The Complexity of CEO Compensation: Incentives and Learning*

Arantxa Jarque
Richmond Fed

(PRELIMINARY AND INCOMPLETE DRAFT)

May 15, 2014

Abstract

I study what are the firm characteristics that may justify the use of options or refresher grants in the compensation packages for CEOs as part of an optimal contract in the presence of moral hazard. I model explicitly the determination of stock prices from the output realizations of the firm: Symmetric learning by all players about the exogenous quality of the firm makes stock prices sensitive to output observations. Compensation packages become an instrument to transform this sensitivity of prices to output into the optimal sensitivity of consumption to output that is dictated by the optimal contract. Heterogeneity in the structure of firm uncertainty implies that some firms are able to implement the optimal contract with very simple schemes that do not contain options, refresher grants, or perks, while others necessarily need to use these more complex and non–transparent instruments.

Journal of Economic Literature Classification Numbers: D80, D82, D86, G30.

Key Words: mechanism design; moral hazard; CEO compensation; stock options; repricing; refresher grants; perks; learning

1 Introduction

It is widely accepted that in order to solve the agency problem between a CEO and the owners of the firm he works in, the compensation of the executive must be tied to the results of the firm. However, it is less clear how the efficient provision of incentives must be implemented in practice. The recent financial crises has revived an on–going debate about compensation practices for CEOs in US big public firms. Although the interest in explaining the level of compensation is still present,[1] there has been a shift towards understanding the *form* of compensation: is the use of certain instruments, like stock options, golden shakehands and parachutes, a sign of captured compensation boards and misaligned incentives? In a similar spirit, concerns that certain pay practices, like the use of options,

*I would like to thank Andy Atkeson, Marco Celentani, Huberto Ennis, Willie Fuchs, Juan Carlos Hatchondo, Hugo Hopenhayn, Ángel Hernando, Nobu Kiyotaki, Ned Prescott and audiences at the University of Iowa, the Feedback group at Carlos III, and Yonsei University for valuable comments.

[1] Academics have recently proposed explanations for the increase of the level of pay in the past decades, mainly based on assortative matching combined with the sharp increase in the size of firms during this period (see Gabaix and Landier (2008) and references therein).

may induce excessive risk taking have prompted regulatory agencies to increase their involvement in overseeing pay practices in the banking sector. Given the state of the debate and the increased will for intervention, a necessary first step is to enhance our understanding of the form of compensation packages that *are* consistent with a correct alignment of the interests of shareholders and the CEO. This is the objective of this paper.

Contract theory informs us about the properties of contracts that implement incentives optimally (Holmstrom (1979), Grossman and Hart (1983), Wang, 1997). But such characterizations are mainly given in a context that allows the transfers to the CEO to depend on signals of performance (accounting measures, stock prices) in a very general way (what I will refer to as "unrestricted" transfers). On the other hand, in real life we observe the use of a fairly limited set of compensation instruments, like bonus programs and stock grants, which depend on accounting measures and stock prices in specific ways. A commonly cited reason for the implementation through these type of instruments is their simplicity, which makes clear the ties of the compensation of the executive to the performance of the firm. This may facilitate both the communication of objectives of stockholders to CEOs, and the transparency of compensation practices to potential outside investors. Hence, especially in times of heavier scrutiny of CEO pay like the ones we have seen recently, compensation boards may find it in their interest to restrict their proposals for CEO pay to commonly used instruments that are clearly tied to firm performance measures. Another reason frequently cited for the proliferation of stock grants is the tax advantages of compensation contingent on performance over base salaries.[2] Whatever the reason, the fact is that most of the compensation given to the CEOs of the largest public firms in the US is given in the form of salary, bonus, restricted stock and option grants. Figure 1 illustrates this regularity for the CEOs of the largest 1,200 firms that are listed in the S&P index.[3]

In spite of this evidence, a common avenue taken in the literature that studies the properties of real life compensation instruments has been to exogenously restrict the class of instruments that are available to the firm, and derive the optimal scheme within that class.[4] The restriction on the number and generality of instruments is necessary for practical purposes, since the complexity of the optimization problem increases very fast with the number of instruments allowed. Although interesting insights on the way particular instruments may work are gained with this approach, it poses a big problem: we are restricting the firm to use an inefficient compensation scheme.

In this paper, I take a different approach: I look at the problem of implementing the optimal contract with a rich enough set of instruments, so that the transfers given to the CEO with the real–life instruments correspond to those derived assuming an unrestricted set of instruments. The issue that I want to study is not whether, for an exogenously given set of compensation instruments,

[2] The Omnibus Budget Reconciliation Act (OBRA) resolution 162(m) of 1992 imposed a one million dollar cap on the amount of non–performance based compensation of the top executives of the firm that qualifies for a tax deduction. Certain restricted stock and option plans are considered performance based pay.

[3] CEOs who owned more than 3% of the total stock of the firm at any point in the sample were considered "owners," i.e., not subject to a morak hazard problem, and they were dropped when constructing Figure 1.

[4] See for example Kandan and Swinkles (2007), Clementi, Cooley and Wang (2006), Assef and Santos (2005), Edmans and Liu (2010), or Bolton, Merah and Shapiro (2010). Alternatively, Edmans et al. (forthcoming) make assumtions on the structure of the model that imply that very simple contracts consisting of cash and stock implement the unrestricted optimal contract.

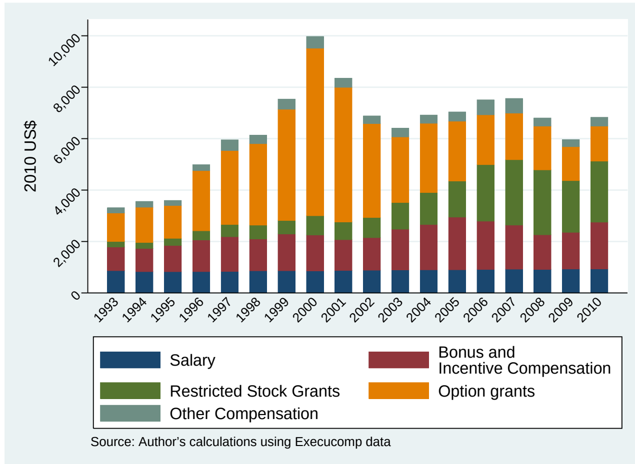


Figure 1: Relative importance of the different components of CEO compensation packages over time. Column height represents average compensation.

the optimal contract is feasible, but rather what are the necessary instruments to implement it. Just as in the data, the stylized compensation package that I consider will potentially include a salary, a bonus program, restricted stock, options and "other compensation" (which I will refer to as "perks").

My framework allows me to evaluate some counterintuitive compensation practices, such as that of issuing "refresher" grants — compensation that is set contingent on some new information on the performance of the firm, typically to substitute option grants that have gone well out of the money. Such practice arguably makes the compensation contract more complex, and potentially less transparent to outside investors. In my analysis, I find that, in some instances, it is optimal to commit to such complexity or lack of transparency: it is the least expensive way of providing incentives for the CEO to work hard in the interest of the shareholders.

I model the moral hazard problem between the owners of the firm and the CEO as a principal –agent problem. I propose a simple two period framework in which a risk averse CEO is asked to exert an unobservable and costly effort in the first period only. The risk neutral owners of the firm coordinate to act as the unique principal who designs the compensation package of the CEO. I assume commitment to the contract for the two periods for both the CEO and the firm owners, and I abstract from firing or quitting decisions. The effort of the CEO determines the distribution of the results of the firm in both the first and the second period. I interpret the first period as an interim stage at which information about performance is revealed (the company announces its earnings in the middle of the fiscal year). New grants may be awarded to the CEO at this interim stage (refresher grants), but no consumption takes place then; the CEO receives and consumes his compensation in period two only, after two outcome observations are available.

Under these assumptions, if potential (risk neutral) buyers of the stock of the firm value it according to the expected stream of future output, stock prices would not change contingent on the value of earnings announced in the interim stage. This is because, in equilibrium, the recommended level of effort is chosen. Hence, the expectations about output in the second period are independent of the first period realization. To capture the fact that stock prices vary in reality with firm results, I augment the model by introducing an exogenous source of uncertainty: a stochastic state that affects the effectiveness of the effort of the CEO. This can be interpreted as the quality of the match of the CEO and the firm, or as idiosyncratic market conditions for the firm. I assume that both the CEO and the owners, as well as potential buyers of the stock, have a prior about this state, which they update through Bayes' rule when they observe a new realization of the output of the firm. This generates contingent stock prices. As an important difference with the literature, these assumptions imply that in my framework the distribution of stock prices contingent on effort is endogenously generated through the learning process.

The interplay of the learning about the exogenous state and the optimal provision of incentives has important implications for the optimal contract, which in turn influence the composition of the compensation package. In particular, the sensitivity of compensation to price movements on the optimal contract may decrease with the cumulative output of the firm, i.e., optimal pay may be a concave function of cumulative output. This means that issuing options or stock grants after a bad sequence of results (refresher grants) is needed to implement a higher sensitivity of compensation following bad results than following good ones. In other words, the transfers implied by real life compensation instruments are (weakly) convex in prices, by their own nature; prices are weakly increasing and sometimes may be convex in output. In these cases, concavity of the optimal contract can only be achieved by granting new stock or options in the interim period, even after a bad earnings realization. What may look, to the uneducated eye, like undoing incentives, and a sign of entrenchment, may in fact arise as part of the optimal provision of incentives.

In section 3.3 I present my conclusions about the form of real life compensation packages. I define three types of pay schemes, according to the instruments that are included in each. My first distinction is for schemes that include "perks," i.e., forms of compensation that are not clearly tied to measures of performance (for example, perquisites, pension payments, life insurance premiums, subsidized loans, discounted share purchases or tax reimbursements). I label schemes that include perks as "non–transparent". In contrast, transparent schemes include only instruments that depend on output or stock prices, i.e. they may include a bonus program and both restricted stock or options. Within this category, I distinguish between simple schemes (including only restricted stock issued before any realization of output is observed,) or complex (which include stock options, or refresher grants issued contingent on the first period realization). I numerically characterize the combinations of parameters that determine whether a firm is able to implement the optimal contract with each type of scheme. In particular, I report the distribution of firms within the parameter space for which a simple scheme is feasible, and those for which a non–transparent one is necessary. I provide examples to illustrate the role of refresher grants in implementing the optimal contract without the use of perks (i.e., with complex but transparent schemes).

In the data publicly available about compensation practices (Execucomp), not every firm seems to have compensation packages of the same complexity. In Fig. 2 I illustrate this fact by presenting the percentage of firms that use stock or options in the compensation of their CEO in a given year.[5]

[5]These are all firms for which the CEO owns less than 3% of the stock, and which have a positive number

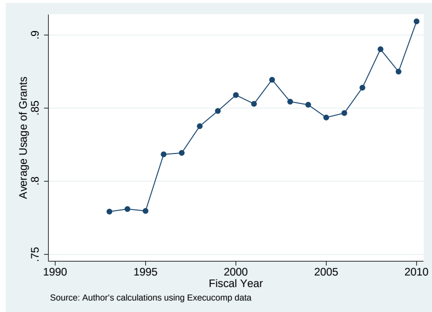


Figure 2: Portion of firms that include stock or options in their compensation package in a given year.

This percentage was at its lowest (77%) in 1993, and it is currently at its highest (90%). I discuss the link of the model to the data in more detail in section 5.

1.1 Related literature

In a related theoretical paper, Acharya, John and Sundaram (2000) present a model that implies that the practice of repricing options (i.e., changing the exercise price of options previously granted, typically to make options that are currently out–of–the–money to be at–the–money) can be optimal in a wide range of circumstances. The practice of repricing is, in practice, equivalent to that of issuing refresher grants.[6] Their model is also a two period problem, but they assume a risk neutral CEO, and both output and consumption occur only at the end of the second period. In their model there is one action every period, and the distribution of output in the second period depends on both actions. Also, there is a signal revealed in the first period, and the agent can make his action contingent on it. Although there is persistence through output, incentives in the second period are independent of actions taken in the first period, and their model reduces formally to a repeated moral hazard with consumption in the final period only. In their framework, the rigidity of options structure that they assume implies the following inefficiency: in some branches of the second period there are no instruments to provide incentives, and the agent does not exert (otherwise efficient) effort. Repricing in these branches allows to implement high effort, without lowering the utility of the agent. However, this repricing is anticipated by the agent in the first period and potentially increases the cost of the contract. The authors show that, in many cases, the benefits of adjusting incentives offset the cost of lack of commitment, and hence repricing is optimal.

In my model, I do not relax the assumption of commitment. Contrary to the case studied in Acharya, John and Sundaram (2000), refresher grants or repricing do not arise to reincentivate the

(regardless of its magnitude) in either stock or option grants in the Execucomp entry for the compensation of their CEO in a given year.

[6] The main difference may be that options granted outside of a shareholder pre–aproved long term option plan are not tax deductible; hence, repricing may be an attractive alternative for companies that have exhausted the options available in their option plans.

manager under lack of commitment. In my model, when repricing is optimal, the principal commits to it ex–ante, and only in the appropriate nodes of the game.[7]

In a related paper studying the choice between stock options or restricted stock as an incentive device, Kadan and Swinkles (2007) model the hidden effort of the CEO as affecting the distribution over stock prices directly. They assume this distribution is such that compensation is monotonic. They restrict compensation instruments to a wage and an option with an exercise price that is derived optimally. They allow for a continuum of effort choices and stock price realizations. Although their model is static, they model and explicitly study the role of previous outstanding grants, which imply a limit on the minimum compensation that the CEO gets for every current price realization. They show that for firms with high enough probability of the firm going into bankruptcy, the optimal compensation (within the restricted class of instruments) implies a zero exercise price, i.e., restricted stock is a better incentive instrument than options.

Despite the different modelling choices and the fact that their restriction on the instruments implies that the implemented contract is not the optimal one, the spirit of the exercise in Kadan and Swinkles (2007) is similar to the exercise in this paper: it studies the characteristics of real life instruments if they are to provide incentives to the CEO. The main contribution of this paper over Kadan and Swinkles (2007) is the fact that prices are endogenous, which helps understand an added difficulty in the use of compensation instruments that are contingent on stock prices: the possibility of non–monotonicities in the optimal compensation contract.

2 The model

I model the moral hazard problem that arises between the CEO and the owners of a firm due to the unobservability of the CEO's effort. I assume the CEO (agent) is risk averse with $u(c) = \ln(c)$. The firm is owned by well diversified shareholders that coordinate perfectly to act as the unique risk neutral principal of the agent. I also assume that there is a competitive stock market that prices the stock of the firm according to its expected per–period output. I assume that the owners of the firm can commit to implement a compensation contract, even if it was signed prior to their ownership of the stock. For simplicity, I also assume that the agent cannot save, and that all players discount the future at a common rate β.

The contract lasts for two periods, $t = \{1, 2\}$. The output of the firm can take two values each period, $y_L = 0$, and $y_H = 1$. The agent affects the probability distribution over output with his effort, which can take two values: e_L, with disutility of effort $v(e_L) = 0$, or e_H with $v(e_H) = e$. This effort is done only once, at the beginning of period 1, but it affects the distribution of output both at period 1 and period 2. Hence, the effect of effort is persistent in time. The agent receives his payment only at the end of the second period. The first period represents an interim stage at which new information is revealed (the first period output realization is observed), but no consumption or effort takes place in it.

[7] It is worth noting that the equilibrium of my model is not robust to allowing for ex–post Pareto improving renegotiations. Once the effort is done, it is optimal to renegotiate to uncontingent payments in the consumtion stage. However, thsi is anticipated by the agent ex–ante, and implementing high effort is no longer feasible.

