C.I.R.P.É.E

Centre Interuniversitaire sur le Risque,
les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **12-07**

The Optimal Timing of CEO Compensation

Pierre Chaigneau

Février/February 2012

Chaigneau: Department of Finance, HEC Montréal and CIRPÉE
pierre.chaigneau@hec.ca

I thank Sudipto Bhattacharya, Amil Dasgupta, Adolfo de Motta, Denis Gromb, Ian Jewitt, David Webb, seminar participants at McGill, the University of Waterloo, and at the Northern Finance Association 2010 annual meeting, for useful comments and suggestions. A previous version of this paper was circulated under the title "The optimal timing of compensation with managerial short-termism". I acknowledge financial support from Institut de Finance Mathématique de Montréal.

Abstract:
This paper extends a standard principal-agent model of CEO compensation by modeling the progressive attenuation of information asymmetries between firm insiders and shareholders in continuous time. In this setting, we show that the optimal timing of compensation results from a tradeoff between the progressive accumulation of noise in the stock price process and the progressive resolution of information asymmetries. Since all points in the stock price process are incrementally informative about the CEO action, we also show that the whole stock price process should a priori be used for compensation purposes. This may however lead CEOs to inefficiently divert resources to repeatedly manipulate the stock price, which is why it might be optimal to use only a few points in the stock price process instead.

Keywords: Executive compensation, principal-agent problem, short-termism, stock-options, deferred compensation, vesting

JEL Classification: G34, J33, M52

The equity-based compensation of CEOs is typically viewed as a mechanism which aligns their interests with shareholders'. Accordingly, CEO compensation contracts have often been interpreted in the light of the principal-agent (shareholders-CEO) model of effort provision. In particular, the literature has established why the stock price is used as a performance measure (e.g. Holmstrom and Tirole (1993)), as well as the functional form which relates a given stock price to the manager's pay (e.g. Innes (1990), Hemmer, Kim, and Verrecchia (2000), Hall and Murphy (2002), Dittmann and Maug (2007), Dittmann, Maug and Spalt (2010), Dittmann and Yu (2011)). However, it is still unclear how firms select the stock price(s) used for compensation purposes. In practice, CEO compensation is usually deferred,[1] vesting schedules differ across firms, and only a few points in the stock price process are used to determine CEO compensation. This paper proposes a model which explains these facts.[2]

To this end, we add two essential ingredients to a standard principal-agent model of CEO pay in the spirit of Holmstrom (1979) and Dittmann and Maug (2007). First, we assume that investors face some initial uncertainty about the firm technology, whether because they are initially asymmetrically informed (relative to firm insiders), or because all agents in the model face this initial uncertainty – we will emphasize the asymmetric information interpretation in what follows. Crucially, we also assume that investors progressively learn about the firm's technology and receive increasingly precise signals about the value of firm profits as time passes. Second, following models of signal-jamming (Narayanan (1985), Stein (1989)), we assume that the manager can divert resources to manipulate the information at the disposal of investors at any point in time.

The first element, namely the progressive attenuation of information asymmetries via a learning process, can explain why it may be optimal to defer CEO pay. Intuitively, this is because the stock price is set endogenously by investors according to their information, so

[1]Deferred compensation is the norm for CEOs: restricted stock grants typically vest after three years, while close to 90% of stock-options granted to S&P500 CEOs in 2006 had a maturity of 6 to 10 years.

[2]Deferred compensation is sometimes justified as a retention tool. However, this hypothesis can only explain the progressive vesting of stocks and stock-options grants, not the fact that they are contingent upon the future, as opposed to the current, stock price.

that progressive learning will raise the informativeness of the stock price process over time.[3] However, the progressive accumulation of noise in the stock price process due to the succession of exogenous shocks to firm value progressively *reduces* its informativeness. The tradeoff that ensues determines the optimal timing of compensation. If only one point in the stock price process is used in the compensation contract, we find that CEO compensation is all the more deferred that there is more learning over time, and that noise is weak – and conversely. In particular, if noise accumulates at a constant rate over time and learning occurs at a decreasing rate (i.e., there is more learning in the early stages of a project than in the latter stages), we find that it is optimal to use the stock price at the point in time when both rates are equal.

