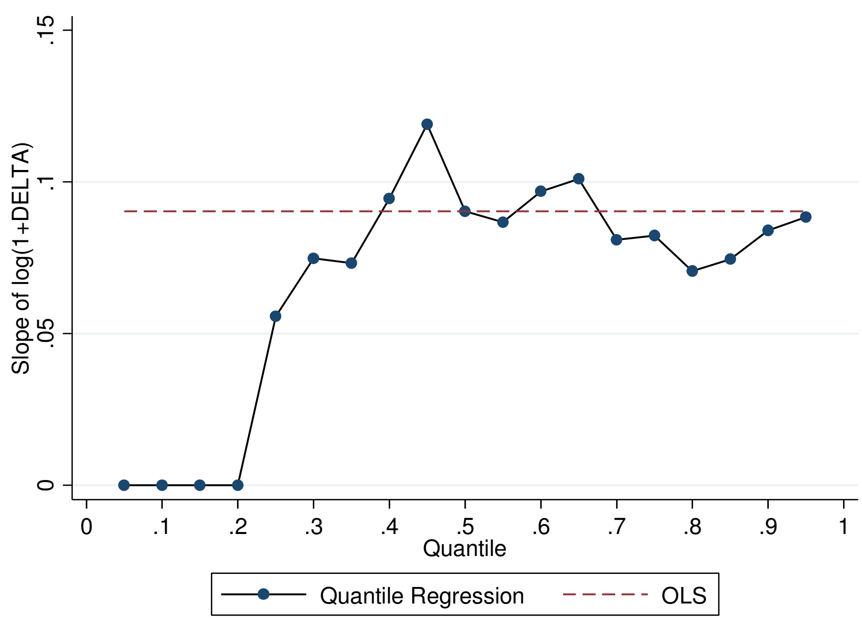
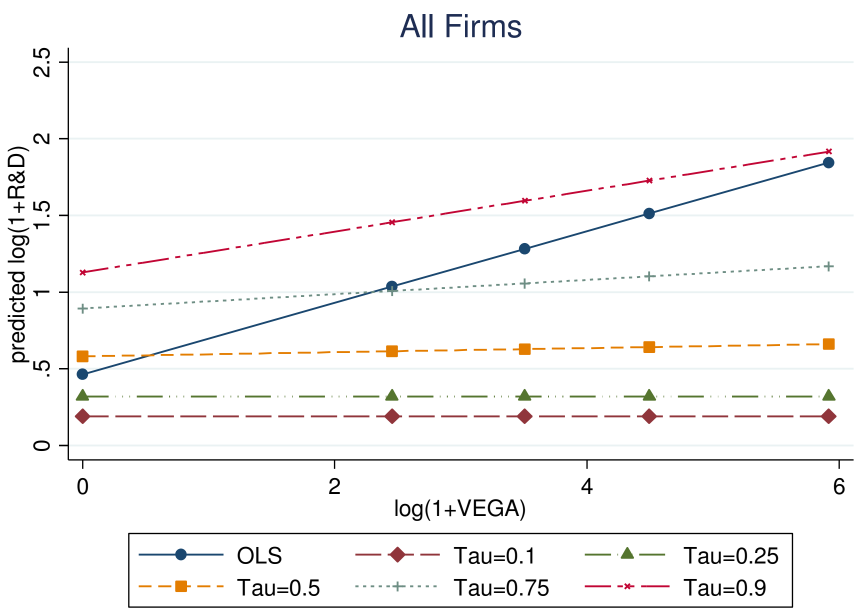


FIGURE 5A: Quantile Regression – R&D Expenditure scaled by Total Asset vs. VEGA

The upper figure uses all the observations in the estimation sample ("All Firms") and the lower one uses only firms in innovative industries in the estimation sample ("Firms in innovative industries"). The y-axis is the predicted value of the corporate innovation variable. The predicted value is computed based on the quantile regression estimate at a given quantile (τ) assuming all the explanatory variables are evaluated at their respective means. The x-axis is the logarithmic transformation of one plus the managerial incentive variable.



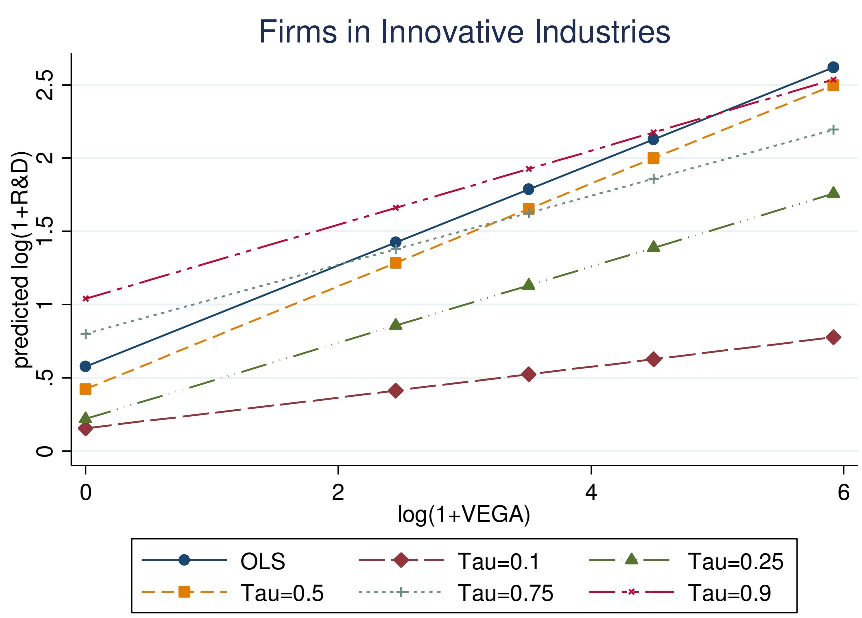
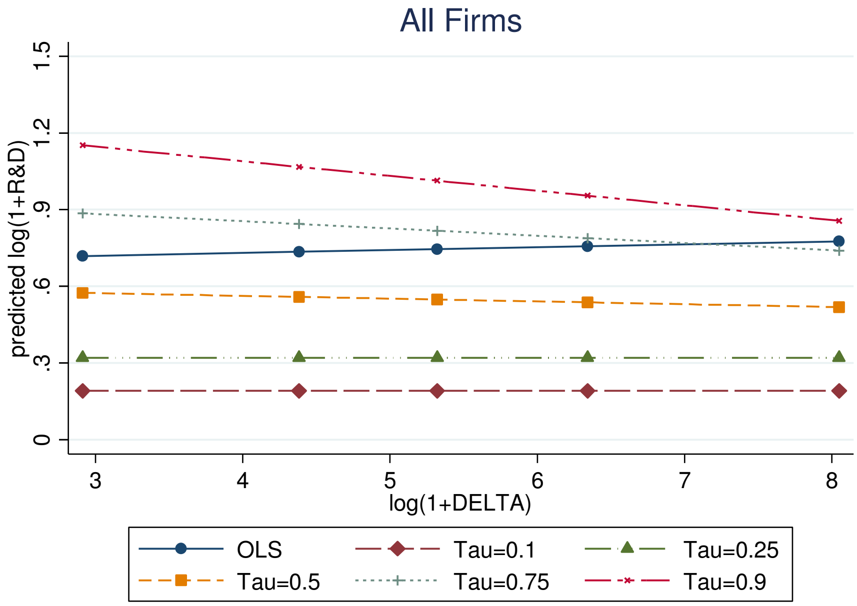


