

Essays on Earning Management and Leading Indicator Variables

by

Wei Li

*A Dissertation Submitted to the Tepper School of Business
in Partial Fulfillment of the Requirement for the Degree of*

Doctor of Philosophy

at

Carnegie Mellon University

September, 2008

Acknowledgements

It has been a long journey to get to the current stage. For sure, I cannot achieve it without the guidance, suggestion, encouragement, and support from my advisors, friends and family.

I am deeply indebted to my advisor, Professor Jonathan Glover, for his endless efforts to guide and encourage me on my work. His knowledge opens my eyes in the research opportunities; his encouragement helps me through my uneasy time; his insightful comments make my dissertation more polished; and, in particular, his personality and research attitude inspire me a lot. I really feel honored being advised by such a wonderful person for my Ph.D. study.

I would like to express my gratitude to Professor Pierre Jinghong Liang and Professor Zhaoyang Gu. The theory reading group initiated by Professor Jinghong Liang has been a wonderful experience to a Ph.D. student. I started my first research project under the supervision of Professor Zhaoyang Gu. It has also been a fruitful experience. In particular, I gratefully thank them for their valuable comments and editorial advice on my dissertation.

I am very grateful to Professor Bryan Routledge for his serving as the outside reader and helpful comments on my dissertation. I am also thankful to Professor Yuji Ijiri, Professor Carolyn Levine, Professor Lin Nan, and Professor John O'Brien for their valuable suggestions.

I appreciate the collaborations and enjoyable time with my fellow graduate students Jeremy Bertomeu, Carl Brousseau, Min Cao, Jie Chen, Tao Chen, Ting Chen, Edwige Cheynel, Chen Li, Haijin Lin, Karen Lin, Min Lin, Jong Chool Park, Hong Qu, Jiong Sun, Xue Sun, Xiaoyan Wen, Chen Xiang, Jian Xue and Yinqing Zhao. I also would like to thank Jackie Cavendish and Lawrence Rapp for their valuable help to our Ph.D. students.

Last but not Least, I am forever indebted to my parents and my sister for their unselfish support and endless encouragement since I was very young. Although we are at the two ends of the earth, I can always feel their constant love. I also gratefully thank my wife, Lang Xiao, for her accompanying me through my most difficult time. Thank her for her unconditional support, patience, encouragement, and love with all my heart!

Abstract

This dissertation consists of two essays. The first essay provides a rational explanation of the mysterious earnings discontinuity phenomenon. It also develops two new empirical predictions which are supported by the empirical tests. The second essay identifies conditions under which leading indicator variables discourage long-term investment.

Essay 1: Discontinuity in Earnings Distribution: A Theory and Evidence

This paper presents a rational model of financial reporting in which investors use reported earnings not only to infer true (pre-managed) earnings but also to update their beliefs about the precision (inverse of the variance) of earnings. In the model, over-reporting earnings has two opposing pricing effects. For example, when earnings are positively auto-correlated, inflating a positive (reported) earnings surprise has a positive pricing effect because investors infer higher (pre-managed) earnings for both current and future periods. However, investors also infer a lower earnings precision from the higher earnings surprise, leading to a lower pricing weight placed on the higher surprise. This is the negative pricing effect of over-reporting earnings. For firms whose earnings are strongly positively auto-correlated, the trade-off between the two opposing effects creates a pooled report right above the prior mean of the earnings distribution and a no-reporting "hole" right below the prior mean in equilibrium (i.e., an earnings discontinuity around the prior mean). The pricing function of reported earnings exhibits an overall "S-shape" and a negative slope for medium (positive and negative) earnings surprises. The above theoretical results are consistent with existing empirical findings. What distinguish the paper are two new empirical predictions: (1) no earnings discontinuity exists for firms whose earnings are negatively or weakly positively auto-correlated, and (2) the earnings discontinuity is more pronounced for firms with more positively auto-correlated earnings (higher auto-covariance or lower variance). The paper also presents empirical evidence supporting the two predictions.

Essay 2: When Leading Indicator Variables reduce Long-term Investment

One apparent advantage of leading indicator variables is that they encourage long-term investments. In this paper, we show that leading indicator variables sometimes increase and sometimes decrease long-term investments. We study a two-period short-term contracting relationship between a principal and an agent in which the agent takes both a short-term and a long-term (investment) action. At the beginning of the second period, the principal can either retain the existing agent or hire a new agent and incur a nontrivial replacement cost. The short-term action increases the first-period outcome, while the long-term investment decreases the first-period outcome, and only the net effect is observed with noise. The long-term investment also increases the second-period outcome. There is no pure strategy equilibrium but instead a mixed strategy equilibrium in which the agent sometimes takes the long-term investment and the principal sometimes retains the existing agent. So far, the possibility of retention is the only means of encouraging the long-term investment. We then introduce a non-contractible binary leading indicator variable made available to both parties at the end of the first period. If the net return on investment is small and/or the leading indicator variable is highly likely to be high when investment is high, there is a unique equilibrium in which the principal retains the existing agent upon seeing a low realization of the leading indicator variable and randomizes upon seeing a high realization. When the principal observes a high realization of the leading indicator variable, her updated belief about the agent's probability of investment (which is higher than the agent's *ex ante*/equilibrium probability of investment) is the same as the agent's equilibrium probability of investment under the no-leading-indicator scenario. Hence, in this case, the agent's *ex ante* expected investment is lowered by the presence of the leading indicator variable. The effect of the smaller expected investment dominates other factors, and the principal is worse off with the leading indicator than without it. We also identify conditions under which the leading indicator variable leads to more investment and the principal is better off.

Contents

Acknowledgements	I
Abstract	III
Essay 1. Discontinuity in Earnings Distribution: A Theory and Evidence	1
1.1 Introduction	1
1.2 The Model and the Benchmark	7
1.3 The Partially Pooling Equilibrium	14
1.4 The Empirical Properties of the Equilibrium	20
1.5 The Empirical Results	28
1.6 Conclusion	37
Appendix A	39
Appendix B	45
Essay 2. When Leading Indicator Variables reduce Long-term Investment	51
2.1 Introduction	51
2.2 The Model	53
2.3 The Scenario with No Leading Indicator Variables	55
2.4 The Scenario with Leading Indicator Variables	61
2.5 Conclusion	68
Appendix	70
References	79

1 Essay 1

Discontinuity in Earnings Distributions: A Theory and Evidence

1.1 Introduction

Prior literature documents that the distribution of reported earnings is discontinuous around a threshold such as zero earnings, previous earnings, or analyst consensus forecasts of earnings (see Hayn 1995, Burgstahler and Dichev 1997, and Degeorge, Patel, and Zeckhauser 1999). The discontinuity exhibits a substantially low (high) frequency of earnings reports in a small interval right below (at/above) the threshold. This empirical phenomenon is interpreted as managers having incentives to avoid reporting losses, earnings decreases, and earnings that will miss analyst consensus forecasts.

In this paper, I provide a rational explanation for the earnings discontinuity phenomenon in a market setting, where the reported-earnings discontinuity arises endogenously even though the distribution of the underlying true (pre-managed) earnings is continuous (a normal distribution).¹ Moreover, this paper shows analytically that the earnings discontinuity depends on the time-series property of firms' earnings and provides empirical evidence consistent with this prediction. In particular, the paper predicts and finds no discontinuity in the earnings distribution of firms whose earnings are negatively or weakly positively auto-correlated. The paper also predicts and finds that the earnings discontinuity is more pronounced for firms with more positively auto-correlated earnings (higher auto-covariance or lower variance).

Specifically, I study a two-period model of financial reporting in which the manager manipulates reported earnings to maximize the firm price just after his first-period earnings report. In a risk-neutral market, the end-of-first-period price consists of three parts: the inferred first-period earnings, the prior mean of the second-period earnings, and the expected second-period earnings

¹True earnings can be thought of as the "shareholders' income" which, according to Sunder (1997), "exist(s) in the form of physical capital." He argues that "translation of this physical-capital income into units of money...by the proprietor or shareholder (instead of the manager) of the firm (referred to as *first-best valuation*)...is not without ambiguity." In this paper, I assume such a "*first-best valuation*" exists. True earnings refer to the earnings before managers' discretionary manipulation (i.e., pre-managed earnings), whereas reported earnings refer to the earnings reports that are subject to manipulation.

innovation (the difference between the true earnings realization and the prior mean of the true earnings).² The expected second-period earnings innovation is proportionate to the (inferred) first-period earnings innovation with the coefficient being the product of the earnings covariance and the (perceived) earnings precision (inverse of the variance). Intuitively, the more correlated the two earnings are (i.e., higher covariance or higher precision), the more likely the same earnings innovation will be expected for the second period.

Following prior literature (e.g., Beyer 2005 and Subramanyam 1996), the precision of the underlying true earnings is assumed to be an unknown random variable. To value the firm, investors use reported earnings *not only to infer the true earnings but also to update their beliefs about the precision of earnings*. As a consequence, over-reporting earnings always has two opposing effects on the market price. For example, when earnings are positively correlated and the first-period earnings innovation is positive, inflating reported earnings would increase the investors' inferred first-period earnings and earnings innovation, potentially increasing the market price. However, manipulating reported earnings upward would also decrease the investors' perceived earnings precision due to a larger deviation of the inferred earnings from the prior mean. The reduced perceived precision would lead investors to place a lower pricing weight on the higher (inferred) first-period earnings innovation, dampening the positive effect of over-reporting. Intuitively, when the earnings surprise becomes larger, investors view the surprise as more transitory and rely less on the larger surprise in valuing the firm.

The manager trades off the two opposing pricing effects in choosing his optimal reporting strategy. The paper's main results are as follows:

1. When true earnings are strongly positively correlated across the two periods (referred to as SPC Case throughout the paper), the positive effect of over-reporting dominates if the magnitude of the first-period earnings innovation is relatively small or large, and the negative effect dominates if the magnitude is medium. In equilibrium, there is a pooled report right above the prior mean of the first-period true earnings (akin to a high frequency of reports above the prior mean), and there are no reports around one point (a no-reporting "hole")

²The difference between the *true* earnings realization and the prior mean of the true earnings is referred to as "earnings innovation," and the difference between the *reported* earnings and the prior mean as "earnings surprise" throughout the paper.

right below the prior mean (akin to a low frequency of reports below the prior mean).³ If we think of the prior mean as the threshold in the empirical earnings distribution, these results are consistent with the empirical earnings discontinuity described above.

2. When true earnings are strongly negatively correlated across the two periods (referred to as SNC Case), the positive effect dominates if the magnitude of the first-period earnings innovation is large, and the negative effect dominates if the magnitude is small. In equilibrium, there is a pooled report (a no-reporting hole) at one point far below (above) the prior mean, opposite to SPC Case. In theory, discontinuities exist at these two points for individual firms. However, the discontinuities may not be statistically significant in a cross-sectional earnings distribution, aligned at the prior mean/threshold, of such firms (e.g., the distribution of analyst forecast errors).⁴ The reasons are twofold. First, if the two points are far away from the prior mean, the probability densities around these two points may be too small for the discontinuities to be statistically significant. Second, these two points are widely dispersed across firms and, thus, each individual discontinuity is smoothed away in a cross-sectional setting.
3. When true earnings are weakly correlated, either positively or negatively, across the two periods (referred to as WKC Case), the positive effect always dominates. Thus, the equilibrium is fully separating, and no discontinuity is expected in the earnings distribution of such firms.

In summary, the model predicts the discontinuity is observable for SPC firms but not for SNC or WKC firms. The dominance of SPC firms may be the reason the discontinuity shows up in the empirical earnings distribution that combines all three types of firms.

The above theoretical prediction allows me not only to empirically re-examine the discontinuity phenomenon but also to validate the theory with empirical data. In my empirical tests, firms are partitioned into two groups according to the sign (positive or negative) of their earnings auto-covariances.⁵ Since the negative auto-covariance group comprises all SNC firms and part of WKC

³ A pooled report means the manager issues the same (pooled) report for all earnings realizations in an earnings interval (referred to as the pooling interval).

⁴ If we could draw a cross-sectional earnings distribution aligned at one of these two points, an empirically observable discontinuity would exist at this point. However, these two points are not directly observable or estimable for individual firms. Therefore, I choose to focus on the earnings distribution aligned at the prior mean/threshold.

⁵ Ideally, I would partition firms into the three groups according to the criteria the theory implies. However, because the cut-off points for partitioning are not directly observable or estimable, I take this indirect approach.

firms, the model predicts no empirically observable discontinuity in the earnings distribution of this group. The positive auto-covariance group consists of all SPC firms and the rest of WKC firms. According to the model, a significant earnings discontinuity will exist in the earnings distribution of this group given the dominance of SPC firms.

The data come from a sample of firms with both quarterly analyst consensus forecasts and quarterly actual earnings available from First Call between 1990 and 2005. The analyst consensus forecast serves as the proxy for the prior mean of the forecasted earnings (Brown and Rozeff 1978). First Call's actual earnings are used to estimate the earnings auto-covariance and variance of each firm.⁶ I construct seven subsamples where the required minimum number of consecutive quarterly earnings of each firm (to estimate auto-covariance) ranges from 8 to 32. The empirical results indicate that the earnings discontinuity in the positive group is significant in all subsamples, whereas the earnings discontinuity in the negative group is not significant in six out of the seven subsamples, consistent with the prediction of the model.

Another important result of the model is that, for SPC firms, the pooled report and the no-reporting hole move toward the prior mean as the auto-covariance of earnings increases or the variance of earnings decreases, leading to a more pronounced discontinuity. The intuition is as follows. When the earnings series becomes more correlated (higher auto-covariance or lower variance), the higher correlation induces more negative pricing effect than before, and the manager has to pool his reports closer to the prior mean where the positive pricing effect is larger.

The empirical implication of the above result is that, *ceteris paribus*, the lower the variance of earnings or the higher the auto-covariance of earnings, the more pronounced the earnings discontinuity. Since earnings variance and auto-covariance are positively correlated, to isolate the auto-covariance effect, I control for the variance effect by dividing the positive group (proxy for SPC firms) into variance deciles. Then, firms in each variance decile are equally divided into two groups based on the ranking (high or low) of their earnings auto-covariances. The empirical results indicate that the discontinuity in the low auto-covariance group is less pronounced than in the high auto-covariance group, consistent with the prediction of the model. I similarly isolate the variance effect after controlling for the auto-covariance effect and find that the discontinuity in the low-variance group is more pronounced than in the high-variance group, also consistent with the

⁶See more discussion on this empirical design issue in Section 5.

prediction. Further robustness tests indicate that the above results hold even after controlling for a possible bias due to more WKC firms being included in the low auto-covariance group or the high variance group.

Other than providing a rational explanation for the earnings discontinuity phenomenon, my model also produces two theoretical results regarding the price-earnings relationship, which are consistent with empirical findings in prior literature. Freeman and Tse (1992) and Skinner and Sloan (2002) find that the price increases more sharply when the earnings surprise (relative to analyst forecasts) is small in magnitude and much less sharply when the surprise is large (i.e., an “S-shape”). Das and Lev (1994) use a nonparametric method and find a negative relationship between the abnormal return and the annual earnings change for medium positive earnings changes as well as some negative earnings changes (see Figure 2 in Das and Lev 1994). Consistent with these empirical findings, in my model, for SPC firms, the pricing function of reported earnings exhibits an overall “S-shape” and a negative slope for medium (positive and negative) earnings surprises.⁷

A few analytical papers have also studied the endogenous earnings discontinuity. Closest to my paper, Kirschenheiter and Melumad (2002) present a rational model in which investors infer the precision of cashflows (which can be thought of as their measure of pre-managed earnings) from reported earnings. The authors assume the manager seeks to maximize only the second-period price. Since the total reported earnings from both periods equals the total true cashflows due to their complete accrual reversal assumption, the manager cares about only the investors’ perceived precision. My model differs in that the manager is assumed to maximize the price before (instead of after) the complete accrual reversal. Thus, the manager in my model cares not only about the investors’ perceived earnings precision but also about the investors’ inferred earnings. My model fits the case of managers with a shorter horizon and their model fits the case of managers with a longer horizon. My model also provides other empirical predictions regarding the cross-sectional variation in the earnings discontinuity phenomenon that are supported by empirical results. Nevertheless, the models are quite similar.

Guttman, Kadan, and Kandel (2006) study a one-period model in which the cost of manipulation plays an important role. They show that, in a partially pooling equilibrium, the earnings

⁷Other analytical papers producing the “S-shaped” price-earnings relationship include Subramanyam (1996) and Liang (2004).

discontinuity results from a trade-off between a higher stock price (by over-reporting) and a simultaneous higher cost of manipulation. Xin (2007) extends their model by exogenously assigning asymmetric “over-reporting costs” and “missing-the-forecast costs,” and shows that the manager has a pooled report at the exogenously given forecast. Instead of using a market setting, Fedyk (2007) presents an agency model in which the discontinuity in earnings reports emerges endogenously.^{8 9}

Although prior literature empirically documents the earnings discontinuity around zero earnings, previous earnings, or analyst consensus forecasts, Durtschi and Easton (2005) argue that the discontinuity at zero earnings or previous earnings is induced by the deflator choice (i.e., using the market value as the deflator instead of the number of shares), because the market price is systematically higher for profit firms than for loss firms across the distribution of EPS (earnings per share). However, Beaver, McNichols, and Nelson (2007) argue that examining the behavior of deflators across the distribution of EPS is likely to be problematic because the number of shares is systematically higher for loss firms than for profit firms across the distribution of undeflated net income, whereas the market value is relatively symmetric at zero net income. Thus, they argue that negative share-deflated earnings are shifted toward zero, which reduces the discontinuity around zero earnings, whereas market-value deflation does not distort the distribution around zero. When discussing the discontinuity at analyst forecasts, Durtschi and Easton (2005) provide an alternative explanation. They argue the discontinuity appears because analyst forecast errors tend to be larger when they are optimistic (i.e., negative errors) than when they are pessimistic (i.e., positive errors). However, this analyst behavior explanation seems unable to account for the empirical predictions/findings of my paper.

On the empirical front, my paper contributes to the literature by predicting variation in the

⁸Other agency models of earnings management include Evans and Sridhar (1996), Demski (1998), Dye (1988), and Arya, Glover, and Sunder (1998).

⁹Although analytical papers attribute the discontinuity phenomenon to earnings management, the empirical results on the cause of discontinuity are mixed. Beaver, McNichols, and Nelson (2003) find that, in the property-casualty insurance industry, small profit firms significantly understate the claim loss reserve accrual compared to small loss firms, consistent with firms manipulating earnings to avoid losses. However, Beaver, McNichols, and Nelson (2007) suggest the discontinuity around zero earnings may be due to asymmetric impacts of income taxes and special items on profit and loss firms. Dechow, Richardson, and Tuna (2003) find both small profit and small loss firms have a similar level of discretionary accruals and a similar proportion of firms with positive discretionary accruals, inconsistent with small loss firms boosting discretionary accruals to avoid reporting a loss. The results in my paper may help explain why Dechow et al. do not find significant difference in discretionary accruals between small profit and small loss firms. According to my model, the pooled report right above the threshold (small profit firms) and the no-reporting hole right below the threshold (small loss firms) result from both earnings-increasing and earnings-decreasing manipulations.

earnings discontinuity phenomenon across firms. The empirical results indicate that firms with strongly positively auto-correlated earnings are the primary contributors to the discontinuity in the cross-sectional earnings distribution. The empirical results also indicate that the earnings discontinuity is more pronounced for firms with more auto-correlated earnings (higher auto-covariance or lower variance).

The rest of the paper proceeds as follows. Section 2 describes the model and discusses a benchmark case where the manager is confined to issue a truthful earnings report. Section 3 derives a partially pooling equilibrium of the model. Section 4 studies the empirical properties of the equilibrium. Section 5 generates empirical hypotheses and conducts empirical tests. Section 6 concludes the paper.

1.2 The Model and the Benchmark

This paper considers a two-period model of financial reporting in a risk-neutral capital market.¹⁰ I assume the following events take place sequentially. The normal ongoing activities of a firm generate correlated stochastic earnings over two periods. At the end of the first period, the owner-manager privately observes the realized (true) earnings, while outside investors cannot. However, outside investors will receive a mandatory earnings report from the manager and then price the firm accordingly. After releasing the report, the manager sells the firm to interested investors due to exogenous liquidity constraints, and the new shareholders are entitled to the earnings (cashflows) from both periods. Since investors value the firm based only on the reported earnings which are not confined to be truthful, the manager has an incentive to strategically manipulate the report to inflate the investors' perceived value of the firm. Below, I present the model in detail.

The earnings in period t ($t = 1, 2$), denoted by \tilde{e}_t , are normally distributed with mean μ_t and precision $\tilde{\tau}_t$ (i.e., inverse of variance). The two earnings, \tilde{e}_1 and \tilde{e}_2 , are assumed to be correlated with a covariance of σ .¹¹ Both the mean μ_t ($t = 1, 2$) and the covariance σ are common knowledge, while the precision of earnings $\tilde{\tau}_t$ is presumed to be an unknown random variable. Neither the owner-manager nor outside investors observe the realized value of $\tilde{\tau}_t$. Following prior literature

¹⁰Suppose the discount rate is zero for simplicity.

¹¹The covariance σ measures the persistence of earnings. All other things being equal, the larger the covariance, the more persistent the earnings.

(e.g., Beyer 2005 and Subramanyam 1996), I assume the prior distribution of $\tilde{\tau}_1$ is Gamma(α, β) with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$, and this distribution is independent of any other distributions in this model.¹² The parameters α and β are common knowledge. The distributions of \tilde{e}_1 and \tilde{e}_2 can be summarized as follows:

$$\begin{bmatrix} \tilde{e}_1 \\ \tilde{e}_2 \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \frac{1}{\tilde{\tau}_1} & \sigma \\ \sigma & \frac{1}{\tilde{\tau}_2} \end{bmatrix} \right). \quad (1)$$

At the end of the first period, the manager privately observes the realized earnings e_1 and issues a mandatory earnings report m to the public. The earnings report may differ from the actual earnings realization since the manager has discretion to manipulate the reported earnings. For tractability, I assume the manager's cost of manipulation is zero if the magnitude of manipulation is less than ω ($\omega > 0$) and infinity otherwise.¹³ Accordingly, the manager's discretion, denoted by δ , is bounded from both above and below (i.e., $\delta \in [-\omega, \omega]$). The size of ω reflects how easily the manager can manipulate the report: the larger the ω , the more freedom the manager has to manipulate the report. The manager's reporting strategy is a real function $M(e_1) : \mathcal{R} \rightarrow \mathcal{R}$ that maps the realized earnings into the reported earnings. It can be expressed as $m = M(e_1) = e_1 + \delta$. The manager seeks to maximize the selling price by optimally selecting his reporting strategy.

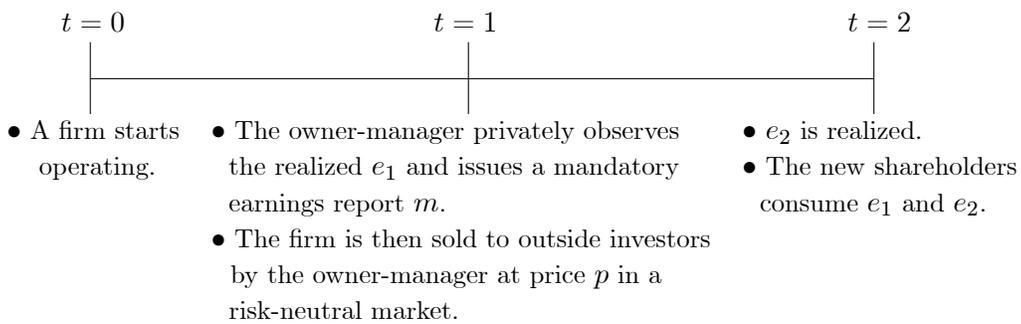


Figure 1. Timeline

Outside investors infer the firm's value (i.e., the expected total earnings from both periods)

¹² A Gamma(α, β) distribution has a support of $(0, +\infty)$ with mean $\frac{\alpha}{\beta}$ and variance $\frac{\alpha}{\beta^2}$.

¹³ Kirschenheiter and Melumad (2002) make a similar assumption about the cost of manipulation. This assumption can be viewed as auditors pulling the trigger only when the manipulation reaches a threshold and/or litigation arising only when the manipulation is severe.

based solely on the reported earnings m with the knowledge of possible managerial manipulation. The investors' pricing function is a real function $P(m) : \mathcal{R} \rightarrow \mathcal{R}$ that maps the reported earnings into a price. It can be expressed as $p = P(m) = E[\tilde{e}_1 + \tilde{e}_2 | m, M(e_1)]$. Figure 1 depicts the timeline of the model.

I define the equilibrium as follows.

