Treatment of Outliers: Managerial Incentives and Corporate Innovation*

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

Abstract

Ordinary least squares estimates are predominately used to quantify the effect of managerial incentives on corporate innovation. However, they are inadequate for two reasons. First, data on corporate innovation is discontinuous with a spike of zero because many firms allocate no resources to such activities. Second, they are sensitive to outliers. To address the discontinuity concern, we use a mixture distribution model to distinguish innovative firms from non-innovative firms. To mitigate the outliers' influence, we first use quantile regression and second use only innovative firms. Our quantile regression results indicate that higher sensitivity of CEO wealth to stock return volatility (vega) induces innovative firms to increase research and development investments and intensity of innovation in patent counts and citations. In contrast, higher CEO pay-for-performance sensitivity (delta) has no material effect on corporate innovation. Between the two well-known managerial incentives for corporate innovation, the vega incentive survives our scrutiny while the delta incentive does not.

[Do not quote or cite without authors' permission]

Current Version: May 6, 2017

Keywords: outlier, vega, delta, winsorization, mixture distribution model, quantile regression

JEL Classifications: G34, G38, J31, J33

---

* We are grateful to Pierre Azoulay, Po-Hsuan Hsu, Jim Ohlson and seminar participants at 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

Ordinary least squares (OLS) regressions are predominantly used in finance. Nevertheless, least squares estimator is notoriously sensitive to outliers. To mitigate the influence of potential outliers, extant studies usually perform data treatment before applying the least squares regression. The leading data treatment is data winsorization for which extreme data values of a variable (or unconditional outliers) are arbitrarily set at a specified percentile. However, this data treatment presents conceptual challenges because outliers are better treated as a conditional concept in light of a reasonable choice of a model and covariates. Extreme data values are not outliers and should not be treated as such if they can be properly explained by a reasonable model and covariates. This implies that data winsorization may distort informative data. Furthermore, the extent of data treatment is subjective and susceptible to confirmation bias. In practice, some winsorize just one variable whereas others all variables. 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 relation is chosen for three reasons. First, corporate innovation is a major decision in many firms and vital to the long-term economic growth of a country (Kogan et al. 2016). Second, nearly each key variable of interest—corporate innovation and managerial incentives—has noticeable extreme values (Hall et al., 2001 & 2005; Coles et al., 2006; and Hirshleifer et al., 2012; Kini and Williams, 2012; Dong et al., 2016). This suggests that the outliers' influence is potentially serious. 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 CEOs rely exclusively on stock incentives but many others 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.

Last, the structure of managerial incentives, in the form of stock- and option-based compensation, has vital implications on corporate innovation (Coles et al, 2006; Hirshleifer and Suh, 1992). However, managerial incentives provided by stock options have been heavily criticized since the drastic growth in CEO compensation in the 1990s and the unfolding of the option backdating scandal (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 stocks (or standard pay-for-performance compensation contracts) is suboptimal for corporate innovation (Ederer and Manso, 2013; Manso, 2011). Thus, these reasons compel us to examine the outliers' influence on this empirical relation.

To identify conventional methods used to treat outliers in the literature, we first searched the JSTOR database for articles published in eight leading journals in Accounting and Finance for keywords "delta" and "vega" from 1989 to 2016. 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 comprehensive 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 winsorization to treat outliers while the remaining do not mention any treatments for outliers. Of the eighteen studies that apply data winsorization, thirteen (72%) winsorize nearly all variables while the rest apply winsorization selectively. Data are predominately winsorized at the 1% level. In sixteen (89%) of these studies, data are winsorized at the first and 99th percentiles.

Another remedy for outliers is data transformation. Logarithmic transformation is often applied to highly right-skewed variables. We are surprised that even though data on managerial incentives are highly skewed, seventeen (65%) studies use raw (unscaled) dollar incentives to measure delta and vega.[1] Nine studies (35%) perform data transformation on equity incentives: six by taking logarithm of one plus delta (vega) while three by taking logarithm of delta (vega). In the latter case, data truncation occurs because observations with a zero value are excluded from the sample.

Occasionally, some use robust estimators. For example, the least absolute deviations method (i.e., median regression) is moderately used for estimation in the executive compensation literature (Guthrie et al., 2012; Garvey and Milbourn, 2006; and Aggrawal and Samwick, 1999). Other robust estimation methods include quantile regression, MM-estimator, and Theil-Sen regression (e.g., Hallock et al., 2010; Adams et al., 2015; and Ohlson and Kim, 2015). Even if robust estimators are used, they are mostly used as supplementary results to demonstrate the

[1] In some cases, dollar incentives are scaled by cash compensation or total compensation in the compensation year. However, this transformation is ineffective for two reasons. First, the scaled dollar incentives remain highly skewed. Second, this approach creates extreme values (or unconditional outliers) for CEOs who receive extremely low compensation, e.g., one dollar in salary (Hamm et al., 2015; Loureiro et al., 2014; and Guthrie et al. 2012).

robustness of the least squares estimates. We find that only one (4%) out of a total of twenty-six studies uses median regression in the robustness test.[2]

To properly estimate the effect of managerial incentives on corporate innovation, we use a mixture distribution model because nearly one-half of our observations allocate no resources on corporate innovation. Our research methodology involves a two-stage regression procedure. In the first-stage, we use logistic regression to separate firms into either innovative or non-innovative. In the second-stage, quantile regression is used to estimate the effects of managerial incentives on corporate innovation for innovative firms only. We exclude non-innovative firms in the second-stage analysis because it is pointless to estimate the aforementioned relation for firms that do not (or rarely) engage in corporate innovation. Quantile regression is used to mitigate the outliers' concern because our key variables of interest have noticeable extreme values. Second, quantile regression allows us to examine the heterogeneity in the relationship between managerial incentives and corporate innovation. This consideration is particularly relevant because some firms rarely engage in corporate innovation while others actively.

Our results indicate 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 innovative firms can meaningfully change the estimated coefficients of delta. We also find that least squares estimates of delta are fragile and vary appreciably to different outlier remedies. In contrast, our quantile regression estimates on vega and delta are robust and vary only slightly across different quantiles of the conditional distribution. Congruent with the literature, our results indicate that higher sensitivity of CEO wealth to stock return volatility (vega) induces innovative firms to increase corporate innovation including R&D investments, number of patent counts, and

[2] In unreported results, we find that robust estimators are significantly less likely to be used in the corporate innovation literature than in the executive compensation literature. For example, approximately 37 percent of empirical studies use median regression to model the level of executive compensation in the robustness tests.

number of patent citations. In contrast, higher CEO pay-for-performance sensitivity (delta) has no material effect on corporate innovation.

To the best of our knowledge, this paper is the first to address the outlier's influence in the corporate innovation literature. As the nature of outliers varies widely by model specifications, we first use the mixture distribution model to formulate the relation of managerial incentives on corporate innovation. Next, we propose to use quantile regressions to investigate the outliers' influence on the estimates.

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 relation between managerial incentives and corporate innovation for Chief Executive Officers, we use the baseline model in Coles et al.:[3]

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

where y_{it} represents corporate innovation of firm i in year t and X_{it} is a vector of control variables including a vector of lagged CEO equity incentives, and u_{it} is an error term. For y_{it}, we take logarithmic transformation of one plus the corporate innovation variable because corporate innovation measures are highly right-skewed with extreme values. Besides, they also have a discrete spike at zero.

1. Measures of Corporate Innovations

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). 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).[4]

The output measures are the number of a firm's patent counts (NumPatent) and the number of a firm's patent citations (NumCites). Data on the firm's patent counts and citations are obtained from the April 2013 edition of the patent database in Kogan, Papanikolaou, Seru,

[3] Coles et al. is widely cited with Google Scholar citations of over 1,400 as of May 2017.
[4] Companies can make a conscious choice of not separating R&D expenses from other reported expenses (McVay, 2006). This implies that firms can report a missing value in R&D expenditure despite they allocate real resources to such activities. Thus, our treatment to missing values in R&D expenditure underestimates the actual R&D investments. Nevertheless, this underestimation is a minor issue in our study because firms reporting a missing value in R&D expenditure usually have no innovation output. For example, Koh and Reeb (2015) find that only 10.5 percent of firms reporting a missing value in R&D expenditure receive patents. This implies that nearly 45 percent of our sample observations have no R&D investments and patent grants[= 50% × (100% −10.5%)].

and Stoffman (see Kogan et al. 2016). This database covers U.S. patent grants and patent citations from 1926 to 2010. Patents are included in the database only if they are eventually granted.

2. Measures of CEO Equity Incentives

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.

3. 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 (ln(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 serving in her current position. All control variables are measured in the current year.

The 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. Appendix 1 contains a detailed description of these variables.

IV. EMPIRICAL RESULTS

A. Descriptive Statistics

Table 2 provides summary statistics of all variables using untreated data. Our sample medians are slightly larger but qualitatively similar to those reported in the literature.[5] 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]

[5] The reported summary statistics in our study could be different from those in the literature because our variables are untreated. For example, some or all variables are winsorized in Coles et al. (2006) and Chava and Punanandam (2010).