FIGURE 5B: Quantile Regression– R&D Expenditure scaled by Total Asset vs DELTA



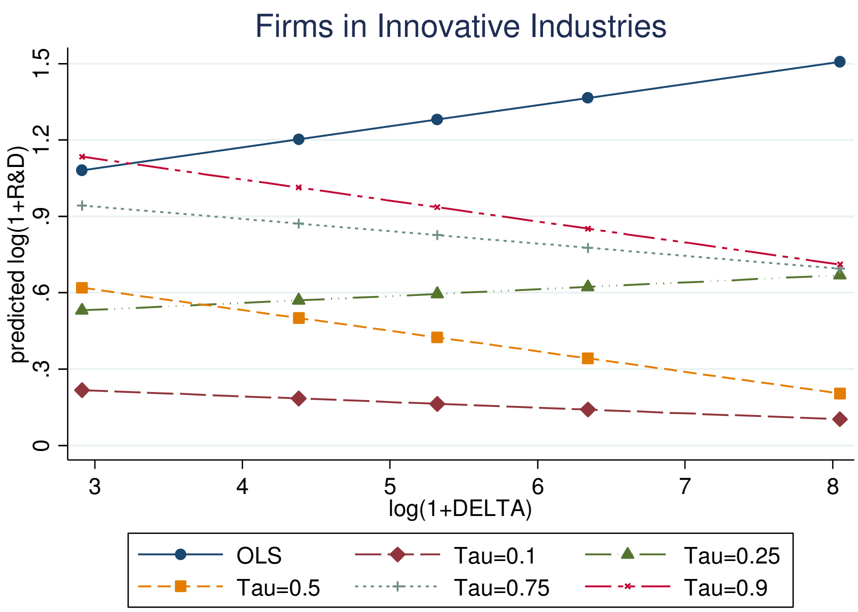
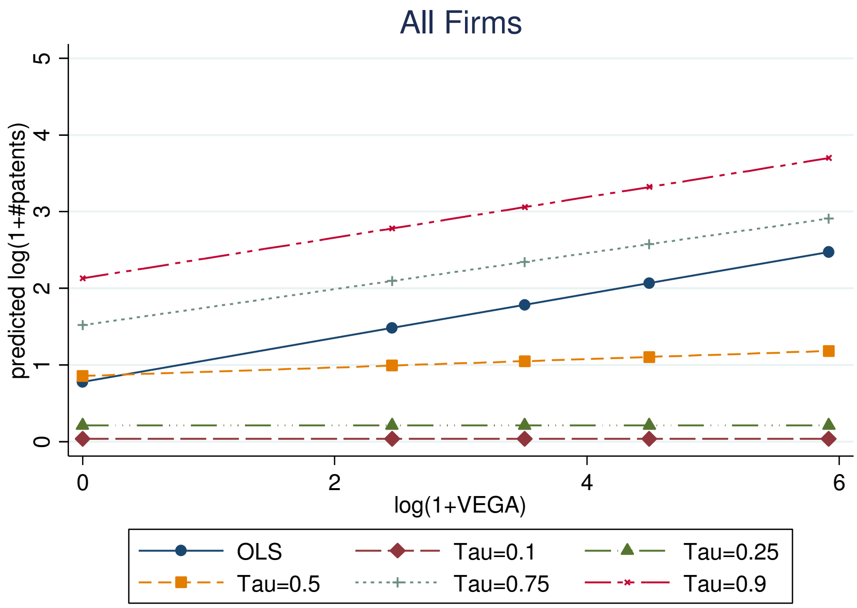


FIGURE 6A: Quantile Regression– Number of Patent Counts vs VEGA



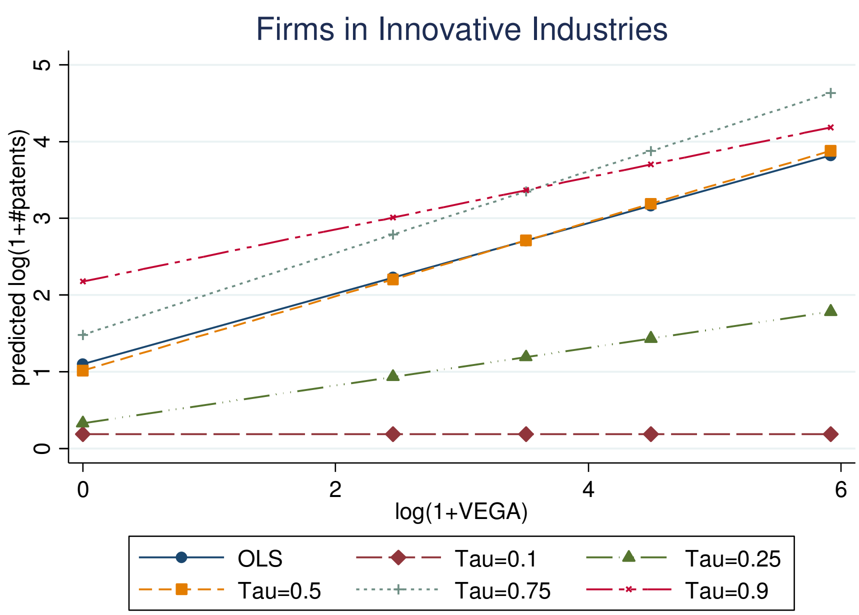
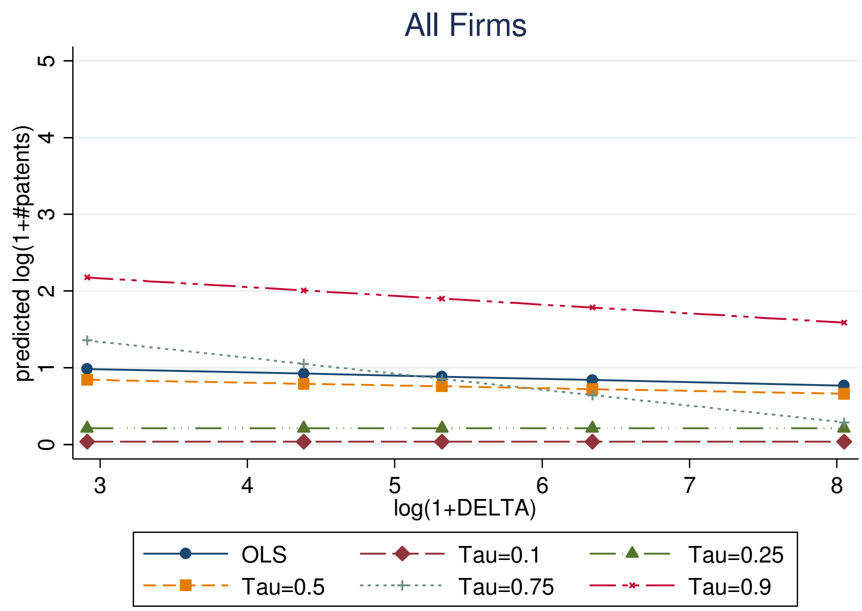