Definition 1 *An equilibrium comprises a reporting strategy $M^*(e_1)$ and a pricing function $P^*(m)$ that satisfy the following conditions¹⁴:*

1. *Given the pricing function $P^*(m)$, the realized first-period earnings e_1 , and the available discretion $\delta \in [-\omega, \omega]$, the earnings report $M^*(e_1)$ maximizes the market price p ;*
2. *The pricing function $P^*(m)$ is consistent with the reporting strategy $M^*(e_1)$ in the sense that:*

$$p = P^*(m) = E[\tilde{e}_1 + \tilde{e}_2 | m, M^*(e_1)]. \quad (2)$$

This definition of equilibrium is straightforward. Anticipating how the market will value the firm afterward, the manager issues a report within the available discretion range to maximize the selling price. Given the conjecture about the manager's reporting strategy, the market prices the firm at a value equal to the expected value of total earnings $\tilde{e}_1 + \tilde{e}_2$ conditional on the reported earnings. In equilibrium, the conjectured reporting behavior coincides with the reporting strategy the manager actually selects.

Before proceeding to derive the equilibrium of the model, I first present a benchmark case where the manager has no discretion in reporting earnings and has to tell the truth, i.e., $\omega = 0$ or $m = e_1$. Outside investors get to know the realized earnings e_1 directly from the manager's earnings report m . To characterize this benchmark case, I need only to figure out the price p investors would like to pay for the firm. In a risk-neutral market, the price p is the expected total earnings from both

¹⁴Generally, we still need to define off-equilibrium beliefs to complete the definition. Following prior literature, I will defer characterizing the off-equilibrium beliefs until I derive the equilibrium of the model.

periods conditional on e_1 :

$$\begin{aligned}
p &= E[\tilde{e}_1 + \tilde{e}_2|e_1] \\
&= e_1 + E[\tilde{e}_2|e_1] \\
&= e_1 + \mu_2 + \sigma E[\tilde{\tau}_1|e_1](e_1 - \mu_1).
\end{aligned} \tag{3}$$

The last equality results from the joint normal distribution assumption (1). According to (3), the market price p consists of the first-period realized earnings e_1 , the prior mean of the second-period earnings μ_2 , and a third term proportionate to the first-period earnings innovation (i.e., $e_1 - \mu_1$). The third term, referred to as “expected innovation term” hereafter, reflects the market’s expectation of the second-period earnings innovation conditional on e_1 . As seen in (3), the expectation is dependent on the expected/perceived earnings precision as well as the earnings covariance. Intuitively, if the two earnings are not closely correlated due to small covariance or small precision, investors perceive the first-period earnings innovation as more transitory and rely less on it in predicting the second-period earnings innovation. Accordingly, a lower pricing weight would be placed on the first-period earnings innovation.

Since $\tilde{e}_1 \sim N(\mu, \frac{1}{\tilde{\tau}_1})$, given the realized e_1 , $\tilde{\tau}_1$ is distributed gamma(α' , β') with $\alpha' = \alpha + \frac{1}{2}$ and $\beta' = \beta + \frac{1}{2}(e_1 - \mu_1)^2$, where α and β are parameters of the prior gamma distribution of $\tilde{\tau}_1$ (see DeGroot 1970). Hence,

$$E[\tilde{\tau}_1|e_1] = \frac{\alpha'}{\beta'} = \frac{\alpha + \frac{1}{2}}{\beta + \frac{1}{2}(e_1 - \mu_1)^2}. \tag{4}$$

Equation (4) implies the capital market updates its perceived earnings precision based on the deviation of e_1 from its prior mean μ_1 (i.e., the first-period earnings innovation). In particular, the perceived precision and the earnings innovation (in magnitude) are negatively related. Earnings with a higher innovation are associated with a lower precision and vice versa. Substituting (4) into (3) yields a closed form expression for the price p , which is summarized in the following Lemma 1.

Lemma 1 *Given no information asymmetry, the selling price of the firm at the end of the first period is*

$$p = P(e_1) = e_1 + \mu_2 + \frac{\sigma h(\alpha + \frac{1}{2})}{\beta + \frac{1}{2}h^2}, \quad \text{where } h = e_1 - \mu_1. \tag{5}$$

Taking the first derivative of the pricing function $P(e_1)$ with respect to e_1 yields

$$P'(e_1) = \frac{\frac{1}{4}h^4 + [\beta - \frac{1}{2}\sigma(\alpha + \frac{1}{2})]h^2 + \beta^2 + \sigma(\alpha + \frac{1}{2})\beta}{(\beta + \frac{1}{2}h^2)^2}. \quad (6)$$

Notice that both numerator and denominator of this derivative are quadratic functions of h^2 (h being the first-period earnings innovation $e_1 - \mu_1$), which makes the model highly tractable. Some algebraic manipulation yields the following characterization of the shape of the pricing function $P(e_1)$.

Lemma 2 *Given no information asymmetry,*

(i) *When $\sigma > \frac{8\beta}{\alpha + \frac{1}{2}}$, the pricing function $P(e_1)$ is*

$$\begin{cases} \text{increasing,} & \text{if } h \leq -r_2, -r_1 \leq h \leq r_1, \text{ or } h \geq r_2 \\ \text{decreasing,} & \text{if } -r_2 < h < -r_1 \text{ or } r_1 < h < r_2 \end{cases}$$

$$\begin{aligned} \text{where } r_1 &= \sqrt{2[-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2}) - \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}]} \\ r_2 &= \sqrt{2[-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2}) + \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}]} \quad ; \end{aligned}$$

(ii) *When $\sigma < \frac{-\beta}{\alpha + \frac{1}{2}}$, the pricing function $P(e_1)$ is*

$$\begin{cases} \text{increasing,} & \text{if } h \leq -r_2 \text{ or } h \geq r_2 \\ \text{decreasing,} & \text{if } -r_2 < h < r_2 \end{cases} \quad ;$$

(iii) *When $\frac{-\beta}{\alpha + \frac{1}{2}} \leq \sigma \leq \frac{8\beta}{\alpha + \frac{1}{2}}$, the pricing function $P(e_1)$ is consistently increasing in e_1 .*

Proof. (The proof is in the Appendix.) ■

As Lemma 2 shows, both the sign and the magnitude of the covariance σ affect the shape of the pricing function:

(i) If the earnings are strongly positively correlated (SPC Case: $\sigma > \frac{8\beta}{\alpha + \frac{1}{2}}$), the pricing function is increasing in e_1 when the magnitude of the first-period earnings innovation is either large or small and decreasing in e_1 when the magnitude is medium;

- (ii) If the earnings are strongly negatively correlated (SNC Case: $\sigma < \frac{-\beta}{\alpha + \frac{1}{2}}$), the pricing function is increasing in e_1 when the magnitude of the first-period earnings innovation is large and decreasing in e_1 when the magnitude is small; and
- (iii) If the earnings are weakly (either positively or negatively) correlated (WKC Case: $\frac{-\beta}{\alpha + \frac{1}{2}} \leq \sigma \leq \frac{8\beta}{\alpha + \frac{1}{2}}$), the pricing function is consistently increasing in e_1 .

Figure 2. The Pricing Function $P(h)$ under the Benchmark Case ($h = e_1 - \mu_1$)

Figure 2a. SPC Case: $\sigma > \frac{8\beta}{\alpha + \frac{1}{2}}$

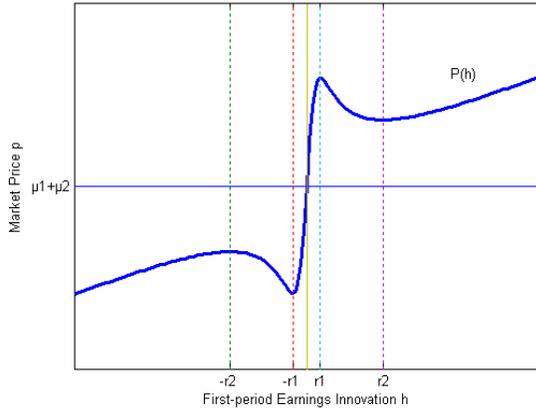


Figure 2b. SNC Case: $\sigma < \frac{-\beta}{\alpha + \frac{1}{2}}$

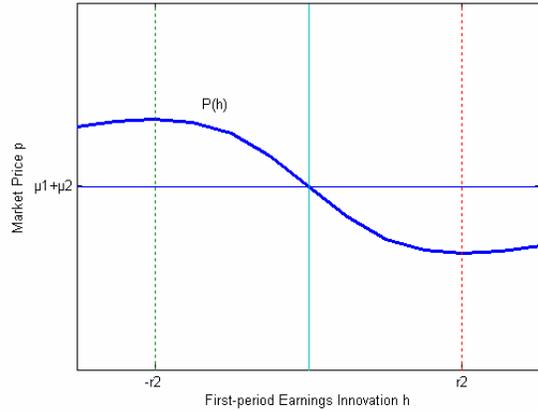
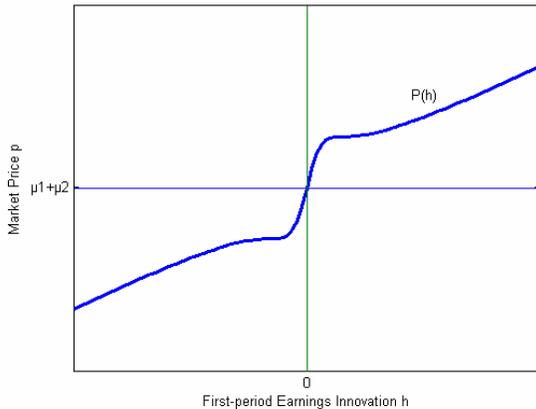


Figure 2c. WKC Case: $\frac{-\beta}{\alpha + \frac{1}{2}} \leq \sigma \leq \frac{8\beta}{\alpha + \frac{1}{2}}$



α, β : parameters of the Gamma distribution of the earnings precision with the expected precision $\frac{\alpha}{\beta}$

σ : earnings covariance

e_1 : first-period realized earnings

μ_1 : prior mean of the first-period true earnings

μ_2 : prior mean of the second-period true earnings

$h = e_1 - \mu_1$: first-period earnings innovation

Figures 2a, 2b, and 2c depict the pricing functions under SPC, SNC, and WKC Cases respectively in the $P(h) - h$ plane ($h = e_1 - \mu_1$). Since each graph is symmetric to the point $(\mu_1 + \mu_2, 0)$,

all three graphs are centered at this point. In Figure 2a, the pricing function attains its local maximum at $h = -r_2$ or r_1 and its local minimum at $h = -r_1$ or r_2 . In Figure 2b, the function attains its local maximum at $h = -r_2$ and its local minimum at $h = r_2$.

To understand the economic intuition of the shape of the pricing function $P(e_1)$, reconsider (3). Any increase in e_1 affects the market price through both the e_1 term (the first term) and the expected innovation term (the third term $\sigma E[\tilde{\tau}_1|e_1](e_1 - \mu_1)$) on the RHS of (3). The e_1 term reflects the “consumption role” of e_1 because it is part of the total consumption investors will enjoy at the end of the second period. It has an unambiguous positive pricing effect. The expected innovation term reflects the “information role” of e_1 since it is the market’s expectation of the second-period earnings innovation conditional on e_1 . It has both positive and negative effects on the market price. Below I discuss its opposing pricing effects by considering two cases, one where the earnings are positively correlated ($\sigma > 0$) and the other where they are negatively correlated ($\sigma < 0$), with more attention focused on the first one.

First, consider the positive correlation case ($\sigma > 0$). To better understand the opposing pricing effects from the expected innovation term, it is helpful to divide this case into two sub-cases based on the sign of the first-period earnings innovation h . If the first-period earnings innovation is positive ($h > 0$), we would also expect a positive earnings innovation for the second period due to the positive correlation. In particular, this expected second-period earnings innovation is proportionate to h with the coefficient being the product of the earnings covariance σ and the investors’ perceived earnings precision $E[\tilde{\tau}_1|e_1]$. Therefore, an increase in e_1 has a positive pricing effect by producing a larger h . However, it also decreases the investors’ perceived earnings precision (due to a larger deviation of e_1 from its prior mean μ_1). The lower perceived precision leads investors to place a lower pricing weight on the larger positive h , which offsets the price-increasing effect from the larger positive h (i.e., a negative pricing effect). This is the basic trade-off managers must face when contemplating inflating the earnings reports.

If the first-period earnings innovation is negative ($h < 0$), an increase in e_1 would decrease its deviation (in magnitude) from the prior mean μ_1 , leading to a higher perceived earnings precision by investors. The higher perceived precision has a *price-decreasing* effect, which seems at odds with the above argument. To see why, first notice that the expected second-period earnings innovation is also negative. A higher perceived precision leads investors to place a higher pricing weight on

the *less negative* first-period earnings innovation h , which offsets the price-increasing effect from the less negative h (i.e., a negative pricing effect).

Second, consider the negative correlation case ($\sigma < 0$). Any increase in e_1 would drag down the expected value of the second-period earnings innovation due to the negative correlation (i.e., negative effect). However, the effect of any increase in e_1 on the investors' perceived earnings precision (as well as the effect from the e_1 term) would offset the negative effect (i.e., positive effect). The detailed argument is similar to the positive correlation case and, thus, omitted.

Given the opposing pricing effects from the expected innovation term as well as the positive pricing effect from the e_1 term, we have the following results:

- (i) If the earnings are strongly positively correlated (SPC Case), the overall positive effect dominates if the magnitude of h is relatively small or large, and the negative effect dominates if the magnitude is medium;
- (ii) If the earnings are strongly negatively correlated (SNC Case), the overall positive effect dominates when the magnitude of h is large, and the negative effect dominates when the magnitude is small; and
- (iii) If the earnings are weakly (either positively or negatively) correlated (WKC Case), the overall positive effect always dominates.

1.3 The Partially Pooling Equilibrium

This section derives the equilibrium of the complete model where the owner-manager is free to manipulate the reported earnings m within a restricted range, that is, $m \in [e_1 - \omega, e_1 + \omega]$.

First, notice that always reporting truthfully cannot be an equilibrium strategy under any circumstances. The reasoning is as follows. Suppose in equilibrium the manager always truthfully reports the realized earnings e_1 to the market (i.e., $m = e_1$) for any $e_1 \in \mathcal{R}$. In response to this reporting strategy, the market values the firm at $p = P(e_1)$ for any $e_1 \in \mathcal{R}$, as described in (5). However, given this naive market belief, the manager has incentives to deviate from reporting the truth for most realized earnings. To see this, observe that, regardless of specific parameter values, there always exist regions where the pricing function $P(e_1)$ is increasing in e_1 and, thus, the

manager benefits by reporting a higher earnings than e_1 (but still within the restricted reporting range). Hence, the manager cannot always report the true e_1 in equilibrium.

Moreover, not only does no truth-reporting equilibrium exist, but no fully separating equilibrium exists when the earnings auto-correlation (either positive or negative) is relatively strong.¹⁵

Proposition 1 *Given that the owner-manager has restricted discretion in issuing the earnings report, no fully separating equilibrium exists when $\sigma > \frac{8\beta}{\alpha+\frac{1}{2}}$ (SPC Case) or $\sigma < \frac{-\beta}{\alpha+\frac{1}{2}}$ (SNC Case).*

Proof. (The proof is in the Appendix.) ■

Notice that $\sigma > \frac{8\beta}{\alpha+\frac{1}{2}}$ corresponds to SPC Case and $\sigma < \frac{-\beta}{\alpha+\frac{1}{2}}$ to SNC Case. Under both cases, the pricing function $P(e_1)$ has at least one local maximum point: SPC Case at the point $e_1 = \mu_1 - r_2$ or $\mu_1 + r_1$ and SNC Case at the point $e_1 = \mu_1 - r_2$. The local maximum point is just where any fully separating equilibrium would break down. Below I use SPC Case to illustrate the idea.

First, observe that there always exists a small neighborhood around $\mu_1 + r_1$ (the local maximum point) such that, for any realized e_1 in this neighborhood, $P(e_1) < P(\mu_1 + r_1)$ and the manager is able to report $M(\mu_1 + r_1)$, the earnings he would report in equilibrium if the realized earnings are $\mu_1 + r_1$. Now, suppose a fully separating equilibrium exists. The market is able to fully unravel the realized e_1 from the manager's report $M(e_1)$ and value the firm at $P(e_1)$. Given this market belief, for any e_1 in the above small neighborhood, the manager has an incentive to report $M(\mu_1 + r_1)$ instead of $M(e_1)$, because, by doing so, the market would simply value the firm at a higher price $P(\mu_1 + r_1)$ rather than $P(e_1)$. This result contradicts the equilibrium assumption and, thus, no fully separating equilibrium exists for the model.

Therefore, a partially pooling equilibrium is of interest, and Proposition 2 characterizes such an equilibrium. The proposition describes how the manager issues his earnings reports and how the market responds and values the firm.

Proposition 2 *Given that the owner-manager has restricted discretion in issuing the earnings report, the following hold:*

¹⁵In a fully separating equilibrium, earnings may be misreported but are perfectly inferred by the market. This result is different from that in Guttman, Kadan, and Kandel (2006), where a fully separating equilibrium coexists with multiple partially pooling equilibria.

(i) Suppose $\sigma > \frac{8\beta}{\alpha + \frac{1}{2}}$. If ω satisfies the following condition A1,

$$A1 \left\{ \begin{array}{l} \omega < \min\{r_1, \frac{1}{2}(r_2 - r_1)\}, \\ E[P(e_1) | r_1 - \omega \leq h \leq r_1 + \omega] \geq \max\{P(\mu_1 + r_1 - \omega), P(\mu_1 + r_1 + \omega)\}, \text{ and} \\ E[P(e_1) | -r_2 - \omega \leq h \leq -r_2 + \omega] \geq \max\{P(\mu_1 - r_2 - \omega), P(\mu_1 - r_2 + \omega)\}, \end{array} \right.$$

where $h = e_1 - \mu_1$, then there exists a partially pooling equilibrium where the manager's reporting strategy is

$$m = M^*(e_1) = \left\{ \begin{array}{l} e_1 - \omega, \text{ if } -r_2 + \omega < h < -r_1 \text{ or } r_1 + \omega < h < r_2 \\ e_1 + \omega, \text{ if } h < -r_2 - \omega, -r_1 \leq h < r_1 - \omega, \text{ or } h \geq r_2 \\ \mu_1 - r_2, \text{ if } -r_2 - \omega \leq h \leq -r_2 + \omega \\ \mu_1 + r_1, \text{ if } r_1 - \omega \leq h \leq r_1 + \omega, \end{array} \right.$$

and the market values the firm at¹⁶

$$p = P^*(m) = \left\{ \begin{array}{l} P(m + \omega), \text{ if } -r_2 < k < -r_1 - \omega \text{ or } r_1 < k < r_2 - \omega \\ P(m - \omega), \text{ if } k < -r_2, -r_1 + \omega \leq k < r_1, \text{ or } k \geq r_2 + \omega \\ E[P(e_1) | -r_2 - \omega \leq h \leq -r_2 + \omega], \text{ if } k = -r_2 \\ E[P(e_1) | r_1 - \omega \leq h \leq r_1 + \omega], \quad \text{if } k = r_1, \end{array} \right.$$

where $k = m - \mu_1$, with the following off-equilibrium pricing rule:

$$p = P^*(m) = \left\{ \begin{array}{l} P(\mu_1 - r_1), \text{ if } -r_1 - \omega \leq k < -r_1 + \omega \\ P(\mu_1 + r_2), \text{ if } r_2 - \omega \leq k < r_2 + \omega. \end{array} \right.$$

(ii) Suppose $\sigma < \frac{-\beta}{\alpha + \frac{1}{2}}$. If ω satisfies the following condition A2,

$$A2 \left\{ \begin{array}{l} \omega < r_2, \text{ and} \\ E[P(e_1) | -r_2 - \omega \leq h \leq -r_2 + \omega] \geq \max\{P(\mu_1 - r_2 - \omega), P(\mu_1 - r_2 + \omega)\}, \end{array} \right.$$

¹⁶Notice that the equilibrium pricing function $P^*(\cdot)$ is expressed using $P(e_1)$, the pricing function (of true earnings) in the benchmark case as defined in equation (5).

then there exists a partially pooling equilibrium where the manager's reporting strategy is

$$m = M^*(e_1) = \begin{cases} e_1 - \omega, & \text{if } -r_2 + \omega < h < r_2 \\ e_1 + \omega, & \text{if } h < -r_2 - \omega \text{ or } h \geq r_2 \\ \mu_1 - r_2, & \text{if } -r_2 - \omega \leq h \leq -r_2 + \omega, \end{cases}$$

and the market values the firm at

$$p = P^*(m) = \begin{cases} P(m + \omega), & \text{if } -r_2 < k < r_2 - \omega \\ P(m - \omega), & \text{if } k < -r_2 \text{ or } k \geq r_2 + \omega \\ E[P(e_1) \mid -r_2 - \omega \leq h \leq -r_2 + \omega], & \text{if } k = -r_2, \end{cases}$$

with the following off-equilibrium pricing rule:

$$p = P^*(m) = P(\mu_1 + r_2), \text{ if } r_2 - \omega \leq k < r_2 + \omega.$$

(iii) Suppose $\frac{-\beta}{\alpha + \frac{1}{2}} \leq \sigma \leq \frac{8\beta}{\alpha + \frac{1}{2}}$. There exists a fully separating equilibrium where the manager's reporting strategy is $m = M^*(e_1) = e_1 + \omega$, and the market values the firm at $p = P^*(m) = P(m - \omega)$.

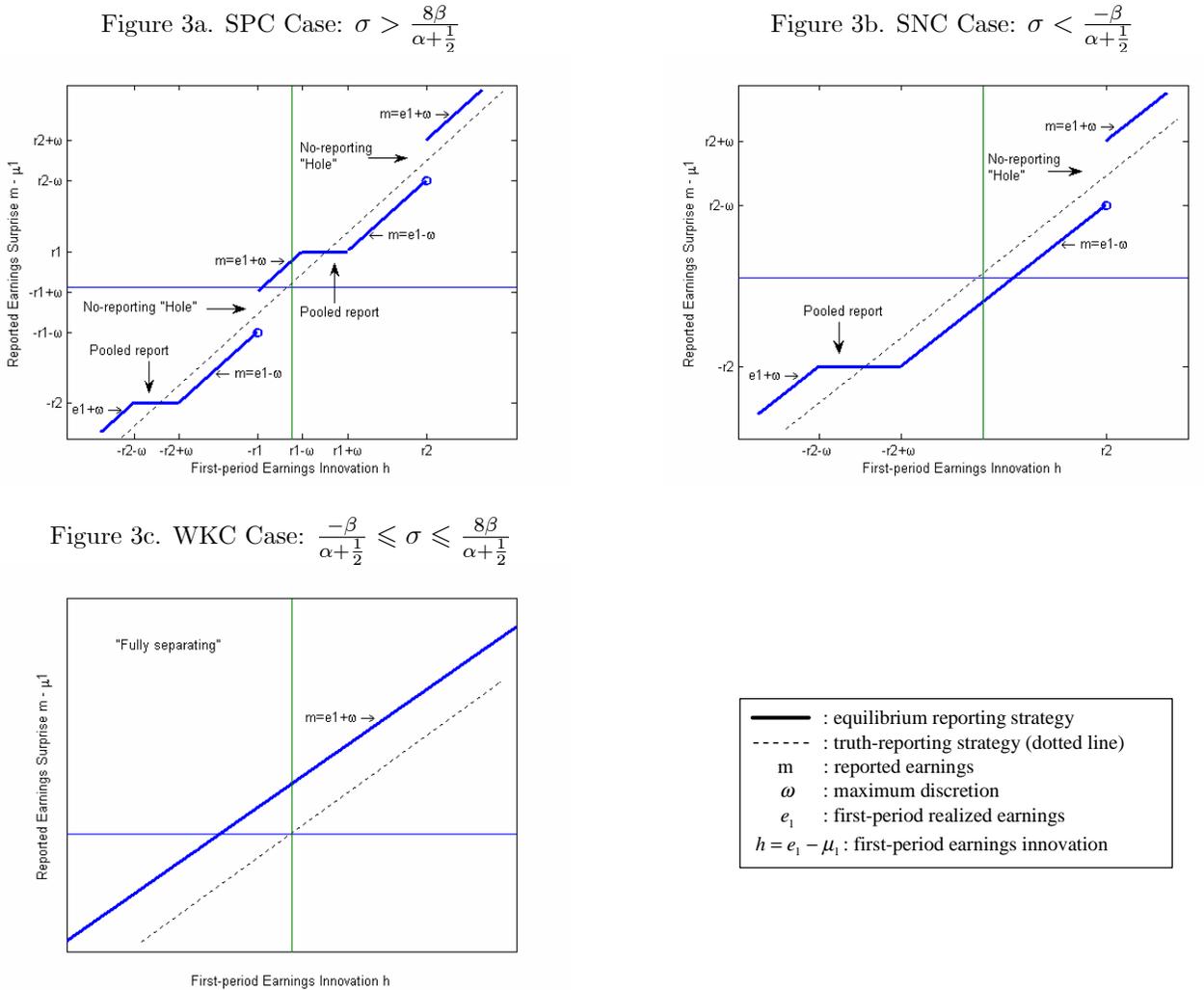
Proof. (The proof is in the Appendix.) ■

The three cases in Proposition 2 correspond to the three cases in Lemma 2. The structure of each specific equilibrium depends on the earnings correlation or the shape of the corresponding pricing function $P(e_1)$. Figures 3a, 3b, and 3c depict the manager's equilibrium reporting strategies for SPC, SNC and WKC Cases respectively with the dotted line describing the truth-reporting strategy.