It is obvious that all the variables used in this study have extreme values, particularly for our key variables of interest. Extreme values 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 than 25. These correspond to a probability of less than 0.0001% of such extreme values occurring assuming these variables are normally distributed. Another indication of extreme values is that the distribution of the key variables of interest has a fat tail as indicated by the large kurtosis. Our key variables of interest are highly right-skewed which suggests that they have extremely large values. This result is anticipated because extant studies show that for these variables their means are usually many times larger than their respective medians.

It is equally obvious that our corporate innovation proxies have a discrete spike of zero. This indicates that some firms never engage in corporate innovation activities. Figures 1A−1C present the histograms of R&D expenditure, the number of patent counts, and the number of patent citations, respectively. Nearly one-half of all observations allocate nothing on research and development. Similarly, more than half of them have no innovation outputs: patent counts and patent citations.

In a univariate context simple data treatment moderates the potential influence of the outliers. If vega is winsorized at the 99^{th} 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). However, the winsorized data value is still extreme and corresponds to a probability of less than 0.02% of such extreme value occurring assuming vega is normally distributed. Similarly, extreme values remain extreme but less skewed if these variables are treated by logarithmic transformation.

B. Mixture Distribution Model

Our corporate innovation measures exhibit a discrete spike of zero, which indicates that some companies never commit resources on corporate innovation. Thus, managerial incentives should play a different role for firms that are engaging in corporate innovation from those that are not.[6] Therefore, we first use a mixture distribution model to separate firms into either innovative or non-innovative. Next, we use only innovative firms to estimate the effect of managerial incentives on corporate innovation reasoning that managerial incentives are consciously structured to achieve corporate innovation initiatives only in firms that engage in corporate innovation activities.

Thus, our research methodology involves a two-stage regression procedure. In the first-stage, we use logistic regression to classify firms into either innovative or non-innovative. In the second-stage, we use a regression method to estimate the effect of managerial incentives on corporate innovation for firms classified as innovative in the first-stage regression (henceforth "innovative firms").

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: $C_i = 1$ if $C_i^* > 0$ and $C_i = 0$ if otherwise. We use R&D expenditure to capture a firm's proclivity to innovate because it captures real corporate resources committed for 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).

[6] This implies that managerial incentives are used for purposes other than innovation in firms that do not engage in corporate innovation activities, e.g., mitigating agency problem (Alchian and Demsetz, 1972; Jensen and Meckling, 1976).

Specifically, C_i is a binary variable and takes the value of zero if a firm's research and development expenditure is consistently zero during the sample period, and one otherwise.

The Z_i is a vector of explanatory variables and v_i is a random error following a standard logistic distribution. Our explanatory variables include firm size and its square term plus industry dummy variables. We measure firm size by taking logarithm of the average sales of the firm during the sample period. As the nature and intensity of corporate innovation vary considerably across industries, we include 48 industry dummies according to the Fama-French methodology (Fama and French, 1997). We estimate equation (2) using the logistic maximum likelihood estimation method.[7]

To obtain out-of-sample prediction of the mixture distribution model, we first split our sample observations randomly into three sub-samples: 60% in Estimation sample (E), 20% in Validation sample (V), and 20% in Testing sample (T).[8] The estimation sample is used in the first-stage to estimate the model specified in equation (2). Next, 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 for each firm in the validation sample (V). 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.[9] Last, in the estimation sample (E) we choose only those firms that are classified as innovative to estimate the effect of

---

[7] In some cases, the logistic regression fails to distinguish between innovative firms and non-innovative firms in some industries. The failure occurs because the R&D expenditure is identical for all the observations within the same industry, i.e. either one or zero. In such circumstances, we classify all the firms in an industry as innovative if over one-half of them report non-zero expenses on R&Ds (i.e., C_i =1), and non-innovative otherwise. Consequently, firms specializing in the tobacco, medical equipment, aerospace, and defense industries are classified as innovative. In contrast, firms specializing in the precious metal, coal, measuring and control equipment, as well as transportation industries are classified as non-innovative.

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

[9] Less desirably, we could have used a naïve classification scheme, i.e., classify a firm 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.

managerial incentives on corporate innovation using the regression model specified in equation (1).

In the testing sample (T), 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:[10]

(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.

We use a single equation model as a benchmark to evaluate the performance of the mixture distribution model. The single equation model uses the baseline model in equation (1) to estimate the effect of managerial incentives on corporate innovation for all observations in the estimation sample (E).[11] Next, we use the estimated coefficients in equations (1) from the estimation sample (E) to compute the predicted value of the intensity of corporate innovation for each observation in the testing sample (T). Last, for each observation in the testing sample (T), 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 (E). Our results indicate that firm size and industry dummies are powerful determinants of a firm's inclination to invest in research and development. Larger firms are significantly more likely to be classified to be innovative than smaller firms. This result is natural because larger firms have more financial slacks than smaller firms. We also find that industry dummies are

[10] 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.
[11] 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 classified as innovative in the first-stage regression.

important determinants of a firm's inclination for R&D investments. In unreported results, the adjusted R-squared is 0.44 if we include only industry dummies in the first-stage regression, compared with 0.47 when we also include firm size in the regression.[12]

[INSERT TABLE 3A HERE]

Table 3B reports the classification accuracy of the first-stage regression in the testing sample (T). Although our explanatory variables include only rudimentary economic variables, they are useful to classify firms into innovative and non-innovative. Out of a total of 356 firms in the testing sample, 84.3% of them are correctly classified.[13] 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 regression specified in equation (1) using the estimation sample (E). 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 model performance, our results indicate that the mixture distribution model is superior to the single-equation model. For example, 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.

---

[12] Similarly, in unreported results we find that firms in R&D intensive industries are more likely to be classified as innovative. For example, firms specializing in computer, electronic equipment, electrical equipment, automobiles and trucks, and aircrafts are more likely to be classified to be innovative than firms specializing in non-R&D intensive industries such as wholesale, retail, and beer and liquor.

[13] 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 (T) is (159+141)/356, or 0.843.

Similarly, the relative goodness of fit statistic also points to the same conclusion.[14] The relative goodness of fit statistics are consistently larger than 0.5 in the mixture distribution model.[15] 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 prediction error in the mixture distribution model is smaller in absolute value than that in the single-equation model for 55.8 percent of the testing sample (T) observations.

[INSERT TABLE 4 HERE]

In the second-stage of the mixture distribution model, we eliminate non-innovative firms, i.e., firms that do not (or rarely) engage in corporate innovation activities. If higher managerial incentives enhance corporate innovation activities, we expect that the relation between managerial incentives and corporate innovation should be stronger in the mixture distribution model than in the singe equation model. In congruent with our expectations, the magnitude of our least squares estimates on vega is significantly larger in the mixture distribution model than those in the single equation model. Besides, they are also statistically more significant than those in the single-equation model. Specifically, least squares estimates on vega are statistically significant at the conventional levels in the mixture distribution model.

Our results also show that all the least squares estimates of delta are positive and statistically significant at the 10% level or better in both models. This implies that higher CEO pay-for-performance sensitivity increases corporate innovation outcomes. However, this conclusion should not be taken seriously because our estimated coefficients of delta vary

[14] 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).

[15] 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 $\sum_{i=1}^{n} I\left[\left|\hat{y}_i^{(2)} - y_i\right| < \left|\hat{y}_i^{(1)} - y_i\right|\right]/n > 0.5$.

appreciably to different outlier remedies (to be discussed in the subsequent Sections). This implies that our estimates on delta are vulnerable to outliers' influence.[16]

C. Influential Observations

To assess the effect of influential observations on least squares estimates using untreated data, we perform two sets of regression for each corporate innovation proxy. In the first set of regression, we use only innovative firms in the estimation sample (E). In the second set, of all the innovative firms in the estimation sample (E), we exclude one firm which has the largest combined influence on the estimated coefficients of delta and vega.[17] To quantify the 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 estimated coefficients 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 is Microsoft in all the regressions.

[INSERT TABLE 5 HERE]

Our results indicate that the outliers' influence on the estimated coefficients of delta is

---

[16] Besides, our least squares estimates may produce a seemingly puzzling result. In particular, our least squares estimates of delta are significantly smaller in the mixture model than those in the single equation model. This result implies that higher CEO pay-for-performance sensitivity lowers corporate innovation for innovative firms relative to non-innovative firms. 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.

[17] 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 innovative firms in the estimation sample (E) 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 combined influence on the estimated coefficients of vega and delta. Besides, the most influential firm can be viewed as an (conditional) outlier.

pronounced. For example, in the R&D regression after excluding Microsoft from the sample of 635 innovative firms, the magnitude of the estimated coefficient of delta decreases to –0.0125 (p-value of 23.83%) from 0.00103 (p-value of 7.67%) in the first set of regression, or by 1,313%. Similarly, after excluding Microsoft from the sample, the magnitude of the estimated coefficient of delta decreases by 506% in the number of patent count regression and by 395% in the number of patent citations regression. Additionally, excluding only Microsoft from the sample renders the estimated coefficient of delta statistically insignificant at the 10% level, compared with statistical significance at the same level in all the first set of regressions.

In contrast, the outliers' influence on the estimated coefficients of vega is much smaller. For example, in the R&D regression after excluding Microsoft from the sample, the magnitude of the coefficient of vega increases slightly to 0.234 (p-value of 1.39%) from 0.218 (p-value of 2.42%) in the first set of regression, or by 7.4%. Similarly, after excluding Microsoft from the sample, the magnitude of the coefficient of vega increases by 5.2% in the number of patent count regression and by 4.8% in the number of patent citations regression. In terms of statistical significance, all the estimated coefficients on vega are statistically significant at the conventional levels, regardless whether the most influential firm is included or excluded from the sample.