FIGURE 6B: Quantile Regression– Number of Patent Counts vs DELTA



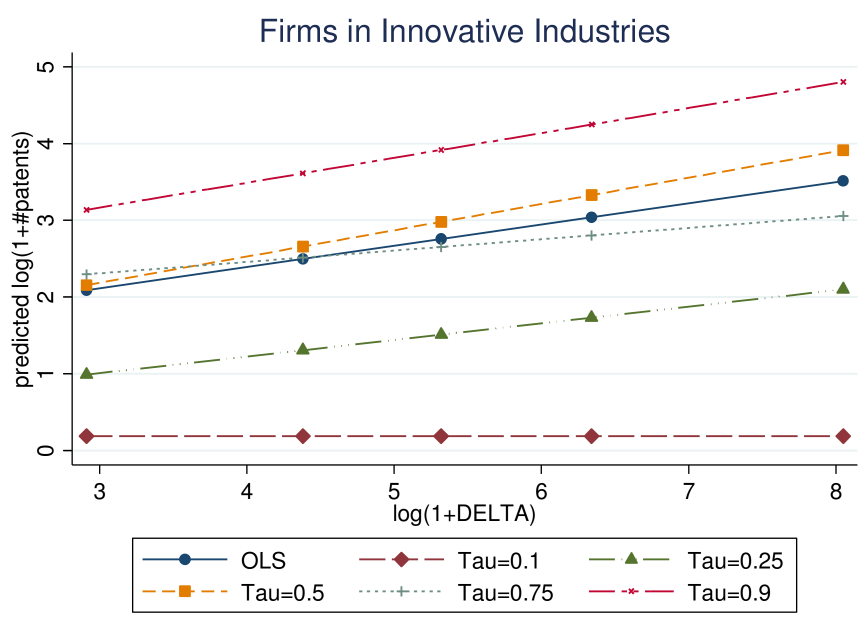
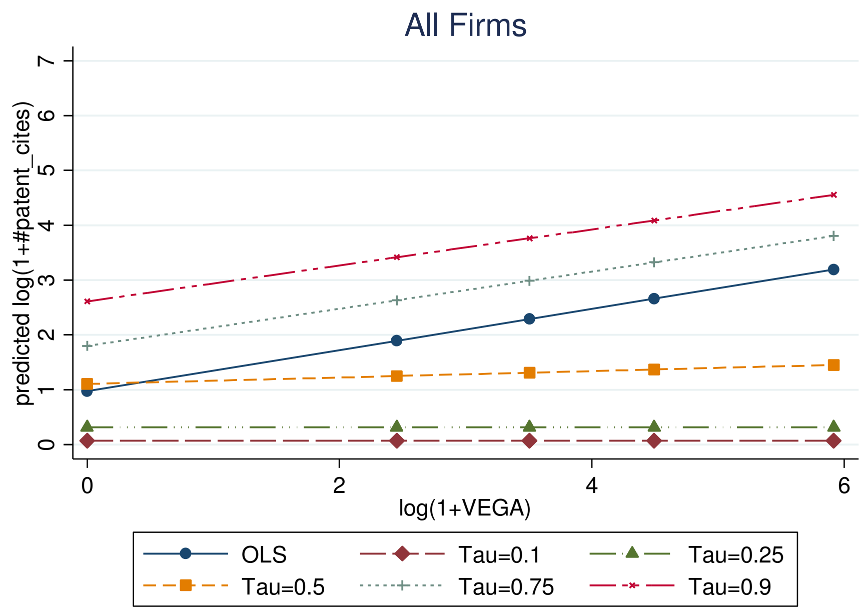


FIGURE 7A: Quantile Regression– Number of Patent Citations vs VEGA



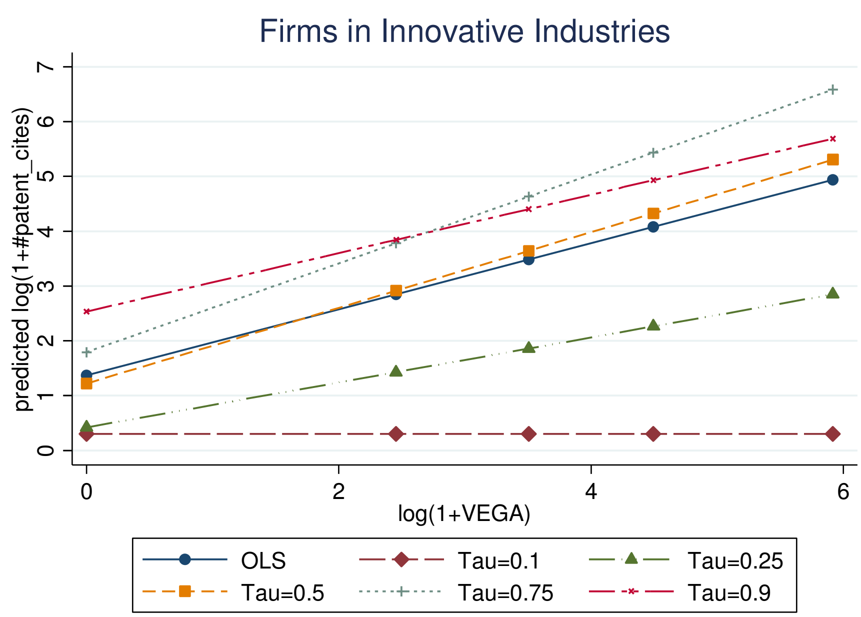
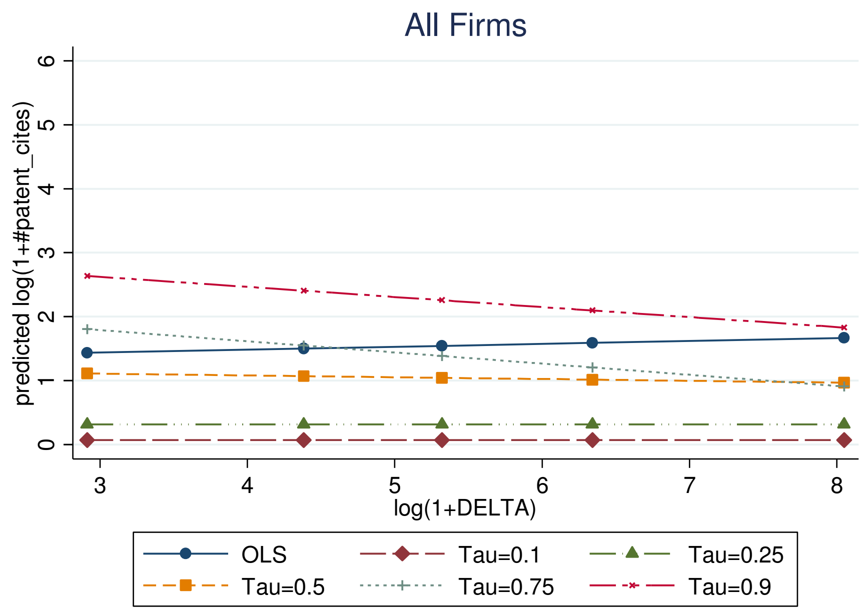