First, when earnings are strongly positively correlated (SPC Case), the equilibrium involves both pooling and separating strategies from the manager. As Figure 3a shows, pooling of reports occurs over two intervals, each centered at one of the two local maximum points of $P(e_1)$ ($\mu_1 - r_2$ and $\mu_1 + r_1$). Intuitively, firms with true earnings around the local maximum point have incentives to mimic the firm right at the local maximum point. For example, firms to the left of $\mu_1 + r_1$ wish to *increase their inferred earnings innovation* h , whereas firms to the right wish to *increase*

their inferred earnings precision. Observing either of the two pooled reports ($m = \mu_1 - r_2$ or $\mu_1 + r_1$), investors cannot infer the true earnings e_1 and have to price the firm at an expected value conditional on e_1 being in the corresponding interval. The reason the manager pools at the local maximum points is that, given condition *A1*, the expected value is higher than the price he can get by deviating to any other feasible report.¹⁷

Figure 3. The Manager's Equilibrium Reporting Strategy



¹⁷Condition *A1* is for tractability (to facilitate the following comparative statics analysis in Section 4). The part of *A1* that appears difficult to relax is " $\omega < \min\{r_1, \frac{1}{2}(r_2 - r_1)\}$." The assumptions on the pricing function seem less essential in that even if the assumptions do not hold, we can always find an earnings interval $[a - \omega, a + \omega]$ including either $\mu_1 + r_1$ or $\mu_1 - r_2$ such that $E[P(e_1) | a - \omega \leq e_1 \leq a + \omega] \geq \max\{P(a - \omega), P(a + \omega)\}$. Thus, there still exists a similar equilibrium with reported earnings pooled at a instead of $\mu_1 + r_1$ or $\mu_1 - r_2$. A similar argument applies also to condition *A2* in SNC Case.

In addition to the two pooled reports, the manager’s reporting strategy also creates two no-reporting intervals in which no earnings reports exist. The manager’s reporting strategy jumps from under-reporting the minimum earnings (i.e., $m = e_1 - \omega$) to over-reporting the maximum earnings (i.e., $m = e_1 + \omega$) right at the two local minimum points of $P(e_1)$ ($\mu_1 - r_1$ and $\mu_1 + r_2$), creating two “holes” centered at these two points. The (assumed) off-equilibrium belief on the two no-reporting intervals/holes is as follows: if investors see a report in either of the two intervals, they would think the true earnings e_1 is at the corresponding local minimum point, and price the firm at the local minimum price.¹⁸

In the remaining separating regions, the manager over-reports the maximum earnings if the observed earnings innovation is relatively large or small in magnitude and under-reports the minimum earnings if the earnings innovation is medium. Under both cases, investors fully infer the true earnings e_1 from the manager’s report m by adjusting for the appropriate discretion ω or $-\omega$.

Second, when earnings are strongly negatively correlated (SNC Case), the equilibrium is still a partially pooling equilibrium. As Figure 3b shows, the manager pools his reports over an interval centered at the local maximum point of $P(e_1)$ ($\mu_1 - r_2$), and no earnings reports exist in an interval centered at the local minimum point of $P(e_1)$ ($\mu_1 + r_2$). In the rest regions, he separates his reports by over-reporting (under-reporting) the maximum (minimum) earnings if the observed earnings innovation is relatively large (small) in magnitude. Although this case is similar to SPC Case in terms of the partially pooling behavior, it differs in two aspects. First, only one pooled report and one no-reporting hole exist in this case rather than two of each as in SPC Case. Second, the pooled report (the no-reporting hole) is below (above) the prior mean μ_1 in this case, opposite to SPC Case.

Third and last, when earnings are weakly (either positively or negatively) correlated (WKC Case), there exists a fully separating equilibrium. The manager always over-reports the maximum earnings in equilibrium, and the market fully unravels the true earnings e_1 by deducting ω from the reported earnings m . Neither pooled reports nor no-reporting holes exist in this case, and the distribution of reported earnings is continuous and smooth.

As a summary, the key intuition for this proposition rests on the information asymmetry of

¹⁸This off-equilibrium belief specification supports the equilibrium as a sequential equilibrium as well as a Perfect Bayesian equilibrium (Fudenberg and Tirole 1991).

the earnings signal and the uncertainty of its precision, which underlie the opposing (positive and negative) effects on the market price, as elaborated in Section 2. When the positive effect from over-reporting (under-reporting) outweighs the concomitant negative effect, the manager has an incentive to over-report (under-report) earnings. The manager’s pooling behavior is a result of the opposing misreporting incentives at the two sides of a local maximum point of $P(e_1)$. Specifically, the manager has an incentive to over-report earnings at the left side of the local maximum point but an opposing incentive to under-report at the right side, leading to a pooled report at the local maximum point. Similarly, the incentive to under-report earnings at the left side of a local minimum point combined with the opposing incentive to over-report at the right side creates a no-reporting hole around the local minimum point.

1.4 The Empirical Properties of the Equilibrium

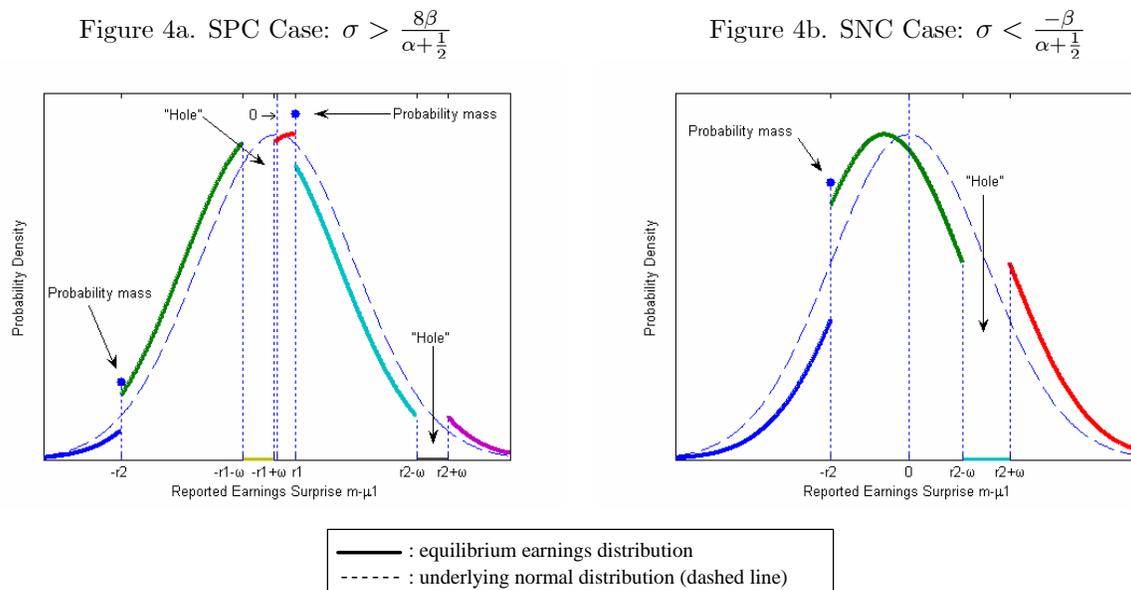
Based on the optimal reporting strategy and the pricing scheme Proposition 2 describes, I now discuss several empirical properties/implications of the partially pooling equilibrium in this section. I first study the distribution of reported earnings implied by the equilibrium (referred to as equilibrium earnings distribution hereafter) and explain how this distribution fits well with the empirical earnings distribution. Next, I conduct comparative statics analysis to show how the location of the probability mass/hole and the degree of discontinuity change as parameter values change. These properties lay the foundations for the subsequent empirical analysis in Section 5 (i.e., hypothesis testing). Last, I elaborate on an “S-shaped” price-reported earnings relationship in equilibrium, which is consistent with existing empirical findings.

1.4.1 The Distribution of Reported Earnings and the Existence of Discontinuity

The first empirically relevant property/implication of the model relates to the existence of discontinuity in the earnings distribution. The underlying (normal) earnings distribution, combined with the manager’s optimal reporting strategy, generates the equilibrium earnings distribution. For WKC Case, since the manager always over-reports as much as possible (uses the maximum discretion ω), the equilibrium distribution is still normal but with a different mean of $\mu_1 + \omega$. For SPC and SNC Cases, most parts of the equilibrium earnings distribution are still continuous

as they are either a shift to the right by ω of the underlying normal distribution or a shift to the left by ω . However, other parts of the equilibrium earnings distribution are no longer continuous due to the existence of pooled reports and no-reporting intervals. Each pooled report produces a probability mass in the earnings distribution, whereas each no-reporting interval creates a (probability density) hole in the distribution. Figures 4a and 4b depict the equilibrium earnings distributions under SPC and SNC Cases respectively with the dashed line describing the underlying normal earnings distribution.¹⁹ In Figure 4a, there are two probability masses at $m = \mu_1 - r_2$ and $\mu_1 + r_1$ and two holes centered at $m = \mu_1 - r_1$ and $\mu_1 + r_2$; whereas in Figure 4b, there exist only one probability mass at $m = \mu_1 - r_2$ and one hole centered at $m = \mu_1 + r_2$.

Figure 4. The Distribution of Reported Earnings in Equilibrium



As mentioned before, the empirical earnings distribution aligned at a threshold (e.g., zero earnings, previous earnings, or analyst forecasts) is discontinuous around the threshold. The frequency of earnings reports in the interval just below (above) the threshold is unusually low (high), creating a visual “dent”/“divot” in the empirical distribution. Below I compare the equilibrium earnings distribution under each case with the empirical earnings distribution and explain how they can be

¹⁹Since the equilibrium earnings distribution under WKC Case is normal, I omit the corresponding graph.

reconciled with each other.

First, for SPC Case, if we think of the prior mean μ_1 as the threshold in the empirical earnings distribution, then the equilibrium earnings distribution around the prior mean fits well with the empirical earnings distribution. We can view the probability mass at $m = \mu_1 + r_1$ as a substantially high frequency of the pooled report right above the prior mean and the no-reporting hole centered at $m = \mu_1 - r_1$ as an extremely low frequency of reports right below the prior mean (assuming r_1 is small). Moreover, this no-reporting hole can be reconciled with the empirically observed “dent”/“divot” on the grounds of cross-sectional firms. The equilibrium earnings distribution describes the distribution of only one single firm with specific parameter values, whereas the empirical earnings distribution results from cross-sectional firms. Since firms generally have different firm-specific parameter values, the location of the no-reporting hole varies across firms. Thus, the holes of different firms do not completely overlap with each other, which would only drag down the frequency of reports below the threshold instead of creating a hole in a cross-sectional earnings distribution (assuming r_1 does not vary much across firms). In this sense, the no-reporting hole is consistent with the “dent”/“divot” we see in the empirical earnings distribution.²⁰

Below we discuss another seeming inconsistency between the two distributions. That is, there exist a second probability mass at $m = \mu_1 - r_2$ and a second no-reporting hole around $m = \mu_1 + r_2$ in the equilibrium earnings distribution that we do not observe in the empirical distribution. A property of normal distribution can help resolve this seeming inconsistency. For normal distribution, the probability density declines quickly away from the mean. Hence, the second probability mass/hole may not be so remarkable due to low densities around $\mu_1 \pm r_2$ (assuming r_2 is relatively large). In addition, the location of the second probability mass/hole also varies across firms and may vary much more than the location of the first one (see Corollary 1 below). Therefore, the second probability mass/hole in the equilibrium earnings distribution seems consistent with not observing the corresponding discontinuities in the empirical earnings distribution.

Corollary 1 below characterizes the difference between r_1 and r_2 in their sensitivities to changes in parameter values.

Corollary 1 *For SPC Case, the following hold:*

²⁰See Guttman, Kadan, and Kandel (2006) for similar arguments with simulation results.

- (1) r_2 is more sensitive to α and σ than r_1 : $\left(\frac{\partial r_1}{\partial \alpha}\right)^2 < \left(\frac{\partial r_2}{\partial \alpha}\right)^2$ and $\left(\frac{\partial r_1}{\partial \sigma}\right)^2 < \left(\frac{\partial r_2}{\partial \sigma}\right)^2$;
- (2) r_2 is less sensitive to β than r_1 : $\left(\frac{\partial r_1}{\partial \beta}\right)^2 > \left(\frac{\partial r_2}{\partial \beta}\right)^2$.

Proof. (The proof is in the Appendix.) ■

r_1 is the distance to the threshold of the first probability mass/hole (the one closer to the threshold), and r_2 is the distance to the threshold of the second probability mass/hole (the one further away from the threshold). Corollary 1 shows that changes in α or σ affect r_2 more than r_1 , and changes in β affect r_1 more than r_2 . If the effect from variations in α and σ across firms is dominant (relative to that from variations in β), the locations of the second probability mass/hole must be less centralized than the locations of the first one. This result provides additional support for not observing the second probability mass/hole in the empirical distribution, because the locations of the second probability mass/hole might be too dispersed across firms to create a significant discontinuity in a cross-sectional earnings distribution aligned at the prior mean/threshold.

Second, for SNC and WKC Cases, both equilibrium earnings distributions are at odds with the empirical distribution. In particular, the equilibrium distribution in WKC Case is smooth and no discontinuity exists. In SNC Case, the equilibrium distribution is not smooth: the only probability mass locates below the threshold (at $m = \mu_1 - r_2$) and the only no-reporting hole locates above the threshold (at $m = \mu_1 + r_2$), opposite to SPC Case. As argued above, since r_2 is likely to be large and vary much across firms, we would not expect to find a significant discontinuity in the empirical distribution of SNC firms that is aligned at the prior mean/threshold.²¹ The reconciliation of these two cases with the empirical findings can also be made on the grounds of cross-sectional firms. If most firms are SPC firms, the effects from SNC and WKC firms would be dominated, and the cross-sectional earnings distribution would exhibit only the characteristics of SPC firms, which is what we observe.

Not only does this section reconcile the theoretical results with the empirical findings, it also indicates an opportunity to test the theory by comparing the empirical earnings distribution of each type of firm with the corresponding equilibrium earnings distribution. I will discuss the details in

²¹A significant discontinuity would exist in a cross-sectional earnings distribution of SNC firms that is aligned at $\mu_1 - r_2$ or $\mu_1 + r_2$ if we could draw such a distribution. However, as argued later, r_2 is not observable or estimable for individual firms. Thus, I choose to focus on the earnings distribution aligned at the prior mean.

Section 5.

1.4.2 The Location of the Probability Mass/Hole and the Degree of Discontinuity

The second empirically relevant property/implication of the model relates to the degree of discontinuity. Here, only SPC Case is of interest because only SPC firms contribute to the discontinuity in the empirical earnings distribution. In SPC Case, the location of the (first) probability mass/hole relative to the prior mean/threshold is dependent on r_1 . The following corollary shows how changes in parameter values affect r_1 .

Corollary 2 *For SPC Case, r_1 is decreasing in σ and α : $\frac{\partial r_1}{\partial \sigma} < 0$ and $\frac{\partial r_1}{\partial \alpha} < 0$, and increasing in β : $\frac{\partial r_1}{\partial \beta} > 0$.*

Proof. (The proof is provided in the Appendix.) ■

Corollary 2 shows that, in SPC Case, the probability mass/hole moves toward the threshold when σ or α increases or β decreases, and vice versa. Given that the expected prior precision of earnings is $\frac{\alpha}{\beta}$, an increase in α or a decrease in β indicates a decrease in the expected volatility of the underlying earnings. Then, the realized volatility of earnings is likely to decrease as well. Roughly speaking, Corollary 2 says that firms with more auto-correlated earnings (higher auto-covariance or lower variance) are associated with a probability mass/hole closer to the threshold.

To understand the result, first recall that the shape of the pricing function $P(e_1)$ results from the trade-off between the two opposing pricing effects induced by any change in e_1 . $\mu_1 \pm r_1$ are where the net negative effect from the expected innovation term is completely offset by the positive effect from the e_1 term (i.e., well balanced). If the parameters change, the balance will be broken as explained below. When σ or α increases, as a coefficient (or part of a coefficient) in the expected innovation term, it directly magnifies the original net negative effect from the expected innovation term. When β decreases, it also increases the net negative effect both directly and indirectly.²²

²²First, as part of the denominator of the expected innovation term, β directly magnifies the original net negative effect as it decreases. Second, a decrease in β indirectly alters the original net negative effect due to the following reason. Notice that the original net negative effect results from the trade-off between the positive effect from h in the numerator and the negative effect from $\beta + \frac{1}{2}h^2$ in the denominator. When β decreases, the positive effect from any increase in h stays the same but the negative effect increases due to a smaller β , which increases the original net negative effect.

Thus, the balance at $\mu_1 \pm r_1$ is broken as the negative effect takes control. A new balance has to be rebuilt at a point closer to the threshold where the positive effect is larger; that is, r_1 has to decrease as α or σ increases or β decreases.

In the following, I study the probability q of true earnings falling into the pooling interval where the manager issues a pooled report. This probability measures the size of the probability mass as well as the size of the no-reporting hole in the earnings distribution. Given a realized precision of earnings, this probability is generally a good measure of how pronounced the discontinuity in the empirical distribution might be: the larger the probability q , the more pronounced the discontinuity. Corollary 3 summarizes how changes in ω and σ affect the probability q .

Corollary 3 *For SPC Case, the following hold:*

- (1) *The width of the pooling interval or the no-reporting interval is 2ω , and the probability q is increasing in ω ;*
- (2) *The probability q is increasing in σ .*

Proof. (A formal proof is omitted.) ■

According to the optimal reporting strategy in SPC Case, the width of each interval is 2ω and, accordingly, the probability q increases as ω increases. Intuitively, with less restriction in earnings manipulation, firms generally manipulate more, and the discontinuity becomes more pronounced.

The covariance σ affects the probability q through its impact on the relative distance of the pooling/no-reporting intervals to the prior mean (i.e., r_1). As Corollary 2 shows, the pooling and the no-reporting intervals move toward the prior mean as σ increases. Thus, the probability q is increasing in σ . In other words, ceteris paribus, the earnings discontinuity is more pronounced for firms with higher earnings auto-covariances.

When α increases or β decreases, the expected volatility of earnings decreases, and the realized volatility is likely to decrease as well. Then, the underlying distribution of true earnings concentrates to the prior mean, and the probability density around the prior mean increases. Thus, q would mechanically increase even if r_1 is held unchanged. This mechanical increase in q would not predict any change in the degree of discontinuity (i.e., more/less pronounced), since the new distribution

approximates to a magnified version of the old distribution around the prior mean if r_1 is held unchanged.²³ However, as Corollary 2 shows, firms with less volatile earnings are associated with smaller r_1 , which also increases the probability q . Unlike the mechanical increase in q , this increase due to smaller r_1 would make the discontinuity more pronounced.

To sum up, the discontinuity in the earnings distribution of firms with more auto-correlated earnings (higher auto-covariance or lower variance) is more pronounced.²⁴ This result is also empirically testable, which I will explore in Section 5.

1.4.3 The Price-Earnings Relationship

The third empirically relevant property of the model relates to the price-reported earnings relationship. Researchers have empirically documented that the price-earnings relationship is nonlinear. For example, Freeman and Tse (1992) and Skinner and Sloan (2002) report evidence that the abnormal stock return and the earnings surprise (relative to analyst forecasts) exhibit an “S-shaped” relationship: the price increases more sharply when the surprise is small in magnitude and much less sharply when the surprise is large. Das and Lev (1994), using a nonparametric method, depict a more detailed graph (i.e., the LWR curve in their Figure 2) that shows not only the overall S-shape but also a negative relationship between the abnormal return and the annual earnings change for medium positive earnings changes as well as some negative earnings changes.

The results from my model fit well with these empirical findings. Figure 2 depicts the graphs of price versus earnings innovation given the earnings reports are truthful. When the manager is allowed to manipulate the reported earnings, Figure 2 no longer represents the price-reported earnings relationship. Based on Proposition 2 as well as Figure 2, Figures 5a and 5b depict the price-reported earnings graphs for SPC and SNC Cases respectively.²⁵ As the graphs show, most parts of Figure 5a (Figure 5b) are a shift to the right by ω or to the left by ω of the corresponding parts in Figure 2a (Figure 2b). To see how the marginal price response (i.e., the slope) changes with the earnings surprise in Figure 5a (Figure 5b), we can resort to Figure 2a (Figure 2b), which is smooth and continuous. Corollary 4 characterizes how the slope in Figures 2a

²³Please refer to Section 5.3 for the empirical measure of the degree of discontinuity, which is a ratio measuring how the distribution is irregular at the prior mean relative to its vicinity.

²⁴On the contrary, the discontinuity is less pronounced and might even *disappear* for firms with less auto-correlated earnings due to larger r_1 .

²⁵The graph for WKC Case is a shift to the right by ω of the graph in Figure 2c and is, thus, omitted.

and 2b changes with the earnings innovation, which applies to Figures 5a and 5b accordingly.

Figure 5. The Pricing Function of Reported Earnings in Equilibrium

Figure 5a. SPC Case: $\sigma > \frac{8\beta}{\alpha + \frac{1}{2}}$

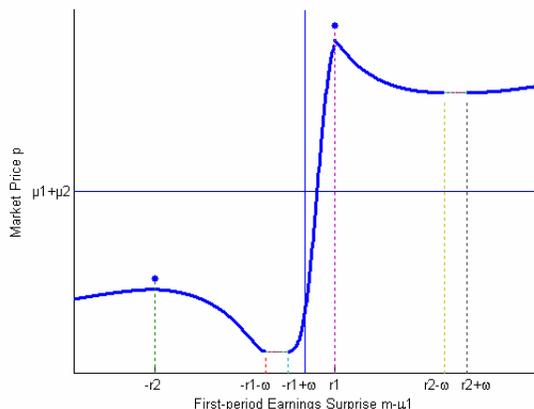


Figure 2 in Das and Lev (1994), p. 362

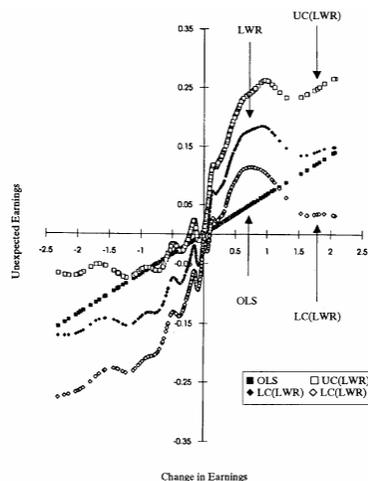
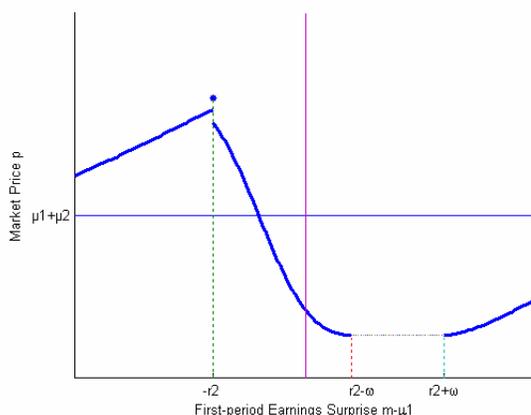


Figure 5b. SNC Case: $\sigma < \frac{-\beta}{\alpha + \frac{1}{2}}$



Corollary 4 *The following hold:*

- (1) *For SPC Case, the first derivative of the pricing function $P(e_1)$ is increasing in e_1 when $-\sqrt{6\beta} < h < 0$ or $h > \sqrt{6\beta}$ and decreasing in e_1 otherwise, where $r_1 < \sqrt{6\beta} < r_2$;*
- (2) *For SNC Case, the first derivative of the pricing function $P(e_1)$ is increasing in e_1 when $h < -\sqrt{6\beta}$ or $0 < h < \sqrt{6\beta}$ and decreasing in e_1 otherwise, where $\sqrt{6\beta} > r_2$.*

The first part of Corollary 4 (based on Figure 2a but applied to Figure 5a accordingly) shows that the marginal price response/slope is decreasing in the magnitude of the earnings innovation when the magnitude is less than r_1 for SPC firms (i.e., an “S-shaped” price-earnings curve). Moreover, Figure 5a fits well with Figure 2 in Das and Lev (1994). In both graphs, the pricing curve on the right side of the zero-earnings surprise is increasing in earnings for small and large positive earnings surprises and decreasing in earnings for medium positive surprises (i.e., a negative price-earnings relationship). Although the left sides are not a perfect match, both exhibit negative slopes in some regions.

The second part of Corollary 4 (based on Figure 2b but applied to Figure 5b accordingly) presents a “reversed S-shaped” price-earnings graph for SNC firms. The reconciliation with the empirical findings is also based on the argument of cross-sectional firms. If most firms are SPC firms, the effects from SNC and WKC firms are dominated.