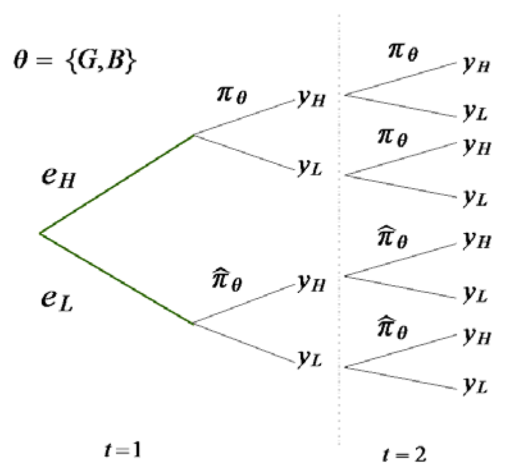
The distribution over output is also affected by another parameter: a state that determines the effectiveness of the effort of the CEO, denoted $\theta \in \{A, B\}$. The true state is unknown by both the agent, the principal, and the stock market, and all players attach a prior probability of q_0 to $\theta = A$. The probability of observing a high output contingent on an effort level and a realization of the match is as follows, for every t:

$\Pr(y_t = y_H \mid e, \theta)$	(q_0) A	$(1 - q_0)$ B
e_H	π_A	π_B
e_L	$\widehat{\pi}_A$	$\widehat{\pi}_B$

(1)

where I assume $\pi_A = \pi$, $\widehat{\pi}_A = \widehat{\pi}$, and $\pi_B = \widehat{\pi}_B = 1$. In such a firm, when effort is not effective the firm always produces high output. This is a simplifying assumption that I will relax in section 4, when I will consider firms for which output is always low in state B ($\pi_B = \widehat{\pi}_B = 0$), and firms for which effort is also effective in state B, but it implements different probabilities than in state A. To distinguish this type of firm in matrix 1 from other types introduced later, I will refer to it as a type $\mathcal{H}$ firm.

All probabilities are common knowledge. I assume that the prior over $\theta = A$ satisfies $0 < q_0 < 1$. Also, higher effort (e_H) implies higher probability of observing y_H in $\theta = A$, that is, $1 > \pi > \widehat{\pi} > 0$. The timing and the stochastic structure are depicted in Fig. 1. The above assumptions on the probabilities imply that, at time 0, all the nodes of the tree have positive probability of being reached under both levels of effort.



Timing and probabilities of output realizations.

Matrix 1 constitutes a very stylized model of a firm's technology. However, moral hazard and learning are present in this technology, complicating the analysis of compensation that is the objective of this paper.

2.1 Learning about the quality of the match

Each period, after observing the output realization, both the principal, the agent, and the stock market update their prior about the quality of the match. The updating is done using Bayes' rule, for a given e choice. I denote the posteriors in the first period when the agent chooses e_H as q_i, and those in the second period as q_{ij}, for $i = L, H$, $j = L, H$. Similarly, the posteriors when the agent chooses e_L are denoted $\widehat{q}_i$ and $\widehat{q}_{ij}$. To simplify the exposition, I introduce the following notation:

$$\begin{aligned}\pi_0 &= q_0\pi_A + (1 - q_0)\,\pi_B \\ \pi_i &= q_i\pi_A + (1 - q_i)\,\pi_B, \quad i = L, H,\end{aligned}$$

which denotes the probability attached by all players to observing a high output realization in the first period (π_0), and in the second period, contingent on the first period realization (π_L and π_H). Similarly, for e_L, the corresponding probabilities are denoted by $\widehat{\pi}_0$, $\widehat{\pi}_L, \widehat{\pi}_H$.

2.2 Valuation of the firm by outside investors

I assume that there is a large number of investors in the economy who are willing to buy the stock of the firm. Investors value the stock of the firm as a claim to the expected stream of output that the firm will generate in all future periods. I also assume a large number of shareholders (sellers of the stock,) so no individual deviation affects the equilibrium price. Competition implies a price equal to the expected output. Investors and shareholders update their beliefs about θ, for a given effort e. Hence, the market price for the stock varies as output realizations become available.

In order to simplify my analysis of the compensation problem, and without loss of generality, I introduce the following normalization: I assume the firm produces output for an infinite number of periods, and I normalize the price to the expected per–period value of the firm:[8]

$$\begin{aligned}p\left(y^t, e\right) &\equiv (1 - \beta)\, E_t\left[\sum_{t=1}^{\infty} \beta^{t-1} y_t \mid e\right] \\ &= E\left[y_t \mid e, y^t\right],\end{aligned}$$

where $E_t\left[\cdot\right]$ denotes the expectation taken with all information known at t, which is summarized by the history of realizations, y^t. Since I limit compensation schemes to be contingent only on the first two period stock prices, the stochastic process that drives them after period two is irrelevant for my analysis. For consistency,[9] I assume that the probability of observing a high output after period two is fixed to the updated probability at the end of period 2, i.e.

$$\Pr\left(y_t = y_H | e, y^{t-1}\right) = \Pr\left(y_3 = y_H | e, y^2\right) \;\; \forall t > 2.$$

[8] If instead I were to match the life of the firm to that of the contract (i.e., if the firm lived only for two periods,) the price in the first period would represent a claim to two output realizations, while the price in the first period would be a claim only to one output. This would imply a difference in the level of prices that is irrelevant for the economic problem of interest in this paper, but complicates the algebra.

[9] Exogenously limiting the contract to not use prices after $t = 2$ is qualitatively equivalent to assuming $\Pr\left(y^t | e, y^{t-1}\right) = \Pr\left(y^t | e, y^2\right)$ for all $t > 2$, and letting the firm choose the future prices that the compensation contract can be contingent on.

I introduce the following notation for prices. The price of the stock corresponds to the expected value of the firm given the history of realizations, and a given effort choice:

$$p_0 \equiv p(\emptyset, e_H) = q_0\pi_A + (1 - q_0)\pi_B, \tag{2}$$

$$p_i \equiv p(y_i, e_H) = q_i\pi_A + (1 - q_i)\pi_B \quad i = L, H, \tag{3}$$

$$p_{ij} \equiv p(y_i, y_j, e_H) = q_{ij}\pi_A + (1 - q_{ij})\pi_B \quad i = L, H, \ j = L, H, \tag{4}$$

under high effort, and similarly under low effort:

$$\begin{aligned} \widehat{p}_0 &\equiv p(e_L) = q_0\widehat{\pi}_A + (1 - q_0)\widehat{\pi}_B \quad i = L, H, \\ \widehat{p}_{ij} &\equiv p(y_i, e_L) = \widehat{q}_i\widehat{\pi}_A + (1 - \widehat{q}_i)\widehat{\pi}_B \quad i = L, H, \\ \widehat{p}_{ij} &\equiv p(y_i, y_j, e_L) = \widehat{q}_{ij}\widehat{\pi}_A + (1 - \widehat{q}_{ij})\widehat{\pi}_B \quad i = L, H, \ j = L, H. \end{aligned}$$

2.3 Compensation Packages

In this section I define the compensation instruments available to the firm. I allow the compensation package to include the following elements: an annual wage, a bonus plan, perks, and long term performance–based plans that include both stock and option grants. With these elements I try to capture the most important features of real–life compensation practices.[10] In the year 2010, data in Execucomp for the CEOs of the 1,500 largest public companies in the US shows that the average pay was \$4,371,060, with a minimum of \$200,000 and a maximum of \$25,761,432. The median pay of the highly skewed distribution of pay was of \$3,022,00.[11] Of this total pay, the salary represented an average of a 25% (or a median 19%), the bonus and incentive program represented an average of a 25% (or a median of 23%), stock grants a 28% (median of 25%), option grants an average of 19% (median of 13%), and perks and other compensation an average of 3% (median of 1%).

I now present a brief description of each instrument and how it is captured in the model.

Base Salaries

In real life, salaries for CEOs are normally negotiated at the time of signing a contract, based on industry benchmarks. The negotiation usually includes a pre–specified annual increase for the duration of the contract, independently of performance.

In the model the salary is a constant payment given in period 2. I denote the salary as W.

Bonus plans

In real life, virtually all companies offer bonus plans paid annually based on the current year's performance only. They usually specify a performance target, together with a minimum and maximum limit for bonuses and the sensitivity of the bonus to the performance measure. These performance measures consist mainly on objective measures such as net–income, revenue, pre–tax income,

[10] See Murphy (1999) for a detailed description of compensation instruments based on compensation surveys. See Clementi and Cooley (2009) for a recent and careful description of the main facts related to the level and structure of compensation of the executives of the largest public US firms in the last two decades.

[11] This calculation excludes CEOs that at some point in their tenure with their firms owned more than 3% of the total stock of their company; in picking this threshole I follow Clementi and Cooley (2010), who argue that such a high ownership is not consistent with a moral hazard problem.

or other accounting figures. About a 25% of the total measures used in the evaluation are labeled as "individual performance" measures, which are subjective evaluations.

In the model, I summarize these characteristics by making the bonus plan depend only on accumulated annual output, $(y_1 + y_2)$, in two possible variations, as follows: $B = B(y_1, y_2) = \min\{b, b(y_1 + y_2)\}$. This mimics the structure of bonus programs in real life, were a "pool" available for bonuses determines a "cap" (b, in this case) for annual payments.[12]

Perks

This instrument includes categories such as personal benefits, pension payments, perks, life insurance premiums, subsidized loans, discounted share purchases or tax reimbursements. This fraction of compensation is not clearly tied to any objective of performance.

In the model all these payments are simply a transfer that is contingent potentially in the whole history of realizations. I simply refer to this category as "perks" throughout the paper, and denote them as k_{ij}, where $i, j = L, H$ represent the first and second period output realizations.

Long term plans

In real life, compensation plans include long term payments in the form of (i) stock of the company and (ii) options to buy stock at a pre–determined price (the "exercise" price or "strike" price.) Both come with selling restrictions: they cannot be traded before their "vesting" time. Also, the manager cannot take these grants with him if he leaves the company, and he is not allowed to hedge against the risk in his compensation package.

In the model I assume that all grants vest in period 2, and they are exercised immediately by the CEO. Consistently with real life practices, I assume an exercise price equal to the market price of the stock at the time of granting. I denote the long–term plans as follows:

- r_0: restricted stock issued in period 0
- r_i, for $i = L, H$: restricted stock issued in period 1, contingent on realization i being observed in period 1
- s_0 : stock option grant issued at time 0, with exercise price p_0
- s_i, for $i = L, H$: stock option grant issued at time 1, contingent on realization i being observed in period 1, with exercise price p_i

I denote the set of compensation packages as:

$$\mathbb{P} = \left\{ \begin{array}{c} \left(W, B, r_0, s_0, \{r_i\}_{i=L,H}, \{s_i\}_{i=L,H}, \{k_{ij}\}_{i,j=H,L}\right) \\ s.t. \;\; W, b, r_0, s_0, \{r_i\}_{i=L,H}, \{s_i\}_{i=L,H}, \{k_{ij}\}_{i,j=H,L} \in \mathbb{R}_+, \\ and \; B = \min\{b, b(y_1 + y_2)\} \end{array} \right\}.$$

I denote an arbitrary element of $\mathbb{P}$ as P.

I assume that the principal can perfectly control the savings of the agent, and force him to sell all stock and options when they are profitable, and consume all income generated from it. Hence,

[12] Results for an alteraive linear bonus program, were $B = B^L(y_1, y_2) = b(y_1 + y_2)$, are similar and are available upon request.

I introduce the following notation to denote the consumption of the agent as a function of the compensation package: $\mathcal{C}(P) = \{c_{ij}\}_{i,j=L,H}$. For any $P \in \mathbb{P}$, $\mathcal{C}(P)$ takes the following form:

$$\begin{aligned}
c_{HH} &= W + b + r_0 p_{HH} + s_0 (p_{HH} - p_0) + r_H p_{HH} + s_H (p_{HH} - p_H) + k_{HH} \\
c_{HL} &= W + b + r_0 p_{HL} + \max\{s_0 (p_{HL} - p_0)\} + r_H p_{HL} + k_{HL} \\
c_{LH} &= W + b + r_0 p_{HL} + \max\{s_0 (p_{HL} - p_0)\} + r_L p_{HL} + s_L (p_{HL} - p_L) + k_{LH} \\
c_{LL} &= W + r_0 p_{LL} + r_L p_{LL} + k_{LL}.
\end{aligned} \tag{5}$$

It is important to note that the function $\mathcal{C} : \mathbb{P} \to \mathbb{R}_+^4$ is not injective, that is, different compensation packages may imply the same contingent consumption vector. Using the function $\mathcal{C}$ we can calculate the expected utility of the agent. For a given compensation package $P \in \mathbb{P}$ and high effort, this expected utility is:

$$\begin{aligned}
U(P, e_H) &= \beta^2 \pi_0 [\pi_H \ln(c_{HH}) + (1 - \pi_H) \ln(c_{HL})] \\
&\quad + \beta^2 (1 - \pi_0) [\pi_L \ln(c_{LH}) + (1 - \pi_L) \ln(c_{LL})] - e.
\end{aligned}$$

In the same way, if the agent were to choose low effort, his expected utility would be:

$$\begin{aligned}
U(P, e_L) &= \beta^2 \widehat{\pi}_0 [\widehat{\pi}_H \ln(c_{HH}) + (1 - \widehat{\pi}_H) \ln(c_{HL})] \\
&\quad + \beta^2 (1 - \widehat{\pi}_0) [\widehat{\pi}_L \ln(c_{LH}) + (1 - \widehat{\pi}_L) \ln(c_{LL})].
\end{aligned}$$

Finally, the cost to the principal of a contract P that implements e_H is

$$\begin{aligned}
K(P, e_H) &= \beta^2 \pi_0 [\pi_H c_{HH} + (1 - \pi_H) c_{HL}] \\
&\quad + \beta^2 (1 - \pi_0) [\pi_L c_{LH} + (1 - \pi_L) c_{LL}].
\end{aligned}$$

The cost $K(P, e_L)$ is constructed in a similar manner, changing the probabilities to those corresponding to low effort.

2.4 Incentive problem

With the compensation packages and the consumption function in hand, we are now ready to write the optimization problem of the principal. I assume throughout that parameters are such that it is always profitable to implement e_H. Hence, the optimal compensation package P^* is the solution to the following cost minimization problem, where $\underline{U}$ represents the outside utility the agent would obtain if he were not to participate in the contract:

$$V(P) = \min_{P \in \mathbb{P}} \quad K(P, e_H) \qquad s.to \tag{\textbf{P1}}$$

$$\underline{U} \leq U(P, e_H) \tag{PC}$$

$$U(P, e_H) \geq U(P, e_L) \tag{IC}$$

$$W, b, r_0, r_L, r_H, s_0, s_L, s_H, k_{LL}, k_{LH}, k_{HL}, k_{HH} \geq 0. \tag{NNC}$$

Problem P1 is difficult to solve in general, due to the large amount of non negativity constraints in (NNC). I propose, instead, to solve a simplified problem in which the principal chooses directly a tuple $C = \{c_{ij}\}_{i,j=L,H}$ of transfers contingent on the history of output realizations, as follows:

$$V(C) = \min_C \quad K(C, e_H) \tag{\textbf{PS}}$$
$$s.to$$
$$\underline{U} \leq U(C, e_H) \tag{PC'}$$
$$U(C, e_H) \geq U(C, e_L) \tag{IC'}$$
$$c_{LL}, c_{LH}, c_{HL}, c_{HH} \geq 0. \tag{NNC'}$$

I denote the solution to PS as $C^* \equiv \left\{c^*_{ij}\right\}_{i,j=L,H}$. The standard arguments valid for a static moral hazard problem (see Grossman and Hart, 1983) justify that both the PC and the IC constraints bind in the optimum. Note that the agent has logarithmic utility, so the non–negativity constraints (which are now in terms of consumption levels) will never bind. Also, with a simple change of choice variables to utility levels, the objective function is linear and the constraint set is compact and convex, so the solution to PS exists and is unique.

Lemma 1 *Any solution P^* to problem P1 implements the same consumption for the agent as the solution C^* to the simplified problem PS.*

It is easy to see that the set of available compensation instruments $\mathbb{P}$ is rich enough to implement any (positive) transfer scheme contingent on the history of output realizations, i.e., any value for the tuple $\{c_{ij}\}_{i,j=L,H}$. This result implies that I can study the problem of choosing the instruments separately from the determination of contingent consumption in the optimal contract. However, since the function $\mathcal{C}(P)$ is not invertible there might be several compensation packages that solve problem $P1$ and satisfy $\mathcal{C}(P) = C^*$

3 Equilibrium

Recall from section 2.2 that individual deviations of the shareholders and the investors do not affect the stock prices in the equilibrium of the stock market. This implies that there are only two pricing rules that may appear in any equilibrium: one for any contract that implements e_L and one for any contract that implements e_H. By changing his effort, the CEO can only affect the probability distribution over prices, but not the prices themselves.

An equilibrium of the above game between the principal, the agent and the stock market is defined next.

Definition A Perfect Bayesian Equilibrium of this game in which effort e_H is implemented consists of a compensation contract P^*, and stock prices such that *a)* $\mathcal{C}(P^*) = C^*$, where $C^* = \arg\min_{\{C\}} K(C, e_H)$ and ,

b) The utility of the agent choosing e_H is higher than if choosing e_L, and is as large as his outside utility $\underline{U}$,

c) Market prices and the beliefs of the stockmarket participants about θ are consistent with the agent choosing e_H, as defined by $p_0, p_L, p_{LL}, p_{LH}, p_{HL}, p_{HH}$ in Equations (2) -(4)

d) Beliefs about θ are updated according to Bayes' rule.