FIGURE 7B: Quantile Regression– Number of Patent Citations vs DELTA



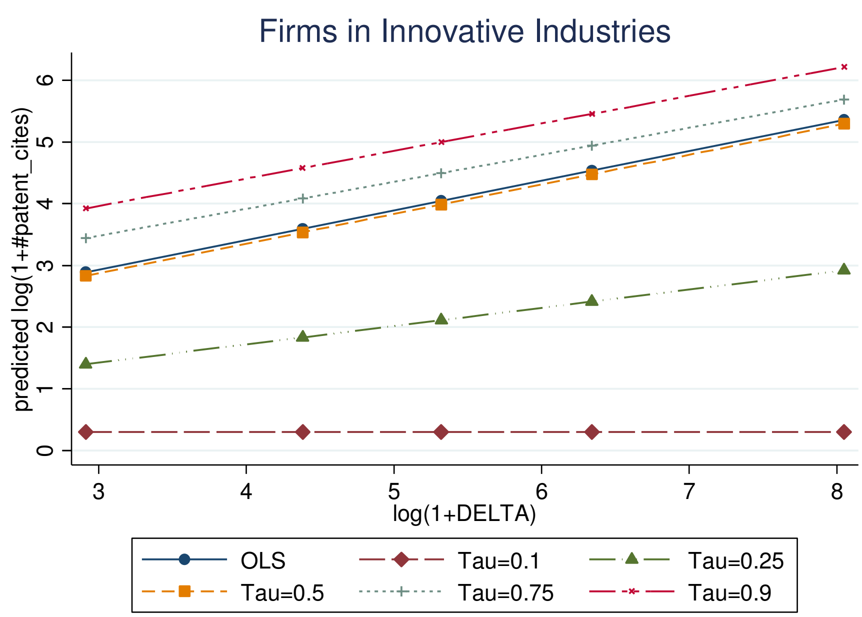


TABLE 2: Summary Statistics

This table reports summary statistics of all variables used in this study. All variables are measured in the current year except for VEGA and DELTA, which are measured in the previous year. Adjusted skewness is the cube root of the skewness as follows: $\left[\frac{\sum_i (x_i-\bar{x})^3}{(\sum_i (x_i-\bar{x})^2)^{3/2}}\right]^{1/3}$ and adjusted kurtosis is the 4th root of the kurtosis as follows: $\left[\frac{\sum_i (x_i-\bar{x})^4}{(\sum_i (x_i-\bar{x})^2)^2}\right]^{1/4}$.

Variable	Mean	SD	Min	1st Percentile	50th Percentile	99th Percentile	Max	Adj. Skewness	Adj. Kurtosis
R&D	3.635	10.806	0.000	0.000	0.222	31.062	849.659	3.459	7.440
#patent_cites	54.738	308.781	0.000	0.000	0.000	1073.781	9209.707	2.521	4.352
#patents	22.685	119.571	0.000	0.000	0.000	422.000	3396.000	2.460	4.206
$VEGA_{t-1}$	95.633	248.409	0.000	0.000	32.336	1021.409	10840.440	2.337	4.325
$DELTA_{t-1}$	1211.224	11612.710	0.000	2.917	203.394	14252.970	709829.700	3.406	6.657
CASH	1244.607	1364.179	0.000	51.039	895.542	6260.769	43511.530	1.961	3.402
SALE	4385.605	13081.060	0.047	19.293	1096.270	51760.000	286103.000	2.133	3.392
M/B	2.117	1.853	0.393	0.720	1.612	8.841	78.562	2.191	4.148
SURCASH	0.075	0.112	-2.573	-0.217	0.069	0.373	0.944	-1.304	2.595
SALEGRW	0.090	0.289	-6.092	-0.696	0.081	0.907	4.111	-1.465	3.013
RET	0.157	0.682	-0.991	-0.822	0.071	2.470	19.719	1.889	3.369
LEVERAGE	0.231	0.203	0.000	0.000	0.217	0.806	6.605	1.611	3.103
CEOTenure	8.031	7.496	0.082	0.501	5.999	36.997	54.995	1.268	1.698

TABLE 3A: Mixture-distribution Model — First-stage Logistic Regression

This table reports the first-stage results of applying logistic regression to classify industries into innovative or non-innovative industries. The dependent variable is a binary variable which takes the value of zero if a firm's research and development expenditure is always zero throughout the sample period, and one otherwise. The control variables in the first stage logit regression are constant and 47 industry dummies with the baseline industry being agriculture, which reflects Fama-French's 48 industries. The numbers in parentheses are robust standard errors, clustered at the industry level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels. The estimated innovation intensity is the probability of an industry being classified as innovative, derived from the estimated coefficient. An industry is classified as innovative if the estimated innovation intensity is greater than 0.44 which is the optimal threshold as determined by grid search algorithm using the validation sample.

	Industry Name	Estimated Coefficient	Standard Error	Estimated Innovation Intensity
Innovative Industries	Electronic Equipment	4.736***	(1.360)	0.987
	Pharmaceutical Products	4.167***	(1.363)	0.977
	Automobiles & Trucks	3.541***	(1.371)	0.958
	Machinery	3.314***	(1.089)	0.948
	Computers	3.080***	(1.050)	0.935
	Chemicals	2.944***	(1.093)	0.927
	Miscellaneous	2.197	(1.415)	0.857
	Electrical Equipment	2.110*	(1.194)	0.846
	Shipping Containers	1.792	(1.444)	0.800
	Shipbuilding & Railroad Equipment	1.504	(1.473)	0.750
	Recreation	1.322	(1.239)	0.714
	Consumer Goods	1.099	(1.011)	0.667
	Fabricated Products	1.099	(1.528)	0.667
	Rubber & Plastic Products	0.916	(1.170)	0.625
	Business Supplies	0.875	(0.998)	0.615
	Business Services	0.654	(0.927)	0.562
	Steel Works Etc	0.647	(0.998)	0.560
	Beer & Liquor	0.405	(1.355)	0.500
	Construction Materials	0.405	(0.994)	0.500
Non-innovative Industries	Food Products	0.143	(1.006)	0.435
	Textiles	0.069	(1.085)	0.417
	Candy & Soda	-0.693	(1.473)	0.250
	Non-Metallic & Industrial Metal Mining	-0.693	(1.473)	0.250
	Petroleum & Natural Gas	-0.847	(0.970)	0.222
	Printing & Publishing	-1.135	(1.113)	0.176
	Construction	-1.135	(1.113)	0.176
	Entertainment	-1.44	(1.105)	0.136
	Communication	-1.451	(1.032)	0.135
	Apparel	-1.674	(1.182)	0.111
	Wholesale	-1.819*	(1.054)	0.098
	Personal Services	-2.079	(1.385)	0.077
	Healthcare	-2.269*	(1.170)	0.065
	Restaurants, Hotels, Motels	-2.813**	(1.369)	0.038
	Retail	-3.390***	(1.160)	0.022
	Intercept	-0.405	(0.913)	
	N	1035		
	Pseudo R^2	0.440		

TABLE 3B: Classification Accuracy of the First-Stage Logistic Regression

This table reports the classification accuracy of the first-stage logistic regression based on the actual and predicted firm type (innovative or non-innovative) in the testing sample. The classification accuracy is the ratio of the sum of the number of firms correctly classified as innovative plus the number of firms correctly classified as non-innovative to the total number of sample firms; the classification sensitivity is the ratio of the number of firms correctly classified as innovative to the total number of actual innovative firms; and the classification precision is the ratio of the number of firms correctly classified as innovative to the total number of predicted innovative firms.