1.5 The Empirical Results

The above analysis generates a couple of empirically testable predictions. This section is devoted to developing and testing these predictions. The empirical evidence sheds light not only on the earnings discontinuity phenomenon but also on the validity of the theory developed above.

1.5.1 Hypotheses

As argued in Section 4.1, the model suggests the earnings discontinuity is observable for SPC firms but not for SNC or WKC firms. To empirically partition firms into the three types, I need to calculate the cut-off points for SPC, SNC, and WKC firms, which involves estimating α and β . Recall that α and β are parameters of the presumed Gamma distribution of earnings precision. Given the limited number of observations of each firm, estimating α and β with sufficient efficiency is impossible. Hence, I take an indirect approach by partitioning firms into two groups based on the sign (positive or negative) of their earnings auto-covariances. The positive auto-covariance group consists of all SPC firms and part of WKC firms, whereas the negative auto-covariance group consists of all SNC firms and the rest of WKC firms. The model predicts we would be able to observe an earnings discontinuity in the positive group given the dominance of SPC firms but no

earnings discontinuity in the negative group. This prediction leads to the first empirical hypothesis:

H1: *A discontinuity exists in the earnings distribution of firms with positively auto-correlated earnings, whereas no discontinuity exists in the earnings distribution of firms with negatively auto-correlated earnings.*

Section 4.2 predicts that, for SPC firms, the earnings discontinuity is more pronounced for firms with more auto-correlated earnings (higher auto-covariance or lower variance). This prediction leads to the second empirical hypothesis:

H2: *For SPC firms, ceteris paribus, the lower the variance of true earnings or the higher the auto-covariance of true earnings, the more pronounced the earnings discontinuity.*

1.5.2 Sample Selection and Research Design

Although the three empirical thresholds around which the discontinuity occurs (i.e., zero earnings, previous earnings, and analyst forecasts) can all be explained as proxies for the prior mean of the current earnings to some extent, analyst forecasts are the most accurate proxy among them (Brown and Rozeff 1978). To have the best fit with the model, I choose to use analyst consensus forecasts to conduct the empirical tests. My sample consists of all firms with quarterly analyst consensus forecasts and quarterly actual earnings available from First Call between 1990 and 2005.²⁶ The reason to use the actual earnings from First Call instead of the earnings reported by firms (from Compustat) is that the earnings in First Call have been adjusted to reflect the basis on which the majority of the analysts make earnings forecasts (Gu and Chen 2004). Thus, the analyst forecasts serve as the expectations of such actual earnings.

Prior Mean/Expectation: Although consensus forecasts are chosen as the proxy for the prior expectation of earnings, the timing of picking the forecasts is worth some discussion. Analysts often update their forecasts when new information regarding the forecasted earnings arrives. Specifically, after the start of the forecasted period, uncertainty about earnings gets resolved gradually as time elapses, and such new information leads analysts to update their forecasts to reflect the

²⁶First Call's consensus forecasts are calculated using the most recent estimate made by each broker. According to First Call, "New statistics are generated each time a broker begins or ends coverage of a security, revises an estimate, or begins or ends participation in First Call's database" (First Call 1999, p.10).

partial resolution of uncertainty. In this sense, consensus forecasts calculated after the start of the forecasted period comprise both the partially realized earnings and the expectation of the remaining uncertain earnings. These forecasts are no longer expectations of the uncertain earnings in entirety. Thus, I do not use consensus forecasts after the start of the forecasted period.²⁷ Instead, I pick the most current consensus forecast *before* the start of the forecasted period, which is the best proxy for the expectation of the entire uncertain earnings. Operationally, I choose the most current consensus forecast in a time period from 180 days to 90 days before the end of the forecasted period.²⁸ The forecast error is the difference between the First Call actual earnings and this consensus forecast.

Auto-covariance and Variance: To test hypotheses H1 and H2, I need to estimate the auto-covariance/variance of each firm’s true earnings series. However, only the managers’ earnings reports are observable. These reports already incorporate managed components and are not the true earnings per se. If one believes in discretionary accrual models such as the modified Jones model (e.g., Dechow, Sloan, and Sweeney 1995), one can estimate the discretionary accruals and back out the “true/pre-managed earnings.” However, these models have been widely criticized for producing noisy results (e.g., Guay, Kothari, and Watts 1996).²⁹ Thus, I choose to estimate the auto-covariance/variance directly using the First Call’s actual earnings series. Although these estimates may contain measurement error, the error would bias against finding evidence supporting the hypotheses, which I will discuss in the next Section 5.3 where appropriate.

There are a total of 111,360 firm-quarter observations for 6,684 firms in my sample. I first replicate the prior findings (e.g., Figure 6 in Degeorge, Patel, and Zeckhauser 1999) using my sample of First Call data. Figure 6 depicts the frequency distribution of forecast error in 1-penny intervals in a range from -\$0.30 to \$0.30 (notice that the \$0 interval represents observations with

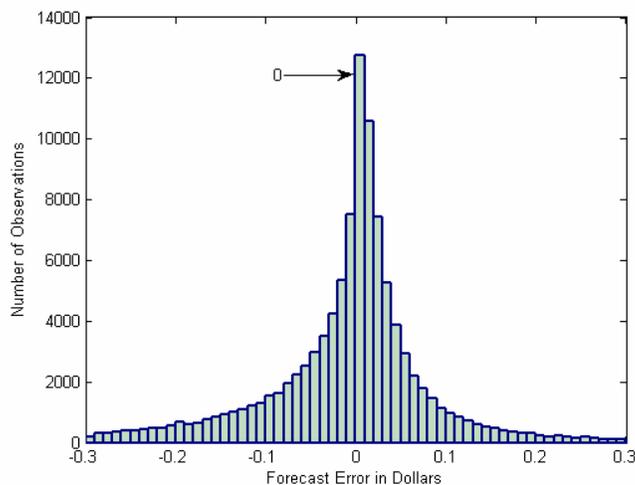
²⁷If we could separate the realized part from the expectation part for these commonly used consensus forecasts, we can conduct the tests by focusing on the remaining uncertain earnings. In fact, the difference between the realization and the expectation of the remaining uncertain earnings coincides with the traditionally defined forecast error (i.e., the entire realized earnings minus the forecast), since the early partial realization incorporated in both the forecast and the entire realized earnings cancels in the calculation. The auto-covariance and variance estimates should be based on the realized values of the remaining uncertain earnings instead of on the entire realized earnings. My conjecture is that all the empirical results would still hold if we could do so.

²⁸In First Call, these forecasts are usually associated with a “Forecast Period” of 2. The “Forecast Period” indicates how far in advance of the fiscal period an estimate is made. A value of “2” means the forecast is made one period ahead of the forecasted period.

²⁹Moreover, since I use First Call’s actual earnings to run the tests, to back out the “*pre-managed actual earnings*,” the estimation of the discretionary accruals has to use the First Call’s actual earnings (instead of the “income before extraordinary items” from Compustat). Given the relatively small coverage of firms by First Call, estimating the discretionary accruals based on First Call firms may bring additional noise into the estimation.

zero forecast errors).³⁰ It shows that the distribution of forecast error drops sharply below zero with the $-\$0.01$ ($-\$0.02$) interval containing 7,504 (5,354) observations, and the $\$0$ ($\$0.01$) interval containing 12,759 (10,592) observations.³¹ This observation is consistent with prior findings that there are much fewer observations to the left of zero forecast error than to the right.

Figure 6. The Empirical Distribution of Forecast Error



This figure depicts the frequency distribution of forecast error in 1-penny intervals in a range from $-\$0.30$ to $\$0.30$. The sample consists of all firms with quarterly analyst consensus forecasts and quarterly actual earnings available from First Call between 1990 and 2005 (fiscal year). There are a total of 111,360 firm-quarter observations for 6,684 firms in the sample. 7.72% (8,597) of the observations lie outside the range $[-\$0.30, \$0.30]$.

Following Burgstahler and Dichev (1997) and Beaver, McNichols, and Nelson (2007), I construct a statistic to test the existence of earnings discontinuity under the assumption that the expected percentage of observations in the $-\$0.01$ interval is the average of the actual percentages of observations in the two adjacent intervals, namely, the $-\$0.02$ and $\$0$ intervals. The test statistic is the difference between the actual and expected percentages of observations in the $-\$0.01$ interval

³⁰7.72% (8,597) of the observations lie outside the range $[-\$0.30, \$0.30]$. Although part of the observations are not graphed for presentation purpose, all observations are used to calculate the test statistics.

³¹In addition, the graph also shows that the distribution has a fat left tail and a thin right tail, which is consistent with the fact that the mean forecast error is $-\$0.138$, while the median forecast error is $\$0$.

divided by the estimated standard deviation of the difference.³² If the distribution is smooth, this standardized difference will be distributed approximately Normal with mean 0 and standard deviation 1. In Figure 6, the standardized difference at the -\$0.01 interval is -12.54 (significant at the 0.01 level), consistent with prior findings.

1.5.3 Empirical Results

A. Test of H1

To test H1, I divide the sample firms into two groups based on the sign (positive or negative) of their earnings auto-covariances. In general, firms need to have enough consecutive earnings in order to get relatively accurate estimates of their earnings auto-covariances. However, requiring more consecutive earnings would bring survivorship bias and reduce the sample size, which may reduce the power of the test. To address this concern, I separately conduct the test for subsamples that require different minimum numbers of consecutive earnings. In particular, I require firms to have at least 8, 12, 16, 20, 24, 28, or 32 consecutive quarterly earnings (corresponding to 2, 3, 4, 5, 6, 7, or 8 years) to be included into Subsamples (1), (2), (3), (4), (5), (6), or (7) respectively.³³ The minimum requirement of two years' consecutive earnings is to assure a minimum level of accuracy of estimation. The maximum requirement of eight years' consecutive earnings is to confine the survivorship bias and to assure there are enough firms to conduct the test. For each subsample, firms are first divided into the positive and negative auto-covariance groups. Then, for each group, the standardized difference at the -\$0.01 interval is calculated to test whether the earnings discontinuity is significant. Hypothesis H1 predicts the earnings discontinuity would not be significant for the negative group.

As mentioned above, using reported earnings to estimate the auto-covariance of true earnings induces estimation error. If enough SPC firms that belong to the positive group are, thus, misclassified into the negative group, the earnings distribution of the negative group may exhibit a

³²Operationally, let n , n_1 , n_2 , and n_3 denote the total number of observations and the number of observations in the -\$0.02, -\$0.01 and \$0 intervals respectively, and let $p = \frac{n_2}{n}$ and $p_0 = \frac{n_1+n_3}{2n}$. Notice that p and p_0 represent the actual and expected percentages of observations falling into the -\$0.01 interval respectively. The test statistic is

$$\frac{(p - p_0)}{\sqrt{[(p_0(1 - p_0) + p(1 - p))/n]}}$$

³³For each firm, all available consecutive earnings are used to estimate the auto-covariance. Thus, the subsamples are overlapping. For example, firms with at least 32 consecutive earnings are included in every subsample.

significant discontinuity, which works against H1.³⁴ Even so, I show below that the evidence is consistent with the prediction.

[Table 1 about here]

Table 1 tabulates the main results. First, the total number of firms in each subsample decreases from 2,992 to 399 when the minimum number of consecutive earnings increases from 8 to 32 (see Column (1)). The requirement of more consecutive earnings increasingly reduces the sample size. Second, firms with negative auto-covariances are generally rare with the percentage of firms in the negative group ranging from 14.9 percent to 1.8 percent. The percentage is decreasing in the minimum number of consecutive earnings. This phenomenon is likely due to firm misclassification, since firms that belong to the positive group are less likely to be misclassified into the negative group if more consecutive earnings are available and, thus, the estimation is more accurate.

More importantly, from Column (6), I find the earnings discontinuity in the positive group is consistently significant for all subsamples. However, for six out of the seven subsamples, the earnings discontinuity in the negative group is not significant with the standardized difference ranging from 0.06 to -1.22. The only exception is Subsample (1) with the shortest consecutive earnings requirement (i.e., two years). In this subsample, the standardized difference in the negative group is -2.07, significant at the 0.05 level.³⁵ This significant discontinuity is likely due to firm misclassification resulting from the smallest number of consecutive earnings required. As more consecutive earnings are required in other subsamples, the incidence of misclassification becomes smaller and the significant earnings discontinuity disappears.

To sum up, the empirical evidence supports the prediction that no discontinuity exists in the earnings distribution of SNC/WKC firms, namely, firms with negatively or weakly positively auto-correlated earnings.

B. Test of H2

To test H2, I use the positive group of firms in the above Subsample (4), which requires at least five years' consecutive quarterly earnings. The choice of Subsample (4) is due to the trade-off between a higher estimation accuracy and a larger sample size. I use the positive group of

³⁴Misclassifying WKC firms with positive auto-covariances into the negative group does not affect the empirical test since these firms are not expected to have earnings discontinuity, similar to the firms with negative auto-covariances.

³⁵Even though the discontinuity in the negative group is significant in Subsample (1), the discontinuity in the positive group is much more pronounced with a standardized difference of -11.25.

firms because hypothesis H2 is based on SPC firms. Given that entirely separating SPC firms is impossible, the positive group serves as a good proxy due to the dominance of SPC firms in this group.

Hypothesis H2 is a prediction regarding the auto-covariance and variance effects on the degree of earnings discontinuity. To test H2, I first construct a statistic to measure the *degree* of earnings discontinuity. At the -\$0.01 interval, this degree of discontinuity measure is the difference between the actual and expected percentages of observations falling into the -\$0.01 interval, divided by the expected percentage of observations in this interval.³⁶ The deflation is intended to make this degree (of discontinuity) measure comparable across groups/firms with different earnings variances. For example, a 5 percent difference between the expected and actual percentages of observations in the -\$0.01 interval does not indicate the same degree of discontinuity between a distribution with an expected 40 percent of observations in the -\$0.01 interval and a distribution with an expected 10 percent in the interval. In fact, the discontinuity in the former distribution is less pronounced (with a smaller degree measure of 12.5 percent) than in the latter one (with a larger degree measure of 50 percent).

The auto-covariance of earnings is usually positively correlated with the variance of earnings. That is, firms with high (low) auto-covariances also tend to have high (low) variances. For all sample firms in this test, the Spearman rank (Pearson) correlation coefficient between the auto-covariance and variance is 0.9 (0.23). Thus, to test the auto-covariance effect on the degree of discontinuity, it is necessary to control for the concomitant variance effect, and vice versa. Accordingly, I first divide the sample into variance deciles such that, within each decile, firms have comparable variances. Next, I divide firms in each variance decile equally into two groups based on the ranking of their earnings auto-covariances, that is, the high auto-covariance group and the low auto-covariance group. Then, the degree (of discontinuity) measure at the -\$0.01 interval is calculated for each group. Finally, a comparison between the degree measures of the two groups in each variance decile is conducted to test H2. According to H2, we expect the high auto-covariance group to have a higher degree of discontinuity than the low auto-covariance group in each variance decile. A similar technique is also used to test the variance effect after controlling for the auto-covariance effect.

³⁶ Following the notation in footnote (32), the degree of discontinuity measure is $\frac{p-p_0}{p_0}$. The larger the magnitude of the measure, the more pronounced the discontinuity.

The above test technique is dependent on the relative ranking of firms' earnings variances/auto-covariances, but not on their true values. If all firms manipulate their earnings in similar ways, although there exist errors in the variance/auto-covariance estimates, the relative ranking of variances/auto-covariances may still be preserved. In this sense, the estimation errors would not bias the empirical test.

[Table 2 about here]

Table 2 presents the number of firms in each auto-covariance and variance decile combination. Clearly, firms cluster on the diagonal line, suggesting the auto-covariance and variance move in the same direction in general. This observation is consistent with the high Spearman rank correlation between them. A potential problem associated with this high correlation is that there may not be enough auto-covariance (variance) variation in each variance (auto-covariance) decile, which makes finding evidence supporting H2 difficult even if H2 is valid. However, even under such stringent conditions, I show below that the evidence still supports the hypothesis.

[Table 3 about here]

Panel A of Table 3 provides the test results for the auto-covariance effect within variance deciles. Specifically, rows (6) and (7) report the degree (of discontinuity) measures at the -\$0.01 interval for the low and high auto-covariance groups respectively, and row (8) reports the difference. Most notably, in every decile, the degree measure is more negative (i.e., larger in magnitude) in the high auto-covariance group than in the low auto-covariance group. The difference is significantly less than 0 with a Wilcoxon signed rank statistic of -27.5 (row (8)). These results suggest the earnings discontinuity in the high auto-covariance group is more pronounced than in the low auto-covariance group, supporting H2. Similarly, Panel B of Table 3 tabulates the test results for the variance effect within auto-covariance deciles. A quick glance at row (8) reveals that, in seven out of the 10 deciles, the degree measure is less negative (i.e., smaller in magnitude) in the high variance group than in the low variance group. The difference is significantly greater than 0 with a Wilcoxon signed rank statistic of 21.5 (row (8)). These results indicate the earnings discontinuity in the high variance group is less pronounced than in the low variance group, as H2 predicts.

Rows (2), (3), and (4) in Panel A (Panel B) report the test results for the existence of earnings discontinuity in each variance (auto-covariance) decile group as well as the low and high auto-covariance (variance) groups within each variance (auto-covariance) decile respectively.

Notice that the discontinuity is consistently significant for all high auto-covariance groups (row (4) in Panel A) and all low variance groups (row (3) in Panel B). However, row (3) in Panel A (row (4) in Panel B) shows that only three out of the 10 low auto-covariance groups (two out of the 10 high variance groups) are associated with significant earnings discontinuities. There are two possible explanations for this result. First, due to smaller auto-covariance (higher variance), r_1 of most SPC firms in the low auto-covariance (high variance) groups becomes larger, and the discontinuity disappears in a cross-sectional earnings distribution as argued in Sections 5.1 and 5.2. This explanation is the premise underlying H2. A second explanation, however, is that WKC firms in each variance (auto-covariance) decile could fall more into the low auto-covariance (high variance) group, leading to the disappearance of discontinuity.³⁷ Below I provide additional tests on whether the second explanation rather than the first one (i.e., H2) is driving the results.

WKC firms are generally firms with low auto-covariances or high variances (see footnote (37)). To possibly eliminate WKC firms from each decile, a reasonable choice is to remove the firms with small auto-covariance-to-variance ratios. Thus, I rank firms in each decile based on their auto-covariance-to-variance ratios, delete those in the lowest 10 percent of the ratios, and then redo the above tests for the remaining firms in each decile.³⁸ Panels A and B of Table 4 report the test results for the auto-covariance and variance effects respectively.

[Table 4 about here]

Table 4 indicates that there is little impact from WKC firms. Specifically, for the low auto-covariance groups (row (3) in Panel A), again, only three out of the 10 are associated with significant earnings discontinuities. For the high variance groups (row (4) in Panel B), only four out of the 10 are associated with significant earnings discontinuities (2 more compared to Table 3). Thus, the impact from WKC firms appears weak, and the nonexistence of earnings discontinuity is more likely due to higher value of r_1 .

Additional results in Table 4 still support hypothesis H2. In Panel A, for nine out of the 10 deciles, the degree (of discontinuity) measure is more negative in the high auto-covariance groups

³⁷Recall that the positive group consists of all SPC firms and part of WKC firms. The partition criterion for SPC and WKC firms in the positive group is: SPC firms are firms with auto-covariances greater than $\frac{8\beta}{\alpha+1/2}$ and the rest are WKC firms. Since the expected earnings precision is $\frac{\alpha}{\beta}$, roughly speaking, high $\frac{8\beta}{\alpha+1/2}$ implies high variance. Thus, in general, WKC firms are firms with relatively low auto-covariances or high variances.

³⁸The choice of 10% is arbitrary. However, given the dominance of SPC firms in the sample, 10% seems to be a reasonable choice. To be more conservative, I have redone the tests using 25% in place of 10%. Most results are qualitatively similar.

than in the low auto-covariance groups. The difference is significant at the 0.01 level (row (8)). In Panel B, for seven out of the 10 deciles, the degree measure is less negative in the high variance groups than in the low variance groups. The difference is also significant at the 0.05 level (row (8)).

To sum up, the above empirical evidence indicates that the earnings discontinuity in the high auto-covariance (low variance) group is more pronounced than in the low auto-covariance (high variance) group, consistent with H2.³⁹

1.6 Conclusion

This paper develops a rational model of financial reporting in which investors use reported earnings not only to infer true earnings but also to update their beliefs about the precision of earnings. In the model, inflating reported earnings has two opposing effects on the market price. For instance, when earnings are positively correlated and the first-period earnings innovation is positive, a higher value of reported earnings leads to a higher level of inferred earnings, potentially increasing the market price. However, the higher earnings surprise also reduces the investors' perceived earnings precision, dampening the positive pricing effect of over-reporting. The manager's optimal reporting strategy depends on the trade-off between the two opposing effects. This paper shows that, for firms whose earnings are strongly positively auto-correlated, the manager's optimal reporting strategy leads to a discontinuity around the prior mean in the distribution of reported earnings. The pricing function of reported earnings exhibits an overall "S-shape" and a negative slope for medium (positive and negative) earnings surprises. These theoretical results are consistent with empirical findings in prior literature. However, for firms whose earnings are negatively or weakly positively auto-correlated, the model predicts no discontinuity in their cross-sectional earnings distribution. In six out of the seven subsamples, no significant discontinuity is found for the negative auto-covariance group, consistent with the theoretical prediction. In addition, the model also predicts that, for firms whose earnings are strongly positively auto-correlated, the lower the variance of earnings or the higher the auto-covariance of earnings, the more pronounced the discontinuity. Empirical results also support the prediction.

However, people may have other alternative explanations for these empirical results although

³⁹In addition to testing the empirical predictions from the theory, Appendix B provides a relatively direct assessment of the fit of the theory with the data.

they are consistent with my theory. What other explanations could be is still an open question and worth exploring.

My model is simple but illustrates a trade-off capable of explaining the already documented discontinuity in earnings distributions and provides new empirical predictions that are supported by the data. Future works may benefit from analyzing models with multiple earnings management vehicles, a trade-off between the manager's long-term and short-term incentives, and/or endogenous analyst forecasts. Analysts may strategically bias their ex ante forecasts (e.g., make them easy to beat) in order to affect the manager's pooling behavior and make the forecasts seem more accurate ex post.

Appendix A

Proof. (of Lemma 2)

From the first derivative expression (6) of the pricing function, the denominator is always positive. Thus, the sign of the first derivative is determined by the numerator, which is a quadratic function of h^2 . Let $f(h^2)$ denote this function, namely,

$$f(h^2) = \frac{1}{4}h^4 + [\beta - \frac{1}{2}\sigma(\alpha + \frac{1}{2})]h^2 + \beta^2 + \sigma(\alpha + \frac{1}{2})\beta.$$

There are two potential roots, R_1 and R_2 , for $f(h^2) = 0$ as long as they are non-negative (notice $h^2 \geq 0$) and valid (i.e., $\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2}) \geq 0$):

$$\begin{aligned} R_1 &= 2[-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2}) - \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}] \\ R_2 &= 2[-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2}) + \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}]. \end{aligned}$$

Notice $R_t = r_t^2, t = 1, 2$. It can be easily shown that

(1) when $\sigma > \frac{8\beta}{\alpha + \frac{1}{2}}$, R_1 & $R_2 > 0$, and we have

$$\begin{cases} f(h^2) \geq 0 & \text{if } h^2 \leq R_1 \text{ or } h^2 \geq R_2 \\ f(h^2) < 0 & \text{if } R_1 < h^2 < R_2 \end{cases} ;$$

(2) when $0 < \sigma < \frac{8\beta}{\alpha + \frac{1}{2}}$, both R_1 and R_2 are invalid and, thus, $f(h^2) > 0$ for any h^2 . When

$$\sigma = \frac{8\beta}{\alpha + \frac{1}{2}}, R_1 = R_2 > 0, \text{ and } f(h^2) \geq 0 \text{ with } 0 \text{ achieved only at } h^2 = R_1;$$

(3) when $\frac{-\beta}{\alpha + \frac{1}{2}} < \sigma \leq 0$, $R_1 < 0$ & $R_2 < 0$ and, thus, $f(h^2) > 0$ for any h^2 . When $\frac{-\beta}{\alpha + \frac{1}{2}} = \sigma$, $R_1 < 0$

$$\text{\& } R_2 = 0, \text{ and } f(h^2) \geq 0 \text{ with } 0 \text{ achieved only at } h^2 = R_2 = 0;$$

(4) when $\sigma < \frac{-\beta}{\alpha + \frac{1}{2}}$, $R_1 < 0$ & $R_2 > 0$, and we have

$$\begin{cases} f(h^2) \geq 0 & \text{if } h^2 \geq R_2 \\ f(h^2) < 0 & \text{if } h^2 < R_2. \end{cases}$$

Lemma 2 is an equivalent representation of the above results. ■

Proof. (of Proposition 1)

It is proven from a contradiction.

Suppose a fully separating equilibrium exists. Then, the manager's equilibrium reporting strategy $m = M(e_1)$ is fully invertible, that is, $e_1 = M^{-1}(m)$. Thus, the market would value the firm at $P(e_1)$ correctly.