Since the probability of observing any history is positive under the equilibrium level of effort, Bayesian updating provides consistent beliefs, and no refinement is necessary.

In the next subsections I describe the properties of the equilibrium.

3.1 Equilibrium Stock prices

All stock traders anticipate that, in equilibrium, the agent chooses the recommended level of effort, e_H. Hence, they update their beliefs using the probabilities in the above matrix corresponding to e_H. The equilibrium price of the stock corresponds to the expected value of the firm given the history of realizations, given by p_i and p_{ij} in Equations (2) -(4). For the rest of the analysis in the paper, it will be useful to keep in mind the following property of stock prices:

Lemma 2 *Stock prices are monotonic in the period's output:* $p_L \leq p_H$, *and* $p_{LL} \leq p_{LH} = p_{HL} \leq p_{HH}$.

In particular, for any histories containing at least one y_L, the updated believes put probability one on $\theta = A$, i.e., we have that $q_{ij}^{\mathcal{H}} = 1$ if i or j equals L. If the observed history does not contain any y_L, instead, $\theta = B$ has still positive probability. This is the case for histories y_H and (y_H, y_H). That is,

$$\begin{aligned}
q_H &= q_0 \frac{\pi}{q_0 \pi + 1 - q_0}, \\
q_{HH} &= q_0 \frac{\pi^2}{q_0 \pi^2 + 1 - q_0}, \\
q_L &= q_{LH} = q_{LL} = 1.
\end{aligned}$$

A direct implication of this learning is that the stock prices take the simple form:

$$\begin{aligned}
p_0 &= q_0 \pi + 1 - q_0, \qquad (6) \\
p_H &= q_H \pi + 1 - q_H, \\
p_{HH} &= q_{HH} \pi + 1 - q_{HH}, \\
p_L &= p_{HL} = p_{LL} = \pi.
\end{aligned}$$

3.2 Equilibrium Consumption

Problem **PS** is a particular example of a static moral hazard problem, with i.i.d. output and exogenous uncertainty about the probability distribution implemented by each effort level. The characterization of the optimal contract with unrestricted instruments follows easily from the standard first order conditions of problem **PS**. Define the likelihood ratio (LR) of a history of realizations as the ratio of the expected probabilities of that history under low and high effort:

$$\begin{aligned}
LR_{HH} &= \frac{\widehat{\pi}_0}{\pi_0}\frac{\widehat{\pi}_H}{\pi_H} = \frac{q_0\widehat{\pi}^2 + 1 - q_0}{q_0\pi^2 + 1 - q_0}, \\
LR_{HL} &= \frac{\widehat{\pi}_0}{\pi_0}\frac{1-\widehat{\pi}_H}{1-\pi_H} = \frac{(1-\widehat{\pi})\,\widehat{\pi}}{(1-\pi)\,\pi}, \\
LR_{LH} &= \frac{1-\widehat{\pi}_0}{1-\pi_0}\frac{\widehat{\pi}_L}{\pi_L} = \frac{(1-\widehat{\pi})\,\widehat{\pi}}{(1-\pi)\,\pi}, \\
LR_{LL} &= \frac{1-\widehat{\pi}_0}{1-\pi_0}\frac{1-\widehat{\pi}_L}{1-\pi_L} = \frac{(1-\widehat{\pi})^2}{(1-\pi)^2}.
\end{aligned} \tag{7}$$

Proposition 1 *Consumption levels in the optimal contract C^* are ranked by likelihood ratios:*

$$c_m^* > c_n^* \Leftrightarrow LR_m < LR_n \quad , \textit{ for } n, m \in \{LL, LH, HL, HH\}\,.$$

Moreover, consumption is linear in the LR:

$$c_m^* = \lambda + \mu\,(1 - LR_m)\,, \tag{8}$$

where λ is the multiplier of the constraint PC and μ that of the constraint IC.

It is worth noting that, if θ were A for sure, the above characterization would always imply the same ranking for consumptions for all combinations of parameters, as the next proposition states. Define $\Delta_L \equiv c_{LH}^* - c_{LL}^*$ and $\Delta_H \equiv c_{HH}^* - c_{HL}^*$.

Proposition 2 *In the absence of learning about θ (certainty case with $\theta = A$), the optimal consumption is monotonic in output and it satisfies:*

$$0 < \Delta_H < \Delta_L.$$

With uncertainty about θ, however, the posterior evolves differently under e_L than e_H, changing the weight of each probability in the numerator and denominator of the LR. As stated in the next proposition, this can create non–monotonicities. The intuition for the existence of non–monotonicities in the optimal contract is that learning shifts the weights given to certain histories in the provision of incentives, with respect to the benchmark case characterized in Prop. 2. Since the value of θ is not controllable by the agent, the principal would like to insure him against this risk. However, under such a contract, the agent would shirk and blame poor performance on a bad realization of θ. The optimal contract, hence, demands exposing the agent to some θ-related risk. This, in turn, may lead to non–monotonicities in consumption, since the principal evaluates the relative likelihood of effort and learns about the realization of θ at the same time. For example, a high outcome following a low one may increase the likelihood that the first observed (low) output was the result of low effort; hence, a high second period outcome is "bad news" for the agent. In other words, the weights given to each θ's probability distribution (i.e., the posteriors,) are different for high and low efforts, so the ordering of each θ's probabilities is not preserved in the probability unconditional on θ.

As pointed out previously, for our benchmark firm we have $q_L = q_{LH} = q_{LL} = 1$. This may mean that, when the first period output has been y_H, the agent's wage may be higher if we observe y_L in the second period than if we observe y_H. This is because observing y_L in the second period reveals that $\theta = A$. In state A, the first period observation y_H makes the history's likelihood ratio be much lower (it is a much less likely history under low effort than it would be if $\theta = B$).

We can use the likelihood ratios in 7 to establish the following properties of consumption in the optimal contract:

Proposition 3 *When the firm is of type $\mathcal{H}$, consumption spread always satisfies $\Delta_H < \Delta_L$. Also, non-monotonicities never arise in the lower consumption, i.e., $\Delta_L > 0$ always. Moreover, we have*

(i) *whenever $\pi + \widehat{\pi} > 1$, for all $q_0 \in (0,1)$, $0 < \Delta_H$,*

(ii) *whenever $\pi + \widehat{\pi} < 1$,*

$$\begin{aligned} \text{for } q_0 &\in (0, q^*),\ 0 < \Delta_H, \\ \text{for } q_0 &\in (q^*, 1),\ \Delta_H < 0, \end{aligned}$$

where $q^* = \frac{\pi\widehat{\pi}}{1-\pi-\widehat{\pi}+\pi\widehat{\pi}}$.

The fact that non-monotonicities only arise in Δ_H, and only when $\pi + \widehat{\pi} < 1$, is related to the interaction of the informativeness of the signal (y_L, y_H) (or, equivalently, (y_H, y_L)) and the learning about the true state. On one hand, $\pi + \widehat{\pi} < 1$ implies that LR_{LH} is less than one, and hence the optimal contract seeks to reward the agent when observing LH, as well as when observing HH. Punishments are reserved for LL. However, observing LH reveals perfectly that the true state is A, making y_H a more informative signal about effort than if there were still positive probability on state B (which is the case when we observe HH). This tends to make c_{LH} large, but not c_{HH}, making Δ_L large and Δ_H small. For high enough q_0, the relative informativeness of LH and HH may be reversed and we may get $\Delta_H < 0$.

We conclude this section summarizing the properties of the optimal contract derived from the above propositions:

1. Contingent consumption is ranked by the LR of output realization histories.
2. Since $LR_{HL} = LR_{LH}$, we have $c_{HL} = c_{LH}$.
3. Consumption in the optimal contract may be non monotonic in output in the second period: $c_{HL} > c_{HH}$
4. We have $\Delta_H < \Delta_L$ always.

3.3 Equilibrium compensation packages

The analysis of the properties of equilibrium consumption in the previous section was based on the solution to problem **PS**, with contingent consumption transfers C^*. In this section, we use the $\mathcal{C}(C^*)$ mapping, together with the properties of C^*, to analyze the characteristics of the solution to the original problem **P1** in terms of compensation packages P in $\mathbb{P}$.

The first thing to note is that given the richness of the elements of $\mathbb{P}$, the optimal contract C^* characterized in section 3.2 is always feasible in a trivial way: because of the availability of perk payments, the firm can simply set $k_{ij} = c^*_{ij}$ for all ij pairs. However, there may be other combinations of compensation instruments that implement a given optimal contract. To solve the indeterminacy of the compensation package, I assume that the principal, when presented with several choices to implement a given contingent consumption scheme, chooses the simplest possible. That is, I study the properties of the compensation packages that implement the optimal contract C^* with the *simplest possible compensation package*, and the *most transparent* one. I define simplicity and transparency next.

Consider the following strict subsets of $\mathbb{P}$:

$$
\begin{aligned}
\mathbb{S} &= \{P \in \mathbb{P} \text{ such that } s_0, r_i, s_i, k_{ij} = 0 \ \forall i, j\}, \\
\mathbb{C} &= \{P \in \mathbb{P} \backslash \mathbb{S} \text{ such that } k_{ij} = 0 \ \forall i, j\}.
\end{aligned}
$$

Definition A compensation package P is classified as:

- **Transparent** if it does not include any perks, i.e. $P \in \mathbb{C} \cup \mathbb{S}$,
  - **Simple**: if it is transparent and it includes only a wage, bonus scheme and restricted stock granted at time 0, i.e., $P \in \mathbb{S}$.
  - **Complex**: if it is transparent and it includes at least one option grant or a refresher stock grant, but no perks, i.e., $P \in \mathbb{C}$.
- **Non-transparent**: if it includes at least one contingent perk payment, which dependence on performance is not transparent to an outsider, i.e., $P \notin \mathbb{C} \cup \mathbb{S}$.

In the rest of the paper, I ask the following questions: What types of firms are not able to use a transparent scheme, and which are? Which can do with just a simple scheme?

To answer these questions, I use the following strategy. First, I spell out $\mathcal{C}(P)$ under the restrictions implied by a simple and a complex scheme, with each of the two types of bonus programs that I consider. Then, I analyze the system of equations resulting from equating $\mathcal{C}(P) = C^*$.

Definition Consider a firm defined by a probability structure of the form of matrix (1). A compensation scheme P is feasible if, for the C^* corresponding to the parameter values that describe the firm, the system of equations resulting from equating $\mathcal{C}(P) = C^*$ has a solution and this solution satisfies the non–negativity constraint in (NNC).

Although restricting to schemes in the subsets $\mathbb{S}$ and $\mathbb{C}$ simplify the system $\mathcal{C}(P) = C^*$, it is difficult to characterize the solution (or even to check constraint NNC) in general. In what follows, I present a series of results that constitute a partial characterization of the choice of compensation packages between simple, complex or non–transparent. I complement my analysis with a complete numerical characterization of this choice.

A simple benchmark: no learning

As a preview of the method that I use to establish feasibility of the different compensation packages, and for benchmark purposes, I analyze first the no–learning case discussed in Prop. 2).

Proposition 4 *If there is no learning, compensation packages must be non–transparent.*

The proof is simple so I include it in the text. It is easy to see that no P in $\mathbb{S}$ or $\mathbb{C}$ is feasible by looking at the system $C^* = \mathcal{C}(P)$, given that all prices are equal to p. For a linear bonus,

$$\begin{aligned} c^*_{LL} &= W + r_0 p, \\ c^*_{HL} &= W + b + r_0 p, \\ c^*_{HH} &= W + 2b + r_0 p. \end{aligned}$$

It is useful to write these equations as a function of the consumptions spreads:

$$\begin{aligned} c^*_{LL} &= W + r_0 p, \\ \Delta_L &= b, \\ \Delta_H &= b. \end{aligned}$$

Any solution implies $b = \Delta_L$, but also $\Delta_H = \Delta_L$, which is never true because $\Delta_H < \Delta_L$ by Prop. 2. If a capped bonus were used instead, the system would read:

$$\begin{aligned} c^*_{LL} &= W + r_0 p, \\ \Delta_L &= b, \\ \Delta_H &= 0. \end{aligned}$$

In this case, we would need $\Delta_H = 0$, which, again by Prop. 2, is not possible.

Figure 3 illustrates an example without learning in which $q_0 = 1,\ \pi = 0.8$, and $\widehat{\pi} = 0.4$. It is easy to see graphically that having only one price implies that any difference between c^*_{LL}, c^*_{LH} and c^*_{HH} needs to be implemented with the bonus program exclusively, and this is not feasible if it takes one of the standard forms (linear or capped).

3.3.1 Non–transparent schemes

By simple inspection of the system $C^* = \mathcal{C}(P)$, we can see that payments to the agent are necessarily monotonic in output, and hence monotonic in prices, since prices are themselves monotonic in output. As the following proposition describes, monotonicity of the optimal consumption is both a necessary and a sufficient condition for a complex scheme to be feasible

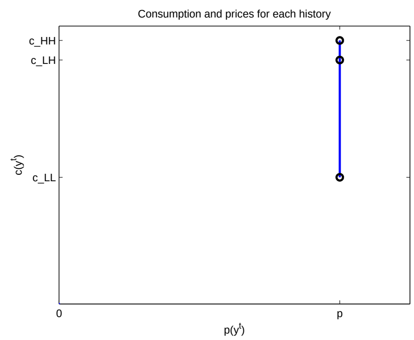


Figure 3:

Proposition 5 *For a firm of type $\mathcal{H}$, compensation schemes must be non–transparent if and only if non–monotonicities arise ($\Delta_H < 0$).*

Figure 4 presents graphically the analytical characterization in Prop. 3 of the parameters that imply a non–monotonicity. For this, we assume that, for each feasible combination of π and $\widehat{\pi}$, there is a mass one of firms whose prior q_0 is distributed uniformly between 0 and 1. The vertical axis, then, represents the proportion of firms for which non–monotonicity is present in the optimal contract. We see that, when π is high enough, or the difference $\pi - \widehat{\pi}$ is large enough (both cases that lead to $\pi + \widehat{\pi} > 1$), consumption is monotonic. For the combinations that imply non–monotonicities, with $\pi + \widehat{\pi} < 1$, it is the firms with the largest priors, $q_0 > q^*$ as defined in Prop. 3, that need to use non–transparent schemes.

3.3.2 Simple schemes

In this section, I provide necessary and sufficient conditions for the feasibility of a simple compensation package in the presence of learning. One caveat is that the conditions will not be, in general, in terms of the primitive parameters of the firm. I complement the analytical derivations with numerical characterizations.[13]

Proposition 6 *A simple scheme is feasible if and only if*

a) $c^*_{LL} > \frac{\pi_B}{q_{HH}(1-\pi_B)} \Delta_H$, *and*

b) $\Delta_H > 0$, or, equivalently, $q_0 > q^* = \frac{\pi\widehat{\pi}}{1-\pi-\widehat{\pi}+\pi\widehat{\pi}}$ *(see Prop. 3).*

The conditions in this proposition are restrictions that the structure of compensation instruments imposes on the sensitivity of the consumption of the agent to changes in stock prices. This sensitivity is sort of a reduced form for the composition of the sensitivity of consumption to signals (i.e., output realizations), which is dictated by the likelihood ratios, and the sensitivity of prices to

[13] The proof of Proposition 6 follows from the proof of a more general statement, Prop. 15, included in Appendix 2.

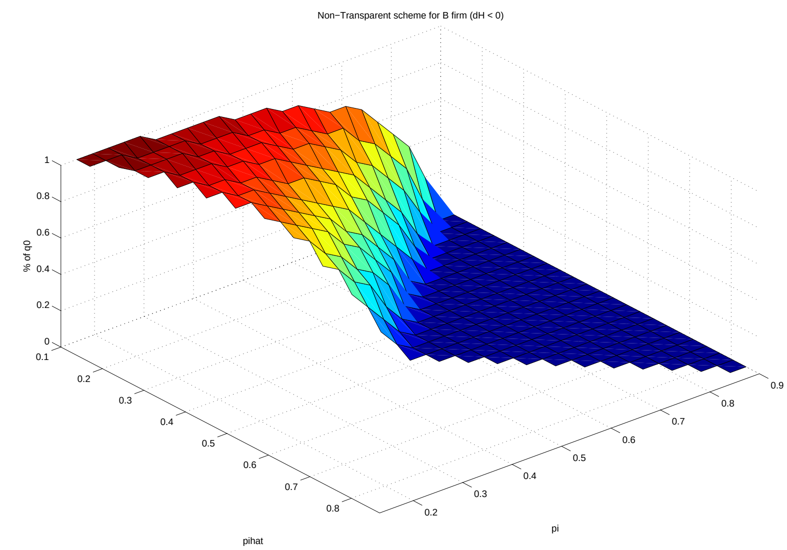


Figure 4: Type $\mathcal{H}$ firm. Probability that non–monotonicities arise and a Non-transparent scheme is necessary.

signals, which is in turn dictated by the pricing rule of outside investors. The proposition shows that a limited set of compensation instruments like $\mathbb{S}$ puts severe restrictions on the relationship of these two sensitivities.