		Actual		
		Innovative	Non-innovative	Total
Predicted	Innovative	171	46	217
	Non-innovative	21	118	139
	Total	192	164	356

Classification Accuracy: 0.812 = (171 +118)/ 356
Classification Sensitivity: 0.891 = 171/ 192
Classification Precision: 0.788 = 171 / 217

TABLE 4: Mixture-distribution Model — Second-stage OLS Regression

This table presents the least squares estimates of vega and delta in the mixture-distribution model and in the single-equation model. The single-equation model uses all observations in the estimation sample and the mixture-distribution model only uses those firms in innovative industries. The dependent variable is the natural logarithm of one plus each proxy of corporate innovation as follows: log(1+R&D), log(1+#patents) and log(1+#patent_cites). The explanatory variables include a set of control variables, year fixed effects, and industry fixed effects classified based on the Fama-French 48 industry classification. Definitions of each variable are available in Appendix 1. The mean absolute residual (MAR) is the simple average of the prediction error. The relative goodness of fit statistics (Rel. GOF) is the proportion of observations in the testing sample in which the absolute prediction error in the mixture-distribution model is smaller than that in the single-equation model. The numbers in brackets are elasticity evaluated at the means of the corresponding variables using the full sample comprising of 12,379 observations. The numbers in parentheses are robust standard errors, clustered at industry level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels.

Dependent	R&D	R&D	#patents	#patents	#patent_cites	#patent_cites
	(1) Single-equation	(2) Mixture-distribution	(3) Single-Equation	(4) Mixture-distribution	(5) Single-equation	(6) Mixture-distribution
VEGA#	0.192*	0.309**	0.307	0.621***	0.423	0.834***
	(0.103)	(0.139)	(0.229)	(0.182)	(0.294)	(0.262)
	[0.0234]	[0.0377]	[0.0307]	[0.0620]	[0.0412]	[0.0812]
DELTA#	0.00193***	0.00211**	0.00646***	0.00705***	0.00875***	0.00915***
	(0.000645)	(0.000836)	(0.00236)	(0.00159)	(0.00250)	(0.00179)
	[0.00298]	[0.00326]	[0.00817]	[0.00892]	[0.0108]	[0.0113]
CASH#	0.0000207	0.0000516	0.0000279	0.000103	0.0000264	0.000102
	(0.0000136)	(0.0000398)	(0.0000292)	(0.0000663)	(0.0000337)	(0.0000746)
log(SALE)	-0.103**	-0.168***	0.397***	0.492***	0.432***	0.523***
	(0.0451)	(0.0596)	(0.0758)	(0.115)	(0.0862)	(0.131)
M/B	0.0571***	0.0535***	0.0508**	0.0536**	0.0871***	0.0893***
	(0.0163)	(0.0184)	(0.0189)	(0.0214)	(0.0202)	(0.0238)
SURCASH	1.197***	1.669***	0.371**	0.259	0.472**	0.391
	(0.359)	(0.350)	(0.173)	(0.248)	(0.211)	(0.307)
SALEGRW	-0.122***	-0.170***	-0.376***	-0.403***	-0.395***	-0.407***
	(0.0413)	(0.0545)	(0.0673)	(0.116)	(0.0761)	(0.120)
RET	-0.0683***	-0.0897***	-0.0428*	-0.0381	-0.0657**	-0.0567
	(0.0146)	(0.0214)	(0.0224)	(0.0392)	(0.0256)	(0.0455)
LEVERAGE	-0.0537	-0.0987	-0.232	-0.238	-0.314*	-0.364
	(0.199)	(0.301)	(0.155)	(0.252)	(0.179)	(0.273)
CEOTenure	-0.00664**	-0.0126***	-0.00925*	-0.0179**	-0.0108*	-0.0209**
	(0.00267)	(0.00335)	(0.00475)	(0.00814)	(0.00579)	(0.00996)
Intercept	1.529***	2.468***	-1.295**	-1.374*	-1.196*	-1.119
	(0.317)	(0.389)	(0.536)	(0.789)	(0.602)	(0.904)
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
N	7,370	4,380	7,370	4,380	7,370	4,380
MAR	0.460	0.433	0.909	0.791	1.119	0.997
Rel. GOF	N/A	0.609	N/A	0.646	N/A	0.634
Adj. R^2	0.595	0.554	0.391	0.366	0.380	0.342

All coefficient estimates on VEGA, DELTA, and CASH should be deflated by 1,000.

TABLE 5: The Effect of Influential Firms on Estimates of Managerial Incentives — Second-stage OLS Regression

This table presents the effect of influential observations on the estimated coefficients of managerial incentives on each proxy of corporate innovation. The dependent variable is the natural logarithm of one plus each proxy of corporate innovation as follows: log(1+R&D), log(1+#patents) and log(1+#patent_cites). Columns (1), (3) and (5) report the results using firms in innovative industries in the estimation sample whereas columns (2), (4) and (6) those after excluding one firm which has the largest combined influence on the estimated coefficients of DELTA and VEGA from this sample. The most influential firm is Microsoft in all the regressions. The relative goodness of fit statistic compares how well the baseline model fits the observations using the innovative-firms sample after we exclude the most influential firm, compared with that using the innovative-firm sample. The numbers in brackets are elasticity evaluated at the means of the corresponding variables using the full sample comprising of 12,379 observations. The numbers in parentheses are robust standard errors, clustered at the industry level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels.

Dependent	R&D	R&D	#patents	#patents	#patent_cites	#patent_cites
	(1) Firms in innovative industries	(2) Drop Microsoft	(3) Firms in innovative industries	(4) Drop Microsoft	(5) Firms in innovative industries	(6) Drop Microsoft
VEGA#	0.309**	0.348**	0.621***	0.709***	0.834***	0.921***
	(0.139)	(0.145)	(0.182)	(0.240)	(0.262)	(0.322)
	[0.0377]	[0.0424]	[0.0620]	[0.0708]	[0.0812]	[0.0897]
DELTA#	0.00211**	-0.0102	0.00705***	-0.0195	0.00915***	-0.0164
	(0.000836)	(0.00873)	(0.00159)	(0.0180)	(0.00179)	(0.0209)
	[0.00326]	[-0.0158]	[0.00892]	[-0.0247]	[0.0113]	[-0.0202]
CASH#	0.0000516	0.0000519	0.000103	0.000104	0.000102	0.000104
	(0.0000398)	(0.0000414)	(0.0000663)	(0.0000685)	(0.0000746)	(0.0000774)
log(SALE)	-0.168***	-0.168**	0.492***	0.491***	0.523***	0.521***
	(0.0596)	(0.0630)	(0.115)	(0.120)	(0.131)	(0.137)
M/B	0.0535***	0.0554***	0.0536**	0.0574**	0.0893***	0.0925***
	(0.0184)	(0.0197)	(0.0214)	(0.0233)	(0.0238)	(0.0255)
SURCASH	1.669***	1.652***	0.259	0.221	0.391	0.348
	(0.350)	(0.348)	(0.248)	(0.245)	(0.307)	(0.303)
SALEGRW	-0.170***	-0.164***	-0.403***	-0.390***	-0.407***	-0.393***
	(0.0545)	(0.0550)	(0.116)	(0.119)	(0.120)	(0.122)
RET	-0.0897***	-0.0936***	-0.0381	-0.0461	-0.0567	-0.0637
	(0.0214)	(0.0220)	(0.0392)	(0.0430)	(0.0455)	(0.0503)
LEVERAGE	-0.0987	-0.103	-0.238	-0.245	-0.364	-0.368
	(0.301)	(0.297)	(0.252)	(0.248)	(0.273)	(0.267)
CEOTenure	-0.0126***	-0.0121***	-0.0179**	-0.0170*	-0.0209**	-0.0200*
	(0.00335)	(0.00343)	(0.00814)	(0.00849)	(0.00996)	(0.0104)
Intercept	2.468***	2.467***	-1.374*	-1.370	-1.119	-1.103
	(0.389)	(0.418)	(0.789)	(0.831)	(0.904)	(0.952)
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
N	4,380	4,371	4,380	4,371	4,380	4,371
Rel. GOF	N/A	0.514	N/A	0.510	N/A	0.511
Adj. R^2	0.458	0.460	0.435	0.433	0.407	0.404