As Lemma 2 shows, the pricing function $P(e_1)$ has at least one local maximum when $\sigma > \frac{8\beta}{\alpha+2}$ or $\sigma < \frac{-\beta}{\alpha+2}$. Let x denote the point at which the pricing function $P(\cdot)$ attains its local maximum. Hence, $\exists \varepsilon > 0$, such that $\forall e_1 \in [x - \varepsilon, x) \cup (x, x + \varepsilon]$, $P(e_1) < P(x)$.

Then, $\forall e_1 \in ([x - \varepsilon, x) \cup (x, x + \varepsilon]) \cap [M(x) - \omega, M(x) + \omega]$ (this set is not empty since $x - \omega \leq M(x) \leq x + \omega$), the manager is able to report $M(x)$ and has an incentive to do so since the market would simply value the firm at a higher price $P(x)$ rather than $P(e_1)$, which is a contradiction.

Hence, no fully separating equilibrium exists. ■

Proof. (of Proposition 2)

The proof for SPC Case is provided below. The proofs for SNC and WKC Cases are similar and omitted.

For SPC Case, we can see that the equilibrium pricing function $P^*(m)$ given in the proposition is the expected total earnings from both periods conditional on the reported earnings m and the manager's optimal reporting strategy $M^*(e_1)$. The off-equilibrium beliefs are set in such a way that the market believes the true earnings e_1 is at the appropriate local minimum point when observing an earnings report not expected in equilibrium.

Next, to complete the proof, I show that the manager does not have incentives to deviate from the equilibrium reporting strategy $M^*(e_1)$ given the market pricing rule $P^*(m)$. Consider the following cases:

- when $h \in (-\infty, -r_2 - \omega)$, we can see that the manager will over-report the maximum earnings (i.e., $e_1 + \omega$) since the equilibrium price $P^*(m)$ is increasing in m for $m - \mu_1 \in (-\infty, -r_2)$. No deviation is beneficial;
- when $h \in [-r_2 - \omega, -r_2 + \omega]$, only $\mu_1 - r_2$ will be reported for any e_1 in this interval. In fact,

given the market pricing rule, any deviation from $\mu_1 - r_2$ would result in a price lower than $P(\mu_1 - r_2 - \omega)$ or $P(\mu_1 - r_2 + \omega)$, which is lower than the price for a report of $\mu_1 - r_2$ (i.e., $E[P(e_1) | -r_2 - \omega \leq h \leq -r_2 + \omega]$) due to condition *A1*. Again, no deviation is beneficial. A similar argument also applies to the case where $h \in [r_1 - \omega, r_1 + \omega]$;

- when $h \in (-r_2 + \omega, -r_1)$, since the equilibrium price $P^*(m)$ is decreasing in m for $m - \mu_1 \in (-r_2, -r_1 - \omega)$ and the market values the firm at the local minimum price for any $m - \mu_1 \in [-r_1 - \omega, -r_1 + \omega]$, under-reporting the minimum earnings (i.e., $e_1 - \omega$) is optimal for the manager. A similar argument also applies to another case where $h \in (r_1 + \omega, r_2)$;
- when $h \in [-r_1, r_1 - \omega)$, since $P^*(m)$ is increasing in m for $m - \mu_1 \in [-r_1 + \omega, r_1)$ and the market values the firm at the local minimum price for any $m - \mu_1 \in [-r_1 - \omega, -r_1 + \omega]$, over-reporting the maximum earnings (i.e., $e_1 + \omega$) is optimal for the manager. A similar argument also applies to the case where $h \in [r_2, +\infty)$.

Hence, the reporting strategy $M^*(e_1)$, the market pricing rule $P^*(m)$, and the off-equilibrium beliefs described in the proposition constitute a partially pooling equilibrium.

■

Proof. (of Corollary 1 and 2) Recall that

$$r_1 = \sqrt{R_1} = \sqrt{2[-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2}) - \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}]}$$

$$r_2 = \sqrt{R_2} = \sqrt{2[-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2}) + \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}]}$$

Taking the first partial derivative of r_1 and r_2 with respect to α respectively yields

$$\frac{\partial r_1}{\partial \alpha} = \frac{1}{\sqrt{R_1}} \cdot \frac{\sigma[\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})} - (\frac{1}{2}\sigma(\alpha + \frac{1}{2}) - 2\beta)]}{2\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}}$$

$$\frac{\partial r_2}{\partial \alpha} = \frac{1}{\sqrt{R_2}} \cdot \frac{\sigma[\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})} + (\frac{1}{2}\sigma(\alpha + \frac{1}{2}) - 2\beta)]}{2\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}}$$

Then, the following holds for SPC Case where $\sigma > \frac{8\beta}{\alpha + \frac{1}{2}}$:

$$\begin{aligned}
& \left(\frac{\partial r_1}{\partial \alpha}\right)^2 < \left(\frac{\partial r_2}{\partial \alpha}\right)^2 \iff \\
& \frac{[\frac{1}{2}\sigma^2(\alpha + \frac{1}{2})^2 - 4\beta\sigma(\alpha + \frac{1}{2}) + 4\beta^2] - (\sigma(\alpha + \frac{1}{2}) - 4\beta)\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}}{[-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2})] - \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}} < \\
& \frac{[\frac{1}{2}\sigma^2(\alpha + \frac{1}{2})^2 - 4\beta\sigma(\alpha + \frac{1}{2}) + 4\beta^2] + (\sigma(\alpha + \frac{1}{2}) - 4\beta)\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}}{[-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2})] + \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}} \iff \\
& -4\beta\sigma(\alpha + \frac{1}{2}) < -3\beta\sigma(\alpha + \frac{1}{2}).
\end{aligned}$$

The last inequality is obvious and, thus, $(\frac{\partial r_1}{\partial \alpha})^2 < (\frac{\partial r_2}{\partial \alpha})^2$.

For SPC Case, the following also holds:

$$\begin{aligned}
& \frac{\partial r_1}{\partial \alpha} < 0 \iff \\
& \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})} < \frac{1}{2}\sigma(\alpha + \frac{1}{2}) - 2\beta \iff \\
& \frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2}) < \frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2}) + 4\beta^2
\end{aligned}$$

Hence, $\frac{\partial r_1}{\partial \alpha} < 0$.

Next, taking the first partial derivative of r_1 and r_2 with respect to σ respectively yields

$$\begin{aligned}
\frac{\partial r_1}{\partial \sigma} &= \frac{1}{\sqrt{R_1}} \cdot \frac{(\alpha + \frac{1}{2})[\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})} - (\frac{1}{2}\sigma(\alpha + \frac{1}{2}) - 2\beta)]}{2\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}} \\
&= \frac{(\alpha + \frac{1}{2})}{\sigma} \cdot \frac{\partial r_1}{\partial \alpha}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial r_2}{\partial \sigma} &= \frac{1}{\sqrt{R_2}} \cdot \frac{(\alpha + \frac{1}{2})[\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})} + (\frac{1}{2}\sigma(\alpha + \frac{1}{2}) - 2\beta)]}{2\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}} \\
&= \frac{(\alpha + \frac{1}{2})}{\sigma} \cdot \frac{\partial r_2}{\partial \alpha}
\end{aligned}$$

Hence, $(\frac{\partial r_1}{\partial \sigma})^2 < (\frac{\partial r_2}{\partial \sigma})^2$ and $\frac{\partial r_1}{\partial \sigma} < 0$ for SPC Case.

Last, taking the first partial derivative of r_1 and r_2 with respect to β respectively yields

$$\frac{\partial r_1}{\partial \beta} = \frac{1}{\sqrt{R_1}} \cdot \frac{-\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})} + \sigma(\alpha + \frac{1}{2})}{\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}}$$

$$\frac{\partial r_2}{\partial \beta} = \frac{1}{\sqrt{R_2}} \cdot \frac{-\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})} - \sigma(\alpha + \frac{1}{2})}{\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}}.$$

Then, for SPC Case, we have $(\frac{\partial r_1}{\partial \beta})^2 > (\frac{\partial r_2}{\partial \beta})^2$ since

$$\begin{aligned} & \left(\frac{\partial r_1}{\partial \beta}\right)^2 > \left(\frac{\partial r_2}{\partial \beta}\right)^2 \iff \\ & \frac{[\frac{5}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})] - 2\sigma(\alpha + \frac{1}{2})\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}}{[-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2})] - \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}} > \\ & \frac{[\frac{5}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})] + 2\sigma(\alpha + \frac{1}{2})\sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}}{[-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2})] + \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})}} \iff \\ & \frac{5}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2}) > [-\beta + \frac{1}{2}\sigma(\alpha + \frac{1}{2})] \cdot 2\sigma(\alpha + \frac{1}{2}) \iff \\ & \frac{5}{4}\sigma^2(\alpha + \frac{1}{2})^2 > \sigma^2(\alpha + \frac{1}{2})^2, \end{aligned}$$

and we have $\frac{\partial r_1}{\partial \beta} > 0$ as well since

$$\begin{aligned} & \frac{\partial r_1}{\partial \beta} > 0 \iff \\ & \sigma(\alpha + \frac{1}{2}) > \sqrt{\frac{1}{4}\sigma^2(\alpha + \frac{1}{2})^2 - 2\beta\sigma(\alpha + \frac{1}{2})} \iff \\ & \frac{3}{4}\sigma^2(\alpha + \frac{1}{2})^2 > -2\beta\sigma(\alpha + \frac{1}{2}). \end{aligned}$$

■

Proof. (of Corollary 4)

Taking derivative of equation (6) with respect to e_1 yields

$$\begin{aligned}
P''(e_1) &= \frac{[h^3 + 2[\beta - \frac{1}{2}\sigma(\alpha + \frac{1}{2})]h](\beta + \frac{1}{2}h^2)^2}{(\beta + \frac{1}{2}h^2)^4} - \\
&\quad \frac{2[\frac{1}{4}h^4 + [\beta - \frac{1}{2}\sigma(\alpha + \frac{1}{2})]h^2 + \beta^2 + \sigma(\alpha + \frac{1}{2})\beta](\beta + \frac{1}{2}h^2)h}{(\beta + \frac{1}{2}h^2)^4} \\
&= \frac{\frac{1}{2}\sigma(\alpha + \frac{1}{2})h(h^2 - 6\beta)}{(\beta + \frac{1}{2}h^2)^3}.
\end{aligned}$$

Hence, for SPC Case where $\sigma > \frac{8\beta}{\alpha + \frac{1}{2}} > 0$, $P''(e_1) > 0$ when $-\sqrt{6\beta} < h < 0$ or $h > \sqrt{6\beta}$, and $P''(e_1) < 0$ when $h < -\sqrt{6\beta}$ or $0 < h < \sqrt{6\beta}$. And we can easily verify $r_1 < \sqrt{6\beta} < r_2$.

For SNC Case where $\sigma < \frac{-\beta}{\alpha + \frac{1}{2}} < 0$, $P''(e_1) > 0$ when $h < -\sqrt{6\beta}$ or $0 < h < \sqrt{6\beta}$, and $P''(e_1) < 0$ when $-\sqrt{6\beta} < h < 0$ or $h > \sqrt{6\beta}$. And we can easily verify $r_2 < \sqrt{6\beta}$. ■

Appendix B

In the main text, the empirical tests are based on the empirical predictions derived from the theory of this paper. In this part, I attempt to give a relatively direct assessment of the fit of the model with the data.

As argued before, we cannot accurately estimate each firm's α or β (parameters of the presumed Gamma distribution of earnings precision) due to limited number of observations. However, we could estimate α and β by treating the entire sample as a single firm. In other words, we could view each firm's (estimated) earnings precision as an independent draw from a $\text{Gamma}(\alpha, \beta)$. After having these estimates, we can run simulation and draw a simulated (reported) earning distribution based on the equilibrium earnings reporting strategy. Then we can compare how well the simulated distribution fits with the empirical distribution.

Operationally, I pick up the positive auto-covariance group of Subsample (4) (including both SPC and WKC firms) to estimate α and β . The estimated values of α and β are 0.59 and 0.0057 respectively and the mean precision is $\frac{\alpha}{\beta}$ or 104.⁴⁰ In order to run simulation, I also fit each firm's earnings auto-covariance σ with a gamma distribution and get an estimated $\text{Gamma}(0.2, 3.5)$. In each round of the simulation, earnings are drawn from a normal distribution with mean 0 and a precision which is independently drawn from the $\text{Gamma}(0.59, 0.0057)$, and σ is independently drawn from the $\text{Gamma}(0.2, 3.5)$. The reported earnings are determined based on the equilibrium reporting strategy for SPC or WKC firms, whichever applicable. Totally, I run 40,000 rounds of simulation. Figure 7a depicts the distribution of reported earnings from the simulation. It seems that the simulated distribution does not fit with the empirical distribution. However, this does not necessarily mean the theory is not valid. Notice that each firm differs in many aspects, such as business nature, operational environment and management style. Treating the whole sample as a single firm inevitably invites estimation errors into α and β , which may lead to the above result.

The next interesting question would be what values of α and β can deliver a satisfactory simulation result. Figure 7b depicts an earnings distribution from a new simulation which differs from the above one by only a set of new values of α and β . In this simulation, α and β are 0.05 and 0.00004 respectively and the mean precision is $\frac{\alpha}{\beta}$ or 1250, which is about 12 times the mean

⁴⁰The 95% confidence intervals for α and β are [0.55, 0.64] and [0.0051, 0.0063] respectively.

precision in the above simulation. From Figure 7b, we can see this distribution fits well with the empirical distribution. In particular, the standardized difference at the $-\$0.01$ interval is -8.07 , very close to the standardized difference of -9.42 from the empirical distribution (refer to Table 1). Thus, it is the firms with high earnings precision/low earnings variance that contribute to the earnings discontinuity, which is consistent with my theory.

Figure 7. The Distribution of Forecast Error from Simulation

Figure 7a. $\alpha = 0.59, \beta = 0.0057$

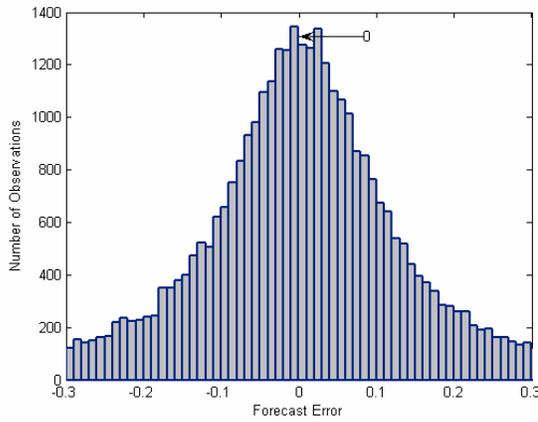
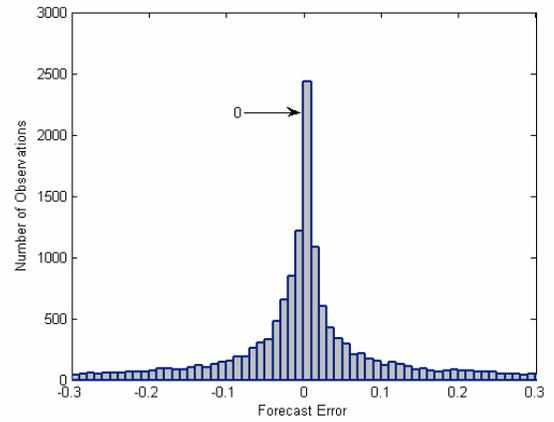


Figure 7b. $\alpha = 0.05, \beta = 0.00004$



Both figures depict the frequency distribution of forecast error from 40,000 rounds of simulation in 1-penny intervals in a range from $-\$0.30$ to $\$0.30$. In each round of the simulation, earnings are drawn from a normal distribution with mean 0 and a precision which is independently drawn from the $\text{Gamma}(\alpha, \beta)$, and σ is independently drawn from the $\text{Gamma}(0.2, 3.5)$.

Table 1
Test for the existence of an earnings discontinuity in the positive and negative auto-covariance groups

Subsamples/ Minimum consecutive earnings requirement ^a		# Of Firms	Total # of observations n	# Of observations in the -\$0.02 Interval n ₁	# Of observations in the -\$0.01 Interval n ₂	# Of observations in the \$0 Interval n ₃	The difference between the actual and expected percentages of observations in the -\$0.01 interval ^c (%)
		(1)	(2)	(3)	(4)	(5)	(6)
(1) 2 years	Positive	2,545 (85.1%) ^b	73,657	3,531	5,219	9,231	-1.58 (-11.25***)
	Negative	447 (14.9%)					11,572
(2) 3 years	Positive	1,910 (89.9%)	62,380	3,027	4,466	8,016	-1.69 (-11.02***)
	Negative	214 (10.1%)					6,629
(3) 4 years	Positive	1,433 (93.6%)	50,705	2,424	3,608	6,509	-1.69 (-9.96***)
	Negative	98 (6.4%)					3,509
(4) 5 years	Positive	1,066 (95.6%)	40,277	1,922	2,878	5,287	-1.80 (-9.42***)
	Negative	49 (4.4%)					1,835
(5) 6 years	Positive	753 (96.4%)	30,546	1,487	2,223	4,069	-1.82 (-8.20***)
	Negative	28 (3.6%)					1,080
(6) 7 years	Positive	548 (97.5%)	23,553	1,151	1,744	3,237	-1.91 (-7.50***)
	Negative	14 (2.5%)					571
(7) 8 years	Positive	392 (98.2%)	17,644	850	1,274	2,387	-1.95 (-6.69***)
	Negative	7 (1.8%)					294

^a Firms are required to have at least 8, 12, 16, 20, 24, 28, or 32 consecutive earnings (corresponding to 2, 3, 4, 5, 6, 7, or 8 years) to be included Subsamples (1), (2), (3), (4), (5), (6), or (7) respectively. In each subsample, firms are divided into the positive and negative auto-covariance groups.

^b I report the percentage of firms in the positive or negative groups of each subsample in parentheses.

^c The statistic to test the earnings discontinuity is the difference between the actual and expected percentages of observations in the -\$0.01 interval, divided by the estimated standard deviation of the difference. Let n, n₁, n₂, and n₃ denote the total number of observations and the number of observations in the -\$0.02, -\$0.01, and \$0 intervals respectively, and let $p = n_2/n$ and $p_0 = (n_1 + n_3)/(2n)$. Notice that p and p₀ represent the actual and expected percentages of observations falling into the -\$0.01 interval respectively. Then, the test statistic (i.e., the standardized difference) is

$$(p - p_0) / \sqrt{((p_0(1 - p_0) + p(1 - p))/n)}$$

If the distribution is smooth, this standardized difference will be distributed approximately Normal with mean 0 and standard deviation 1. The value of this standardized difference is reported in parentheses below the unstandardized difference. ***, **, and * indicate the significance at the 0.01, 0.05, and 0.1 levels respectively, two-sided.

Table 2
The number of firms in each auto-covariance and variance decile combination

# Of firms	Auto-covariance decile									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
1st	75	31	0	0	0	0	0	0	0	0
2nd	12	43	47	3	0	0	1	0	0	0
3rd	9	14	31	49	3	0	0	0	0	1
4th	4	5	9	29	56	3	0	1	0	0
Variance decile	5th	1	4	8	13	24	52	5	0	0
	6th	3	6	4	5	11	28	47	2	0
	7th	1	2	2	4	4	10	38	42	2
	8th	0	1	3	3	4	8	10	44	29
	9th	1	0	3	1	2	4	4	14	55
	10th	0	0	0	0	3	2	2	3	20

Table 3

Test for the auto-covariance and variance effects on the degree of discontinuity in the earnings distribution^a

Panel A: Auto-covariance effect

Variance decile		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	
Total # of observations		(1)	3,647	3,669	4,050	4,026	4,044	4,134	4,034	4,094	4,282	4,269
Difference between the actual and expected percentages of observations in the $-\$0.01$ interval ^b (%)	whole decile	(2)	-3.29 (-4.20***)	-2.45 (-3.34***)	-2.38 (-3.33***)	-2.60 (-3.96***)	-1.47 (-2.36**)	-2.24 (-3.64***)	-2.08 (-3.81***)	-1.23 (-2.43**)	-0.53 (-1.14)	-0.15 (-0.45)
	low auto-cov group	(3)	-3.36 (-2.88***)	-1.34 (-1.33)	-0.78 (-0.81)	-2.05 (-2.19**)	-0.62 (-0.73)	-1.56 (-1.84*)	-0.66 (-0.85)	-0.76 (-1.02)	0.07 (0.10)	0.62 (1.22)
	high auto-cov group	(4)	-3.23 (-3.06***)	-3.59 (-3.38***)	-3.95 (-3.73***)	-3.08 (-3.36***)	-2.31 (-2.54**)	-2.87 (-3.24***)	-3.45 (-4.47***)	-1.73 (-2.54**)	-1.07 (-1.76*)	-1.02 (-2.34**)
Degree of discontinuity measure at the $-\$0.01$ interval ^c (%)	whole decile	(5)	-22.68	-19.87	-18.43	-23.89	-15.76	-23.15	-27.81	-19.84	-10.34	-5.88
	low auto-cov group d_1	(6)	-21.97	-11.90	-7.36	-20.32	-7.55	-18.34	-9.92	-11.43	1.38	22.95
	high auto-cov group d_2	(7)	-23.40	-26.75	-25.88	-26.61	-22.17	-26.68	-41.52	-30.13	-22.02	-41.41
Difference between the degree measures in the two groups, d_2-d_1 (Wilcoxon signed rank statistic) ^d		(8)	-1.43	-14.85	-18.52	-6.29	-14.62	-8.34	-31.60	-18.70	-23.40	-64.36
												(-27.5**)

Panel B: Variance effect

Auto-covariance decile		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	
Total # of observations		(1)	3,615	3,830	3,886	3,987	4,209	4,013	4,175	4,106	4,265	4,163
Difference between the actual and expected percentages of observations in the $-\$0.01$ interval (%)	whole decile	(2)	-2.56 (-3.36***)	-1.83 (-2.56**)	-2.30 (-3.27***)	-1.87 (-2.73***)	-2.03 (-3.31***)	-2.07 (-3.32***)	-2.37 (-4.06***)	-1.85 (-3.78***)	-0.49 (-1.10)	-0.85 (-2.34**)
	low var group	(3)	-4.29 (-3.67***)	-3.03 (-2.85***)	-3.11 (-3.01***)	-3.40 (-3.19***)	-3.09 (-3.30***)	-2.28 (-2.44**)	-2.81 (-3.17***)	-3.06 (-4.09***)	-1.47 (-2.10**)	-1.16 (-1.95*)
	high var group	(4)	-0.96 (-0.97)	-0.58 (-0.62)	-1.54 (-1.60)	-0.26 (-0.31)	-1.02 (-1.28)	-1.85 (-2.26**)	-1.91 (-2.54**)	-0.62 (-0.98)	0.38 (0.66)	-0.54 (-1.30)
Degree of discontinuity measure at the $-\$0.01$ interval (%)	whole decile	(5)	-19.41	-15.32	-19.27	-16.46	-20.88	-21.61	-26.54	-30.28	-10.50	-26.01
	low var group d_3	(6)	-26.85	-21.38	-23.89	-22.46	-26.62	-20.52	-26.32	-39.56	-24.89	-25.93
	high var group d_4	(7)	-9.05	-6.08	-14.02	-3.50	-12.87	-23.23	-26.90	-13.81	10.43	-26.19
Difference between the degree measures in the two groups, d_4-d_3 (Wilcoxon signed rank statistic)		(8)	17.80	15.30	9.87	18.96	13.75	-2.71	-0.58	25.75	35.32	-0.26
												(21.5**)

^a For Panel A, I first divide all firms into variance deciles and, then, divide the firms in each variance decile equally into two groups, i.e., the high auto-covariance group and the low auto-covariance group. For Panel B, I first divide all firms into auto-covariance deciles and, then, divide the firms in each auto-covariance decile equally into two groups, i.e., the high variance group and the low variance group.

^b See Table 1 for the definition of the standardized difference between the actual and expected percentages of observations in the $-\$0.01$ interval. If the distribution is smooth, this standardized difference will be distributed approximately Normal with mean 0 and standard deviation 1. I report the value of this standardized difference in parentheses below the unstandardized difference. ***, **, and * indicate the significance at the 0.01, 0.05, and 0.1 levels respectively, two-sided.

^c The degree of discontinuity measure is the difference between the actual and expected percentages of observations in the $-\$0.01$ interval, divided by the expected percentage of observations in this interval, i.e., $(p-p_e)/(p_e)$ (see Table 1 for the notations). The larger the magnitude of the measure, the more pronounced the discontinuity.

^d One-sided test.