In particular, the proposition can be otherwise stated as a simple scheme being feasible whenever the following solution satisfies the non–negativity constraint for the three instruments:

$$\begin{aligned} W &= c_{LL}^{*} - \pi \frac{\Delta_H}{(1-\pi)\left(1-q_{HH}^{\mathcal{H}}\right)}, \\ b &= \Delta_L, \\ r_0 &= \frac{\Delta_H}{(1-\pi)\left(1-q_{HH}^{\mathcal{H}}\right)}. \end{aligned}$$

For a firm of type $\mathcal{H}$, any number of low outputs is perfectly informative about the state. Hence, there is no variation in prices in the lower range, implying that the spread $c_{LH} - c_{LL}$ needs to be implemented with the bonus payout: $b = \Delta_L$. Setting the bonus b to satisfy this constraint is always feasible, since $\Delta_L > 0$ always for a firm of type $\mathcal{H}$. Since the bonus is capped, the spread $c_{HH} - c_{HL}$ needs to be implemented with restricted stock:

$$\Delta_H = r_0 \left(p_{HH} - p_{LH}\right).$$

This is feasible whenever the optimal consumption is monotonic, so that we have $\Delta_H > 0$ (condition b) in the above corollary). For a firm of type $\mathcal{H}$, high output levels are purely out of luck if we are in state B, which means that no incentives are needed in the upper range of output. Hence,

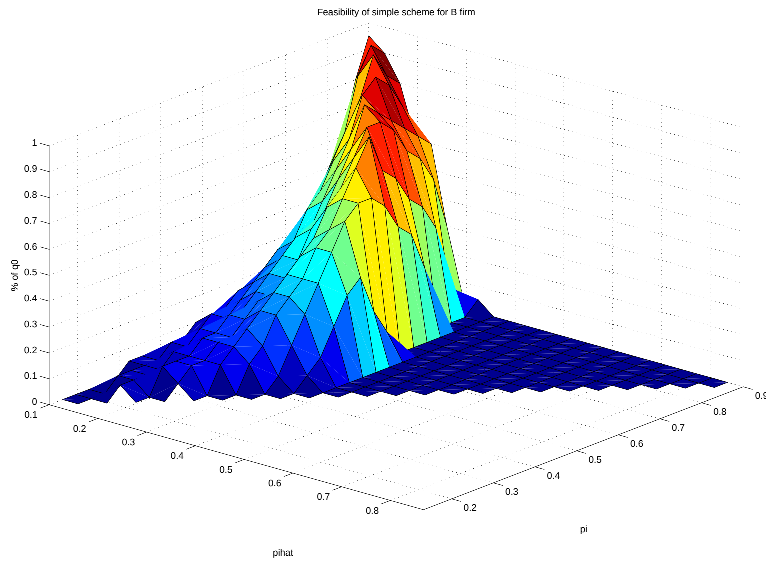


Figure 5: Fraction of firms for which a simple scheme is feasible (type $\mathcal{H}$ firm).

only for high enough prior that we are in state A incentives are high powered enough to make Δ_H positive. Finally, the level of c_{LL} needs to be implementable, given the restricted stock r_0, with a positive wage:

$$W = c_{LL} - r_0 p_{LL}.$$

This is feasible whenever condition a) in the corollary is satisfied. If the r_0 needed to implement Δ_H or p_{LL} are too high, a simple scheme is not feasible.

Unfortunately, condition a) depends, also for this type of firm, in a non trivial way on the primitives of the model (through c^*_{LL}). Although it is not possible to provide an analytical characterization of the combination of primitives for which it is satisfied, it is easy to check it numerically. Fig. 5 plots the probability that both condition a) and b) are satisfied, and hence a simple scheme is feasible for a firm of type $\mathcal{H}$. The values that make it most likely are combinations of high values of π with intermediate values of $\hat{\pi}$. Whenever $\hat{\pi}$ takes values above 0.4, no combination of π and q_0 makes a simple scheme feasible.

3.3.3 Complex schemes

Proposition 7 *A complex scheme is feasible (with a capped bonus) if and only if $\Delta_H > 0$, i.e., if $q_0 > q^* = \frac{\pi\widehat{\pi}}{1-\pi-\widehat{\pi}+\pi\widehat{\pi}}$.*

Hence, condition b) in Prop. 6 is both necessary and sufficient for a complex scheme to be feasible. The proof of this proposition (see Appendix 2) presents the expressions for the solutions to the optimal packages. The system is undetermined, i.e., if it has a solution it has an infinite number of them. However, only solutions that satisfy the non–negativity constraints on all the instruments constitute feasible complex schemes. In the proof of the sufficiency it is shown that whenever $\Delta_H > 0$ we can construct a feasible scheme that uses only a limited set of instruments:

$$\begin{aligned}
W &= c_{LL} \\
b &= \Delta_L \\
s_H &= \frac{\Delta_H}{p_{HH} - p_H} \\
s_0 &= r_0 = r_L = r_H = s_L = 0.
\end{aligned}$$

Hence, a refresher stock grant given to the agent after observing a high output in the first period is instrumental in implementing the optimal scheme. Given that there is no variation in prices following a low realization in the first period, the bonus needs to be used to implement Δ_L. Since $\Delta_H < \Delta_L$, the bonus needs to be capped. Hence, an instrument that will only pay off in the HH history is needed - an option granted when market price is equal to p_H.

Given that condition b) in Prop. 6 is both necessary and sufficient for a complex scheme to be feasible, a natural question is whether there is a non–trivial role for complex instruments. Can they help in the implementation of the optimal contract when a simple scheme is not feasible? The answer to these questions is positive. The next proposition illustrates that there is a role for refresher stock grants and option grants.

Proposition 8 *There is a non empty set of firms (parameter values) for which a simple scheme is not feasible but a complex one is.*

Figure 6 presents graphically the combinations of q_0, π and $\hat{\pi}$ for which *a* simple scheme is not feasible but a complex one is (condition a) in Prop. 6 is violated but condition b) is satisfied).

4 Generalization of the model

A more general firm description than the one I have been using as the leading example (the type $\mathcal{H}$ firm) would be one as in matrix 1, where output in state B is not only stochastic but it also depends on the effort choice of the agent. When considering this general case, I assume that $\pi_\theta \neq \widehat{\pi}_\theta$ for at least one θ, and the prior over $\theta = A$ satisfies $0 < q_0 < 1$. Also, higher effort (e_H) implies higher probability of observing y_H (for any quality of the firm): $\pi_A \geq \widehat{\pi}_A$ and $\pi_B \geq \widehat{\pi}_B$, with at least one being a strict inequality. For a firm of this generality, moral hazard and learning interact in more complicated ways; however, the intuition behind the properties of optimal consumption will rely on the same forces highlighted earlier for a type $\mathcal{H}$ firm.

It is my goal in this section to illustrate the usefulness of the results related to a firm of type $\mathcal{H}$ to understand what drives the feasibility of compensation schemes for a firm of the general case. For that purpose, I now introduce a second special type firm. Let a firm of type $\mathcal{L}$ be described, for all t, by:

$\mathcal{L}$ $\Pr(y_t = y_H \mid e, \theta)$	(q_0) A	$(1-q_0)$ B
e_H	π	0
e_L	$\widehat{\pi}$	0

(9)

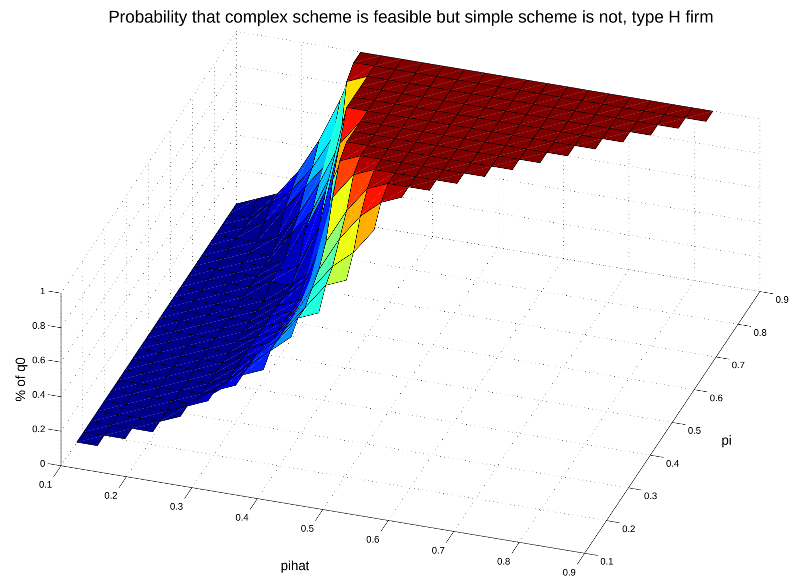


Figure 6: Proportion of firms for which a simple scheme is not feasible but a complex one is (type $\mathcal{H}$ firm).

In such a firm, when effort is not effective the firm always produces low output. In contrast, we may refer to the firm we have analyzed in the core of the paper as a type $\mathcal{H}$ firm, since output is always high in the state when effort is not effective.

Firms of type $\mathcal{H}$ and $\mathcal{L}$ represent particular examples of technologies that we may identify in real life. For example, we may think of $\mathcal{H}$ firms as mature, successful ones in which only the possibility of a bad match triggers bad realizations. Type $\mathcal{L}$ firms, instead, may be younger, struggling firms for which only a good match paired with high effort may improve outcomes. Other interpretations relate to the dependence of the firm's results on exogenous factors, such as uncertain new regulations or R&D developments.

Moreover, these two types of firms may be valuable in understanding the determinants of compensation packages in firms of a more general type, as described next. To see this, consider the following case: When the quality of the firm is B, a high output is realized with probability $\bar{\pi}$, regardless of the effort choice of the agent. Learning interacts with the moral hazard problem in this setting in a more general way than in the type $\mathcal{H}$ and $\mathcal{L}$ firms, since a high output observation is more or less informative about the moral hazard problem depending on the particular value of $\bar{\pi}$. It is easy to see that, for q_1 that satisfies

$$\bar{\pi} = q_1 + (1 - q_0 - q_1)\, 0 = q_1,$$

we can think of this case as a firm whose effort effectiveness varies across three states of the world: in A it is effective, in B_1 output is always high, and in B_2 output is always low. That is, whenever effort is not effective, output is high with probability q_1, and low with probability $1 - q_0 - q_1$.

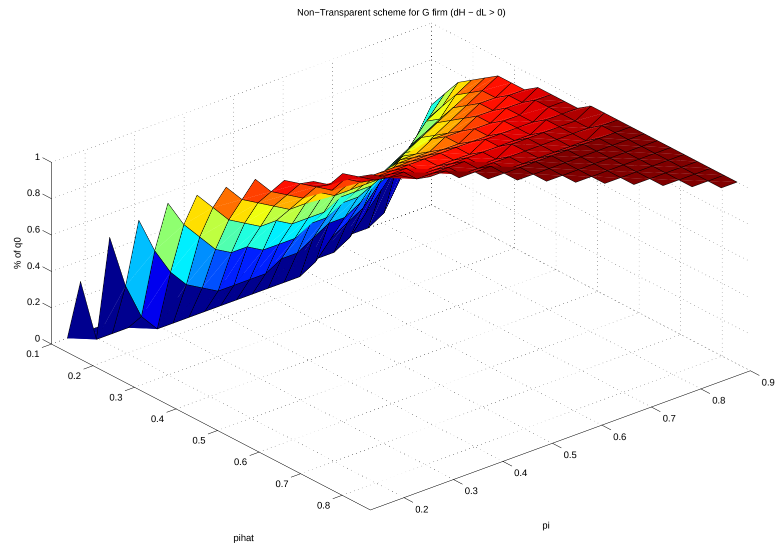


Figure 7: Type $\mathcal{L}$ firm. Probability that $\Delta_H - \Delta_L > 0$ is satisfied and a Non-transparent scheme is necessary.

Next, I summarize the analyses of a type $\mathcal{L}$ firm, the details of which are included in the appendix.

4.1 Firm of type $\mathcal{L}$

Derivations for the results regarding type $\mathcal{L}$ firms, which parallel those of a type $\mathcal{H}$ firm, are included in the appendix. Here, I present the discussion of the results, and compare them to those of a type $\mathcal{H}$ firm.

4.1.1 Non–transparent schemes

For a firm of type $\mathcal{L}$, non–monotonicities never arise following a high realization in the first period ($\Delta_H > 0$ always), but we may have $\Delta_H > \Delta_L$ and non–monotonicities following a low realization in the first period ($\Delta_L < 0$).

Proposition 9 *For a firm of type $\mathcal{L}$, compensation schemes must be non–transparent if and only if $\Delta_L < \Delta_H$.*

Prop. 13 in Appendix 2 characterizes formally the set of parameters for which $\Delta_L < \Delta_H$, and hence a non–transparent scheme is necessary for a type $\mathcal{L}$ firm. Figure 7 presents this characterization graphically.

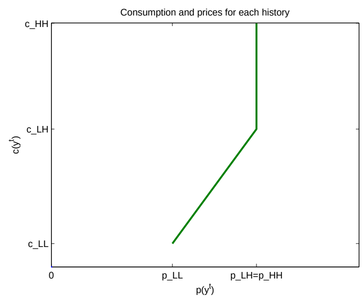


Figure 8: Mapping of consumption to prices for the $\mathcal{L}$ firm in Example 1 (see matrix **??**).

4.1.2 Simple schemes

Figure 8 represents an example of the mapping between consumption and stock prices for a firm of type $\mathcal{L}$. The infeasibility of a capped bonus is straight forward: because for an $\mathcal{L}$ firm, $p_{HH} = p_{LH}$, or $\alpha_H = 0$, we have that, under a capped bonus, Δ_H needs to be zero. This means that the only case in which a simple scheme could implement the optimal scheme is one in which $c^*_{LH} = c^*_{HH}$. But this is never the case for the non-trivial parametrization that we study, with $\pi \neq \widehat{\pi}$ and $q_0 \in (0, 1)$. On the other hand, a linear bonus, defined as $B^L = b\,(y_1 + y_2)$, may be feasible (see Corollary 2 in the Appendix for details). Conditions $a)$ and $b)$ in Corollary 2 summarize when the following solution satisfies the non–negativity constraint for the three instruments:

$$\begin{aligned}
W &= c^*_{LL} - q_{LL}\frac{\Delta_L - \Delta_H}{\left(1 - q^{\mathcal{L}}_{LL}\right)}, \\
b &= \Delta_H, \\
r_0 &= \frac{\Delta_L - \Delta_H}{\left(1 - q^{\mathcal{L}}_{LL}\right)\pi}.
\end{aligned}$$

The form of this solution is intuitive. For a firm of type $\mathcal{L}$, any number of high output realizations is perfectly informative about the state, so there is no variation in the upper range of prices (see Figure 8). This means that the spread $c_{HH} - c_{LH}$ needs to be implemented through a bonus payout: $b = \Delta_H$. Given the linear bonus program, this is not a problem as long as $\Delta_L \geq \Delta_H$. If this is the case, the quantity of restricted stock is determined to satisfy

$$\Delta_L - b = r_0\left(p_{LH} - p_{LL}\right).$$

It is the case that $\Delta_L \geq \Delta_H$ whenever incentives need to be more high powered in the low range of outcomes; since an $\mathcal{L}$ firm is more likely to get low output levels out of luck (when the true state is B,) incentives are high powered in the low range of outcomes only for high enough prior that the state is A. Finally, it must be the case that the implied wage given r_0 is positive (condition $a)$ in Corollary 2):

$$W = c_{LL} - r_0 p_{LL}.$$

If this condition is not met, a simple scheme is not feasible.