All coefficient estimates on VEGA, DELTA, and CASH should be deflated by 1,000.

TABLE 6: The Effect on Estimates of Managerial Incentives to Different Outlier Remedies — Second-stage OLS Regression

This Table reports regression estimates using untreated data under different remedies for outliers: log transformation of one plus each equity incentive proxy in column (2); winsorization of only four variables—delta, vega, cash compensation, and market-to-book ratio— at the first and 99th percentiles ("partially-winsorized") in column (3); winsorization of all variables at the first and 99th percentiles ("fully-winsorized") in column (4); and median regression which is a robust estimation method in column (5). The dependent variable is the natural logarithm of one plus each proxy of corporate innovation as follows: log(1+R&D), log(1+#patents) and log(1+#patent_cites). Our explanatory variables include $VEGA_{t-1}$, $DELTA_{t-1}$, control variables and industry fixed effects classified based on the Fama-French 48 industry classification. Definitions of each variable are available in Appendix 1. The relative goodness of fit statistic compares how well the baseline model fits the observations in the innovative-firms sample after an outlier remedy is applied, compared with the baseline model without this remedy. The numbers in brackets are elasticity evaluated at the means of the corresponding variables using the full sample comprising of 12,379 observations. The numbers in parentheses are robust standard errors, clustered at the industry level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels.

	(1)	(2)	(3)	(4)	(5)
Independent variable	Untreated	Log transform on VEGA and DELTA	Winsorize dependent variable at 1%	Winsorize all variables at 1%	Median regression
PANEL A: log(1+R&D) (N=4,380, with control variables, industry F.E. and year F.E.)					
VEGA#	0.309**	98.4***	0.305**	0.813***	0.356
	(0.139)	(16.1)	(0.138)	(0.220)	(0.396)
	[0.0377]	[0.124]	[0.0372]	[0.0991]	[0.0434]
DELTA#	0.00211**	15.6	0.00202**	-0.0104	0.00187***
	(0.000836)	(38.1)	(0.000815)	(0.0270)	(0.000534)
	[0.00326]	[0.0199]	[0.00312]	[-0.0161]	[0.00289]
Rel. GOF	N/A	0.544	0.546	0.552	0.550
Adj. R^2	0.458	0.470	0.457	0.404	0.422
PANEL B: log(1+#patents) (N=4,380, with control variables, industry F.E. and year F.E.)					
VEGA#	0.621***	131*	0.600***	1.50***	0.528
	(0.182)	(72.7)	(0.166)	(0.408)	(0.330)
	[0.0620]	[0.135]	[0.0599]	[0.150]	[0.0527]
DELTA#	0.00705***	52.1	0.00710***	-0.0361	0.00605***
	(0.00159)	(61.3)	(0.00154)	(0.0634)	(0.00163)
	[0.00892]	[0.0544]	[0.00898]	[-0.0457]	[0.00765]
Rel. GOF	N/A	0.528	0.481	0.541	0.532
Adj. R2	0.435	0.439	0.435	0.281	0.416
PANEL B: log(1+#patent_cites) (N=4,380, with control variables, industry F.E. and year F.E.)					
VEGA#	0.834***	172**	0.817***	2.01***	0.766
	(0.262)	(81.1)	(0.244)	(0. 498)	(0.758)
	[0.0812]	[0.173]	[0.0796]	[0.196]	[0.0746]
DELTA#	0.00915***	90.3	0.00909***	-0. 0245	0.00771***
	(0.00179)	(77.0)	(0.00175)	(0. 0810)	(0.00217)
	[0.0113]	[0.0919]	[0.0112]	[-0.0302]	[0.00951]
Rel. GOF	N/A	0.545	0.476	0.545	0.548
Adj. R^2	0.407	0.413	0.407	0.329	0.377

All coefficient estimates on VEGA and DELTA should be deflated by 1,000.

TABLE 7: Second-stage Quantile Regression — Firms in Innovative Industries

This Table reports the least square estimates in column (1) and estimates of nine quantile regressions (τ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9) in columns (2)‒(10) using only firms in the innovative industries of the estimation sample. The dependent variable is the natural logarithm of one plus each proxy of corporate innovation, log(1+R&D) in Panel A, log(1+#patents) in Panel B, and log(1+#patent cites) in in Panel C. The vega and delta variables are transformed by taking logarithm of one plus delta and vega, log(1+vega) and log(1+delta). The regressions include firm-level control variables as well as industry fixed effects classified based on the Fama-French 48 industry classification. Definition of each variable is available in Appendix 1. The numbers in brackets are elasticity evaluated at the means which are computed using the full sample of 12,379 observations. The numbers in parentheses are robust standard errors, clustered at the industry level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels.