Table 4
Test for the auto-covariance and variance effects on the degree of discontinuity in the earnings distribution after controlling for the possible impact from WKC firms^a

Panel A: Auto-covariance effect

Variance decile		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	
Total # of observations	(1)	3,307	3,341	3,646	3,706	3,708	3,699	3,660	3,709	3,892	3,892	
Difference between the actual and expected percentages of observations in the -\$0.01 interval ^b (%)	whole decile	(2)	-3.63 (-4.38***)	-2.62 (-3.40***)	-2.62 (-3.44***)	-2.59 (-3.78***)	-1.24 (-1.91*)	-2.26 (-3.51***)	-2.32 (-4.06***)	-1.32 (-2.58***)	-0.77 (-1.63)	-0.04 (-0.11)
	low auto-cov group	(3)	-4.77 (-3.81***)	-1.78 (-1.67*)	-0.74 (-0.71)	-1.58 (-1.62)	0.00 (0.00)	-1.59 (-1.85*)	-1.26 (-1.55)	-0.90 (-1.22)	-0.46 (-0.65)	0.87 (1.60)
	high auto-cov group	(4)	-2.62 (-2.38**)	-3.47 (-3.13***)	-4.37 (-3.95***)	-3.48 (-3.63***)	-2.47 (-2.57**)	-2.85 (-3.03***)	-3.34 (-4.14***)	-1.76 (-2.47**)	-1.05 (-1.66*)	-1.08 (-2.40**)
Degree of discontinuity measure at the -\$0.01 interval ^c (%)	whole decile	(5)	-24.34	-20.96	-19.63	-23.70	-13.57	-23.82	-30.69	-22.79	-15.63	-1.51
	low auto-cov group d ₁	(6)	-28.68	-15.23	-6.74	-16.04	0	-20.75	-18.22	-15.32	-9.09	32.14
	high auto-cov group d ₂	(7)	-19.57	-26.08	-28.11	-29.34	-22.89	-25.69	-40.72	-30.77	-21.83	-44.83
Difference between the degree measures in the two groups, d ₂ -d ₁ (Wilcoxon signed rank statistic) ^d	(8)	9.11	-10.85	-21.37	-13.30	-22.89	-4.94	-22.50	-15.45	-12.74	-76.97	(-25.5***)

Panel B: Variance effect

Auto-covariance decile		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	
Total # of observations	(1)	3,223	3,477	3,479	3,626	3,821	3,614	3,800	3,714	3,829	3,735	
Difference between the actual and expected percentages of observations in the -\$0.01 interval (%)	whole decile	(2)	-2.76 (-3.39***)	-1.94 (-2.57**)	-2.73 (-3.65***)	-2.07 (-2.83***)	-2.09 (-3.21***)	-2.41 (-3.62***)	-2.42 (-3.91***)	-1.84 (-3.58***)	-0.78 (-1.65*)	-1.03 (-2.60***)
	low var group	(3)	-4.30 (-3.42***)	-3.00 (-2.66***)	-3.18 (-2.88***)	-3.33 (-2.93***)	-2.95 (-3.02***)	-2.53 (-2.59***)	-2.78 (-2.94***)	-3.05 (-3.83***)	-1.47 (-1.97**)	-1.26 (-1.98**)
	high var group	(4)	-1.32 (-1.26)	-0.87 (-0.86)	-2.31 (-2.28**)	-0.78 (-0.86)	-1.25 (-1.44)	-2.28 (-2.55**)	-2.05 (-2.60***)	-0.60 (-0.93)	-0.15 (-0.25)	-0.79 (-1.72*)
Degree of discontinuity measure at the -\$0.01 interval (%)	whole decile	(5)	-20.32	-15.94	-22.2	-17.36	-21.00	-24.17	-26.51	-30.11	-15.87	-29.06
	low var group d ₃	(6)	-25.97	-21.04	-24.04	-21.63	-25.57	-22.71	-25.42	-38.46	-23.89	-26.97
	high var group d ₄	(7)	-12.22	-8.62	-20.19	-9.33	-14.81	-26.14	-28.21	-14.10	-3.95	-33.33
Difference between the degree measures in the two groups d ₄ -d ₃ (Wilcoxon signed rank statistic)	(8)	13.75	12.42	3.85	12.30	10.76	-3.43	-2.79	24.36	19.94	-6.36	(20.5**)

^a WKC firms are generally firms with low auto-covariances or high variances. To remove WKC firms from each decile, I first rank firms in each decile based on their auto-covariance-to-variance ratios and then delete those in the lowest 10% of the ratios.

^{b,c,d} See Table 3.

2 Essay 2

When Leading Indicator Variables Reduce Long-term Investment

2.1 Introduction

Kaplan and Norton (1992) develop the idea of a “Balanced Scorecard”, which includes not only traditional financial accounting measures but also nonfinancial measures, such as lead time, product quality, and employee skills. Such nonfinancial measures are generally leading indicators of a firm’s long-term profitability. Proponents of leading indicator variables usually argue that they help resolve underinvestment problems by encouraging long-term investments that have a positive net present value but would likely be forgone under more traditional backward-looking performance measures. Critics of leading indicator variables argue that such measures are highly subjective and difficult to incorporate in compensation contracts.

This paper studies whether the common wisdom that leading indicator variables increase long-term investments is correct. We study a two-period agency problem in which contracting is assumed to be short-term (period-by-period) and identify conditions under which leading indicator variables encourage long-term investments and other conditions under which leading indicator variables discourage long-term investment.

Our model is similar to Dutta and Reichelstein (2003) (hereafter DR). In addition to the usual moral hazard problem, the agent has to decide whether to take a long-term investment which decreases the first-period cash flow but increases the second-period cash flow. At the end of the first period, the principal decides whether to retain the existing agent or hire a new agent with a nontrivial replacement cost for the second period.

We find that there is no pure strategy equilibrium but instead a mixed strategy equilibrium. In equilibrium, the agent randomizes between investing and not investing in the first period, while the principal randomizes between retaining the existing agent and hiring a new agent. The intuition is as follows. From the principal’s perspective, if she retains the existing agent, she has to pay an information rent to the agent who invested in the first period and endure a distorted effort from the agent who did not invest in the first period. If the principal hires a

new agent, she has to pay the replacement cost. The retention cost and the replacement cost are equated so that the principal is indifferent between retaining the existing agent and hiring a new one. From the agent's perspective, if she invests in the first period, the information rent she earns in the second period fully offsets her reduced cash flow and compensation in the first period so that she earns exactly two times her single-period reservation utility across the two periods. If she does not invest in the first period, she gets her single-period reservation utility in each period. Thus, the agent is willing to randomize between investing and not investing.

We next introduce a binary leading indicator variable (hereafter LIV) publicly available at the end of the first period, which is unverifiable (i.e., subjective) and thus cannot be contracted on. The LIV is more (less) likely to be high (low) if the agent invests in the first period than if she does not. Hence, the principal can update her beliefs about the agent's investment choice based on the realized LIV.

If the probability of a high LIV given investment is high and/or the net return on investment is small, there exists a unique equilibrium in which the principal retains the existing agent when observing a low LIV and randomizes otherwise. When the principal observes a high LIV, her updated belief about the probability of investment (which is higher than the agent's actual/equilibrium probability of investment) is the same as the agent's equilibrium probability of investment in the no LIV scenario. Thus, the ex ante probability of investment is smaller. The effect from the smaller expected investment dominates other factors, and the principal is worse off with the LIV than without it. Moreover, in this case, the more informative the LIV, the worse off the principal is—the more informative LIV induces an even smaller chance of investment.¹

In contrast, if the probability of a high LIV given investment is low and/or the net return on investment is large, there exists a unique equilibrium in which the principal hires a new agent when observing a high LIV and randomizes otherwise. When the principal observes a low LIV, her updated belief about the probability of investment (which is lower than the agent's actual/equilibrium probability of investment) is the same as the agent's equilibrium

¹The informativeness of the LIV is modeled as the difference between the probability of a high LIV given investment and the probability of a high LIV given no investment. The larger the difference, the more informative the LIV.

probability of investment in the no LIV scenario. Thus, the ex ante probability of investment is higher. The effect from a higher expected investment dominates other factors, and the principal becomes better off. In this case, the more informative the LIV, the better off the principal is—the more informative LIV induces an even higher chance of investment.

In contrast to the mixed results in our paper, DR show contractible leading indicator variables are always helpful in both long-term and short-term contracts.² Our paper differs from DR in two aspects. First, DR assume (for simplicity) that it is not costly to find a new agent to replace the existing agent. As a result, under short-term contracting, there is no equilibrium in which the same agent is retained in the second period. One plausible way to introduce an equilibrium with retention is add a switching cost to hiring a new agent. Such switching costs include but are not limited to severance packages, searching costs, and transition costs, which can be large.³ We assume there is a positive replacement cost in our model. Second, DR assume the LIV is contractible, whereas we assume it is not. The LIVs in our model are subjective measures.

The paper is organized as follows. Section 2 describes the model. Section 3 derives the equilibrium of a benchmark case with no LIVs. Section 4 studies the case with LIVs and characterizes the properties of the equilibrium. Section 5 concludes the paper.

2.2 The Model

In this paper, we study a two-period contracting problem. Both the principal and the agent are risk neutral. The agent is motivated to provide effort a_t , the source of the firm’s operating cash inflow, in both periods $t \in \{1, 2\}$. The cost function of effort $e(a)$ is strictly increasing and convex and has a non-negative third derivative.⁴ In addition, the agent has to make a binary long-term investment decision in the first period, i.e., taking no investments or taking an investment of b . The cash outflow of b is personally costless to the agent and is

²DR show that, under long-term contracts, leading indicator variables help separate the investment problem from the moral-hazard problem, whereas under short-term contracts, leading indicator variables help mitigate the hold-up problem since the agent who makes the long-term investment decision is not retained long enough to benefit from the return on investment.

³For example, in the ouster of HP’s CEO Ms. Fiorina in 2005, she received a severance package valued at about \$21 million (see A8, *Wall Street Journal*, 02/10/2005).

⁴We also require $e'(a) \rightarrow 0 \iff e''(a) \rightarrow 0$.

accompanied with a cash inflow of B in the second period ($B > b$). The net cash flows in the two periods are as follows:

$$\begin{aligned} c_1 &= a_1 - b + \varepsilon_1 \text{ and } c_2 = a_2 + B + \varepsilon_2 \quad \text{if the agent invests, or} \\ c_1 &= a_1 + \varepsilon_1 \text{ and } c_2 = a_2 + \varepsilon_2 \quad \text{if otherwise,} \end{aligned}$$

where ε_1 and ε_2 are independently and normally distributed with mean 0 and variance σ_t , i.e., $\varepsilon_t \sim N(0, \sigma_t)$.

The principal cannot observe the agent's first-period investment choice. Such investments can be thought of as soft investments, such as customer satisfaction improvement and employee training. They are generally embedded in the ordinary operating expenses in the existing reporting system and hard to separate. Further, we assume the principal cannot observe the realized cash flow c_1 until the second-period contract has been signed. Before the second-period contract has been signed, the principal observes only a non-financial performance measure $f \in \{H, L\}$, which is informative about the agent's first-period investment decision.⁵ The probabilistic relationship between this LIV and the investment choice is:

$$\text{prob}(f = H | \text{investment}) = n \text{ and } \text{prob}(f = H | \text{no investment}) = m \quad \text{where } m < n.$$

The principal uses this LIV to update her belief about the agent's investment decision. Due to the subjectiveness of the LIV, we assume it is not verifiable and thus cannot be contracted on.

At the beginning of the second period, the principal is free to contract with the existing agent or a new agent. We assume there is a nontrivial cost of $C(> 0)$ of replacing the existing agent with a new one. Following DR, we also assume linear compensation schemes in this model:

$$s_t = \alpha_t + \beta_t \cdot c_t \quad t \in \{1, 2\}$$

where c_t is the net cash flow in period t . The time line in Figure 1 illustrates the sequence of events.

⁵DR give four real examples of this nonfinancial performance measure, i.e., the measures of customer satisfaction, product quality, on-time delivery, and product awareness.

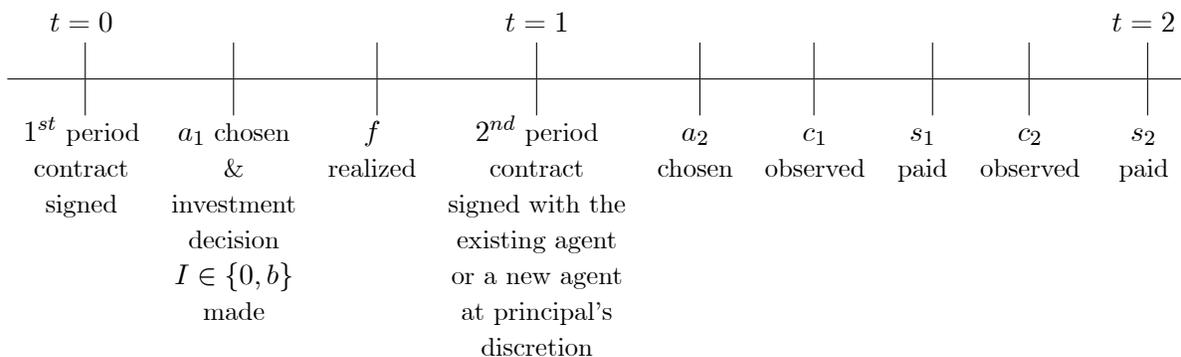


Figure 1. Timeline

2.3 The Scenario with No Leading Indicator Variables

As a benchmark, consider the problem without the LIV. At first, we investigate the possibility of the existence of pure strategy equilibria. Since the first-period cash flow c_1 cannot be observed until the second-period contract has been signed, the principal has two pure strategies in terms of whom to sign the second-period contract with, i.e., “retain the existing agent” and “fire the existing agent/hire a new agent”. Regarding the long-term investment decision in the first period, the agent also has two pure strategies, i.e., “investment of b ” and “no investment.” In the following, we show that no pure strategy equilibria exist.

First, hiring a new agent cannot be an equilibrium strategy. Under the short-term contract setting, both the first-period effort and investment are sunk at the beginning of the second period. Therefore, the principal would be indifferent between contracting with the existing agent and a new agent if there is no replacement cost. However, given the strictly positive replacement cost C , the principal is better off by retaining the existing agent, foregoing the replacement cost.

Second, retaining the existing agent cannot be an equilibrium strategy either. We show this result from a contradiction. Suppose the principal retains the existing agent in equilibrium. Then, the equilibrium must satisfy the following conditions:

- (1) At the beginning of the second period, given the principal’s conjecture about the agent’s first-period investment decision, the second-period contract maximizes the principal’s expected utility subject to the agent’s individual rationality (IR) and incentive compatibility

(IC) constraints. In particular, the conjecture has to be consistent with the actual investment decision made by the agent in the first period; and

- (2) At the beginning of the first period, the agent chooses $(a_1, I \in \{0, b\})$ to maximize her total expected utility under the current contract and the anticipated second-period contract.

Now, suppose taking an investment of b is the agent's equilibrium strategy. Since the agent's second-period expected utility and total expected utility (the first-period plus the second-period) are both held to her reservation utility (i.e., zero) in equilibrium, the agent's first-period expected utility,

$$\alpha_1 + \beta_1 \cdot (a_1 - b) - e(a_1),$$

would also be held to zero in equilibrium. However, the agent could do better by shirking/not investing in the first period and rejecting the second-period contract. In particular, by doing so, the agent can get a more than zero utility (i.e., $\alpha_1 + \beta_1 a_1 - e(a_1) > 0$) as opposed to a zero utility by following the equilibrium strategy. The intuition is that the principal selects a fixed wage α_1 to compensate the agent for her investment. But the agent could take advantage of it by not investing. Thus, investing cannot be an equilibrium strategy for the agent.

Next, if taking no investments is the agent's equilibrium strategy, then the two periods are completely independent. The whole problem simplifies into two separate one-period moral-hazard problems, which leads to same contract coefficients $\{\alpha_t, \beta_t\}$ for both periods (i.e., $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$). Then, the agent could do better by instead investing in the first period, because as long as the net investment return is positive (i.e., $B > b$), she can achieve a more than zero utility by investing,

$$\alpha_1 + \beta_1 \cdot (a_1 - b) - e(a_1) + \alpha_2 + \beta_2 \cdot (a_2 + B) - e(a_2) > 0,$$

as opposed to a zero utility by not investing,

$$\alpha_1 + \beta_1 \cdot a_1 - e(a_1) + \alpha_2 + \beta_2 \cdot a_2 - e(a_2) = 0.$$

Thus, taking no investments cannot be an equilibrium strategy for the agent either.

To sum up, we obtain the following result:

Proposition 1 *There is no pure strategy equilibrium when there is a replacement cost $C(> 0)$ and the investment project is profitable, i.e., $B > b$.*

Hence, a mixed strategy equilibrium is of interest. In any mixed strategy equilibrium, the principal is indifferent between retaining the existing agent and hiring a new one in the second period, and the (existing) agent is indifferent between investing and not investing in the first period.

At the beginning of the second period, the existing agent has private information about her first-period investment choice. If the principal decides to retain the existing agent, by the Revelation Principle she would offer the agent a menu of contracts $\{(\alpha_{21}, \beta_{21}), (\alpha_{22}, \beta_{22})\}$ to motivate the agent to tell the truth. That is, if the agent invests (does not invest) in the first period, she would select the contract $(\alpha_{21}, \beta_{21})$ ($(\alpha_{22}, \beta_{22})$) for the second period. Let p denote the existing agent's probability of investment in the first period. The principal's optimization program is as follows:

$P_{2,old}$ (retaining the existing agent in the second period):

$$U_{2,old}(p) = \max_{\{\alpha_{21}, \beta_{21}, \alpha_{22}, \beta_{22}\}} p(a_{21} + B - (\alpha_{21} + \beta_{21}(a_{21} + B))) + (1 - p)(a_{22} - (\alpha_{22} + \beta_{22}a_{22}))$$

$$\text{s.t.} \quad a_{2i} \in \operatorname{argmax}_{\tilde{a}} \{\beta_{2i}\tilde{a} - e(\tilde{a})\} \quad i = 1, 2 \quad (\text{IC0})$$

$$\alpha_{21} + \beta_{21}(a_{21} + B) - e(a_{21}) \geq 0 \quad (\text{IRI})$$

$$\alpha_{22} + \beta_{22}a_{22} - e(a_{22}) \geq 0 \quad (\text{IRII})$$

$$\begin{aligned} \alpha_{21} + \beta_{21}(a_{21} + B) - e(a_{21}) &\geq \max_{\tilde{a}} \{\alpha_{22} + \beta_{22}(\tilde{a} + B) - e(\tilde{a})\} \\ &= \alpha_{22} + \beta_{22}(a_{22} + B) - e(a_{22}) \end{aligned} \quad (\text{ICI})$$

$$\begin{aligned} \alpha_{22} + \beta_{22}a_{22} - e(a_{22}) &\geq \max_{\tilde{a}} \{\alpha_{21} + \beta_{21}\tilde{a} - e(\tilde{a})\} \\ &= \alpha_{21} + \beta_{21}a_{21} - e(a_{21}) \end{aligned} \quad (\text{ICII})$$

where a_{21} and a_{22} denote the agent's second-period efforts if she invests in the first period and if she does not invest in the first period respectively. Constraint (IC0) motivates the agent to provide the desired efforts, constraints (ICI) and (ICII) induce the agent to report truthfully whether she invested in the first period, and constraints (IRI) and (IRII) ensure

the agent will receive at least her reservation utility. As Lemma 1 implies (shown later), the agent taking the investment b earns a positive information rent, whereas the agent taking no investments earns no information rent.

If the principal decides to hire a new agent, the new agent has no information advantage over the principal since she does not know whether the existing agent invested in the first period. She shares the same belief about the existing agent's first-period investment decision as the principal. The corresponding program is as follows:

$P_{2,new}$ (hiring a new agent in the second period):

$$\begin{aligned} U_{2,new}(p) &= \max_{\{\alpha_2, \beta_2\}} a_2 + pB - (\alpha_2 + \beta_2(a_2 + pB)) - C \\ \text{s.t.} \quad & a_2 \in \operatorname{argmax}_{\tilde{a}} \{\beta_2 \tilde{a} - e(\tilde{a})\} \\ & \alpha_2 + \beta_2(a_2 + pB) - e(a_2) = 0. \end{aligned}$$

Thus, the new agent is held to her reservation utility.

In the first period, anticipating the above second-period programs, the principal solves for the following optimization program:

P_1 (the first period):

$$\begin{aligned} U_1(p) &= \max_{\{\alpha_1, \beta_1\}} a_1 - pb - (\alpha_1 + \beta_1(a_1 - pb)) \\ \text{s.t.} \quad & a_1 \in \operatorname{argmax}_{\tilde{a}} \{\beta_1 \tilde{a} - e(\tilde{a})\} \\ & \alpha_1 + \beta_1 a_1 - e(a_1) = 0. \end{aligned}$$

If the agent invests in the first period, she will earn a less than zero utility in the first period (i.e., a loss of $\beta_1 b$) but earn a positive information rent in the second period if she is retained. If she does not invest in the first period, she will earn the reservation utility in either period.

It is straightforward to simplify P_1 and $P_{2,new}$ into the following two unconstrained programs:

$$P_1: \quad U_1(p) = \max_{a_1} \{a_1 - p \cdot b - e(a_1) + e'(a_1) \cdot p \cdot b\},^6 \text{ and}$$

$$\begin{aligned} P_{2,new}: \quad U_{2,new}(p) &= \max_{a_2} \{a_2 + p \cdot B - e(a_2) - C\} \\ &= pB + \max_{a_2} \{p(a_2 - e(a_2))\} + \max_{a_2} \{(1-p)(a_2 - e(a_2) - C/(1-p))\}. \quad (1) \end{aligned}$$

Let $a_1^*(p)$ and $a_2^*(p)$ denote the maximizers for $U_1(p)$ and $U_{2,new}(p)$ respectively. They are both functions of p .

⁶ $\beta_t = e'(a_t)$, and we directly use $e'(a_t)$ instead of β_t hereafter.

For program $P_{2,old}$, the following Lemma 1 shows that only constraints (IRII) and (ICI) are binding:

Lemma 1 *In program $P_{2,old}$, only constraints (IRII) and (ICI) are binding.*

Proof. (The proof is in the Appendix.) ■

A little algebra yields the following unconstrained program:

$P_{2,old}$:

$$\begin{aligned} U_{2,old} &= \max_{\{a_{21}, a_{22}\}} \{p(a_{21} - e(a_{21})) + pB + (1-p)(a_{22} - e(a_{22}) - pe'(a_{22})B/(1-p))\} \\ &= pB + \max_{a_{21}} \{p(a_{21} - e(a_{21}))\} + \max_{a_{22}} \{(1-p)(a_{22} - e(a_{22}) - pe'(a_{22})B/(1-p))\}. \end{aligned} \quad (2)$$

Let $a_{21}^*(p)$ and $a_{22}^*(p)$ denote the maximizers for $U_{2,old}(p)$. For simplicity, denote

$$\begin{aligned} f_{old}(p) &\equiv a_{22}^* - e(a_{22}^*) - pe'(a_{22}^*)B/(1-p), \\ f_{new}(p) &\equiv a_2^* - e(a_2^*) - C/(1-p), \text{ and} \\ f(p) &\equiv f_{old}(p) - f_{new}(p). \end{aligned}$$

From (1) and (2), $U_{2,old}(p) - U_{2,new}(p) = (1-p)f(p)$ (notice $a_2^*(p) = a_{21}^*(p)$).

To get a mixed strategy equilibrium, we have to make both the principal and the agent indifferent between their own available choices:

1) From the principal's perspective, she would get $U_1(p) + U_{2,old}(p)$ if she retains the existing agent and get $U_1(p) + U_{2,new}(p)$ if she hires a new agent. In equilibrium, we require $U_{2,old} = U_{2,new} (\equiv U_2)$ or, equivalently, there exist a $p \in (0, 1)$ such that

$$(1-p)f_{old}(p) = (1-p)f_{new}(p) \text{ or } f_{old}(p) = f_{new}(p).$$

Lemma 2 shows the existence of such a $p \in (0, 1)$.

Lemma 2 *There exist solutions for $f_{old}(p) = f_{new}(p)$ if the following condition holds:*

(C1) *for \bar{p} such that $C = e'(a_{22}^*(\bar{p}))B$, we have $f_{old}(\bar{p}) < f_{new}(\bar{p})$.*

Proof. (The proof is in the Appendix.) ■

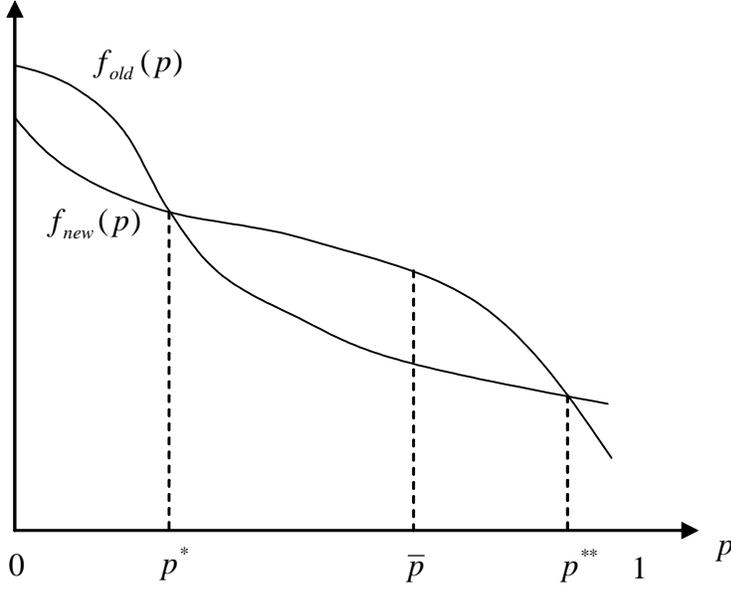


Figure 2.