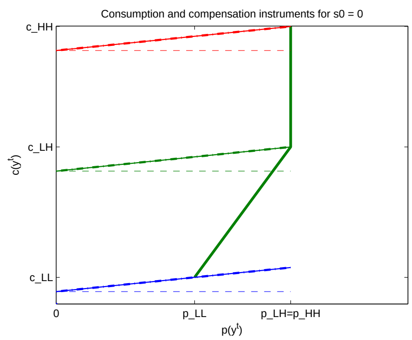


Figure 9: A simple compensation package is feasible for the firm in Example 1. The wage W is where the blue line intersects the vertical axis. The blue line represents the sum of W plus the value of the r_0 stock grants as a function of the stock price. The green line represents the value of the wage, plus the r_0 stock plus the payoff of the bonus scheme if only one y_H is realized (b). The red curve represents the value of the wage, plus that of r_0, plus the payoff of the bonus scheme if two y_H are realized ($2b$).

Example 1 The following parameters describe an example of a firm of type $\mathcal{L}$ for which a simple linear bonus is feasible: $\pi = 0.4$, $\hat{\pi} = 0.3$, $q_0 = .8$. The mapping between optimal consumption and prices is depicted in Figure 8. The the optimal simple package is superposed in Figure 9: the wage is slightly bellow c_{LL}, so that r_0 stock grants (the value of which, as a function of the stock price, is plotted in blue) satisfy $W + r_0 p_{LL} = c_{LL}$. The green line represents the value of the r_0 stock plus the payoff of the bonus scheme if only one y_H is realized, b, and the red curve that of r_0 plus the payoff of the bonus scheme if two y_H are realized, $2b$.

Fig. 10 presents the feasibility of a simple scheme for a firm of type $\mathcal{L}$. Assuming a uniform distribution of q_0 in the population, the probability that conditions a) and b) in Corollary 2 are satisfied for a firm of type $\mathcal{L}$ is depicted. The height of the mountain can be interpreted as the probability that the relationship between q_0, π and $\widehat{\pi}$ characterized in Prop. *13* is met, for a firm of randomly draw q_0.[14] The highest probability is for high values of π combined with low values of $\hat{\pi}$. Small differences $\pi - \hat{\pi}$ are only sustainable for values between 0.3 and 0.4; for values of π close to 0.2 or close to 0.9, even for differences of $\pi - \hat{\pi} = 0.2$ a simple scheme is not feasible for any q_0.

4.1.3 Complex schemes

The analysis parallels that of type $\mathcal{H}$ firms.

Proposition 10 *A complex scheme is feasible (with a capped bonus) if and only if* $\Delta_L > \Delta_H$.

[14] Note that when I present this type of graphics, only the feasible combinations of probabilities are plotted (i.e., pairs that satisfy $\pi > \hat{\pi}$).

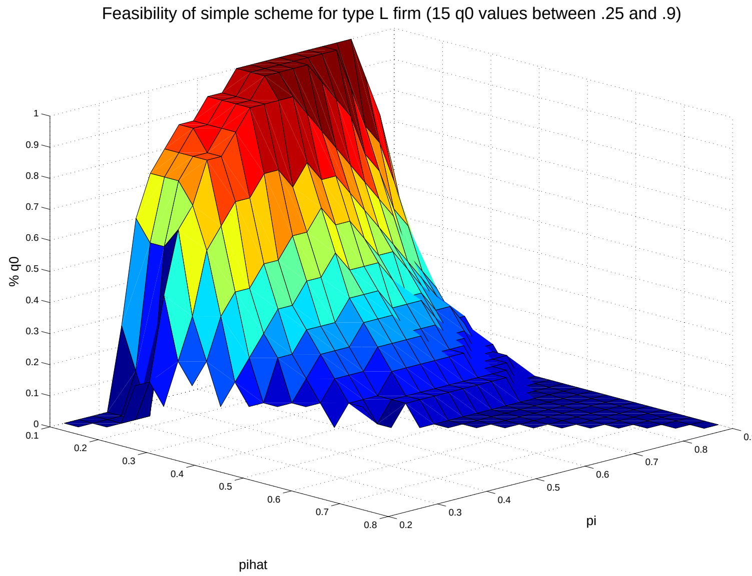


Figure 10: Feasibility of a simple scheme for a type $\mathcal{L}$ firm (with a linear bonus).

Complex schemes that are feasible when $\Delta_L > \Delta_H$ also take a very simple form:

$$\begin{aligned} W &= c_{LL} \\ b &= \Delta_H \\ r_H &= \frac{\Delta_L - \Delta_H}{p_{LH}} \\ s_L &= \frac{\Delta_L - \Delta_H}{p_{LH} - p_L} \\ s_0 &= r_0 = r_L = r_H = 0, \end{aligned}$$

where the bonus is linear in this case.

Again, we may ask whether there is a non–trivial role for complex schemes; the answer is positive as well for type $\mathcal{L}$ firms.

Proposition 11 *There is a non empty set of firms (parameter values) for which a simple scheme is not feasible but a complex one is.*

Figure presents a numerical characterization of the set of parameters. Here I provide an example of the role of refresher grants in implementing the optimal contract without using non–transparent schemes, for a firm of type $\mathcal{L}$.

Example 2 The following parameters describe an example of a firm of type $\mathcal{L}$ for which a simple linear bonus is not feasible: $\pi = 0.4$, $\hat{\pi} = 0.3$, $q_0 = .9$. In this example, condition a) is violated. In figure 12 we see graphically that a simple implementation with a linear bonus would imply

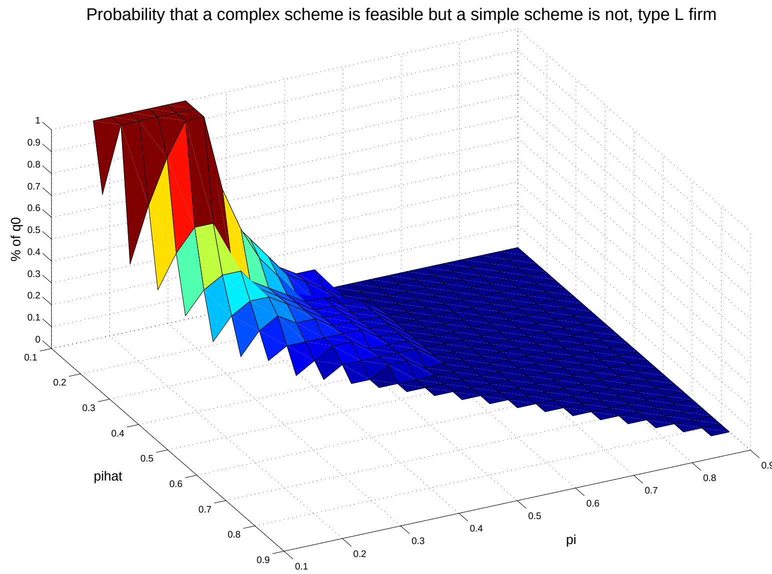


Figure 11: Proportion of firms for whic a simple scheme is not feasible but a complex one is (type $\mathcal{L}$ firm).

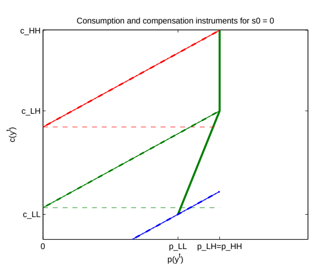


Figure 12: For the firm of type $\mathcal{L}$ in Example 2, described by matrix (**??**), condition a) is violated and hence a simple scheme is not feasible.

$W < 0$.However, I now show that the optimal contract can in this case be implemented with a complex scheme consisting of a linear bonus, $r_0 > 0$, $s_L > 0$ and $r_H > 0$. With $P \in \mathbb{C}$, and under the parametrization of example 2, we have:

$$\begin{aligned}
c_{HH} &= W + 2b + r_0 p_{HH} + r_H p_{HH} \\
c_{HL} &= W + \ b + r_0 p_{HL} + r_H p_{HL} \\
c_{LH} &= W + \ b + r_0 p_{HL} + s_L (p_{LH} - p_L) \\
c_{LL} &= W \qquad + r_0 p_{LL}
\end{aligned}$$

The solution is:

$$\begin{aligned}
W &= c_{LL} - \frac{q_{LL}}{1 - q_{LL}} (\Delta_L - \Delta_H) + A s_L \\
b &= \Delta_H \\
r_0 &= \frac{\Delta_L - \Delta_H}{(1 - q_{LL}) \pi_A} - B s_L \\
r_H &= C s_L
\end{aligned}$$

We draw the following conclusions from comparing our benchmark firm to a type $\mathcal{L}$ firm. For a firm that tends to be more successful when effort is very effective (type $\mathcal{L}$), prices are not as sensitive to output realizations as the optimal incentives; this implies that, although the firm may be able to use a simple scheme consisting only of a wage, restricted stock and a bonus program, the bonus should be linear in output. Instead, for a firm that tends to be more successful for exogenous reasons than for the effect of effort (type $\mathcal{H}$), prices are not as sensitive to output realizations when they are low as the optimal incentives should be; this implies that when a simple compensation scheme is used, the bonus program should be capped.

5 Discussion of testable implications

What do we know about the real life relationship between firm characteristics and the use of different compensation instruments? Here I present some regularities that emerge from my analysis of Execucomp data on CEO compensation, and relate them to the conclusions of my model.

It is a difficult task to identify "refresher" grants in the data. Typically, however, new grants are awarded before the selling restrictions on previous grants has expired. Hence, I interpret all restricted stock grants as refresher grants (r_H and r_L in my model). Based on the usage data presented in figure 2, I construct a classification of "users" and "non-users" that is based on observing repeated option or stock grants by the same firm over time.[15] Because size and industry have been shown to be important determinants of pay level (see Murphy (1999), Gabaix and

[15] All firms are in the sample for at least 6 years. I define "users" as firms that have granted in at least 70% of the total years that they are in the sample. I define "non-users" as firms that have granted 4 years or less, or in less than 50% of the years they are in the sample. Out of 1,457 firms, this generates 145 non–users, 1,186 users, and 126 firms that cannot be classified (I drop them from the analisys). There are 6 firms in the sample that never use any grant, and 116 that use them in every one of the 18 years in the sample.

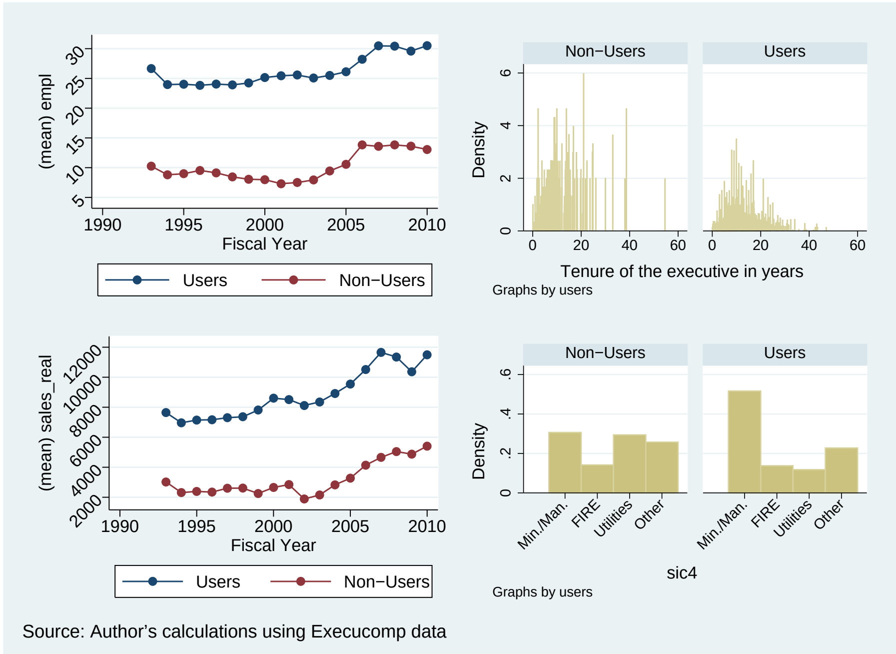


Figure 13: Differences between "users" and "non-users" of grants, in terms of (from left to right, top to bottom): the average number of employees of the firms, the average tenure of the CEOs, the average value of annual sales, and the distribution of the firms across industries.

Landier (2008), and Clementi and Cooley, 2010), I also consider differences across the two groups for measures of size and SIC industry classification.

In Fig. 13 we can see that user firms are larger in terms of employees, and they are larger also in terms of their volume of annual sales. However, we do not have a good a priory explanation of why larger firms should be more likely to be users of grants in compensation. On the other hand, users are more likely to be Mining and Manufacturing firms, which are sectors that tend to pay less than sectors like Finance, Insurance and Real Estate (FIRE) (Clementi and Cooley, 2010). User firms also tend to employ CEOs for a longer tenure, as evidenced by the histogram of the average tenure at the firm level in the upper right hand side corner of Fig. 13. The empirical evidence about the potential complementarity of career concerns and explicit incentives is mixed; also, one may consider the possibility that firms that have a reputation of longer tenures can expose the CEO to higher risk through his compensation package to compensate for the added job security. Hence, the effect on the decision to use grants is not obvious.

SIC classification	**% of users**	**Total firms**	**Users**	**Non–Users**
Utilities	0.80	163	131	32
Other	0.88	313	275	38
FIRE	0.91	197	180	17
Mining and Manufacturing	0.94	639	600	39

Table 1: Proportion of Users and Non-Users in a four group classification of industries.

In table 1 I report the proportion of firms that qualify as "users" in four different groups of firms, according to their SIC industry classification. Firms in "Utilities", which includes Transportation, Communications, Electric, Gas, And Sanitary Services, are the ones who rely less often on grants. On the other hand, firms in Finance, Insurance and Real Estate (FIRE) are not the ones that use them more intensely, as one may conjecture: firms in Manufacturing and in Mining use them more broadly, at a 94% rate.

Small firms seems to be less likely to use grants, i.e. they seem more able to use simple schemes. Because size (employee numbers and sales volume) is correlated with the age of the firm and its reputation, it may be important in determining the effectiveness of the actions of its CEO (whether the firm is of type $\mathcal{L}$ or $\mathcal{H}$, and what is the value of π). The tenure of the CEO may be related to the level of uncertainty that the firm faces (the prior q in the model).

The empirical literature in Finance has also provided some interesting —although scarce— evidence about complex or non–transparent compensation practices: refresher grants, and repricing. Hall and Knox (2004) find evidence that refresher grants are often used by firms both following a stock price decline and following a stock price increase. They interpret refresher grants as a mechanism to restore incentives for the CEO whenever the sensitivity of his compensation to stock price movements decreases. This makes refresher grants following a stock price increase puzzling, since stock price increases tend to increase the sensitivity of pay to firm performance. My model provides a rational for new grants contingent on both good and bad firm performance.

A related compensation practice, option "repricing," consists in lowering the exercise price of options that have gone out–of–the–money (i.e., their exercise price is well above the current stock price in the market.) This can be thought of as a substitute for refresher grants, but with potentially different tax implications. Although this practice is fairly uncommon (Brenner, Sundaram and Yermak (2000), report that 1.3% of the top five officers in a sample of 1500 firms between 1992 and 1995 had options repriced in a given year,) it has received both media and academic attention, perhaps motivated by its reputation as a bad compensation practice. There is a series of empirical papers that presents evidence on the frequency of repricing and the characteristics of the firms that engage in it. Chance, Kumar, and Todd (2000) identify size as the main predictor for firm reprices, with smaller firms repricing more often. Brenner, Sundaram and Yermak (2000) find that higher volatility also significantly rises the probability of repricing. Carter and Lynch (2001) find that young, high technology firms and those whose outstanding options are more out of the money are more likely to reprice. Chen (2004) finds that firms that restrict repricing have a higher probability of loosing their CEO after a decline in their stock price, and that they typically grant new options in those circumstances, possibly in an effort to retain the CEO. With the conclusions of my model in mind, one may suggests that the effectiveness of effort of the CEO may be different for small and more technologically oriented firms; an empirical research along these lines may be of interest to further understand compensation practices.

6 Conclusion

In this paper, I ask the question of what are the firm characteristics that may justify the use of options or refresher grants in the compensation packages for CEOs. I view compensation packages as particular implementations of the optimal contract in the presence of moral hazard. Working with models of asymmetric information and risk averse agents is generally difficult. Here, I present a necessarily stark model of a firm. Its simplicity allows me to enrich it with learning about the effectiveness of the effort of the CEO in enhancing the output of the firm with his or her effort. This provides me with a model that explains stock prices and compensation jointly from primitives. One lesson emerges from the analysis of the model: the level of uncertainty about (and the priors on) the effectiveness of the CEOs' actions is an important factor for the type of compensation instruments that the firm uses. Observable characteristics such as the tenure of the CEO may be correlated with this level of uncertainty. Measures of firm size are correlated with the age of the firm and its reputation, and hence they may be important in determining the effectiveness of the actions of the CEO. With the theoretical analysis in mind, some new empirical questions arise that may lead to a better understanding of what the key explanatory characteristics of compensation practices are; for example, wether the use of certain instruments is persistent over time for a given firm, or wether it is perhaps tied to the tenure of a given CEO.