Our results indicate that the positive relation between higher sensitivity of CEO wealth to stock return volatility and corporate innovation is robust to the outliers' influence. In contrast, the positive relation between higher CEO pay-for-performance sensitivity and corporate innovation is fragile to the outliers' influence. This compels us to investigate whether our results vary meaningfully to different outlier remedies in the next Section.

D. Remedies for Outliers

To investigate the influence of different outlier remedies on our results, we perform three conventional remedies for outliers including (i) log transformation of managerial incentive proxies; (ii) data winsorization; and (iii) median regression. Table 6 reports regression estimates using untreated data under different outlier remedies for innovative firms in the estimation sample (E): log transformation of one plus each proxy of managerial incentive in column (2); winsorization of dependent variable—R&D, NumPatent, and NumCites—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).[18]

[INSERT TABLE 6 HERE]

Our results show that the sign and the statistical significance of the estimated coefficients on delta vary appreciably to different outlier remedies. Specifically, our inference on delta alters significantly depending on whether data are treated or untreated. For instance, in the R&D regression the sign of the estimated coefficient of delta changes from positive using untreated data in column (1) to negative using fully-winsorized data in column (4). Numerically, the estimated coefficient of delta drops from 0.00103 (p-value of 7.67%) using untreated data to −0.0116 (p-value of 75.9%) using fully-winsorized data. The decrease of 1,226% is huge and has economic significance. Statistically, the estimated coefficient of delta is significant at the 10% level using untreated data, compared with statistical insignificance at the same level using fully-winsorized data.[19] Similarly, the inference also changes when we use median regression as

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

[19] Our least squares estimates of delta using untreated data are qualitatively different from those in extant studies. For example, in the R&D regression our estimated coefficients on delta are positive but those in the literature are negative (Coles et al., 2006; Hirshleifer et al., 2012). This discrepancy may arise because extant studies use treated (or winsorized) data but our study uses untreated data. When we fully winsorized our data, the sign of the estimated coefficients of delta in the regression changes to negative from positive.

the estimation method. Specifically, the median regression estimate of delta is 0.000396 (*p*-value of 46.58%) which is positive but statistically insignificant at the conventional levels.

Similarly, our results show that the magnitude of the estimated coefficients of vega also varies significantly to different outlier remedies. For example, in the R&D regression the estimated coefficient of vega increase significantly from 0.218 (*p*-value of 2.41%) using untreated data in column (1) to 0.612 (*p*-value of 0.03%) using fully-winsorized data in column (4). The increase of 180.7% is huge and has economic significance. These estimates are statistically significant at the conventional levels. In contrast, in the R&D regression the median regression estimate of vega is merely 0.177 (*p*-value of 20.93%), or a decrease of 18.8%. The estimated coefficients of vega is significant at the conventional levels using untreated data, compared with statistical insignificance at the same levels using median regression.[20]

As indicated by the results in Table 5, the outliers' influence on the estimated coefficients of vega is small whereas that on the estimated coefficients on delta is large. Thus, we anticipate that our results on vega should be qualitatively similar between median regression (a robust estimator) and least squares regression using untreated data. We are surprised to find that our median regression estimates on vega differ meaningfully from those least squares estimates using untreated data. Specifically, none of the median regression estimates are statistically significant at the conventional level, compared with statistical significance at the same levels using untreated data. This discrepancy compels us to use quantile regression method for the estimation in the following Section. This is because quantile regression method allows us to investigate if the aforementioned relation is different at different points of the conditional distribution rather than at a single quantile point (or median).

[20] Our estimates of vega using treated or untreated data are qualitatively similar to those in extant studies in two major ways, e.g., Coles et al. (2006) and Hirshleifer et al. (2012). First, the sign of our estimates of vega is positive. Second, these estimates are statistically significant at the conventional levels.

Similarly, our results also differ meaningfully if we use logarithmic transformation for our managerial incentive variables, compared with those using untreated data. This also motivates us to examine the role of different model specifications in the following Section.

E. Quantile Regression

In the second-stage procedure, we use quantile regression method for two reasons (e.g., Koenker and Basset, 1978; Koenker and Hallock, 2001). First, their estimates are less vulnerable to the outliers' influence than least squares estimates. Second, quantile regression allows for different estimates to be computed at different quantiles of the conditional distribution. The latter consideration is particularly relevant if the relationship differs meaningfully across different quantiles of the conditional distribution. In practice, many economic relationships differ significantly across firms and individuals. For example, the pay-for-performance sensitivity differs considerably across managers, firm-type, and firm-size (e.g., Hallock et al., 2010; Conyon and Schwalbach, 2000; Baker and Hall, 2004; and Schaefer, 1998). Similarly, the relation of managerial incentives and corporate innovations also differs meaningfully between innovative firms and non-innovative firms.[21]

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

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

where y is the dependent variable, X_{it} is a vector of explanatory variables, β is a vector of regression coefficients, and u_i is an error. The least squares estimates are vulnerable to the

[21] When we use all the observations (innovative and non-innovation firms) of the estimation sample (E) in quantile regressions, our results (available upon request) show that the quantile regression estimates on vega at the lowest quantiles are close to zero and statistically insignificant at the conventional levels. In contrast, those at the highest quantiles are positive and statistically significant at the conventional levels. This indicates that the relation between managerial incentives and corporate innovation is significantly stronger for conditionally more innovative firms than for conditionally less innovative firms.

outlier's influence because they are estimated based on a sum of squared errors. In contrast, quantile regression minimizes the sum of absolute errors (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 vary by quantiles, that is, $\beta(\tau)$ can be different across different quantile points being estimated (τ).

We use the model specified in equation (1) for quantile regressions. The only exception is that vega and delta are transformed by taking logarithm of one plus delta and vega as follows: ln(1+vega) and ln(1+delta). We apply the same transformation to vega and delta as to the dependent variable. In unreported results we find that our quantile regression estimates on vega and delta are significantly more stable across conditional quantiles after the transformation.[22] 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) using a set of identical explanatory variables for only innovative firms in the estimation sample (E).

---

[22] To comprehend the effect of log transformation on regression coefficients, let us use the following simplified log-transformation equation for illustration: $\ln(1+y)= a + b\times\ln(1+x) + u$, where y is R&D expenditure, x is vega, and u is a random error. Note that $b = \partial\ln(1+y)/\partial\ln(1+x) = \varepsilon \times [(1+x)/x) \times (y/(1+y)]$, where ε is the elasticity of y with respect to x defined as $dy/dx\times(x/y)$. Thus, the coefficient b is equal to the elasticity multiplied by a constant depending on the values of x and y. If a higher quantile of y is positively correlated with a higher quantile of x, the coefficient b would be more similar to the elasticity than otherwise.

Table 7 reports the results of the nine quantile regressions of each measure of corporate innovation in the 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).

Our results show that all quantile regression estimates of vega are positive in all corporate innovation regressions. Nevertheless, this positive result is particularly relevant for innovative firms at the middle quantiles (where $0.5 \leq \tau \leq 0.8$) because only those estimates are statistically significant at the conventional levels. In contrast, the quantile regression estimate at the lowest quantile (where $\tau = 0.1$) is statistically insignificant at the conventional levels. For example, the results in Table 7B show that higher sensitivity of CEO wealth to stock volatility (vega) induces a greater effect on patent production for firms which are expected to produce more patent counts (e.g., $\tau = 0.8$) than for firms which are expected to produce less in patent counts (e.g., $\tau = 0.1$). The quantile regression estimates on vega are plotted in Figure 2A, 3A, and 4A for R&D, patent counts, and patent citations regressions, respectively.

The least squares estimates and median regression estimates of vega are similar. For example, in the R&D regression in Table 7A, the least squares estimate is 0.0921 whereas that for the median regression ($\tau = 0.5$) is 0.0888. Both of these estimates are statistically significant at the conventional levels. Additionally, the quantile regression estimates are somewhat similar for the quantile at and greater than 0.2 ($\tau \geq 0.2$). Overall, our quantile regression results are in congruent with the existing findings reported in extant studies in that higher sensitivity of CEO wealth to stock volatility induces firms to increase corporate innovation including R&D investments, the number of patent counts, and the number of patent citations (Coles et al., 2006; Hirshelifer et al., 2016).

[INSERT TABLE 7A-7C HERE]

In contrast, none of the quantile regression estimates of delta are statistically significant at the conventional levels in all the regressions. Our results imply that higher CEO pay-for-performance sensitivity (delta) has no material effect on corporate innovation activities, regardless of the expected intensity of corporate innovation of the firm.

The quantile regression estimates on delta are plotted in Figure 2B, 3B, and 4B for R&D, patent counts, and patent citations regressions, respectively. As indicated clearly in Figure 2B, the least squares estimate of delta in the R&D regression is slightly positive whereas all the quantile regression estimates are negative. For example, the results in Table 7A show that the least squares estimate of delta is 0.000421 whereas that for the median regression is −0.0254. In the R&D regression, the sign of the least square estimate is different from all the quantile regression estimates. These conflicting results are due to the presence of an outlier firm in our estimation sample. Therefore, the least squares estimate is excessively influenced by the single outlier.