	(1) OLS	First (2) τ=0.1	(3) τ=0.2	(4) τ=0.3	(5) τ=0.4	(6) τ=0.5	(7) τ=0.6	(8) τ=0.7	(9) τ=0.8	(10) τ=0.9
PANEL A: log(1+R&D) (N=4,380, with control variables, industry F.E. and year F.E.)										
log(1+VEGA)	0.0984***	0.0300	0.0693***	0.0720***	0.0881***	0.100***	0.0924***	0.0751***	0.0663***	0.0721***
	(0.0161)	(0.0194)	(0.0254)	(0.0159)	(0.0222)	(0.0249)	(0.0282)	(0.0228)	(0.0209)	(0.0153)
	[0.124]	[0.0379]	[0.0874]	[0.0909]	[0.111]	[0.126]	[0.117]	[0.0948]	[0.0837]	[0.0910]
log(1+DELTA)	0.0156	-0.00418	0.00619	-0.00267	-0.00738	-0.0152	-0.0141	-0.00123	-0.0119	-0.0155
	(0.0381)	(0.0111)	(0.0157)	(0.0211)	(0.0290)	(0.0317)	(0.0394)	(0.0360)	(0.0249)	(0.0249)
	[0.0199]	[-0.00533]	[0.00789]	[-0.00340]	[-0.00940]	[-0.0194]	[-0.0180]	[-0.00157]	[-0.0152]	[-0.0197]
Adj. R^2	0.470									
PANEL B: log(1+#patents) (N=4,380, with control variables, industry F.E. and year F.E.)										
log(1+VEGA)	0.131*	0.000	0.000	0.121	0.150**	0.138**	0.154**	0.165***	0.159**	0.0968
	(0.0727)	(0.00931)	(0.0201)	(0.0830)	(0.0598)	(0.0593)	(0.0638)	(0.0634)	(0.0729)	(0.121)
	[0.135]	[0.000]	[0.000]	[0.125]	[0.155]	[0.143]	[0.159]	[0.170]	[0.164]	[0.100]
log(1+DELTA)	0.0521	0.000	0.000	0.0485	0.0530	0.0644	0.0379	0.0425	0.0163	0.0610
	(0.0613)	(0.00909)	(0.0206)	(0.0647)	(0.0753)	(0.0599)	(0.0605)	(0.0675)	(0.0471)	(0.0718)
	[0.0544]	[0.000]	[0.000]	[0.0506]	[0.0553]	[0.0672]	[0.0395]	[0.0443]	[0.0170]	[0.0636]
Adj. R2	0.439									
PANEL C: log(1+#patent_cites) (N=4,380, with control variables, industry F.E. and year F.E.)										
log(1+VEGA)	0.172**	0.000	0.000	0.171	0.201**	0.197***	0.202***	0.221***	0.221***	0.152
	(0.0811)	(0.0154)	(0.0274)	(0.110)	(0.0836)	(0.0691)	(0.0667)	(0.0665)	(0.0709)	(0.138)
	[0.173]	[0.000]	[0.000]	[0.172]	[0.203]	[0.199]	[0.204]	[0.223]	[0.223]	[0.153]
log(1+DELTA)	0.0903	0.000	0.000	0.0748	0.0945	0.0903	0.0969	0.0809	0.0706	0.0840
	(0.0770)	(0.0150)	(0.0269)	(0.0600)	(0.0882)	(0.0767)	(0.0697)	(0.0701)	(0.0708)	(0.0983)
	[0.0919]	[0.000]	[0.000]	[0.0761]	[0.0961]	[0.0919]	[0.0986]	[0.0823]	[0.0718]	[0.0855]
Adj. R2	0.400									

TABLE 8: Quantile Regression — All Firms

This Table reports the least square estimates in column (1) and estimates of nine quantile regressions (τ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9) in columns (2)–(10) using all firms in the estimation sample. The dependent variable is the natural logarithm of one plus each proxy of corporate innovation as follows: log(1+R&D) in Panel A, log(1+#patents) in Panel B, and log(1+#patent_cites) in in Panel C. The vega and delta variables are transformed by taking logarithm of one plus delta and vega as follows: log(1+vega) and log(1+delta). The regressions include all firm-level control variables and industry fixed effects classified based on the Fama-French 48 industry classification. Definitions of each variable are available in Appendix 1. The numbers in brackets are elasticity evaluated at the means of the corresponding variables using the full sample comprising of 12,379 observations. The numbers in parentheses are robust standard errors, clustered at the industry level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels.

Independent Variable	(1) OLS	(2) τ=0.1	(3) τ=0.2	(4) τ=0.3	(5) τ=0.4	(6) τ=0.5	(7) τ=0.6	(8) τ=0.7	(9) τ=0.8	(10) τ=0.9
PANEL A: log(1+R&D) (N=7,370, with control variables, industry F.E. and year F.E.)										
log(1+VEGA)	0.0665***	0.000	0.000	0.000	0.000	0.00383	0.00626	0.00914	0.0168	0.0380**
	(0.0194)	(0.000371)	(0.000646)	(0.000827)	(0.00148)	(0.00516)	(0.00710)	(0.00890)	(0.0152)	(0.0193)
	[0.0839]	[0.000]	[0.000]	[0.000]	[0.000]	[0.00483]	[0.00790]	[0.0115]	[0.0212]	[0.0480]
log(1+DELTA)	0.00209	0.000	0.000	0.000	0.000	-0.00203	-0.00253	-0.00312	-0.00605	-0.0108
	(0.0209)	(0.000242)	(0.000397)	(0.000502)	(0.000793)	(0.00213)	(0.00248)	(0.00263)	(0.00469)	(0.00821)
	[0.00266]	[0.000]	[0.000]	[0.000]	[0.000]	[-0.00259]	[-0.00322]	[-0.00398]	[-0.00771]	[-0.0138]
Adj. R^2	0.625									
PANEL B: log(1+#patents) (N=7,370, with control variables, industry F.E. and year F.E.)										
log(1+VEGA)	0.0817*	0.000	0.000	0.000	0.000	0.0157	0.0472	0.0682	0.0745	0.0756
	(0.0428)	(0.000)	(0.000)	(0.000)	(0.00250)	(0.0282)	(0.0353)	(0.0430)	(0.0495)	(0.0568)
	[0.0844]	[0.000]	[0.000]	[0.000]	[0.000]	[0.0162]	[0.0488]	[0.0705]	[0.0770]	[0.0781]
log(1+DELTA)	-0.00790	0.000	0.000	0.000	0.000	-0.00668	-0.0273	-0.0422*	-0.0349	-0.0215
	(0.0401)	(0.000)	(0.000)	(0.000)	(0.00233)	(0.0122)	(0.0238)	(0.0253)	(0.0307)	(0.0372)
	[-0.00824]	[0.000]	[0.000]	[0.000]	[0.000]	[-0.00697]	[-0.0285]	[-0.0440]	[-0.0364]	[-0.0224]
Adj. R^2	0.474									
PANEL C: log(1+#patent_cites) (N=7,370, with control variables, industry F.E. and year F.E.)										
log(1+VEGA)	0.107**	0.000	0.000	0.000	0.000	0.0167	0.0621	0.100*	0.104**	0.0938*
	(0.0502)	(0.000)	(0.000)	(0.000)	(0.00299)	(0.0263)	(0.0466)	(0.0540)	(0.0529)	(0.0538)
	[0.108]	[0.000]	[0.000]	[0.000]	[0.000]	[0.0168]	[0.0626]	[0.101]	[0.105]	[0.0945]
log(1+DELTA)	0.00847	0.000	0.000	0.000	0.000	-0.00526	-0.0265	-0.0347	-0.0364	-0.0296
	(0.0515)	(0.000)	(0.000)	(0.000)	(0.00257)	(0.0108)	(0.0253)	(0.0315)	(0.0370)	(0.0426)
	[0.00862]	[0.000]	[0.000]	[0.000]	[0.000]	[-0.00535]	[-0.0270]	[-0.0353]	[-0.0370]	[-0.0301]
Adj. R^2	0.466									

APPENDIX 1 – Variable Definitions

Variable	Description
R&D	Research and development expenditure scaled by book asset
#patents	Number of patent counts
#patent_cites	Number of patent citations
VEGA	Sensitivity of CEO wealth to stock return volatility, which is the change in the dollar value of the CEO's wealth for a 0.01 change in the annualized standard deviation of stock returns (When log transformation is not applied, this variable is scaled down by a factor of 1000)
DELTA	Sensitivity of CEO wealth to stock price which is the change in the dollar value of the CEO's wealth for a one percentage point change in stock price (When log transformation is not applied, this variable is scaled down by a factor of 1,000)
CASH	Cash compensation which is the sum of salary and bonus for the CEO in the current year (Scaled down by a factor of 1,000)
SALE	The net annual sales as reported by the company (in millions)
M/B	Market-to-book ratio which is the ratio of market value of assets to book value of assets
SURCASH	Surplus cash scaled by book asset
SALEGRW	Sales growth which is the logarithm of the ratio of sales in the current year to the sales in the previous year
RET	One year total return to shareholders (in percentage)
LEVERAGE	Book leverage which is the ratio of total book value of debt to book value of total assets
CEOTenure	The length of time (in year) since the executive takes the CEO position in the firm

References

Adams, John C., Darren K. Hayunga, and Vincenzo Verardi, 2015, Outliers in Finance Research, Working Paper.