7 Appendix 1: Proofs

Proof of Prop. 1. With $u(c) = \ln(c)$, the non–negativity constraints NNC' will not bind. With λ as the multiplier for the binding PC, and μ for the binding IC, the first order conditions of the problem are

$$\frac{1}{u'(c_{ij})} = \lambda + \mu(1 - LR_{ij}).$$

These simplify to equation 8 in the case of $u(c) = \ln(c)$. ■

Proof of Prop. 2. It is easy to see this by looking at the likelihood ratios in this particular case of our framework, which simplify to:

$$\begin{aligned}
LR_{LL} &= \frac{1-\widehat{\pi}}{1-\pi}\frac{1-\widehat{\pi}}{1-\pi} \\
LR_{LH} &= \frac{1-\widehat{\pi}}{1-\pi}\frac{\widehat{\pi}}{\pi} \\
LR_{HL} &= \frac{\widehat{\pi}}{\pi}\frac{1-\widehat{\pi}}{1-\pi} \\
LR_{HH} &= \frac{\widehat{\pi}}{\pi}\frac{\widehat{\pi}}{\pi},
\end{aligned}$$

where $\frac{\widehat{\pi}}{\pi} < 1 < \frac{1-\widehat{\pi}}{1-\pi}$ and hence $LR_{HH} < LR_{HL} = LR_{LH} < LR_{LL}$, which implies $c_{HH} > c_{HL} =$

$c_{LH} > c_{LL}$. Moreover,

$$\begin{aligned}
\Delta_H &= \mu\left(LR_{LH} - LR_{HH}\right) = \mu\frac{\widehat{\pi}}{\pi}\frac{\pi - \widehat{\pi}}{\pi(1-\pi)}, \\
\Delta_L &= \mu\left(LR_{LL} - LR_{LH}\right) = \mu\frac{1-\widehat{\pi}}{1-\pi}\frac{\pi - \widehat{\pi}}{\pi(1-\pi)},
\end{aligned}$$

which implies the second result in the proposition. ■

Proof of Prop. 3. In the first part of the proposition we want to show that, for a firm of type $\mathcal{H}$, $\Delta_H < \Delta_L$, and also that $\Delta_L > 0$. We first show the second inequality holds, and then we use it to prove the first inequality. From the expressions for the likelihood ratios in 7, we have

$$\begin{aligned}
\Delta_L^{\mathcal{H}} &= \mu\left(LR_{LL}^{\mathcal{H}} - LR_{LH}^{\mathcal{H}}\right) \\
&= \mu\frac{1-\widehat{\pi}}{1-\pi}\frac{\pi - \widehat{\pi}}{\pi(1-\pi)}.
\end{aligned}$$

Since μ is the Lagrange multiplier of the incentive constraint in problem (**P1**), it will satisfy $\mu \geq 0$. Since $\pi > \widehat{\pi}$ by assumption, $\Delta_L > 0$ follows. Now to establish $\Delta_H < \Delta_L$ we compare $LR_{LL}^{\mathcal{H}} - LR_{LH}^{\mathcal{H}}$ to an upper and a lower bound for the difference $LR_{HH}^{\mathcal{H}} - LR_{LH}^{\mathcal{H}}$. First, note that $LR_{LH}^{\mathcal{H}} = \frac{\widehat{\pi}(1-\widehat{\pi})}{\pi(1-\pi)}$ is independent of q_0. In turn, we can show that $LR_{HH}^{\mathcal{H}}$ decreases monotonically for $q_0 \in (0,1)$:

$$\begin{aligned}
\frac{\partial LR_{HH}^{\mathcal{H}}}{\partial q_0} &= \frac{\left(\widehat{\pi}^2 - 1\right)\left[q_0\pi^2 + 1 - q_0\right] - \left(\pi_B^2 - 1\right)\left[q_0\widehat{\pi}_B^2 + 1 - q_0\right]}{\left(q_0\pi_B^2 + 1 - q_0\right)^2} \\
&= \frac{\widehat{\pi}_B^2 - \pi_B^2}{\left(q_0\pi_B^2 + 1 - q_0\right)^2} < 0.
\end{aligned}$$

Then, by taking the limit of $LR_{HH}^{\mathcal{H}}$ with respect to q_0 we can bound $\Delta_H^{\mathcal{H}}$ and compare it to Δ_L in both extreme cases. When q_0 approaches 0, $LR_{HH}^{\mathcal{H}}$ goes to its maximum, 1, and we have:

$$\lim_{q_0 \to 0} \Delta_H^{\mathcal{H}} = \mu\left(\frac{\left(1-\widehat{\pi}_B\right)\widehat{\pi}_B}{\left(1-\pi_B\right)\pi_B} - 1\right),$$

which determines the minimum possible $\Delta_H^{\mathcal{H}}$. When q_0 approaches 1, instead, LR_{HH} approaches its minimum, which coincides with the expression for $LR_{HH}^{\mathcal{H}}$ in the no learning case, $\frac{\widehat{\pi}_B^2}{\pi_B^2}$:

$$\lim_{q_0 \to 1} \Delta_H^{\mathcal{H}} = \mu\frac{\widehat{\pi}_B}{\pi_B}\frac{\pi_B - \widehat{\pi}_B}{\pi_B(1-\pi_B)}.$$

This is the maximum possible value of $\Delta_H^{\mathcal{H}}$. It is easy to see that this implies $\Delta_H < \Delta_L$ for any value of q_0. Note that μ depends on q_0, but the comparison is for a given common μ in both Δ_H and Δ_L. For the second part of the proposition, note that we have that $\Delta_H^{\mathcal{H}} < 0$ if and only if

$$\frac{\hat{\pi}\left(1-\hat{\pi}\right)}{\pi\left(1-\pi\right)} < \frac{q_0 + \left(1-q_0\right)\hat{\pi}^2}{q_0 + \left(1-q_0\right)\pi^2}.$$

Rearranging, this condition becomes

$$q_0 > \frac{\pi\hat{\pi}}{1 - \hat{\pi} - \pi + \pi\hat{\pi}}.$$

Whenever $\pi + \widehat{\pi} \geq 1$, the denominator is smaller than the numerator, and hence there is no q_0 for which $\Delta_H^{\mathcal{H}} < 0$. Whenever $\pi + \widehat{\pi} < 1$, instead, we can define

$$q_1^{\mathcal{H}} = \frac{\pi\hat{\pi}}{1 - \hat{\pi} - \pi + \pi\hat{\pi}},$$

with $0 < q_1^{\mathcal{H}} < 1$, and we have that $\Delta_H^{\mathcal{H}} < 0$ for all $q_0 \geq q_1^{\mathcal{H}}$. ■

8 Appendix 2: Generalizations

8.1 Equilibrium Consumption Properties for a general firm

Proposition 12 *Optimal consumption is not necessarily monotonic in output, i.e. we may have $\Delta_L < 0$ or $\Delta_H < 0$. Also, both $\Delta_H < \Delta_L$ and $\Delta_H > \Delta_L$ may occur.*

The proof for this result, included in Miller (1999), simply analyzes the possible ranking of likelihood ratios for different combinations of probabilities in matrix 1.[16]

8.2 Type $\mathcal{L}$ firm prices and consumption properties

Consider instead a type $\mathcal{L}$ firm, described by matrix (9). The analysis of this case parallels that of case $\mathcal{H}$. For any histories containing at least one y_H, the updated believes put probability one on $\theta = A$, i.e., we have that $q_{ij}^{\mathcal{L}} = 1$ if i or j equals H. If the observed history does not contain any y_H, instead, $\theta = B$ has still positive probability. This is the case for histories y_L and (y_L, y_L). That is,

$$\begin{aligned}
q_L^{\mathcal{L}} &= q_0 \frac{(1-\pi)}{q_0(1-\pi) + 1 - q_0}, \\
q_{LL}^{\mathcal{L}} &= q_0 \frac{(1-\pi)^2}{q_0(1-\pi)^2 + 1 - q_0}, \\
q_H^{\mathcal{L}} &= q_{LH}^{\mathcal{L}} = q_{HH}^{\mathcal{L}} = 1.
\end{aligned}$$

The stock prices take the simple form:

$$\begin{aligned}
p_0^{\mathcal{L}} &= q_0\pi, \\
p_L^{\mathcal{L}} &= q_L^{\mathcal{L}}\pi, \\
p_{LL}^{\mathcal{L}} &= q_{LL}^{\mathcal{L}}\pi, \\
p_H^{\mathcal{L}} &= p_{HL}^{\mathcal{L}} = p_{HH}^{L} = \pi.
\end{aligned} \tag{10}$$

For this type of firm we have that $q_H^{\mathcal{L}} = q_{LH}^{\mathcal{L}} = q_{HH}^{\mathcal{L}} = 1$. Learning may in this case also give rise to non–monotonicities. In particular, when the first period output has been y_L, the agent's wage may be lower if we observe y_H in the second period than if we observe y_L. This is because observing y_H in the second period reveals that $\theta = A$. In state A, the first period observation y_L

[16] See also Celentani and Loveira (2006).

makes the history's likelihood ratio be much higher (it is a much more likely history under low effort than it would be if $\theta = B$). Formally, the likelihood ratios are:

$$LR^{\mathcal{L}}_{LL} = \frac{q_0 (1-\widehat{\pi})^2 + 1 - q_0}{q_0 (1-\pi)^2 + 1 - q_0} \tag{11}$$

$$LR^{\mathcal{L}}_{LH} = \frac{(1-\widehat{\pi})\,\widehat{\pi}}{(1-\pi)\,\pi}$$

$$LR^{\mathcal{L}}_{HH} = \frac{\widehat{\pi}^2}{\pi^2}. \tag{12}$$

The likelihood ratios $LR^{\mathcal{L}}_{HH}$ and $LR^{\mathcal{L}}_{LH}$ coincide in this case with the corresponding ones in the benchmark case of no learning characterized in Prop. **??**, when the true state is known to be A. We can use these likelihood ratios to establish the following properties of consumption in the optimal contract:

Proposition 13 *When the firm is of type $\mathcal{L}$ as described by matrix (9), consumption spread satisfies:*

(i) *whenever $\pi + \widehat{\pi} > 1$,*

$$\begin{aligned} \textit{for } q_0 &\in (0, q_1^{\mathcal{L}}],\ \Delta_L < 0 < \Delta_H, \\ \textit{for } q_0 &\in \left[q_1^{\mathcal{L}}, q_2^{\mathcal{L}}\right], 0 < \Delta_L < \Delta_H, \\ \textit{for } q_0 &\in [q_2^{\mathcal{L}}, 1),\ 0 < \Delta_H < \Delta_L, \end{aligned}$$

(ii) *whenever $\pi + \widehat{\pi} < 1$,*

$$\begin{aligned} \textit{for } q_0 &\in (0, \max\left\{q_2^{\mathcal{L}}, 0\right\}],\ 0 < \Delta_L < \Delta_H, \\ \textit{for } q_0 &\in [\max\left\{q_2^{\mathcal{L}}, 0\right\}, 1),\ 0 < \Delta_H < \Delta_L. \end{aligned}$$

where

$$\begin{aligned} q_1^{\mathcal{L}} &= \frac{\pi + \widehat{\pi} - 1}{\pi\widehat{\pi}}, \\ q_2^{\mathcal{L}} &= \frac{2\pi\widehat{\pi} - \pi\widehat{\pi}^2 - \widehat{\pi}^2 - \pi^2 (1-\pi)}{2\pi\widehat{\pi}\,(\pi - \widehat{\pi})}. \end{aligned}$$

(The proof is included after this discussion.) Again, we find a difference depending on wether $LR^{\mathcal{L}}_{LH}$ is greater or smaller than one. When $\pi + \widehat{\pi} > 1$, we have that $LR^{\mathcal{L}}_{LH}$ is greater than one and hence the contract seeks to punish the agent at LH as well as at LL. When the firm is of type $\mathcal{L}$, however, observing a high realization implies that the state is A with probability one. This makes a low realization a very valuable (and negative) signal about performance, making c_{LH} but not c_{LL} low. This implies a small Δ_L. For low enough prior of being in state A, this can lead to $\Delta_L < 0$. Higher and higher priors diminish the relative informativeness of LH with respect to LL and hence reestablish the relationship $0 < \Delta_H < \Delta_L$ of the no learning benchmark.

When $\pi + \widehat{\pi} < 1$, we have that $LR^{\mathcal{L}}_{LH}$ is smaller than one and hence the contract seeks to reward the agent at LH as well as at HH. The posterior becomes one under either realization, so the standard ranking of $c_{LH} < c_{HH}$ prevails, implying $\Delta_H > 0$ always. However, since the only state in which there is punishment is LL and this is a very likely outcome when the true state is B, unrelated to effort choice, a small q_0 (a high prior that we are in B) implies a very high cost of utility imposed on the agent with very little incentive benefit (he consumes very low very often but his incentives are little changed if the state is A, since LL is very unlikely in this state). This tends to keep c_{LL} not too far from c_{LH}, and hence it can lead to $\Delta_L < \Delta_H$. However, for $\pi + \widehat{\pi}$ smaller than one but with $(\pi - \widehat{\pi})$ small the informativeness of signals in state A decreases a lot; this makes c_{HH} close to c_{LH} as well, and more so than c_{LH} close to c_{LL} for the same reasons as in the benchmark case without learning. This is reflected in the case $q^{\mathcal{L}}_2 < 0$, which is more likely when $(\pi - \widehat{\pi})$ is small and implies $0 < \Delta_H < \Delta_L$ always.

Proof of Prop. 13. From the expressions for the likelihood ratios in 11, we have

$$\begin{aligned} \Delta^{\mathcal{L}}_H &= \mu\left(LR^{\mathcal{L}}_{LH} - LR^{\mathcal{L}}_{HH}\right) \\ &= \mu\frac{\widehat{\pi}}{\pi}\frac{\pi - \widehat{\pi}}{\pi(1-\pi)}. \end{aligned}$$

The expression for $\Delta^{\mathcal{L}}_L$ cannot be easily simplified. However, by taking its limit with respect to q_0 we can bound it and compare it to Δ_H in both extreme cases. The only term in $\Delta^{\mathcal{L}}_L$ that depends on q_0 is LR_{LL}, and it does so monotonically for $q_0 \in (0,1)$:

$$\begin{aligned} \frac{\partial LR^{\mathcal{L}}_{LL}}{\partial q_0} &= \frac{\left[(1-\widehat{\pi})^2 - 1\right]\left[q_0(1-\pi)^2 + 1 - q_0\right] - \left[(1-\pi)^2 - 1\right]\left[q_0(1-\widehat{\pi})^2 + 1 - q_0\right]}{\left[q_0(1-\pi)^2 + 1 - q_0\right]^2} \\ &= \frac{(1-\widehat{\pi})^2 - (1-\pi)^2}{\left[q_0(1-\pi)^2 + 1 - q_0\right]^2} > 0. \end{aligned}$$

When q_0 approaches 1, $LR^{\mathcal{L}}_{LL}$ approaches its maximum, which coincides with the expression for LR_{LL} in the no learning case, $\frac{(1-\widehat{\pi})^2}{(1-\pi)^2}$,so it is easy to see that:

$$\lim_{q_0 \to 1} \Delta^{\mathcal{L}}_L = \mu\frac{1-\widehat{\pi}}{1-\pi}\frac{\pi - \widehat{\pi}}{\pi(1-\pi)} > \Delta_H > 0.$$

When q_0 approaches 0 instead, $LR^{\mathcal{L}}_{LL}$ goes to its minimum, 1, and we have:

$$\lim_{q_0 \to 0} \Delta^{\mathcal{L}}_L = \mu\left(1 - \frac{(1-\widehat{\pi})\,\widehat{\pi}}{(1-\pi)\,\pi}\right).$$

This is not conclusive; we need to analyze two cases separately, according to whether π_A and $\widehat{\pi}_A$ are such that *(i) or (ii)* is satisfied. First, we have that $\Delta^{\mathcal{L}}_L < 0$ if and only if

$$\frac{q_0(1-\hat{\pi})^2 + (1-q_0)}{q_0(1-\pi)^2 + (1-q_0)} < \frac{\hat{\pi}(1-\hat{\pi})}{\pi(1-\pi)}.$$