Specifically, the outlier firm is Microsoft. Statistically, Microsoft is a high leverage point and can exert excessive influence on least square estimates because it invests heavily in R&D. Specifically, Microsoft is one of the biggest spenders on R&D expenditure in the U.S. for the recent decades. Furthermore, William Gates and Steve Ballmer—their CEOs—never received stock option grants during our sample period because they have substantial equity ownership in the firm. Thus, their equity holdings in the company have provided sufficient incentives to enhance corporate innovations. In other words, for Microsoft's CEO the vega is consistently zero whereas the delta is very large during the sample period. It is worthy to note that this scenario is not rare in the U.S. and especially relevant for those firms that are led by the founder CEOs.[23]

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 activities 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 for firms that truly care about corporate innovation.

Second, OLS regression method is inappropriate to estimate the relation between managerial incentives and corporate innovation. This is because least squares estimates are particularly susceptible to the outliers' influence given that all key variables of interest have noticeable outliers. Conventional remedies for outliers (e.g., data winsorization) are inadequate to treat outliers for two reasons. First, least squares estimates are fragile and vary appreciably to different data treatment remedies. Second, data treatment is subjective and vulnerable to the confirmation bias.

To mitigate the outliers' influence, we use quantile regressions for estimation. Our quantile regression results indicate that higher sensitivity of CEO wealth to stock return volatility (vega) induces firms to increase corporate innovation outcomes including R&D investment, the number of patent counts, and the number of patent citations. In addition, the abovementioned effect is 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

[23] For example, the compensation is primarily cash-based for Steve Jobs, the founder of Pixar and Apple. Additionally, Jobs had never received a single stock option grant when he was the CEO of Pixar.

corporate innovation. Specifically, 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

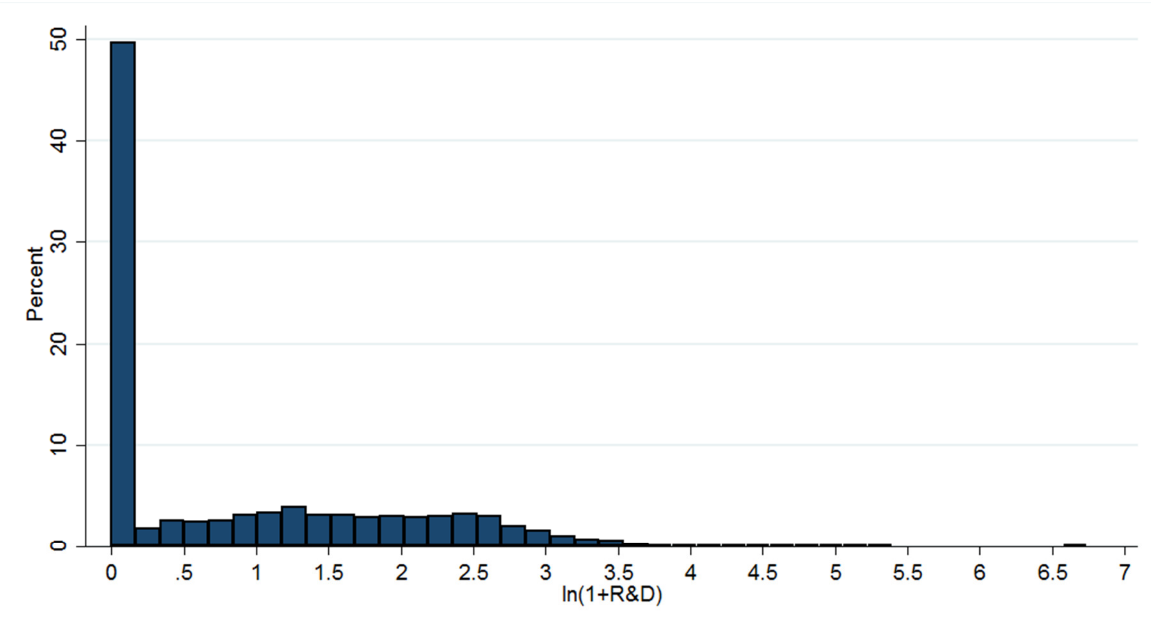

FIGURE 1B: Histogram of the Number of Patent Counts

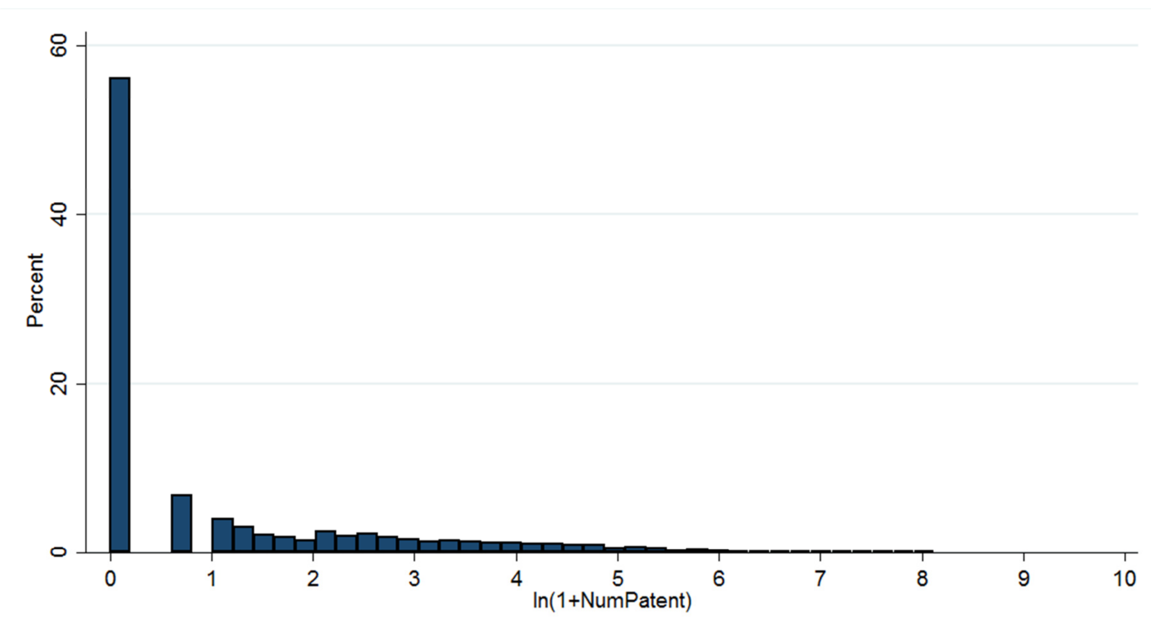

FIGURE 1C: Histogram of the Number of Patent Citations

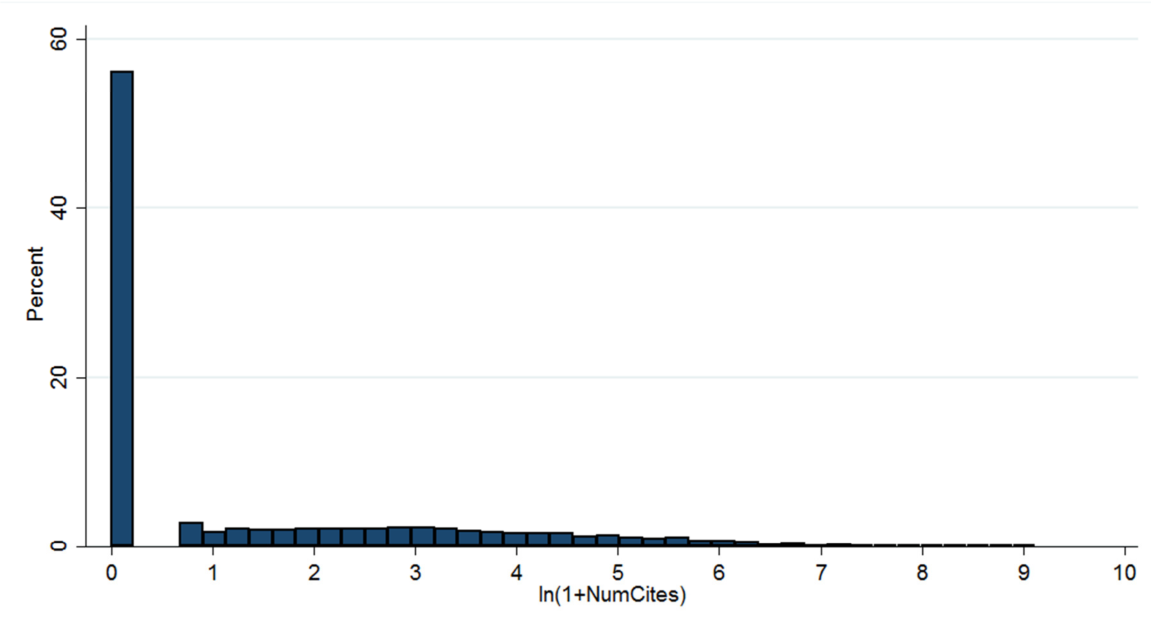

FIGURE 2A: Quantile Regression Estimates of Vega — R&D Expenditure scaled by Total Asset