Figure 2 depicts the relationship between $f_{old}(p)$ and $f_{new}(p)$ or $U_{2,old}(p)$ and $U_{2,new}(p)$ over $p \in (0, 1)$. When $p = p^*$ or p^{**} , $U_{2,old}(p) = U_{2,new}(p)$ and the principal is indifferent between retaining the existing agent and hiring a new one. The economic intuition for the randomization is as follows. If the principal retains the existing agent, she has to pay information rent ($e'(a_{22}^*)B$) to the existing agent who invested in the first period, and the desired effort a_{22}^* from the existing agent who did not invest in the first period is distorted (refer to the proof of Lemma 2). Given the agent randomizes between investing and not investing, the total cost from retention is a weighted average of the costs from the information rent and the distorted effort. If the principal hires a new agent, she has to pay the replacement cost C . When $p = p^*$ or p^{**} , the replacement cost and the retention cost (from the information rent and the distorted effort) are so matched that the principal can mix between retaining the existing agent and hiring a new one.

2) From the agent's perspective, her total utility (across the two periods) is zero if she does not invest in the first period, and her expected utility is as follows if she does invest in the first period:

$$\begin{aligned}
& \alpha_1^* + e'(a_1^*)(a_1^* - b) - e(a_1^*) + q(\alpha_{21}^* + \beta_{21}^*(a_{21}^* + B) - e(a_{21}^*)) \\
& = -e'(a_1^*)b + q \cdot e'(a_{22}^*)B.
\end{aligned} \tag{3}$$

where q denotes the principal's probability of retaining the existing agent in the second period. The equality comes from the relevant binding constraints in the corresponding programs. To make the agent indifferent between the two choices, we need a $q \in (0, 1)$ such that equation (3) is equal to zero, or $e'(a_1^*)b = qe'(a_{22}^*)B$. Lemma 3 shows the existence of such a $q \in (0, 1)$ when $p = p^*$.

Lemma 3 *There exists a $q^* \in (0, 1)$ such that $e'(a_1^*(p))b = q^*e'(a_{22}^*(p))B$ when $p = p^*$, and such a q does not exist when $p = p^{**}$, if the following condition holds:*

$$(C2) \quad be'''(a) \leq e''(a) \quad \forall a, \quad e'(a_1^*(1))b \leq e'(a_{22}^*(\bar{p}))B, \quad \text{and} \quad (e'(a_1^*(0))b =)b \geq e'(a_{22}^*(p^{**}))B.^7$$

Proof. (The proof is in the Appendix.) ■

Given the principal retains the existing agent with probability q^* , the expected information rent the existing agent will earn in the second period (if she invests in the first period) completely offsets her loss in the first period. Thus, in equilibrium, the existing agent invests in the first period with probability p^* .

Proposition 2 summarizes the above results:

Proposition 2 *Given (C1) and (C2), there exists a unique mixed strategy equilibrium with a replacement cost $C (> 0)$, in which the existing agent invests with probability p^* and the principal retains the existing agent with probability q^* .*

In the following, we introduce a LIV to investigate whether it can encourage long-term investments or the agent's equilibrium probability of investment would increase.

2.4 The Scenario with Leading Indicator Variables

Consider the setting where there exists a LIV $f \in \{H, L\}$ with

$$\text{prob}(f = H | \text{investment}) = n \quad \text{and} \quad \text{prob}(f = H | \text{no investment}) = m, \quad m < n.$$

⁷From Lemma 4 (in the Appendix), as p goes to 1, $e'(a_{22}^*(p))$ goes to zero. Hence, the last inequality in (C2) can be valid when p^{**} is close to 1. From $f_{old}(p)$ and $f_{new}(p)$, applying implicit function theorem, we can see p^{**} is increasing in B and decreasing in C (opposite for p^*). Hence, p^{**} is close to 1 when B is large or C is small.

That is, the LIV is more likely to be H if the agent invests in the first period. After observing the realized LIV, the principal updates her belief about the existing agent's first-period investment decision as follows:⁸

$$\begin{aligned} h &= \text{prob}\{I = b|f = H\} = \frac{np}{np + m(1-p)} > p \text{ and} \\ l &= \text{prob}\{I = b|f = L\} = \frac{(1-n)p}{(1-n)p + (1-m)(1-p)} < p, \end{aligned}$$

where p is the existing agent's (actual/equilibrium) probability of investment. The updated belief h (l) after observing H (L) is higher (lower) than the probability p .

After introducing the LIV, the existing agent still has two pure strategies: "invest" and "not invest", whereas the principal's strategy space expands to be {always retain the existing agent, always hire a new agent, retain the existing agent only when seeing H , and hire a new agent only when seeing H }. For simplicity, let {old/ H , old/ L }, {new/ H , new/ L }, {old/ H , new/ L }, and {new/ H , old/ L } denote the four strategies respectively.

With the same argument as in Section 3, the existing agent is held to her reservation utility (i.e., zero) in both periods if she does not invest in the first period. If she invests, she can use the expected information rent earned in the second period to compensate her loss in the first period such that her total expected utility is still zero.

The principal's first-period program stays the same as P_1 , and the principal has an expected utility of $U_1(p)$. The second-period programs depend on the realized LIV. When the realized LIV is H (L), the programs for retaining the existing agent and hiring a new agent are the same as $P_{2,old}$ and $P_{2,new}$ respectively, except that p is replaced with h (l) in both programs, and the principal has an expected utility of $U_{2,old}(h)$ and $U_{2,new}(h)$ ($U_{2,old}(l)$ and $U_{2,new}(l)$) respectively.

The principal's total expected utility U is:

$$\begin{aligned} U &= U_1(p) + p[nU_{2H} + (1-n)U_{2L}] + (1-p)[mU_{2H} + (1-m)U_{2L}] \\ &= U_1(p) + [pn + (1-p)m]U_{2H} + [p(1-n) + (1-p)(1-m)]U_{2L}, \end{aligned} \quad (4)$$

where p is the existing agent's (equilibrium) probability of investment and $U_{2H}(U_{2L})$ denotes

⁸Use Bayes' rule.

the principal's expected second-period utility upon seeing a high (low) LIV. $U_{2H}(U_{2L})$ can be either $U_{2,old}(h)$ or $U_{2,new}(h)$ (either $U_{2,old}(l)$ or $U_{2,new}(l)$) depending on which one is larger. Therefore, to figure out the principal's equilibrium strategy upon observing $H(L)$, we just need to compare $U_{2,old}(h)$ with $U_{2,new}(h)$ ($U_{2,old}(l)$ with $U_{2,new}(l)$). For instance, if $U_{2,old}(h) > U_{2,new}(h)$ and $U_{2,old}(l) > U_{2,new}(l)$, then the principal's equilibrium strategy would be "always retain the existing agent" or $\{\text{old}/H, \text{old}/L\}$.

Before characterizing the equilibria for this scenario, we first introduce two conditions regarding m and n ($m < n$):

$$(C3) \quad b \geq (1 - n)B + ne'(a_{22}^*(p^{**}))B, \text{ and}$$

$$(C4) \quad e'(a_1^*(1))b \leq (1 - n)e'(a_{22}^*(\bar{p}))B \text{ and } h < p^{**}.$$

Conditions (C3) and (C4) are mutually exclusive in the sense that if m and n satisfy (C3), they don't satisfy (C4), and vice versa. In particular, (C3) holds when n is large (notice $b \geq e'(a_{22}^*(p^{**}))B$ from (C2)) and/or the net return on investment is small (i.e., $B - b$ is small). In contrast, (C4) holds when n is small (notice $e'(a_1^*(1))b \leq e'(a_{22}^*(\bar{p}))B$ from (C2)) and/or the net return on investment is large (i.e., $B - b$ is large).⁹

Proposition 3 characterizes the mixed strategy equilibrium under this case with the LIV f :

Proposition 3 *Given (C1) and (C2), after introducing the LIV f :*

- i) there exists a **unique** mixed strategy equilibrium E1 for any (m, n) ($m < n$) satisfying (C3). In equilibrium E1, the principal randomizes between her two choices when she observes $f = H$ and retains the existing agent otherwise. The agent's equilibrium probability of investment is smaller and the principal is **worse off** in equilibrium E1 than in the no LIV case; and*
- ii) there exists a **unique** mixed strategy equilibrium E2 for any (m, n) ($m < n$) satisfying (C4). In equilibrium E2, the principal randomizes between her two choices when she observes $f = L$ and hires a new agent otherwise. The agent's equilibrium probability of investment is higher and the principal is **better off** in equilibrium E2 than in the no LIV case.*

⁹The economic intuition of (C3) and (C4) will be discussed later.

Proof. (The proof is in the Appendix.) ■

In Case (i) where the probability of a high LIV conditional on investment (i.e., n) is high and/or the net investment return is small (i.e., condition (C3)), the principal randomizes between $\{\text{old}/H, \text{old}/L\}$ and $\{\text{new}/H, \text{old}/L\}$ in equilibrium. When the LIV is H , the principal's updated belief h is the same as the agent's equilibrium probability of investment in the no LIV scenario (i.e., $h = p^*$). Thus, as argued in Section 3, the replacement cost C from hiring a new agent equals the retention cost associated with the information rent $e'(a_{22}^*(h))B$ and the distorted effort $a_{22}^*(h)$. Hence, the principal is indifferent between retaining the existing agent and hiring a new one when she observes a high LIV.

When the LIV is otherwise L , the principal believes it is less likely the existing agent invested in the first period than when the LIV is H (i.e., $h > l$). Given the smaller belief l , the retention cost associated with the distorted effort $a_{22}^*(l)$ and the information rent $e'(a_{22}^*(l))B$ becomes smaller compared to when the LIV is H . However, the replacement cost C stays the same as when the LIV is H . Therefore, the principal would retain the existing agent when the LIV is L (refer to Figure 2 with $l < h = p^*$). In summary, the principal mixes between $\{\text{old}/H, \text{old}/L\}$ and $\{\text{new}/H, \text{old}/L\}$ in equilibrium.

The existing agent mixes between investing and not investing in the first period because she gets her reservation utility (zero) regardless of her investment decision. If she does not invest, she gets a zero utility in both periods. If she invests, she earns information rent in the second period which completely offsets her loss in the first period. Given the principal mixes between $\{\text{old}/H, \text{old}/L\}$ and $\{\text{new}/H, \text{old}/L\}$, the existing agent always stays in the second period when $f = L$ and possibly stays otherwise. If the chance of $f = L$ is high or n (as well as m) is small, the existing agent's chance of staying is high. Accordingly, her expected information rent in the second period might be so high that the expected rent would exceed the agent's loss in the first period, destroying the incentive to mix. Similarly, if the net return on investment is large, the expected rent from the second period may also exceed the loss in the first period, motivating deviation. Thus, in order for the agent to mix between investing and not investing given the principal's equilibrium strategy, n needs to be large and/or the net return on investment needs to be small (i.e., condition (C3)).

Below we explore how the principal's total expected utility changes relative to the no LIV

scenario. In Case (i), the updated belief h on seeing a H LIV is the same as the existing agent's equilibrium probability of investment p^* in the no LIV scenario. Thus, in Case (i), the existing agent's equilibrium probability of investment \tilde{p} is smaller ($\tilde{p} < h$), or the agent is less likely to invest in the first period. Accordingly, the principal's first-period expected utility decreases. The intuition is that, due to the smaller probability of investment \tilde{p} , the principal has a smaller expected investment cost $\tilde{p}b$ as well as a smaller expected gain $\tilde{p}e'(a_1^*(\tilde{p}))b$ which is the loss the agent who invests in the first period has. It turns out that the smaller expected gain dominates and, thus, the principal's first-period expected utility decreases.¹⁰

The principal's second-period expected utility decreases as well. Recall from equation (4) that, in Case (i), the second-period expected utility is a weighted average of the expected utility when $f = L$ (i.e., $U_{2L}(h) = U_{2,old}(l)$) and the expected utility when $f = H$ (i.e., $U_{2H}(h) = U_{2,old}(h) = U_{2,new}(h)$). Since $h = p^*$, the expected utility when $f = H$ is the same as the principal's second-period expected utility in the no LIV scenario ($U_2(p^*) = U_{2,old}(p^*) = U_{2,new}(p^*)$). Therefore, to compare the second-period expected utility in Case (i) with that in the no LIV scenario, we need to only compare $U_{2,old}(l)$ with $U_2(p^*)$ or $U_{2,old}(h)$. As argued above, when $f = L$, the retention cost associated with the distorted effort $a_{22}^*(l)$ and the information rent $e'(a_{22}^*(l))B$ is smaller than when $f = H$. However, the expected investment return is smaller as well (i.e., $lB < hB$). It turns out that the smaller investment return dominates and $U_{2,old}(l) < U_{2,old}(h)$. Thus, the principal's second-period expected utility also decreases.¹¹ To sum up, the principal becomes worse off after introducing a LIV as in Case (i). The presence of a LIV decreases the agent's expected investment in the first period, which jeopardizes the principal's welfare.

In Case (ii), the principal randomizes between $\{\text{new}/H, \text{new}/L\}$ and $\{\text{new}/H, \text{old}/L\}$. The economic intuition is similar to that of Case (i) and, thus, omitted. However, a couple of points are worth some discussion. First, Case (ii) is valid when n (as well as m) is small and/or the net return on investment is large. In equilibrium, the principal possibly retain the existing agent only when the LIV is L . In order for the existing agent to get enough information rent in the second period to offset her loss in the first period from investing, the

¹⁰Refer to Lemma 5 in the Appendix: $U_1(p)$ is increasing in p .

¹¹Refer to Lemma 4 in the Appendix: $U_{2,old}(p)$ is increasing in p .

chance of a low LIV has to be large (i.e., n has to be small) or the net investment return has to be large. Second, in Case (ii), the existing agent's equilibrium probability of investment is higher than that in the no LIV scenario, which makes the principal better off. The LIV helps the principal in this case by motivating a higher expected investment.

Next, we study an interesting question of how the informativeness of the LIV affects the welfare of the principal. In general, the larger the difference between n and m , the more likely to distinguish investment from no investment, and the more informative the LIV. The following corollary presents the comparative statics results.

Corollary 1

- i) In equilibrium E1, the principal's total expected utility is increasing in m ; and*
- ii) In equilibrium E2, the principal's total expected utility is increasing in n .*

Proof. (The proof is in the Appendix.) ■

This corollary shows that, in Case (i), the principal gets worse off as m gets smaller (given n). In other words, the more informative the LIV, the worse off the principal, which is counterintuitive. The rationale is that, in Case (i), the LIV distorts the contracting problem by inducing a smaller chance of investment in the first period than in the no LIV scenario. Then, the “more informative” LIV would induce an even smaller chance of investment. Thus, the principal becomes worse off.

In Case (ii), the LIV induces a larger chance of investment than in the no LIV scenario. A “more informative” LIV (larger n given m) would induce an even larger chance of investment. Thus, the principal becomes better off.

To illustrate, consider the following numerical example. Suppose the cost function of effort is $e(a_t) = \exp(ra_t)$ where $r = .01$, the investment cost is $b = 1.8$, the investment return is $B = 6$, and the replacement cost is $C = 2$.

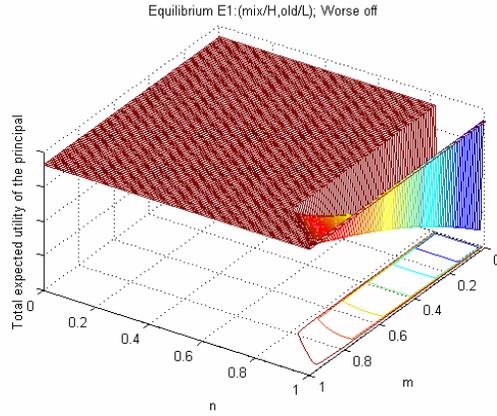


Figure 3. The total expected utility of the principal for equilibrium E1

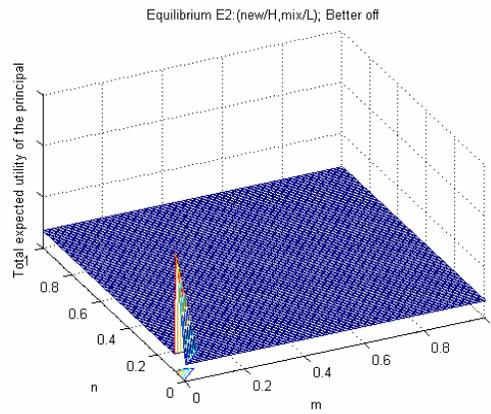


Figure 4. The total expected utility of the principal for equilibrium E2

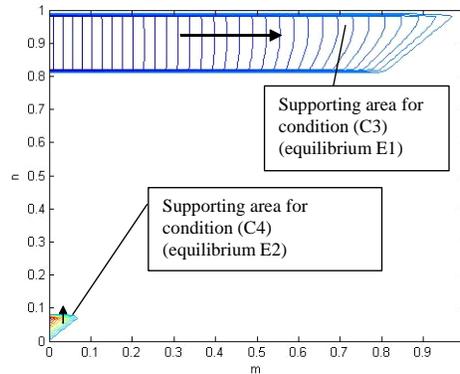


Figure 5. The contours of the utility function for both Equilibria E1 and E2

Figure 3 (Figure 4) depicts the 3-D graph of the principal's total expected utility vs. (m, n) for equilibrium E1 (equilibrium E2). The flat plane in Figure 3 (Figure 4) indicates where condition (C3) (condition (C4)) for equilibrium E1 (equilibrium E2) does not hold

and represents the principal's total expected utility in the no LIV scenario. From Figure 3 (Figure 4), we can see the principal is worse off (better off) after introducing the LIV. Figure 3 (Figure 4) also shows that the total expected utility is increasing in m (increasing in n) under equilibrium E1 (equilibrium E2), which is clearer in Figure 5.

Figure 5 draws the contours of the utility function on the $m - n$ plane for both equilibria E1 and E2 (notice $m < n$). From this graph, we can see the supporting areas for conditions (C3) and (C4) do not overlap with each other, which indicates only one equilibrium exists for any specific (m, n) in the supporting areas. Pointing to the ascending directions, the arrows in the graph also indicates that the utility is increasing in m for equilibrium E1 and increasing in n for equilibrium E2, confirming that the larger the difference between m and n , the smaller the principal's total expected utility in equilibrium E1 and the larger the principal's total expected utility in equilibrium E2.

2.5 Conclusion

In this paper, we develop a two-period short-term agency model in which the agent has to make a long-term investment decision at the beginning of the first period in addition to providing efforts. At the beginning of the second period, the principal can retain the existing agent or hire a new agent with a positive replacement cost. Our analysis shows that there is no pure strategy equilibrium but a mixed strategy equilibrium in which the agent sometimes takes the long-term investment and sometimes does not, whereas the principal sometimes retains the existing agent and sometimes hires a new agent.

We next introduce a publicly available binary LIV at the end of the first period which is not verifiable or contractible. Investment would increase the likelihood of the LIV being high. Our analysis finds that, if the LIV is highly likely to be high with investment and/or the net investment return is small, there is a unique equilibrium in which the principal retains the existing agent when observing a low realization of the LIV and randomizes otherwise. When the principal observes a high LIV, her updated belief about the agent's probability of investment (which is higher than the agent's actual/equilibrium probability of investment) is the same as the agent's equilibrium probability of investment in the no LIV case. Thus,

the agent's ex ante probability of investment is smaller in this case. Because of the dominant effect from the smaller expected investment, the principal becomes worse off with the presence of the LIV. In contrast, we find that, if the LIV is highly likely to be low with investment and/or the net investment return is large, the principal becomes better off with the presence of the LIV by increasing the agent's expected investment in the first period.

This paper provides a guideline on the selection/application of subjective nonfinancial measures in reality. It also has a couple of empirical implications. First, firms use subjective nonfinancial measures mostly on highly profitable projects. Second, such nonfinancial measures are generally tough in the sense that it is hard to get a high score for these measures even if the manager takes the desired investment. It would be worthwhile to test these empirical implications in future.

Appendix

Proof. (of Lemma 1)

From (IC0), $\beta_{2i} = e'(a_{2i})$, $i = 1, 2$. Constraints (IRII) and (ICI) imply constraint (IRI).

Hence, $P_{2,old}$ can be simplified into the following program:

$$U_{2,old}(p) = \max_{\{\alpha_{21}, \beta_{21}, \alpha_{22}, \beta_{22}\}} p(a_{21} + B - (\alpha_{21} + \beta_{21}(a_{21} + B))) + (1 - p)(a_{22} - (\alpha_{22} + \beta_{22}a_{22}))$$

$$\text{s.t.} \quad \alpha_{22} + e'(a_{22})a_{22} - e(a_{22}) \geq 0 \quad (\text{IRII})$$

$$\alpha_{21} + e'(a_{21})(a_{21} + B) - e(a_{21}) \geq \alpha_{22} + e'(a_{22})(a_{22} + B) - e(a_{22}) \quad (\text{ICI})$$

$$\alpha_{22} + e'(a_{22})a_{22} - e(a_{22}) \geq \alpha_{21} + e'(a_{21})a_{21} - e(a_{21}). \quad (\text{ICII})$$

Let λ_1 , λ_2 , and λ_3 denote the Lagrangian multipliers for constraints (IRII), (ICI), and (ICII) respectively. The Kuhn-Tucker conditions can be written as follows:

$$-p + \lambda_2 - \lambda_3 = 0$$

$$-(1 - p) + \lambda_1 - \lambda_2 + \lambda_3 = 0$$

$$p[1 - e'(a_{21}) - (a_{21} + B)e''(a_{21})] + \lambda_2(a_{21} + B)e''(a_{21}) - \lambda_3a_{21}e''(a_{21}) = 0$$

$$(1 - p)[1 - e'(a_{22}) - a_{22}e''(a_{22})] + \lambda_1a_{22}e''(a_{22}) - \lambda_2(a_{22} + B)e''(a_{22}) + \lambda_3a_{22}e''(a_{22}) = 0$$

along with the complementary slackness conditions for constraints (IRII), (ICI), and (ICII).

These conditions can be simplified into (A.1):

$$\lambda_1 = 1 > 0$$

$$\lambda_2 = p + \lambda_3$$

$$p(1 - e'(a_{21})) + \lambda_3Be''(a_{21}) = 0 \quad (\text{A.1})$$

$$(1 - p)(1 - e'(a_{22})) - \lambda_2Be''(a_{22}) = 0.$$

Now we show $\lambda_3 = 0$. Suppose $\lambda_3 > 0$, then $\lambda_2 = p + \lambda_3 > 0$. Hence, both (ICI) and (ICII) are binding, which implies $a_{21} = a_{22}$. Then, adding the last two conditions in (A.1) together yields

$$1 - e'(a_{21}) = pBe''(a_{21}). \quad (\text{A.2})$$

Substituting (A.2) into the third condition in (A.1) yields $(p^2 + \lambda_3)Be''(a_{21}) = 0$, which contradicts the fact that every term on the left side is greater than zero. Thus, $\lambda_2 > \lambda_3 = 0$, and only constraints (IRII) and (ICI) are binding. ■

Proof. (of Lemma 2)

We first prove Lemma 4.

Lemma 4 For $U_{2,old}(p)$ where $1 > p > 0$, $e'(a_{22}^*) < 1$, $a_{22}^* < a_{21}^*$, a_{22}^* is decreasing in p , $U_{2,old}(p)$ is increasing in p , and $e''(a_{22}^*)$ ($e'(a_{22}^*)$ as well) goes to zero as p goes to one.¹²

Proof. (of Lemma 4) FOC gives

$$e'(a_{21}^*) = 1, \quad (\text{A.3})$$

$$\text{and } 1 - e'(a_{22}^*) - pBe''(a_{22}^*)/(1 - p) = 0. \quad (\text{A.4})$$

From (A.3) and (A.4), We can see that $e'(a_{22}^*) < 1 = e'(a_{21}^*)$ or $a_{22}^* < a_{21}^*$. We can also see from (A.4) that $e''(a_{22}^*)$ goes to zero as p goes to 1. So does $e'(a_{22}^*)$ by assumption.

Differentiating both sides of (A.4) with respect to p gives us

$$-e''(a_{22}^*)a_{22p}^* - pBe'''(a_{22}^*)a_{22p}^*/(1 - p) - Be''(a_{22}^*)/(1 - p)^2 = 0.$$

Hence, $a_{22p}^* < 0$ or a_{22}^* is decreasing in p .

$$\begin{aligned} \text{By envelope theorem, } U_{2,old}'(p) &= a_{21}^* - e(a_{21}^*) + B - (a_{22}^* - e(a_{22}^*)) - e'(a_{22}^*)B \\ &= u(a_{21}^*) - u(a_{22}^*) + B(1 - e'(a_{22}^*)) \end{aligned}$$

where $u(a) \equiv a - e(a)$.