Rearranging, the threshold value for q_0 becomes

$$q_1^{\mathcal{L}} = \frac{\hat{\pi} + \pi - 1}{\pi \hat{\pi}}.$$

When $\pi + \widehat{\pi} < 1$, there is no value of q_0 for which $\Delta_L^{\mathcal{L}} < 0$. Note also that $q_1^{\mathcal{L}} < 1$ for any values of π and $\widehat{\pi} \geq 1$, since:

$$\begin{aligned} \frac{\hat{\pi} + \pi - 1}{\pi \hat{\pi}} &< 1 \\ \hat{\pi} + \pi - 1 &< \pi \hat{\pi} \\ \pi \left(1 - \hat{\pi}\right) &< 1 - \hat{\pi} \\ \pi &< 1, \end{aligned}$$

which is always true. For the second threshold, we have that $\Delta_L^{\mathcal{L}} \leq \Delta_H^{\mathcal{L}}$ if and only if

$$\frac{q_0 \left(1 - \hat{\pi}\right)^2 + \left(1 - q_0\right)}{q_0 \left(1 - \pi\right)^2 + \left(1 - q_0\right)} - \frac{\hat{\pi} \left(1 - \hat{\pi}\right)}{\pi \left(1 - \pi\right)} \leq \frac{\widehat{\pi}}{\pi} \frac{\pi - \widehat{\pi}}{\pi (1 - \pi)}.$$

Simplifying, we get:

$$q_2^{\mathcal{L}} = \frac{2\pi\widehat{\pi} - \widehat{\pi}^2 \left(\pi + 1\right) - \pi^2 \left(1 - \pi\right)}{2\pi\widehat{\pi} \left(\pi - \widehat{\pi}\right)}.$$

Note that $q_2^{\mathcal{L}} < 1$ for any combination of probabilities, since the numerator can be bound by an expression smaller than the denominator:

$$2\pi\widehat{\pi} - \widehat{\pi}^2 \left(\pi + 1\right) - \pi^2 \left(1 - \pi\right) < 2\pi\widehat{\pi} - \pi^2 \left(\pi + 1\right) - \pi^2 \left(1 - \pi\right),$$

and

$$\begin{aligned} 2\pi\widehat{\pi} - \pi^2 \left(\pi + 1\right) + \pi^2 \left(\pi - 1\right) &< 2\pi\widehat{\pi} \left(\pi - \widehat{\pi}\right) \\ 2\pi\widehat{\pi} - \pi^2 \left(\pi + 1 - \pi + 1\right) &< 2\pi\widehat{\pi} \left(\pi - \widehat{\pi}\right) \\ \left(\widehat{\pi} - 2\right) &< \widehat{\pi} \left(\pi - \widehat{\pi}\right). \end{aligned}$$

However, whether $q_2^{\mathcal{L}}$ is strictly positive depends on the sum of π and $\widehat{\pi}$:

$$\begin{aligned} q_2^{\mathcal{L}} &> 0 \\ 2\pi\widehat{\pi} - \widehat{\pi}^2 \left(\pi + 1\right) - \pi^2 \left(1 - \pi\right) &> 0 \\ 2\pi\widehat{\pi} - \widehat{\pi}^2 - \pi^2 &> \widehat{\pi}^2 \pi - \pi^3 \\ \left(\pi^2 - \widehat{\pi}^2\right) \pi &> \left(\pi - \widehat{\pi}\right)^2 \\ \pi \left(\pi - \widehat{\pi}\right) \left(\pi + \widehat{\pi}\right) &> \left(\pi - \widehat{\pi}\right)^2 \\ \pi \left(\pi + \widehat{\pi}\right) &> \pi - \widehat{\pi}. \end{aligned}$$

If $\pi + \widehat{\pi} \geq 1$, then the last inequality is always satisfied, and hence $0 < q_2^{\mathcal{L}} < 1$. If $\pi + \widehat{\pi} < 1$, however, for some pairs of π and $\widehat{\pi}$ there will be no q_0 for which $\Delta_L^{\mathcal{L}} \leq \Delta_H^{\mathcal{L}}$. ■

8.3 Generalization of results for feasibility of simple and complex schemes

In this appendix I present the feasibility results for a general firm as defined, at all t, by matrix 1, where I assume that $\pi_\theta \neq \widehat{\pi}_\theta$ for at least one θ, and the prior over $\theta = A$ satisfies $0 < q_0 < 1$. Also, higher effort (e_H) implies higher probability of observing y_H (for any quality of the firm): $\pi_A \geq \widehat{\pi}_A$ and $\pi_B \geq \widehat{\pi}_B$, with at least one being a strict inequality. I also allow for a different type of bonus program: a linear bonus $B = B^L(y_1, y_2) = b\,(y_1 + y_2)$. To differentiate it from the capped bonus, I denote the capped one as B^C in this appendix.

8.3.1 Simple scheme

First, I study the case of a simple scheme with a linear bonus.

Proposition 14 *A simple scheme with a **linear bonus** is feasible if and only if:*

a) $\frac{c^*_{LL}}{p_{LL}} > \frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H}$,

b) $\frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H} > 0$,

c) $\frac{\Delta_H}{\alpha_H} > \frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H}$.

Proof of Prop. 14. With a linear bonus, the function $\mathcal{C}\,(P)$ implies:

$$\begin{aligned} c^*_{HH} &= W + 2b + r_0 p_{HH} \\ c^*_{HL} &= W + b + r_0 p_{HL} \\ c^*_{LH} &= W + b + r_0 p_{HL} \\ c^*_{LL} &= W + r_0 p_{LL}. \end{aligned}$$

It is useful to write these equations as a function of the consumption differences Δ_L and Δ_H defined in the previous section. To simplify, I also introduce the following notation for the differences in prices: $\alpha_L \equiv p_{LH} - p_{LL}$ and $\alpha_H \equiv p_{HH} - p_{HL}$.

$$\begin{bmatrix} \Delta_H \\ \Delta_L \\ c^*_{LL} \end{bmatrix} = \begin{bmatrix} 0 & 1 & \alpha_H \\ 0 & 1 & \alpha_L \\ 1 & 0 & p_{LL} \end{bmatrix} \begin{bmatrix} W \\ b \\ r_0 \end{bmatrix}.$$

The solution to this system is:

$$\begin{aligned} W &= c^*_{LL} - p_{LL}\frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H}, \\ b &= \Delta_H - \alpha_H \frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H}, \\ r_0 &= \frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H}. \end{aligned}$$

The conditions a)-c) are then derived from the non–negativity constraints imposed on W, b and r_0. ■

Corollary 1 *For a type $\mathcal{H}$ firm, a simple scheme with a linear bonus is never feasible.*

Proof. It is straight forward to see that a linear bonus is infeasible. For a firm of type $\mathcal{H}$, we have $p_{LL} = p_{LH}$, or $\alpha_L = 0$. This means that the spread Δ_L must be implemented with the bonus b. However, a linear bonus implies $\Delta_H \geq b$, while we saw in Prop. 3 that for this type of firm it is always the case that $\Delta_H < \Delta_L$. ■

A similar analysis can be pursued for the case of a capped bonus, for a general firm as described in matrix 1.

Proposition 15 *A simple scheme with a* ***capped bonus*** *is feasible if and only if:*

a) $\frac{c_{LL}}{\Delta_H} > \frac{p_{LL}}{\alpha_H}$,

b) $\frac{\Delta_L}{\Delta_H} > \frac{\alpha_L}{\alpha_H}$,

c) $\Delta_H > 0$.

Proof of Prop. 15. The function $\mathcal{C}(P)$ is, in the case of a capped bonus:

$$\begin{aligned} c^*_{HH} &= W + b + r_0 p_{HH} \\ c^*_{HL} &= W + b + r_0 p_{HL} \\ c^*_{LH} &= W + b + r_0 p_{HL} \\ c^*_{LL} &= W + r_0 p_{LL}, \end{aligned}$$

and hence,

$$\begin{bmatrix} \Delta_H \\ \Delta_L \\ c^*_{LL} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \alpha_H \\ 0 & 1 & \alpha_L \\ 1 & 0 & p_{LL} \end{bmatrix} \begin{bmatrix} W \\ b \\ r_0 \end{bmatrix}.$$

The solution to this system is

$$\begin{aligned} W &= c^*_{LL} - p_{LL}\frac{\Delta_H}{\alpha_H}, \\ b &= \Delta_L - \alpha_L\frac{\Delta_H}{\alpha_H}, \\ r_0 &= \frac{\Delta_H}{\alpha_H}. \end{aligned}$$

■

Corollary 2 *For a type $\mathcal{L}$ firm, a simple scheme with a capped bonus is never feasible. A simple scheme with a linear bonus is feasible if and only if:*

a) $c^*_{LL} > \frac{q_{LL}}{1-q_{LL}}(\Delta_L - \Delta_H)$, *and*

b) $\Delta_L > \Delta_H$, or $q_0 > q_1^{\mathcal{L}} = \frac{\pi^2(1-\pi)+2\pi\widehat{\pi}-\pi\widehat{\pi}^2-\widehat{\pi}^2}{2\pi\widehat{\pi}(\pi-\widehat{\pi})}$ *(see Prop. 13).*

8.3.2 Complex scheme

Prop. 8 is proved here for the general case. The corresponding corollaries for a firm of type $\mathcal{H}$ and $\mathcal{L}$ follow.

Proof of Prop. 15. I consider the case of a linear bonus and the case of a capped bonus. Also, for each bonus scheme, I need to consider two cases separately: whenever $p_{LH} < p_0$, options granted at time 0, s_0, are not exerciseable; however, $p_{LH} < p_0$ is also possible for some parameter combinations, and in this case they are. Hence, there are four cases to consider.

1. Linear bonus and $p_{LH} > p_0$. The implementation of the optimal contract with instruments $\mathbb{C}$ and a linear bonus translates into a system of equations:

$$\begin{aligned} c_{HH} &= W + 2b + r_0 p_{HH} + s_0 (p_{HH} - p_0) + r_H p_{HH} + s_H (p_{HH} - p_H) \\ c_{HL} &= W + b + r_0 p_{HL} + \max\{s_0 (p_{HL} - p_0)\} + r_H p_{HL} \\ c_{LH} &= W + b + r_0 p_{HL} + \max\{s_0 (p_{HL} - p_0)\} + r_L p_{HL} + s_L (p_{HL} - p_L) \\ c_{LL} &= W + r_0 p_{LL} + r_L p_{LL}. \end{aligned}$$

   It is useful to write these equations as a function of the consumption differences Δ_L and Δ_H defined in the previous section, as well as using the notation for the price differences, where $\alpha_L = p_{LH} - p_{LL}$ and $\alpha_H = p_{HH} - p_{LH}$:

$$\begin{aligned} c_{LL} &= W + r_0 p_{LL} + r_L p_{LL} \\ \Delta_H &= b + r_0 \alpha_H + s_0 \alpha_H + r_H \alpha_H + s_H (p_{HH} - p_H) \\ \Delta_L &= b + r_0 \alpha_L + s_0 (p_{HL} - p_0) + r_L \alpha_L + s_L (p_{HL} - p_L) \\ 0 &= r_H p_{HL} - r_L p_{HL} - s_L (p_{HL} - p_L). \end{aligned}$$

   The third restriction comes from the property of the optimal contract that states that $c_{LH} = c_{HL}$. In matrix form:

$$\begin{bmatrix} c^*_{LL} \\ \Delta_H \\ \Delta_L \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & p_{LL} & 0 & 0 & 0 & p_{LL} & 0 \\ 0 & 1 & \alpha_H & \alpha_H & \alpha_H & p_{HH} - p_H & 0 & 0 \\ 0 & 1 & \alpha_L & p_{HL} - p_0 & 0 & 0 & \alpha_L & p_{LH} - p_L \\ 0 & 0 & 0 & 0 & p_{LH} & 0 & -p_{LH} & -(p_{LH} - p_L) \end{bmatrix} \begin{bmatrix} W \\ b \\ r_0 \\ s_0 \\ r_H \\ s_H \\ r_L \\ s_L \end{bmatrix}.$$

   The solution for the case $p_{LH} > p_0$ and a linear bonus, in which the variables s_0, s_H, r_L and s_L are undetermined, is as follows:

$$\begin{aligned} W &= A_0 + A_1 s_0 + A_2 s_H + A_3 s_L \\ b &= B_0 + B_1 s_0 + B_2 s_H + B_3 s_L \\ r_0 &= C_0 + C_1 s_0 + C_2 s_H + C_3 s_L - r_L \\ r_H &= D_3 s_L + r_L, \end{aligned}$$

where

$$
\begin{aligned}
A_0 &= c_{LL} - \frac{p_{LL}(\Delta_L - \Delta_H)}{\alpha_L - \alpha_H}; A_1 = \frac{p_{LL}(p_{LH} - p_0 - \alpha_H)}{\alpha_L - \alpha_H};\\
A_2 &= -\frac{p_{LL}(p_{HH} - p_H)}{\alpha_L - \alpha_H}; A_3 = \frac{p_{LL}(p_{LH} - p_L)}{\alpha_L - \alpha_H}\left(1 - \frac{\alpha_H}{p_{LH}}\right)\\
B_0 &= \frac{\alpha_L \Delta_H - \alpha_H \Delta_L}{\alpha_L - \alpha_H}; B_1 = \frac{p_{LL}(p_{LH} - p_0 - \alpha_L)}{\alpha_L - \alpha_H};\\
B_2 &= -\frac{\alpha_L(p_{HH} - p_H)}{\alpha_L - \alpha_H}; B_3 = -\frac{\alpha_H(p_{LH} - p_L)}{p_{LH}}\left(1 - \frac{p_{LH} - \alpha_H}{\alpha_L - \alpha_H}\right)\\
C_0 &= \frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H}; C_1 = -\frac{(p_{LH} - p_0 - \alpha_H)}{\alpha_L - \alpha_H};\\
C_2 &= \frac{p_{HH} - p_H}{\alpha_L - \alpha_H}; C_3 = -\frac{p_{LH} - p_L}{\alpha_L - \alpha_H}\left(1 - \frac{\alpha_H}{p_{LH}}\right)\\
D_3 &= \frac{p_{LH} - p_L}{p_{LH}}.
\end{aligned}
$$

It is easy to find parameters and values of s_0, s_H, r_L and s_L for which the implied W, b, r_0, r_H are positive.