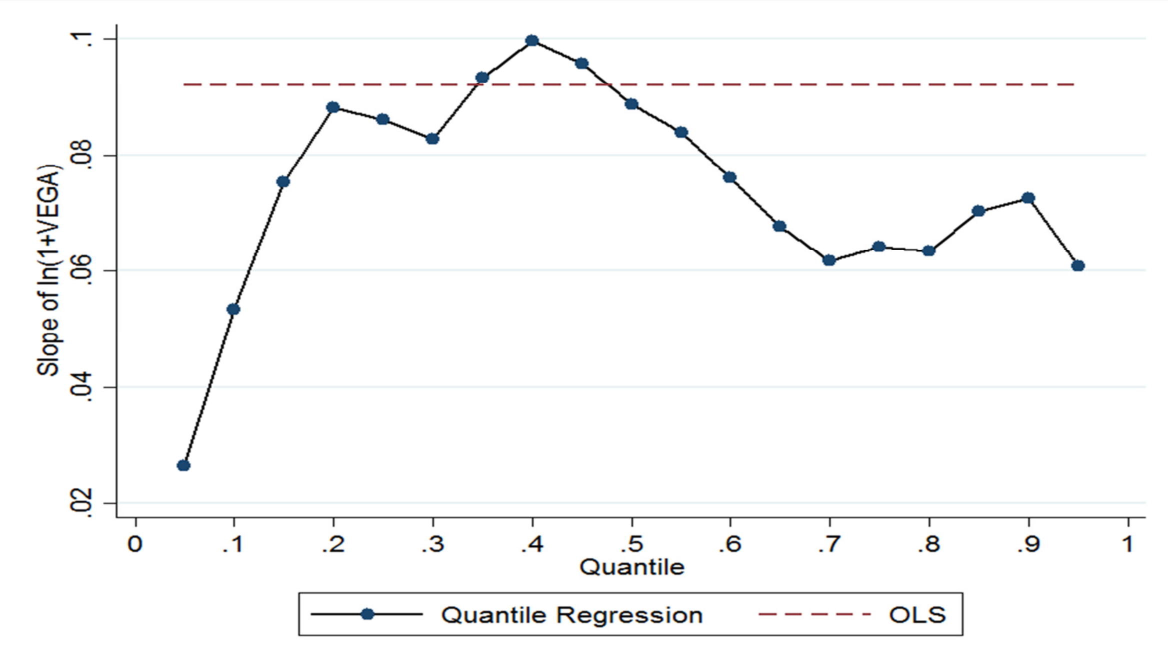

FIGURE 2B: Quantile Regression Estimates of Delta — R&D Expenditure scaled by Total Asset