Alchian, Armen A., and Harold Demsetz, 1972, Production, Information Costs, and Economic Organization, *American Economic Review* 62(5), 777–95.

Aggarwal, R.K., Samwick, A.A., 1999, The other side of the trade-off: the impact of risk on executive compensation, *Journal of Political Economy* 107, 65–105.

Chava, S., Purnanandam, A., 2010, CEOs versus CFOs: incentives and corporate policies, *Journal of Financial Economics* 97, 263–278.

Baker, G. and B. Hall, 2004. CEO incentives and firm size, *Journal of Labor Economics* 22, 767–798.

Bebchuk, Lucian, Jesse Fried, and David Walker, 2002, Managerial power and rent extraction in the design of executive compensation. *University of Chicago Law Review* 69, 751−846.

Bebchuk, Lucian A., Yaniv Grinstein, and Urs Peyer, 2010, Lucky CEOs and Lucky Directors, *Journal of Finance* 65: 2363–2401

Coles, J.L., Daniel, N.D., Naveen, L., 2006, Managerial incentives and risk- taking, *Journal of Financial Economics* 79, 431–468.

Conyon, M. and J. Schwalbach, 2000. Executive compensation: Evidence from the UK and Germany, *Long Range Planning* 33, 504–526.

Core, J., Guay, W., 2002, Estimating the value of employee stock option portfolios and their sensitivities to price and volatility, *Journal of Accounting Research* 40, 613–630.

Dong, Ming, D. Hirshleifer, and S. H. Teoh. 2016. Stock Market Overvaluation, Moon Shots, and Corporate Innovation. Working Paper.

Ederer, Florian, and G. Manso, 2013, Is Pay for Performance Detrimental to Innovation, *Management Science* 59, 1496–1513.

Fama, Eugene F., and Kenneth R. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153–193.

Feng, Changyong, H. Wang, N. Lu, and X.M. Tu, 2012, Log Transformation: Application and Interpretation in Biomedical Research, *Statistics in Medicine* 32, 230–239.

Garvey, G.T., Milbourn, T.T., 2006, Asymmetric benchmarking in compensation: executives are rewarded for good luck but not penalized for bad, *Journal of Financial Economics* 82, 197–225.

Gneezy, U., and A. Rustichini, 2000, Pay Enough or Don't Pay at All, *Quarterly Journal of Economics* 115, 791–810.

Guthrie, Katherine, Jan Sokolowsky, and Kam-Ming Wan, 2012, CEO Compensation and Board Structure Revisited, *Journal of Finance* 67(3), 1149–1168.

Hall, B. H., A. Jaffe, and M. Trajtenberg, 2001, The NBER patent citations data file: Lessons, insights and methodological tools, *NBER Working Paper No.8498, University of California at Berkeley*.

Hall, B. H., A. Jaffe, and M. Trajtenberg, 2005, Market value and patent citations, *RAND Journal of Economics* 36, 16–38.

Hall, Brian. J., and Kevin J. Murphy, 2003, The Trouble with Stock Options, *The Journal of Economic Perspectives* 17, 49–70.

Hallock, K., Madaloozo, R. and Reck, C., 2010, CEO pay-for-performance heterogeneity using quantile regression, *The Financial Review* 45, 1–19.

Hamm, Sophia J.W., Jung, M. and Wang, C., 2015, Making Sense of One Dollar CEO Salaries, *Contemporary Accounting Research* 32, 941–972.

Hayes, R.M., Lemmon, M., Qiu, M., 2012, Stock options and managerial incentives for risk-taking: Evidence from FAS 123R, *Journal of Financial Economics* 105, 174–190.

Hirshleifer, D., A. Low, and S. H. Teoh, 2012, Are overconfident CEOs better innovators? *Journal of Finance* 67, 1457–98.

Hirshleifer, D., and Y. Suh, 1992, Risk, Managerial Effort, and Project Choice, *Journal of Financial Intermediation* 2, 308–345.

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

Jensen, Michael C., and Kevin Murphy, 1990, Performance Pay and Top-Management Incentives, *Journal of Political Economy* 98, 225–264.

Kini, Omesh, and Ryan Williams, 2012, Tournament incentives, firm risk, and corporate policies, *Journal of Financial Economics* 103, 350–376

Koenker, R. and G. Bassett, 1978, Regression quantiles, *Econometrica* 46, 33–50.

Koenker, R. and K.F. Hallock, 2001, Quantile regression, *Journal of Economic Perspectives* 15, 143–156.

Kogan, Leonid, D. Papanikolaou, A. Seru, and N. Stoffman, 2016, Technological Innovation, Resource Allocation and Growth, *Quarterly Journal of Economics*.

Koh, P.-S., and D.M. Reeb, 2015, Missing R&D, *Journal of Accounting and Economics* 60, 73–94.

Kortum, S., 1993, Equilibrium R&D and the Patent-R&D Ratio: U.S. Evidence, *American Economic Review* 83, 450–457.

Kothari, S.P., Sabino, J., Zach, T., 2005, Implications of Survival and Data Trimming for Tests of Market Efficiency, *Journal of Accounting and Economics* 39, 129–161.

Lie, Erik, 2005, On the timing of CEO stock option awards, *Management Science* 51, 802–812.

Lie, Erik, and Randall Heron, 2007, Does backdating explain the stock price pattern around executive stock option grants?, *Journal of Financial Economics* 83, 271–295.

Lien, D. and Balakrishnan, N.,2005, On regression analysis with data cleaning via trimming,winsorization, and dichotomization, *Communications in Statistics – Simulation and Computation*34, 839–849.

Loureiro, Gilberto R., A.K., Makhija, and D. Zhang, 2014, The Ruse of a One-Dollar CEO Salary, Charles A. Dice Working Paper No. 2011-7.

Low, A., 2009, Managerial risk-taking behavior and equity-based compensation, *Journal of Financial Economics* 92, 470–490.

Manso G., 2011, Motivating innovation, *Journal of Finance* 66, 1823–1860.

McVay, S., 2006, Earnings Management Using Classification Shifting: An Examination of Core Earnings and Special Items, *The Accounting Review* 81, 501–531.

Ohlson, James A., and Seil Kim, 2015, Linear Valuation without OLS: the Theil-Sen Estimation Approach, *Review of Accounting Studies* 20, 395–435.

Schaefer, S., 1998. The dependence of pay-performance sensitivity on the size of the firm, *Review of Economics and Statistics* 80, 436–443.

Wan, Kam-Ming, 2014, Incentive Contracts are not Rigged by Powerful CEOs, *Critical Finance Review* 3(1), 99−152.

Yermack, David, 1997, Good Timing: CEO Stock Option Awards and Company News Announcements, *Journal of Finance* 52, 449–476.