Since $u'(a) \equiv 1 - e'(a) \geq 0$ if $a \leq a_{21}^*$, $u(a)$ is increasing in a if $a \leq a_{21}^*$.

Given $e'(a_{22}^*) < 1$ and $a_{22}^* < a_{21}^*$, $U_{2,old}'(p) > 0$ or $U_{2,old}(p)$ is increasing in p . ■

To prove Lemma 2, first notice that when $p = 1$ or 0 , $U_{2,old}(p) - U_{2,new}(p) = C > 0$, because retaining the existing agent would incur no information rent when the principal knows whether the agent invested in the first period ($p = 1$ or 0), whereas hiring a new agent would bring an additional replacement cost C (there is no distortion in the desired efforts under both cases). Hence, we have

$$f_{old}(p) > f_{new}(p) \text{ or } f(p) > 0 \text{ as } p \rightarrow 1 \text{ or } 0. \quad (\text{A.5})$$

¹²By Lemma 4, we can see that there exists a unique \bar{p} satisfying condition (C1) given $C < B$.

By envelope theorem, $f'(p) = f'_{old}(p) - f'_{new}(p) = (C - Be'(a_{22}^*)) / (1 - p)^2$.

Condition (C1) given in this lemma is just,

$$f(\bar{p}) < 0 \text{ and } f'(\bar{p}) = 0. \quad (\text{A.6})$$

By Lemma 4, a_{22}^* is decreasing in p . Hence, we have

$$f'(p) < 0 \text{ if } \bar{p} > p > 0 \text{ and } f'(p) > 0 \text{ if } \bar{p} < p < 1. \quad (\text{A.7})$$

By virtue of (A.5), (A.6) and (A.7), there must exist p^* and p^{**} with $0 < p^* < \bar{p}$ and $\bar{p} < p^{**} < 1$ such that $f_{old}(p^*) = f_{new}(p^*)$ and $f_{old}(p^{**}) = f_{new}(p^{**})$. Or, when $p = p^*$ or p^{**} ,

$U_{2,old}(p) = U_{2,new}(p)$. In addition, we have

$f_{old}(p) > f_{new}(p)$ or $U_{2,old}(p) > U_{2,new}(p)$, if $0 < p < p^*$ or $p^{**} < p < 1$, and

$f_{old}(p) < f_{new}(p)$ or $U_{2,old}(p) < U_{2,new}(p)$, if $p^* < p < p^{**}$. ■

Proof. (of Lemma 3)

We first prove Lemma 5.

Lemma 5 *When $1 > p > 0$, $a_1^*(p)$ and $U_1(p)$ are both increasing in p .*

Proof. (of Lemma 5)

By FOC, we have $1 - e'(a_1^*) + e''(a_1^*)pb = 0$. Differentiating both sides with respect to p yields

$$-e''(a_1^*)a_{1p}^* + e'''(a_1^*)pba_{1p}^* + e''(a_1^*)b = 0.$$

Given $be'''(a) \leq e''(a)$ (from (C2)) and $1 > p > 0$, we have $a_{1p}^* > 0$ or $a_1^*(p)$ is increasing in p . When $p = 0$, $e'(a_1^*(0)) = 1$. Hence, $e'(a_1^*(p)) > 1$ if $1 > p > 0$. By envelope theorem, $U_1'(p) = (e'(a_1^*(p)) - 1)b > 0$ if $1 > p > 0$. Hence, $U_1(p)$ is increasing in p . ■

From Lemma 4 and Lemma 5, $a_1^*(p)$ is increasing in p and $a_{22}^*(p)$ is decreasing in p . Hence, $e'(a_1^*(p^*))b < e'(a_1^*(1))b \leq e'(a_{22}^*(\bar{p}))B < e'(a_{22}^*(p^*))B$, given $0 < p^* < \bar{p} < 1$ and (C2).

There exists a $q^* \in (0, 1)$ such that $e'(a_1^*(p^*))b = q^*e'(a_{22}^*(p^*))B$.

We also have $e'(a_1^*(p^{**}))b > e'(a_1^*(0))b = b \geq e'(a_{22}^*(p^{**}))B$ (from (C2)). Hence, there does not exist any $q \in (0, 1)$ such that $e'(a_1^*(p^{**}))b = qe'(a_{22}^*(p^{**}))B$. ■

Proof. (of Proposition 3)

Recall that Figure 2 depicts the comparison between $U_{2,old}(p)$ and $U_{2,new}(p)$ over $p \in (0, 1)$. In this proof, we need to refer to Figure 2 frequently when we compare $U_{2,old}(p)$ with

$U_{2,new}(p)$.

For Case (i), consider the following cases where \tilde{p} denotes the existing agent's equilibrium probability of investment:

1. If $0 < l < \tilde{p} < h < p^*$, then $U_{2,old}(h) > U_{2,new}(h)$ and $U_{2,old}(l) > U_{2,new}(l)$. The principal's strategy should be "always retain the existing agent." From the existing agent's perspective, we require

$$-e'(a_1^*(\tilde{p}))b + (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B = 0.$$

Since $e'(a_1^*(1))b \leq e'(a_{22}^*(\bar{p}))B$, given $l < \tilde{p} < h < p^* < \bar{p}$, we have $e'(a_1^*(\tilde{p}))b < e'(a_{22}^*(l))B$ and $e'(a_1^*(\tilde{p}))b < e'(a_{22}^*(h))B$.

Hence, no equilibrium exists for this case;

2. If $0 < l < \tilde{p} < h = p^*$, then $U_{2,old}(h) = U_{2,new}(h)$ and $U_{2,old}(l) > U_{2,new}(l)$. The principal's strategy should be "retain the existing agent when $f = L$ and randomize when $f = H$." Denote the probability of retaining the existing agent when $f = H$ by q . From the existing agent's perspective, we need a $q \in (0, 1)$ such that

$$-e'(a_1^*(\tilde{p}))b + (1-n)e'(a_{22}^*(l))B + qne'(a_{22}^*(h))B = 0.$$

Or, equivalently, we require

$$e'(a_1^*(\tilde{p}))b > (1-n)e'(a_{22}^*(l))B \text{ and } e'(a_1^*(\tilde{p}))b < (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B. \quad (\text{A.8})$$

Since $e'(a_1^*(1))b \leq e'(a_{22}^*(\bar{p}))B$, given $0 < l < \tilde{p} < h = p^* < \bar{p}$, we have

$$e'(a_1^*(\tilde{p}))b < e'(a_{22}^*(l))B \text{ and } e'(a_1^*(\tilde{p}))b < e'(a_{22}^*(h))B.$$

Hence, the second inequality of (A.8) is satisfied.

Given $b \geq (1-n)B + ne'(a_{22}^*(p^{**}))B$ and $e'(a_1^*(0)) = e'(a_{22}^*(0)) = 1$, we have

$$e'(a_1^*(\tilde{p}))b > e'(a_1^*(0))b = b > (1-n)B = (1-n)e'(a_{22}^*(0))B > (1-n)e'(a_{22}^*(l))B.$$

Thus, there exists a mixed strategy equilibrium for this case. The principal's total expected utility is:

$$U_1(\tilde{p}) + [\tilde{p}n + (1-\tilde{p})m]U_{2,old}(h) + [\tilde{p}(1-n) + (1-\tilde{p})(1-m)]U_{2,old}(l).$$

Recall that, in the no LIV scenario, the principal's total expected utility is $U_1(p^*) + U_{2,new}(p^*) = U_1(p^*) + U_{2,old}(p^*)$, and $U_1(p)$ and $U_{2,old}(p)$ are all increasing in p . Hence, in equilibrium the principal becomes worse off after introducing the LIV ($l < \tilde{p} < h = p^*$);

3. If $0 < l < p^* < h < p^{**}$, then $U_{2,old}(h) < U_{2,new}(h)$ and $U_{2,old}(l) > U_{2,new}(l)$. The principal's strategy should be "retain the existing agent when $f = L$ and hire a new agent

when $f = H$." From the existing agent's perspective, we require

$$-e'(a_1^*(\tilde{p}))b + (1-n)e'(a_{22}^*(l))B = 0.$$

However, given $b \geq (1-n)B + ne'(a_{22}^*(p^{**}))B$, we have

$$e'(a_1^*(\tilde{p}))b > (1-n)e'(a_{22}^*(l))B.$$

Hence, no equilibrium exists for this case;

4. If $0 < l < p^* < h = p^{**}$, then $U_{2,old}(h) = U_{2,new}(h)$ and $U_{2,old}(l) > U_{2,new}(l)$. As in case 2, the principal's strategy should be "retain the existing agent when $f = L$ and randomize when $f = H$." From the existing agent's perspective, we also require the two inequalities in (A.8) to be held.

However, given $b \geq (1-n)B + ne'(a_{22}^*(p^{**}))B$, we have

$$e'(a_1^*(\tilde{p}))b > (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(p^{**}))B \text{ (notice } h = p^{**}\text{)}.$$

Hence, there is no equilibrium for this case either;

5. If $0 < l < p^* < p^{**} < h < 1$, then as in case 1, the principal's strategy should be "always retain the existing agent." We require

$$-e'(a_1^*(\tilde{p}))b + (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B = 0.$$

However, given $b \geq (1-n)B + ne'(a_{22}^*(p^{**}))B$, we have

$$e'(a_1^*(\tilde{p}))b > (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B.$$

Hence, no equilibrium exists for this case;

6. If $p^* = l < h < p^{**}$, then $U_{2,old}(h) < U_{2,new}(h)$ and $U_{2,old}(l) = U_{2,new}(l)$. The principal's strategy should be "randomize when $f = L$ and hire a new agent when $f = H$." Denote the probability of retaining the existing agent when $f = L$ by q . We need a $q \in (0, 1)$ such that

$$-e'(a_1^*(\tilde{p}))b + q(1-n)e'(a_{22}^*(l))B = 0.$$

However, given $b \geq (1-n)B + ne'(a_{22}^*(p^{**}))B$, we have

$$e'(a_1^*(\tilde{p}))b > (1-n)e'(a_{22}^*(l))B.$$

Thus, no equilibrium exists for this case;

7. If $p^* = l < h = p^{**}$, then $U_{2,old}(h) = U_{2,new}(h)$ and $U_{2,old}(l) = U_{2,new}(l)$. The principal's strategy should be "randomize both when $f = L$ and when $f = H$." Denote the probability of retaining the existing agent by q_1 when $f = L$ and by q_2 when $f = H$. We require a $q_1 \in (0, 1)$ and a $q_2 \in (0, 1)$ such that

$$-e'(a_1^*(\tilde{p}))b + q_1(1-n)e'(a_{22}^*(l))B + q_2ne'(a_{22}^*(h))B = 0, \text{ or}$$

$$e'(a_1^*(\tilde{p}))b < (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B.$$

However, given $b \geq (1-n)B + ne'(a_{22}^*(p^{**}))B$, we have

$$e'(a_1^*(\tilde{p}))b > (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(p^{**}))B.$$

Hence, no equilibrium exists for this case;

8. If $p^* = l < p^{**} < h$, then $U_{2,old}(h) > U_{2,new}(h)$ and $U_{2,old}(l) = U_{2,new}(l)$. The principal's strategy should be "randomize when $f = L$ and retain the existing agent when $f = H$." We need a $q \in (0, 1)$ such that

$$-e'(a_1^*(\tilde{p}))b + q(1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B = 0.$$

However, given $b \geq (1-n)B + ne'(a_{22}^*(p^{**}))B$, we have

$$e'(a_1^*(\tilde{p}))b > (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B, \text{ since } p^{**} < h.$$

Hence, no equilibrium exists for this case;

9. If $p^* < l < h < p^{**}$, then $U_{2,old}(h) < U_{2,new}(h)$ and $U_{2,old}(l) < U_{2,new}(l)$. The principal's strategy should be "always hire a new agent." We require

$$-e'(a_1^*(\tilde{p}))b + 0 = 0.$$

However, $-e'(a_1^*(\tilde{p}))b < 0$. Hence, no equilibrium exists for this case;

10. If $p^* < l < h = p^{**}$, then $U_{2,old}(h) = U_{2,new}(h)$ and $U_{2,old}(l) < U_{2,new}(l)$. The principal's strategy should be "hire a new agent when $f = L$ and randomize when $f = H$." We need a $q \in (0, 1)$ such that

$$-e'(a_1^*(\tilde{p}))b + qne'(a_{22}^*(h))B = 0.$$

However, given $e'(a_1^*(0))b = b \geq e'(a_{22}^*(p^{**}))B$, we have

$$e'(a_1^*(\tilde{p}))b > ne'(a_{22}^*(p^{**}))B \text{ (} h = p^{**}\text{)}.$$

Hence, no equilibrium exists for this case;

11. If $p^* < l < p^{**} < h$, then $U_{2,old}(h) > U_{2,new}(h)$ and $U_{2,old}(l) < U_{2,new}(l)$. The principal's strategy should be "hire a new agent when $f = L$ and retain the existing agent when $f = H$." We require

$$-e'(a_1^*(\tilde{p}))b + ne'(a_{22}^*(h))B = 0.$$

However, given $e'(a_1^*(0))b = b \geq e'(a_{22}^*(p^{**}))B$, we have

$$e'(a_1^*(\tilde{p}))b > ne'(a_{22}^*(p^{**}))B > ne'(a_{22}^*(h))B.$$

Hence, no equilibrium exists for this case;

12. If $p^{**} = l < h$, then $U_{2,old}(h) > U_{2,new}(h)$ and $U_{2,old}(l) = U_{2,new}(l)$. The principal's

strategy should be “randomize when $f = L$ and retain the existing agent when $f = H$.” We need a $q \in (0, 1)$ such that

$$-e'(a_1^*(\tilde{p}))b + q(1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B = 0.$$

However, given $b \geq (1-n)B + ne'(a_{22}^*(p^{**}))B$, we have

$$e'(a_1^*(\tilde{p}))b > (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B, \text{ since } p^{**} < h.$$

Hence, no equilibrium exists for this case; and

13. If $p^{**} < l < h$, then $U_{2,old}(h) > U_{2,new}(h)$ and $U_{2,old}(l) > U_{2,new}(l)$. The principal’s strategy should be “always retain the existing agent.” We require

$$-e'(a_1^*(\tilde{p}))b + (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B = 0.$$

However, given $b \geq (1-n)B + ne'(a_{22}^*(p^{**}))B$, we have

$$e'(a_1^*(\tilde{p}))b > (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B.$$

Hence, no equilibrium exists for this case.

Until now, we have exhausted all possible cases for Case (i) and shown that only case 2 supports a unique mixed strategy equilibrium where the principal is worse off.

For Case (ii) (notice $h < p^{**}$), consider the following cases:

1. If $0 < l < \tilde{p} < h < p^*$, then $U_{2,old}(h) > U_{2,new}(h)$ and $U_{2,old}(l) > U_{2,new}(l)$. The principal’s strategy should be “always retain the existing agent.” From the existing agent’s perspective, we require

$$-e'(a_1^*(\tilde{p}))b + (1-n)e'(a_{22}^*(l))B + ne'(a_{22}^*(h))B = 0.$$

However, given $e'(a_1^*(1))b \leq (1-n)e'(a_{22}^*(\tilde{p}))B$, we have

$$e'(a_1^*(\tilde{p}))b < e'(a_1^*(1))b \leq (1-n)e'(a_{22}^*(\tilde{p}))B < (1-n)e'(a_{22}^*(l))B.$$

Hence, no equilibrium exists for this case;

2. If $0 < l < p^* < h < p^{**}$, then $U_{2,old}(h) < U_{2,new}(h)$ and $U_{2,old}(l) > U_{2,new}(l)$. The principal’s strategy should be “retain the existing agent when $f = L$ and hire a new agent when $f = H$.” We require

$$-e'(a_1^*(\tilde{p}))b + (1-n)e'(a_{22}^*(l))B = 0.$$

However, as argued above, we have

$$e'(a_1^*(\tilde{p}))b < e'(a_1^*(1))b \leq (1-n)e'(a_{22}^*(\tilde{p}))B < (1-n)e'(a_{22}^*(l))B.$$

Hence, no equilibrium exists for this case;

3. If $p^* = l < h < p^{**}$, then $U_{2,old}(h) < U_{2,new}(h)$ and $U_{2,old}(l) = U_{2,new}(l)$. The principal’s

strategy should be “randomize when $f = L$ and hire a new agent when $f = H$.” Denote the probability of retaining the existing agent when $f = L$ by q . We need a $q \in (0, 1)$ such that

$$-e'(a_1^*(\tilde{p}))b + q(1 - n)e'(a_{22}^*(l))B = 0.$$

As argued above, we have

$$e'(a_1^*(\tilde{p}))b < (1 - n)e'(a_{22}^*(l))B.$$

Thus, there exists a mixed strategy equilibrium for this case. The principal’s total expected utility is:

$$U_1(\tilde{p}) + [\tilde{p}n + (1 - \tilde{p})m]U_{2,new}(h) + [\tilde{p}(1 - n) + (1 - \tilde{p})(1 - m)]U_{2,old}(l).$$

Recall that, in the no LIV scenario, the principal’s total expected utility is $U_1(p^*) + U_{2,new}(p^*) = U_1(p^*) + U_{2,old}(p^*)$, and $U_1(p)$, $U_{2,new}(p)$, and $U_{2,old}(p)$ are all increasing in p . Hence, in equilibrium the principal becomes better off after introducing the LIV ($p^* = l < \tilde{p} < h$); and

4. If $p^* < l < h < p^{**}$, then $U_{2,old}(h) < U_{2,new}(h)$ and $U_{2,old}(l) < U_{2,new}(l)$. The principal’s strategy should be “always hire a new agent.” We require

$$-e'(a_1^*(\tilde{p}))b + 0 = 0.$$

However, $-e'(a_1^*(\tilde{p}))b < 0$. Hence, no equilibrium exists for this case.

For Case (ii), only case 3 supports a unique mixed strategy equilibrium where the principal is better off. ■

Proof. (of Corollary 1)

For Case (i), recall from (4) that the principal’s utility is

$$U = U_1(\tilde{p}) + [\tilde{p}n + (1 - \tilde{p})m]U_{2,old}(h) + [\tilde{p}(1 - n) + (1 - \tilde{p})(1 - m)]U_{2,old}(l)$$

where $h = p^*$.

Differentiating U with respect to m , we have

$$\begin{aligned} U'_m &= (U_1(\tilde{p}))'_m + [\tilde{p}'_m(n - m) + (1 - \tilde{p})](U_{2,old}(h) - U_{2,old}(l)) \\ &\quad + [\tilde{p}(1 - n) + (1 - \tilde{p})(1 - m)](U_{2,old}(l))'_m. \end{aligned}$$

Notice that

1) $h (= p^*)$ and $U_{2,old}(h)$ are fixed values and $U_{2,old}(h) > U_{2,old}(l)$;

2) $\tilde{p} = hm/(n - nh + mh)$ and $l = (1 - n)mh/(n - nh - mn + mh)$. Thus, both \tilde{p} and l are increasing in m . Since $U_1(p)$ and $U_{2,old}(p)$ are both increasing in p , $(U_1(\tilde{p}))'_m$ and $(U_{2,old}(l))'_m$ must be positive.

Hence, $U'_m > 0$ or U is increasing in m .

For Case (ii), the principal's utility is

$$U = U_1(\tilde{p}) + [\tilde{p}n + (1 - \tilde{p})m]U_{2,old}(h) + [\tilde{p}(1 - n) + (1 - \tilde{p})(1 - m)]U_{2,old}(l)$$

where $l = p^*$.

Differentiating U with respect to m , we have

$$\begin{aligned} U'_m &= (U_1(\tilde{p}))'_m + [\tilde{p}'_m(n - m) + \tilde{p}](U_{2,old}(h) - U_{2,old}(l)) \\ &\quad + [\tilde{p}n + (1 - \tilde{p})m](U_{2,old}(h))'_m. \end{aligned}$$

Notice that

3) $l (= p^*)$ and $U_{2,old}(l)$ are fixed values and $U_{2,old}(h) > U_{2,old}(l)$;

4) $\tilde{p} = l(1 - m)/(1 - n + nl - ml)$ and $h = (1 - m)nl/(m - ml - mn + nl)$. Thus, both \tilde{p} and h are increasing in n . Since $U_1(p)$ and $U_{2,old}(p)$ are both increasing in p , $(U_1(\tilde{p}))'_n$ and $(U_{2,old}(l))'_n$ must be positive.

Hence, $U'_n > 0$ or U is increasing in n . ■

References

- [1] Arya, A., J. Glover, and S. Sunder. 1998. Earnings management and the revelation principle. *Review of Accounting Studies* 3: 7-34.
- [2] Beaver, W. H., M. F. McNichols, and K. K. Nelson. 2003. Management of the loss reserve accrual and the distribution of earnings in the property-casualty insurance industry. *Journal of Accounting and Economics* 35: 347-76.
- [3] Beaver, W. H., M. F. McNichols, and K. K. Nelson. 2007. An alternative interpretation of the discontinuity in earnings distributions. *Review of Accounting Studies* 12: 525-56.
- [4] Beyer, A. 2005. Capital market prices, management forecasts, and earnings management. Working paper, Northwestern University.
- [5] Brown, L., and M. Rozeff. 1978. The superiority of analysts' forecasts as measures of expectations: evidence from earnings. *Journal of Finance* 33: 1-6.
- [6] Burgstahler, D., and I. Dichev. 1997. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics* 24: 99-126.
- [7] Das, S., and B. Lev. 1994. Nonlinearity in the returns-earnings relation: test of alternative specifications and explanations. *Contemporary Accounting Research* 11: 353-79.
- [8] Dechow, P. M., S. A. Richardson, and I. Tuna. 2003. Why are earnings kinky? An examination of the earnings management explanation. *Review of Accounting Studies* 8: 355-84.
- [9] Dechow, P. M., R. G. Sloan, and A. P. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70: 193-225.
- [10] DeGeorge F., J. Patel, and R. Zeckhauser. 1999. Earnings management to exceed thresholds. *Journal of Business* 72: 1-33.
- [11] DeGroot, M. H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill.
- [12] Demski, J. 1998. Performance Measure Manipulation. *Contemporary Accounting Research* 15: 261-85.

- [13] Durtschi C., and P. Easton. 2005. Earnings Management? The shapes of the frequency distributions of earnings metrics are not evidence ipso facto. *Journal of Accounting Research* 43: 557-92.
- [14] Dutta, S. and S. Reichelstein. 2003. Leading indicator variables, performance measurement, and long-term versus short-term contracts. *Journal of Accounting Research* 41: 837-66.
- [15] Dye, R. 1988. Earnings Management in an Overlapping Generations Model. *Journal of Accounting Research* 26: 195-235.
- [16] Evans, J., and S. Sridhar. 1996. Multiple Control Systems, Accrual Accounting, and Earnings Management. *Journal of Accounting Research* 34: 45-65.
- [17] Fedyk, T. 2007. Discontinuity in earnings reports and managerial incentives. Working paper, UC Berkeley.
- [18] First Call, 1999. First Call historical database (FCHD) user guide. Boston: First Call Corporation.
- [19] Freeman, R. N., and S. Y. Tse. 1992. A nonlinear model of security price responses to unexpected earnings. *Journal of Accounting Research* 30: 185-209.
- [20] Fudenberg, D., and J. Tirole. 1991. Perfect Bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory* 53: 236-60.
- [21] Gu, Z., and T. Chen. 2004. Analysts' treatment of nonrecurring items in street earnings. *Journal of Accounting and Economics* 38: 129-70.
- [22] Guay, W., S. P. Kothari, and R. Watts. 1996. A market-based evaluation of discretionary accrual models. *Journal of Accounting Research* 34: 83-105.
- [23] Guttman, I., O. Kadan, and E. Kandel. 2006. A rational expectations theory of kinks in financial reporting. *The Accounting Review* 81: 811-48.
- [24] Hayn, C. 1995. The information content of losses. *Journal of Accounting and Economics* 28: 125-53.
- [25] Kaplan, R. S., and D. P. Norton. 1992. The balance scorecard: measures that drive performance. *Harvard Business Review* 70: 71-9.

- [26] Kirschenheiter, M., and N. D. Melumad. 2002. Can “big bath” and earnings smoothing co-exist as equilibrium financial reporting strategies? *Journal of Accounting Research* 40: 761-96.
- [27] Liang, P. J. 2004. Equilibrium earnings management, incentive contracts, and accounting standards. *Contemporary Accounting Research* 21: 685-717.
- [28] Skinner, D. J., and R. G. Sloan. 2002. Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. *Review of Accounting Studies* 7: 289-312.
- [29] Subramanyam, K. R. 1996. Uncertain precision and price reactions to information. *The Accounting Review* 71: 207-20.
- [30] Sunder, S. 1997. *Theory of Accounting and Control*. Cincinnati: South-Western Publishing.
- [31] Xin, B. 2007. Earnings forecast, earnings management, and asymmetric price response. Working paper, University of Minnesota.