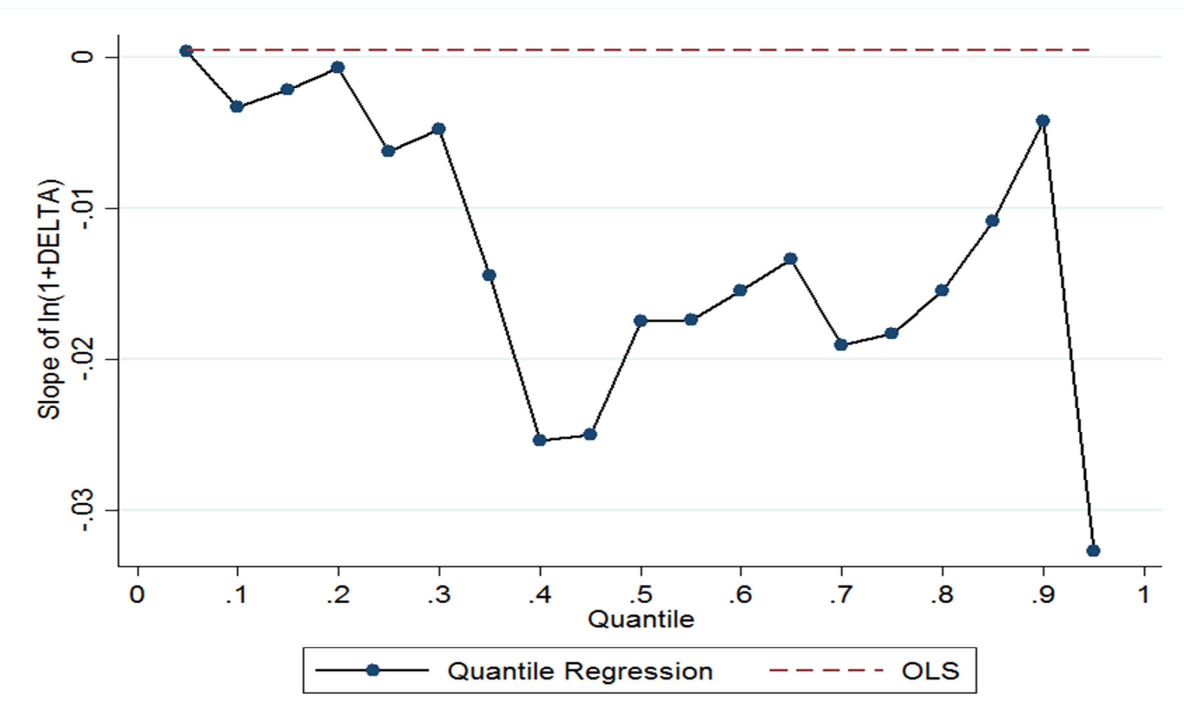

FIGURE 3A: Quantile Regression Estimates of Vega — Number of Patent Counts

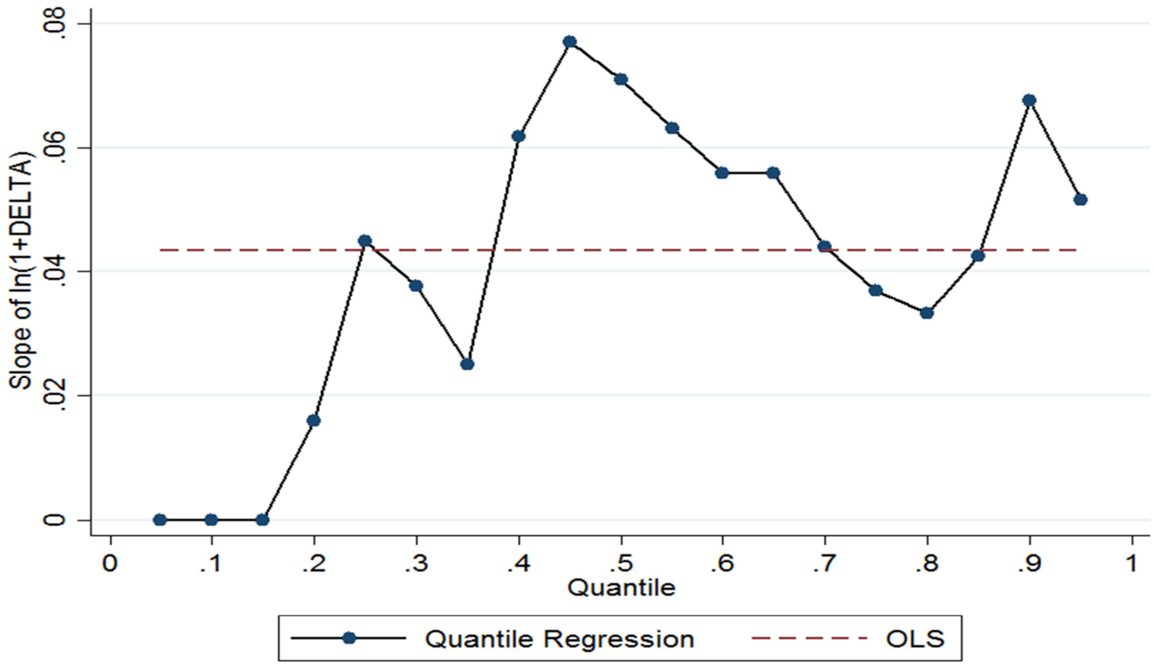

FIGURE 3B: Quantile Regression Estimates of Delta — Number of Patent Counts

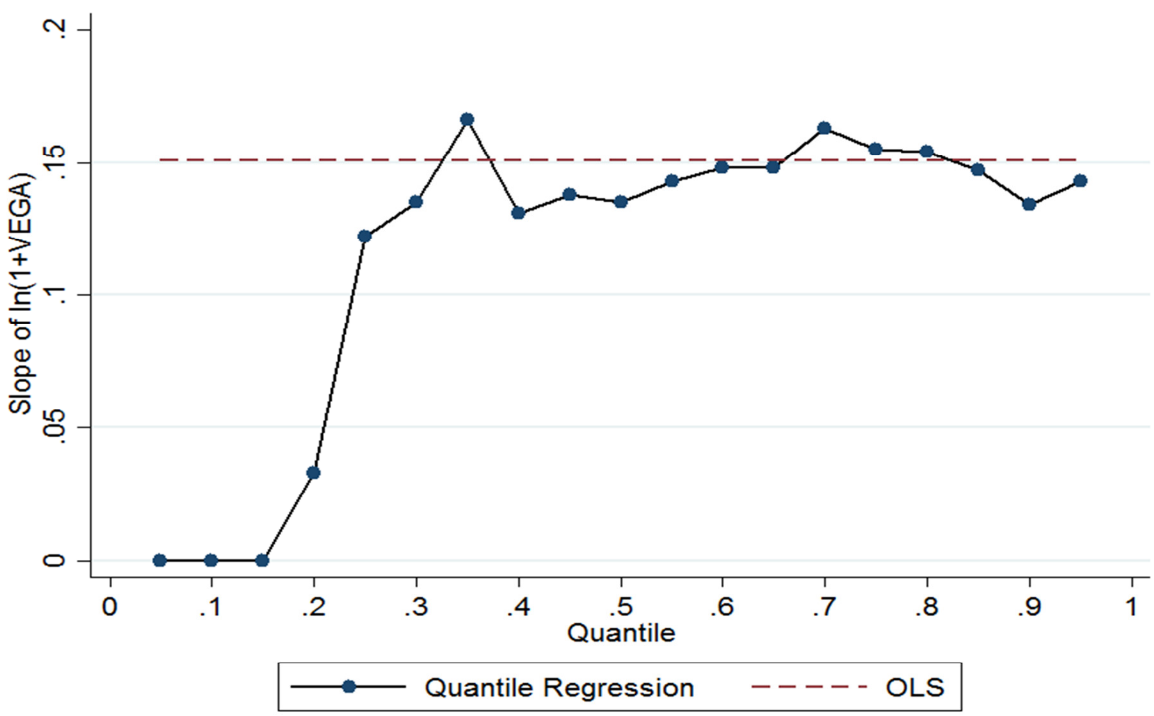

FIGURE 4A: Quantile Regression Estimates of Vega — Number of Patent Citations

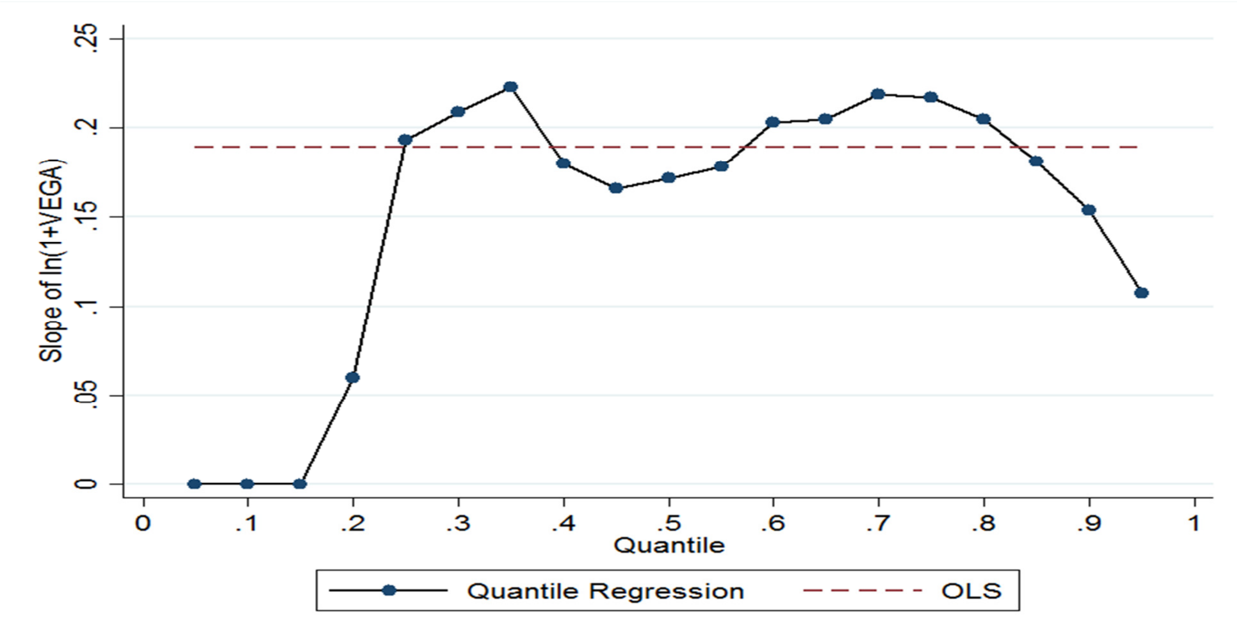

FIGURE 4B: Quantile Regression Estimates of Delta — Number of Patent Citations

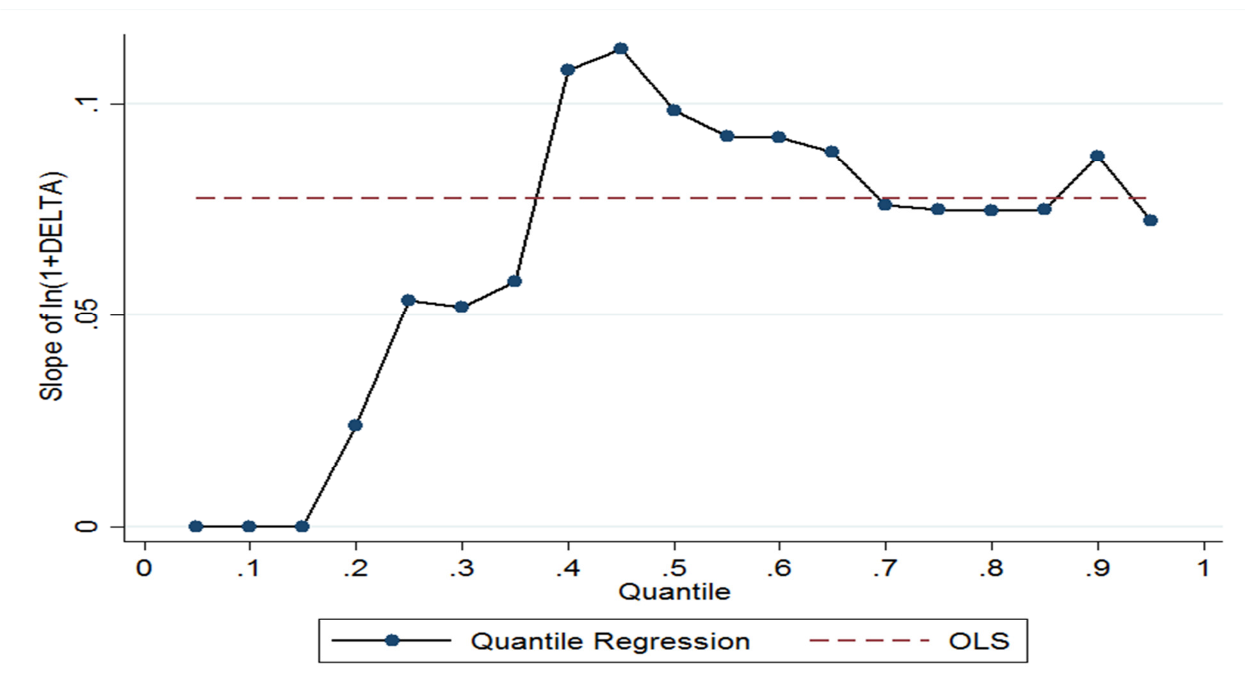

TABLE 2: Summary Statistics

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
NumCites	54.738	308.781	0.000	0.000	0.000	1073.781	9209.707	2.521	4.352
NumPatent	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 firms into innovative or non-innovative firms. The dependent variable is a binary variable and takes the value of zero if a firm's research and development expenditure is always zero during our sample period, and one if otherwise. The control variables include firm size (ln(Sales) and its square) and industry dummies according to the Fama-French 48 industry classification. The numbers in parentheses are robust standard errors, clustered at the industry level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels.

	Estimated coefficient	Odds ratio
ln(SALE)	-3.212***	0.040***
	(0.571)	(0.023)
$\ln(\text{SALE})^2$	0.227***	1.255***
	(0.039)	(0.050)
Intercept	10.484***	
	(2.221)	
Industry F.E.	Yes	Yes
N	1,035	1,035
Pseudo R^2	0.470	
Pseudo R^2 (Industry FE only)	0.440	

TABLE 3B: Classification Accuracy of the First-Stage 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 (T). 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	159	23	182
	Non-innovative	33	141	174
	Total	192	164	356

Classification Accuracy: 0.843 = (159 +141)/ 356
Classification Sensitivity: 0.828 = 159/ 192
Classification Precision: 0.874 = 159 / 182