- For a type $\mathcal{L}$ firm, $\alpha_H = 0$, and $p_{HH} = p_{LH} = p_H$, so a complex schemes is feasible if

$$W^{\mathcal{L}} = c_{LL} - \frac{q_{LL}}{1 - q_{LL}}\left[\Delta_L - \Delta_H - \pi(1 - q_0)s_0 - \pi(1 - q_L)s_L\right] \geq 0$$

$$b^{\mathcal{L}} = \Delta_H + \frac{q_{LL}}{1 - q_{LL}}\pi(q_0 - q_{LL})s_0 \geq 0$$

$$r_0^{\mathcal{L}} = \frac{1}{\alpha_L}\left[\Delta_L - \Delta_H - (p_{LH} - p_0)s_0 - r_L - (p_{LH} - p_L)s_L\right] \geq 0$$

$$r_H^{\mathcal{L}} = r_L + \frac{p_{LH} - p_L}{p_{LH}}s_L \geq 0.$$

- For a type $\mathcal{H}$ firm, $\alpha_L = 0$, and $p_{LL} = p_{LH} = p_L$, so a complex scheme is feasible if

$$
\begin{aligned}
W^{\mathcal{H}} &= c_{LL} + \frac{p_{LL}}{\alpha_H}\left[\Delta_L - \Delta_H - (p_{LH} - p_0 - \alpha_H)s_0 + (p_{HH} - p_H)s_H\right] \geq 0\\
b^{\mathcal{H}} &= \frac{1}{\alpha_H}\left[\alpha_H \Delta_L + p_{LL}(p_0 - p_{LL})s_0\right] \geq 0\\
r_0^{\mathcal{H}} &= \frac{-1}{\alpha_H}\left[\Delta_L - \Delta_H - (p_{LH} - p_0 - \alpha_H)s_0 + (p_{HH} - p_H)s_H - r_L\right] \geq 0\\
r_H^{\mathcal{H}} &= r_L \geq 0.
\end{aligned}
$$

2. Linear bonus and $p_{LH} \leq p_0$. The two first equations vary slightly, and the system becomes:

$$
\begin{aligned}
c_{LL} &= W + r_0 p_{LL} + r_L p_{LL}\\
\Delta_H &= b + r_0\alpha_H + s_0(p_{HH} - p_0) + r_H\alpha_H + s_H(p_{HH} - p_H) && (13)\\
\Delta_L &= b + r_0\alpha_L + r_L\alpha_L + s_L(p_{HL} - p_L)\\
0 &= r_H p_{HL} - r_L p_{HL} - s_L(p_{HL} - p_L). && (14)
\end{aligned}
$$

In matrix form, the case $p_{LH} < p_0$ becomes

$$\begin{bmatrix} c^*_{LL} \\ \Delta_H \\ \Delta_L \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & p_{LL} & 0 & 0 & 0 & p_{LL} & 0 \\ 0 & 1 & \alpha_H & p_{HH}-p_0 & \alpha_H & p_{HH}-p_H & 0 & 0 \\ 0 & 1 & \alpha_L & 0 & 0 & 0 & \alpha_L & p_{LH}-p_L \\ 0 & 0 & 0 & 0 & p_{LH} & 0 & -p_{LH} & -(p_{LH}-p_L) \end{bmatrix} \begin{bmatrix} W \\ b \\ r_0 \\ s_0 \\ r_H \\ s_H \\ r_L \\ s_L \end{bmatrix}.$$

This implies the following solution for the case $p_{LH} \leq p_0$ and a linear bonus, in which the variables s_0, s_H, r_L and s_L are undetermined:

$$\begin{aligned} W &= c_{LL} - \frac{p_{LL}(\Delta_L - \Delta_H)}{\alpha_L - \alpha_H} - \frac{p_{LL}(p_{HH}-p_0)}{\alpha_L - \alpha_H} s_0 - \frac{p_{LL}(p_{HH}-p_H)}{\alpha_L - \alpha_H} s_H + \frac{p_{LL}(p_{LH}-p_L)}{\alpha_L - \alpha_H}\left(1 - \frac{\alpha_H}{p_{LH}}\right) s_L \\ b &= \frac{\alpha_L \Delta_H - \alpha_H \Delta_L}{\alpha_L - \alpha_H} - \frac{\alpha_L(p_{HH}-p_0)}{\alpha_L - \alpha_H} s_0 - \frac{\alpha_L(p_{HH}-p_H)}{\alpha_L - \alpha_H} s_H + \frac{\alpha_H(p_{LH}-p_L)}{\alpha_L - \alpha_H}\left(1 - \frac{\alpha_L}{p_{LH}}\right) s_L \\ r_0 &= \frac{\Delta_L - \Delta_H}{\alpha_L - \alpha_H} + \frac{(p_{HH}-p_0)}{\alpha_L - \alpha_H} s_0 + \frac{p_{HH}-p_H}{\alpha_L - \alpha_H} s_H - r_L - \frac{p_{LH}-p_L}{\alpha_L - \alpha_H}\left(1 - \frac{\alpha_H}{p_{LH}}\right) s_L \\ r_H &= r_L + \frac{p_{LH}-p_L}{p_{LH}} s_L. \end{aligned}$$

It is easy to find parameters and values of s_0, s_H, r_L and s_L for which the implied W, b, r_0, r_H are positive.

- For a type $\mathcal{L}$ firm, $\alpha_H = 0$, and $p_{HH} = p_{LH} = p_H$, so this solution simplifies to:

$$\begin{aligned} W^{\mathcal{L}} &= c_{LL} - \frac{p_{LL}}{\alpha_L}\left[\Delta_L - \Delta_H + (p_{HH}-p_0)\, s_0 - (p_{LH}-p_L)\, s_L\right] \\ b^{\mathcal{L}} &= \Delta_H - (p_{HH}-p_0)\, s_0 \\ r_0^{\mathcal{L}} &= \frac{1}{\alpha_L}\left[\Delta_L - \Delta_H + (p_{HH}-p_0)\, s_0 - \alpha_L r_L - (p_{LH}-p_L)\, s_L\right] \\ r_H^{\mathcal{L}} &= r_L + \frac{p_{LH}-p_L}{p_{LH}} s_L \end{aligned}$$

- For a type $\mathcal{H}$ firm, $\alpha_L = 0$, and $p_{LL} = p_{LH} = p_L$, so this solution simplifies to:

$$\begin{aligned} W^{\mathcal{H}} &= c_{LL} + \frac{p_{LL}}{\alpha_H}\left[\Delta_L - \Delta_H + (p_{HH}-p_0)\, s_0 + (p_{HH}-p_H)\, s_H\right] \\ b^{\mathcal{H}} &= 0 \\ r_0^{\mathcal{H}} &= \frac{-1}{\alpha_H}\left[\Delta_L - \Delta_H + (p_{HH}-p_0)\, s_0 + (p_{HH}-p_H)\, s_H + \alpha_H r_L\right] \\ r_H^{\mathcal{H}} &= r_L \end{aligned}$$

3. Capped bonus and $p_{LH} > p_0$. The system becomes:

$$\begin{aligned}
c_{LL} &= W + r_0 p_{LL} + r_L p_{LL} \\
\Delta_H &= r_0 \alpha_H + s_0 \alpha_H + r_H \alpha_H + s_H \left(p_{HH} - p_H\right) \\
\Delta_L &= b + r_0 \alpha_L + s_0 \left(p_{HL} - p_0\right) + r_L \alpha_L + s_L \left(p_{HL} - p_L\right) \\
0 &= r_H p_{HL} - r_L p_{HL} - s_L \left(p_{HL} - p_L\right).
\end{aligned}$$

In matrix form:

$$\begin{bmatrix} c^*_{LL} \\ \Delta_H \\ \Delta_L \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & p_{LL} & 0 & 0 & 0 & p_{LL} & 0 \\ 0 & 0 & \alpha_H & \alpha_H & \alpha_H & p_{HH} - p_H & 0 & 0 \\ 0 & 1 & \alpha_L & p_{HL} - p_0 & 0 & 0 & \alpha_L & p_{LH} - p_L \\ 0 & 0 & 0 & 0 & p_{LH} & 0 & -p_{LH} & -\left(p_{LH} - p_L\right) \end{bmatrix} \begin{bmatrix} W \\ b \\ r_0 \\ s_0 \\ r_H \\ s_H \\ r_L \\ s_L \end{bmatrix}.$$

This implies the following solution for the case $p_{LH} > p_0$ and a capped bonus,, in which the variables s_0, s_H, r_L and s_L are undetermined:

$$\begin{aligned}
W &= c_{LL} - \frac{p_{LL}\Delta_H}{\alpha_H} + p_{LL} s_0 + p_{LL}\frac{p_{HH} - p_H}{\alpha_H} s_H + p_{LL}\frac{p_{LH} - p_L}{p_{LH}} s_L \\
b &= \Delta_L - \frac{\alpha_L}{\alpha_H}\Delta_H - \left(p_{LL} - p_0\right) s_0 + \alpha_L \frac{p_{HH} - p_H}{\alpha_H} s_H - \left(p_{LH} - p_L\right)\left(1 - \frac{\alpha_L}{p_{LH}}\right) s_L \\
r_0 &= \frac{\Delta_H}{\alpha_H} - s_0 - \frac{p_{HH} - p_H}{\alpha_H} s_H - r_L - \frac{p_{LH} - p_L}{p_{LH}} s_L \\
r_H &= r_L + \frac{p_{LH} - p_L}{p_{LH}} s_L.
\end{aligned}$$

It is easy to find parameters and values of s_0, s_H, r_L and s_L for which the implied W, b, r_0, r_H are positive.

- For a type $\mathcal{L}$ firm, $\alpha_H = 0$, and $p_{HH} = p_{LH} = p_H$, so this solution simplifies to:

$$\begin{aligned}
W^{\mathcal{L}} &= c_{LL} - p_{LL}\left[\frac{\Delta_H}{0} - s_0 - \frac{p_{LH} - p_L}{p_{LH}} s_L\right] \\
b^{\mathcal{L}} &= \Delta_L - \alpha_L\left[\frac{\Delta_H}{0} + \frac{\left(p_{LL} - p_0\right)}{\alpha_L} s_0 + \left(p_{LH} - p_L\right)\left(\frac{1}{\alpha_L} - \frac{1}{p_{LH}}\right) s_L\right] \\
r_0^{\mathcal{L}} &= \frac{\Delta_H}{0} - s_0 - r_L - \frac{p_{LH} - p_L}{p_{LH}} s_L \\
r_H^{\mathcal{L}} &= r_L + \frac{p_{LH} - p_L}{p_{LH}} s_L
\end{aligned}$$

It is clear that a complex package with a capped bonus is not feasible in this case either.

- For a type $\mathcal{H}$ firm, $\alpha_L = 0$, and $p_{LL} = p_{LH} = p_L$, so this solution simplifies to:

$$\begin{aligned} W^{\mathcal{H}} &= c_{LL} - p_{LL}\left[\frac{\Delta_H}{\alpha_H} - s_0 - \frac{p_{HH} - p_H}{\alpha_H} s_H\right] \\ b^{\mathcal{H}} &= \Delta_L \\ r_0^{\mathcal{H}} &= \frac{\Delta_H}{\alpha_H} - s_0 - \frac{p_{HH} - p_H}{\alpha_H} s_H - r_L \\ r_H^{\mathcal{H}} &= r_L \end{aligned}$$

4. Capped bonus and $p_{LH} \leq p_0$. The system becomes:

$$\begin{aligned} c_{LL} &= W + r_0 p_{LL} + r_L p_{LL} \\ \Delta_H &= r_0 \alpha_H + s_0 (p_{HH} - p_0) + r_H \alpha_H + s_H (p_{HH} - p_H) \\ \Delta_L &= b + r_0 \alpha_L + r_L \alpha_L + s_L (p_{HL} - p_L) \\ 0 &= r_H p_{HL} - r_L p_{HL} - s_L (p_{HL} - p_L). \end{aligned}$$

In matrix form:

$$\begin{bmatrix} c_{LL}^* \\ \Delta_H \\ \Delta_L \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & p_{LL} & 0 & 0 & 0 & p_{LL} & 0 \\ 0 & 0 & \alpha_H & p_{HH} - p_0 & \alpha_H & p_{HH} - p_H & 0 & 0 \\ 0 & 1 & \alpha_L & 0 & 0 & 0 & \alpha_L & p_{LH} - p_L \\ 0 & 0 & 0 & 0 & p_{LH} & 0 & -p_{LH} & -(p_{LH} - p_L) \end{bmatrix} \begin{bmatrix} W \\ b \\ r_0 \\ s_0 \\ r_H \\ s_H \\ r_L \\ s_L \end{bmatrix}.$$

This implies the following solution for the case $p_{LH} > p_0$ and a capped bonus,, in which the variables s_0, s_H, r_L and s_L are undetermined:

$$\begin{aligned} W &= c_{LL} - \frac{p_{LL}\Delta_H}{\alpha_H} + \frac{p_{HH} - p_0}{\alpha_H} p_{LL} s_0 + p_{LL}\frac{p_{HH} - p_H}{\alpha_H} s_H - \frac{p_{LH} - p_L}{p_{LH}} p_{LL} s_L \\ b &= \Delta_L - \frac{\alpha_L}{\alpha_H}\Delta_H + \frac{p_{HH} - p_0}{\alpha_H}\alpha_L s_0 + \alpha_L \frac{p_{HH} - p_H}{\alpha_H} s_H + (p_{LH} - p_L)\left(1 - \frac{\alpha_L}{p_{LH}}\right) s_L \\ r_0 &= \frac{\Delta_H}{\alpha_H} - \frac{p_{HH} - p_0}{\alpha_H} s_0 - \frac{p_{HH} - p_H}{\alpha_H} s_H - r_L - \frac{p_{LH} - p_L}{p_{LH}} s_L \\ r_H &= r_L + \frac{p_{LH} - p_L}{p_{LH}} s_L. \end{aligned}$$

It is easy to find parameters and values of s_0, s_H, r_L and s_L for which the implied W, b, r_0, r_H are positive.

- For a type $\mathcal{L}$ firm, $\alpha_H = 0$, and $p_{HH} = p_{LH} = p_H$, so this solution simplifies to:

$$\begin{aligned} W^{\mathcal{L}} &= c_{LL} - \frac{p_{LL}}{0}\left[\Delta_H - (p_{HH} - p_0)\, s_0\right] \\ b^{\mathcal{L}} &= \Delta_L - \frac{\alpha_L}{0}\left[\Delta_H - (p_{HH} - p_0)\, s_0\right] \\ r_0^{\mathcal{L}} &= \frac{1}{0}\left[\Delta_H - (p_{HH} - p_0)\, s_0\right] \\ r_H^{\mathcal{L}} &= r_L + \frac{p_{LH} - p_L}{p_{LH}} s_L \end{aligned}$$

  So a complex package with a capped bonus is never feasible in this case.

- For a type $\mathcal{H}$ firm, $\alpha_L = 0$, and $p_{LL} = p_{LH} = p_L$, so this solution simplifies to:

$$\begin{aligned} W^{\mathcal{H}} &= c_{LL} - \frac{p_{LL}}{\alpha_H}\left[\Delta_H - (p_{HH} - p_0)\, s_0 - (p_{HH} - p_H)\, s_H\right] \\ b^{\mathcal{H}} &= \Delta_L \\ r_0^{\mathcal{H}} &= \frac{1}{\alpha_H}\left[\Delta_H - (p_{HH} - p_0)\, s_0 - (p_{HH} - p_H)\, s_H - \alpha_H r_L\right] \\ r_H^{\mathcal{H}} &= r_L \end{aligned}$$

■

References

[1] Acharya, Viral V., Kose John, and Rangarajan K. Sundaram. 2000. "Contract Renegotiation and the Optimality of Resetting Executive Stock Options." Journal of Financial Economics 57: 65–101.

[2] Jorge G. Aseff and Manuel S. Santos, "Stock options and managerial optimal contracts," Economic Theory 26, 813–837 (2005)

[3] Brenner, M., R. Sundaram and D. Yermak, "Altering the terms of executive stock options," *Journal of Financial Economics* **57**, (2000) 103-128

[4] Bolton, P., Hamid Mehran, and Joel Shapiro. "Executive Compensation and Risk Taking", Mimeo (2010),

[5] Carter, Mary Ellen, and Luann J. Lynch. 2001. "An Examination of Executive Stock Option Repricing." Journal of Financial Economics 61 (August): 207–25.

[6] Celentani, M. and R. Loveira, "A Simple Explanation of the Relative Performance Evaluation Puzzle," Review of Economic Dynamics, vol. 9(3), pages 525-540, July (2006).

[7] Chance, D., R. Kumar, R. Todd, "The 'repricing' of executive stock options". *Journal of Financial Economics* **57** 129-154 (2000)

[8] Chen, M.A. "Executive Option Repricing, Incentives, and Retention". *The Journal of Finance*, Vol. 59, No. 3, pp. 1167-1199 (2004)

[9] Gian Luca Clementi & Thomas F. Cooley, 2009. "Executive Compensation: Facts," NBER Working Papers 15426, National Bureau of Economic Research, Inc.

[10] Clementi, Gian Luca & Cooley, Thomas F. & Wang, Cheng, 2006. "Stock grants as a commitment device," Journal of Economic Dynamics and Control, Elsevier, vol. 30(11), pages 2191-2216, November.

[11] Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, Yuliy Sannikov, "Dynamic CEO Compensation," forthcoming, Journal of Finance

[12] Edmans, A. and Q. Liu, "Inside Debt", Review of Finance (2010), 1–28

[13] Gabaix, X. and Augustin Landier, "Why Has CEO Pay Increased So Much?", Quarterly Journal of Economics, vol. 123(1), 2008, p. 49-100,

[14] Grossman, Sanford and Oliver D. Hart. "An Analysis of the Principal–Agent Problem." *Econometrica* **51**, Issue 1 (Jan.,1983), 7-46.

[15] Brian J. Hall & Thomas A. Knox. "Underwater Options and the Dynamics of Executive Pay-to-Performance Sensitivities," Journal of Accounting Research, Blackwell Publishing, vol. 42(2), pages 365-412, 05. (2004)

[16] Holmström, B. "Moral Hazard and Observability," *Bell Journal of Economics*, Vol. **10** (1) pp. 74-91. (1979)

[17] Kadan, O. and J. Swinkels. "Stocks or Options? Moral Hazard, Firm Viability, and the Design of Compensation." Review of Financial Studies (2007)

[18] Murphy, Kevin J. 1999. "Executive Compensation." In Handbook of Labor Economics, edited by Orley Ashenfelter and David Card. NewYork: Elsevier Science North Holland, 2485–563.

[19] Wang, Cheng (1997), "Incentives, CEO Compensation and Shareholder Wealth in a Dynamic Agency Model," Journal of Economic Theory 76: 72-105.