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

This table presents the least squares estimates of $VEGA_{t-1}$ and $DELTA_{t-1}$ in the mixture distribution model and in the single equation model. The single equation model uses OLS use all observations in the estimation sample. The mixture distribution model uses only firms classified as innovative in the first-stage logistic regression. The dependent variable is the natural logarithm of one plus each proxy of corporate innovation as follows: ln(1+R&D), ln(1+NumPatent) and ln(1+NumCites). 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 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	NumPatent	NumPatent	NumCites	NumCites
	(1) Single-equation	(2) Mixture-distribution	(3) Single-Equation	(4) Mixture-distribution	(5) Single-equation	(6) Mixture-distribution
VEGA#	0.192*	0.218**	0.307	0.711***	0.423	0.854***
	(0.103)	(0.0967)	(0.229)	(0.247)	(0.294)	(0.299)
DELTA#	0.00193***	0.00103*	0.00646***	0.00573***	0.00875***	0.00749***
	(0.000645)	(0.000582)	(0.00236)	(0.00127)	(0.00250)	(0.00141)
CASH#	0.0207	0.0431	0.0279	0.0717	0.0264	0.0708
	(0.0136)	(0.00325)	(0.0292)	(0.0561)	(0.0337)	(0.0637)
ln(SALE)	-0.103**	-0.147***	0.397***	0.513***	0.432***	0.549***
	(0.0451)	(0.0471)	(0.0758)	(0.103)	(0.0862)	(0.117)
M/B	0.0571***	0.0890***	0.0508**	0.0923**	0.0871***	0.139***
	(0.0163)	(0.0240)	(0.0189)	(0.0341)	(0.0202)	(0.0423)
SURCASH	1.197***	1.428***	0.371**	0.160	0.472**	0.253
	(0.359)	(0.229)	(0.173)	(0.257)	(0.211)	(0.332)
SALEGRW	-0.122***	-0.163***	-0.376***	-0.412***	-0.395***	-0.410***
	(0.0413)	(0.0563)	(0.0673)	(0.115)	(0.0761)	(0.118)
RET	-0.0683***	-0.122***	-0.0428*	-0.0781	-0.0657**	-0.109*
	(0.0146)	(0.0349)	(0.0224)	(0.0538)	(0.0256)	(0.0602)
LEVERAGE	-0.0537	0.0585	-0.232	-0.231	-0.314*	-0.330
	(0.199)	(0.220)	(0.155)	(0.277)	(0.179)	(0.298)
CEOTenure	-0.00664**	-0.0139***	-0.00925*	-0.0191**	-0.0108*	-0.0224*
	(0.00267)	(0.00401)	(0.00475)	(0.00921)	(0.00579)	(0.0114)
Intercept	1.529***	2.372***	-1.295**	-1.320*	-1.196*	-1.074
	(0.317)	(0.276)	(0.536)	(0.706)	(0.602)	(0.811)
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
N	7,370	4,025	7,370	4,025	7,370	4,025
MAR	0.460	0.427	0.909	0.763	1.119	0.966
Rel. GOF	N/A	0.558	N/A	0.661	N/A	0.645
Adj. R^2	0.632	0.633	0.406	0.459	0.395	0.428

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: ln(1+R&D), ln(1+NumPatent) and ln(1+NumCites). Columns (1), (3) and (5) report the results using all innovative firms from 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 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	NumPatent	NumPatent	NumCites	NumCites
	(1) Innovative-firm sample	(2) Drop Microsoft	(3) Innovative-firm sample	(4) Drop Microsoft	(5) Innovative-firm sample	(6) Drop Microsoft
VEGA#	0.218**	0.234**	0.711***	0.748***	0.854***	0.895***
	(0.0967)	(0.0951)	(0.247)	(0.261)	(0.299)	(0.314)
DELTA#	0.00103*	-0.0125	0.00573***	-0.0233	0.00749***	-0.0221
	(0.000582)	(0.0106)	(0.00127)	(0.0229)	(0.00141)	(0.0276)
CASH#	0.0431	0.0447	0.0717	0.0759	0.0708	0.0759
	(0.0325)	(0.0337)	(0.0561)	(0.0602)	(0.0637)	(0.0683)
ln(SALE)	-0.147***	-0.146***	0.513***	0.514***	0.549***	0.548***
	(0.0471)	(0.0499)	(0.103)	(0.107)	(0.117)	(0.122)
M/B	0.0890***	0.0910***	0.0923**	0.0957***	0.139***	0.141***
	(0.0240)	(0.0225)	(0.0341)	(0.0320)	(0.0423)	(0.0379)
SURCASH	1.428***	1.409***	0.160	0.116	0.253	0.203
	(0.229)	(0.231)	(0.257)	(0.255)	(0.332)	(0.325)
SALEGRW	-0.163***	-0.157***	-0.412***	-0.399***	-0.410***	-0.397***
	(0.0563)	(0.0561)	(0.115)	(0.117)	(0.118)	(0.119)
RET	-0.122***	-0.126***	-0.0781	-0.0859	-0.109*	-0.116*
	(0.0349)	(0.0324)	(0.0538)	(0.0536)	(0.0602)	(0.0586)
LEVERAGE	0.0585	0.0536	-0.231	-0.238	-0.330	-0.337
	(0.220)	(0.219)	(0.277)	(0.275)	(0.298)	(0.297)
CEOTenure	-0.0139***	-0.0133***	-0.0191**	-0.0178*	-0.0224*	-0.0211*
	(0.00401)	(0.00417)	(0.00921)	(0.00978)	(0.0114)	(0.0122)
Intercept	2.372***	2.362***	-1.320*	-1.328*	-1.074	-1.071
	(0.276)	(0.301)	(0.706)	(0.742)	(0.811)	(0.851)
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
N	4,025	4,016	4,025	4,016	4,025	4,016
Rel. GOF	N/A	0.518	N/A	0.522	N/A	0.519
Adj. R^2	0.477	0.477	0.425	0.422	0.397	0.393

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: ln(1+R&D), ln(1+NumPatent), and ln(1+NumCites). 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 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)
Dependent variable	Independent variable	Untreated	Log transform on VEGA and DELTA	Winsorize dependent variable at 1%	Winsorize all variables 1%	Median regression
ln(1+R&D)	VEGA#	0.218**	92.1***	0.215**	0.612***	0.177
		(0.0967)	(19.4)	(0.0952)	(0.170)	(0.141)
	DELTA#	0.00103*	0.421	0.00100*	-0.0116	0.000396
		(0.000582)	(31.8)	(0.000577)	(0.0378)	(0.000543)
	N	4,025	4,025	4,025	4,025	4,025
	Rel. GOF	N/A	0.551	0.555	0.552	0.544
	Adj. R^2	0.477	0.488	0.475	0.422	0.451
ln(1+NumPatent)	VEGA#	0.711***	151.0*	0.684***	1.59***	0.467
		(0.247)	(77.6)	(0.222)	(0.469)	(0.300)
	DELTA#	0.00573***	43.4	0.00580***	-0.0584	0.00454***
		(0.00127)	(59.8)	(0.00122)	(0.0872)	(0.00117)
	N	4,025	4,025	4,025	4,025	4,025
	Rel. GOF	N/A	0.538	0.464	0.541	0.539
	Adj. R2	0.425	0.430	0.425	0.090	0.401
ln(1+NumCites)	VEGA#	0.854***	189.0**	0.833***	1.96***	0.448
		(0.299)	(90.7)	(0.277)	(0.549)	(0.298)
	DELTA#	0.00749***	77.5	0.00748***	-0.0525	0.00571***
		(0.00141)	(71.9)	(0.00138)	(0.111)	(0.00129)
	N	4,025	4,025	4,025	4,025	4,025
	Rel. GOF	N/A	0.547	0.466	0.542	0.548
	Adj. R^2	0.397	0.405	0.397	0.182	0.359
	Control vars.	Yes	Yes	Yes	Yes	Yes
	Ind. F.E.	Yes	Yes	Yes	Yes	Yes
	Year F.E.	Yes	Yes	Yes	Yes	Yes

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

TABLE 7A: Second-stage Quantile Regression — Research and Development

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 innovative firms classified in the first-stage logit regression. The dependent variable is the natural logarithm of one plus each proxy of corporate innovation as follows: ln(1+R&D) in Panel A, ln(1+NumPatent) in Panel B, and ln(1+NumCites) in in Panel C. The vega and delta variables are transformed by taking logarithm of one plus delta and vega as follows: ln(1+vega) and ln(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 parentheses are robust standard errors, clustered at the industry level. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels.

	(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
ln(1+VEGA)	0.0921***	0.0534	0.0882***	0.0828***	0.0996***	0.0888***	0.0760***	0.0617***	0.0635***	0.0725***
	(0.0194)	(0.0350)	(0.0267)	(0.0216)	(0.0245)	(0.0287)	(0.0274)	(0.0210)	(0.0153)	(0.0134)
ln(1+DELTA)	0.000421	-0.00335	-0.000679	-0.00479	-0.0254	-0.0175	-0.0155	-0.0191	-0.0155	-0.00424
	(0.0318)	(0.0231)	(0.0243)	(0.0241)	(0.0273)	(0.0304)	(0.0377)	(0.0286)	(0.0200)	(0.0212)
CASH#	0.0328	0.0222**	0.000721	-0.00290	0.00325	0.0123	0.0226	0.0155	0.0186	0.0239
	(0.0245)	(0.00869)	(0.00791)	(0.00681)	(0.0404)	(0.0358)	(0.0300)	(0.0219)	(0.0227)	(0.0280)
ln(SALE)	-0.171***	-0.00775	-0.0264	-0.0448	-0.0898	-0.121**	-0.146***	-0.160***	-0.178***	-0.200***
	(0.0424)	(0.0296)	(0.0312)	(0.0392)	(0.0634)	(0.0524)	(0.0498)	(0.0355)	(0.0475)	(0.0463)
M/B	0.0891***	0.0464***	0.0542	0.0708	0.0847**	0.0759*	0.0626	0.0815*	0.0914***	0.0968***
	(0.0269)	(0.0112)	(0.0360)	(0.0472)	(0.0364)	(0.0439)	(0.0501)	(0.0467)	(0.0264)	(0.0220)
SURCASH	1.400***	1.340*	2.004***	1.972***	2.051***	1.848***	1.800***	1.449***	1.314***	0.936***
	(0.219)	(0.691)	(0.438)	(0.479)	(0.713)	(0.460)	(0.665)	(0.305)	(0.150)	(0.298)
SALEGRW	-0.150***	-0.0476	-0.158**	-0.204***	-0.160	-0.173*	-0.123	-0.173***	-0.189***	-0.251**
	(0.0473)	(0.0506)	(0.0780)	(0.0738)	(0.108)	(0.0933)	(0.0751)	(0.0533)	(0.0570)	(0.104)
RET	-0.114***	-0.0863*	-0.0797*	-0.0784**	-0.131***	-0.123***	-0.107**	-0.0958*	-0.106***	-0.0775**
	(0.0361)	(0.0509)	(0.0445)	(0.0399)	(0.0465)	(0.0377)	(0.0427)	(0.0535)	(0.0250)	(0.0378)
LEVERAGE	0.0386	-0.0687	-0.258	-0.256	-0.174	-0.312	-0.217	-0.150	-0.0449	0.0814
	(0.214)	(0.155)	(0.214)	(0.287)	(0.372)	(0.294)	(0.395)	(0.320)	(0.303)	(0.0939)
CEOTenure	-0.0111**	-0.0000519	-0.00518**	-0.00977***	-0.0116***	-0.0134***	-0.0132***	-0.0129**	-0.0107*	-0.00709*
	(0.00451)	(0.00257)	(0.00243)	(0.00303)	(0.00382)	(0.00421)	(0.00395)	(0.00559)	(0.00632)	(0.00382)
Intercept	2.292***	2.902***	3.138***	3.150***	3.244***	3.586***	3.773***	3.809***	3.954***	4.073***
	(0.221)	(0.173)	(0.202)	(0.198)	(0.204)	(0.202)	(0.165)	(0.150)	(0.209)	(0.144)

Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	4,025	4,025	4,025	4,025	4,025	4,025	4,025	4,025	4,025	4,025
R^2	0.451									

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

TABLE 7B: Second-stage Quantile Regression — Number of Patent Counts

	(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
ln(1+VEGA)	0.151*	0.000	0.0328	0.135	0.131*	0.135**	0.148**	0.163**	0.154*	0.134
	(0.0776)	(0.0155)	(0.0339)	(0.0856)	(0.0774)	(0.0687)	(0.0715)	(0.0689)	(0.0929)	(0.106)
ln(1+DELTA)	0.0434	0.000	0.0159	0.0377	0.0619	0.0710	0.0558	0.0439	0.0332	0.0676
	(0.0598)	(0.0150)	(0.0303)	(0.0556)	(0.0635)	(0.0525)	(0.0579)	(0.0643)	(0.0710)	(0.101)
CASH#	0.0000736	0.000	0.0509	0.0725	0.0602	0.0807	0.0895	0.0700	0.0433	0.156
	(0.0000509)	(0.0164)	(0.0945)	(0.107)	(0.111)	(0.0770)	(0.0726)	(0.0734)	(0.0580)	(0.102)
ln(SALE)	0.469***	0.000	0.0358	0.378*	0.536***	0.544***	0.550***	0.556***	0.578***	0.523***
	(0.0843)	(0.0173)	(0.0507)	(0.196)	(0.148)	(0.134)	(0.114)	(0.128)	(0.107)	(0.0967)
M/B	0.0908**	0.000	0.0244	0.0751**	0.0923**	0.0927**	0.111***	0.128***	0.130**	0.114**
	(0.0388)	(0.00793)	(0.0227)	(0.0347)	(0.0399)	(0.0387)	(0.0357)	(0.0313)	(0.0608)	(0.0533)
SURCASH	0.126	0.000	0.124	0.440	0.0409	-0.290	-0.192	-0.548**	-0.718	-0.353
	(0.247)	(0.0607)	(0.158)	(0.269)	(0.244)	(0.337)	(0.449)	(0.274)	(0.477)	(0.340)
SALEGRW	-0.409***	0.000	-0.0541	-0.348*	-0.476***	-0.544***	-0.491**	-0.529**	-0.481**	-0.412**
	(0.111)	(0.0186)	(0.0527)	(0.183)	(0.137)	(0.140)	(0.218)	(0.234)	(0.235)	(0.184)
RET	-0.0576	0.000	-0.0164	-0.0407	-0.00765	-0.00892	-0.0329	-0.0369	-0.0768	-0.0372
	(0.0627)	(0.0157)	(0.0271)	(0.0721)	(0.0757)	(0.0685)	(0.0551)	(0.0384)	(0.0528)	(0.0992)
LEVERAGE	-0.256	0.000	-0.0134	-0.101	-0.112	-0.234	-0.127	-0.330*	-0.354*	-0.380***
	(0.256)	(0.0425)	(0.0787)	(0.273)	(0.284)	(0.453)	(0.122)	(0.177)	(0.203)	(0.134)
CEOTenure	-0.0169	0.000	-0.00472	-0.0147	-0.0193	-0.0226*	-0.0254**	-0.0249**	-0.0241	-0.0104
	(0.0125)	(0.00247)	(0.00401)	(0.00967)	(0.0118)	(0.0129)	(0.0123)	(0.0125)	(0.0154)	(0.0162)
Intercept	-1.623**	2.303***	2.136***	0.848	0.409	0.498	0.513	0.444	0.335	0.433
	(0.616)	(0.298)	(0.343)	(1.028)	(0.726)	(0.592)	(0.542)	(0.471)	(0.420)	(0.369)
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	4,025	4,025	4,025	4,025	4,025	4,025	4,025	4,025	4,025	4,025
R^2	0.432									

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

TABLE 7C: Second-stage Quantile Regression — Number of Patent Citations

	(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
ln(1+VEGA)	0.189**	0.000	0.0595	0.209*	0.180*	0.172**	0.203***	0.219**	0.205**	0.154
	(0.0907)	(0.0224)	(0.0515)	(0.118)	(0.101)	(0.0862)	(0.0773)	(0.0866)	(0.0925)	(0.104)
ln(1+DELTA)	0.0775	0.000	0.0239	0.0519	0.108	0.0984	0.0921	0.0760	0.0747	0.0876
	(0.0719)	(0.0220)	(0.0457)	(0.0763)	(0.0792)	(0.0693)	(0.0619)	(0.0852)	(0.0842)	(0.0787)
CASH#	0.0000674	0.000	0.0853	0.0396	0.0209	0.0647	0.0392	0.0204	0.0237	0.104
	(0.0000557)	(0.0237)	(0.131)	(0.137)	(0.109)	(0.0825)	(0.0928)	(0.0750)	(0.0563)	(0.108)
ln(SALE)	0.485***	0.000	0.0638	0.493**	0.607***	0.602***	0.574***	0.585***	0.561***	0.525***
	(0.0947)	(0.0253)	(0.0784)	(0.231)	(0.172)	(0.143)	(0.152)	(0.123)	(0.107)	(0.0871)
M/B	0.131***	0.000	0.0476	0.116**	0.110**	0.159***	0.155**	0.185**	0.185***	0.156***
	(0.0471)	(0.0114)	(0.0401)	(0.0483)	(0.0537)	(0.0574)	(0.0654)	(0.0738)	(0.0649)	(0.0517)
SURCASH	0.215	0.000	0.244	0.562	0.212	-0.227	-0.244	-0.536	-0.259	0.0212
	(0.319)	(0.0890)	(0.265)	(0.399)	(0.407)	(0.464)	(0.632)	(0.376)	(0.376)	(0.463)
SALEGRW	-0.415***	0.000	-0.100	-0.478***	-0.547***	-0.628***	-0.543*	-0.487*	-0.440**	-0.470***
	(0.115)	(0.0273)	(0.0855)	(0.173)	(0.130)	(0.177)	(0.304)	(0.277)	(0.212)	(0.164)
RET	-0.0706	0.000	-0.0290	-0.0639	-0.00638	-0.0270	-0.0229	-0.0319	-0.0995	-0.0157
	(0.0715)	(0.0230)	(0.0411)	(0.0860)	(0.106)	(0.0949)	(0.0614)	(0.0841)	(0.0633)	(0.0800)
LEVERAGE	-0.352	0.000	-0.0357	-0.232	-0.247	-0.261	-0.252	-0.363**	-0.443**	-0.558***
	(0.278)	(0.0624)	(0.115)	(0.355)	(0.354)	(0.461)	(0.162)	(0.180)	(0.211)	(0.192)
CEOTenure	-0.0211	0.000	-0.00805	-0.0174	-0.0242	-0.0253	-0.0290*	-0.0289	-0.0303*	-0.0167
	(0.0153)	(0.00361)	(0.00612)	(0.0117)	(0.0153)	(0.0172)	(0.0152)	(0.0181)	(0.0163)	(0.0117)
Intercept	-1.478**	3.079***	2.798***	1.265	0.683	0.870	0.983	0.856*	0.932*	1.070***
	(0.710)	(0.436)	(0.476)	(1.022)	(0.819)	(0.652)	(0.656)	(0.471)	(0.478)	(0.368)
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	4,025	4,025	4,025	4,025	4,025	4,025	4,025	4,025	4,025	4,025
R^2	0.400									

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

APPENDIX 1 – Variable Definitions

Variable	Description
R&D	Research and development expenditure scaled by book asset
NumPatent	Number of patent counts
NumCites	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 1000)
CASH	Cash compensation which is the sum of salary and bonus for the CEO in the current year (Scaled down by a factor of 1000)
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.

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.

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.

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.