This being said, in a model with a risk averse manager and a risk neutral principal in which the stock price is affected both by noise and by learning, we show that it would be optimal to use the whole stock price process for compensation purposes – as opposed to the stock price at one or a few points in time only. This is because, in general, no subset of stock prices is a sufficient statistic for the whole stock price process. Intuitively, using the whole stock price process enables the manager to benefit from "temporal diversification." For example, if the manager were only compensated at one point in time, say τ, then he would be fully exposed to the mispricing due to incomplete learning at time τ, and to the noise accumulated until time τ, but he would *not* be at all exposed to the extent of the mispricing eliminated by learning from time 0 to time τ, or to the noise which accumulates after time τ. Compensating the manager at different points in time enables to decrease his exposure to certain shocks, and to expose him (if only slightly) to other shocks. Since risk aversion is second-order for small risks, we know that it is optimal to reduce a significant exposure to certain shocks at the cost of (at least slightly) exposing the manager to other shocks.

[3] We assume that the degree of information asymmetries is exogenous. However, in Holmstrom and Tirole (1993), the degree of liquidity in the secondary market for the firm's stocks determines the incentives of investors to gather information. In Edmans (2009), large blockholders have strong incentives to gather information on the firm's fundamental value, which reduces information asymmetries and mitigates managerial short-termism. Endogeneizing the degree of information asymmetries is a natural extension of the model. This being said, our result that the timing of compensation depends on the equilibrium degree of information asymmetries relative to the degree of noise would be robust to this extension.

This raises the question as to why managers are only compensated based on the stock price at a few points in time rather than on the whole stock price process. In that regard, practitioners often mention the undesirable side effects that would be (and to some extent already are) created by a constant managerial attention to changes in the stock price. It is often alleged that managers would spend too much time and resources "managing" the stock price rather than focusing on the fundamentals of the company if their compensation depended on the day-to-day changes in the stock price. To incorporate this important limitation of such a compensation scheme, we assume that, at any point in time, a manager can spend resources to temporarily lift the stock price (this is similar to "manipulation" in Goldman and Slezak (2006) and to "short-termism" in Edmans, Gabaix, Sadzik, and Sannikov (2011)).[4] This is wasteful and inefficient, but it may be in the manager's interests, even if investors anticipate this behavior. As a consequence, it may be optimal to use only one or a few points in the stock price process for compensation purposes, so as to minimize the diversion of firm resources for stock price manipulation. Thus, the model highlights an implication of short-termism that is not captured by discrete time models. Whereas these models can typically only consider the effect of short-termism on the timing of compensation, we emphasize that the possibility of short-termism (in the form of signal jamming) need not affect the timing of compensation per se. However, it will have an effect on the number of performance measures (i.e., the number of points in the stock price process) to be included in the compensation contract.

Our main results are driven by the coexistence of the accumulation of noise and of the learning process. This sets this paper apart from the rest of the literature. For example, in the model of Dittmann and Maug (2007) in which there is no learning over time but only an accumulation of noise, the stock price at any given time is a sufficient statistic for future stock prices. This type of model can therefore explain why only one stock price is used for compensation purposes, but it cannot explain deferred executive compensation. By contrast, a

[4] In the same vein, Peng and Roell (2008) study the implications of uncertainty on the equilibrium degree of manipulation for managerial compensation contracts. When ex-post efficiency conflicts with ex-ante efficiency, Axelson and Baliga (2009) emphasize that it may be optimal to allow for manipulation to prevent renegotiation. They also derive implications for the timing of compensation.

number of papers acknowledge that managers have private information (e.g., Narayanan (1985), von Thadden (1995)). We adopt the same premise, but we also model learning about the firm technology in continuous time, which may for example take the form of a progressive attenuation of information asymmetries. This is crucial, as persistent uncertainty or persistent information asymmetries would not be sufficient to explain deferred compensation in our setting. Indeed, what matters is the variability of the stock price at a given point in time from the ex-ante perspective of a risk averse CEO. The accumulation of noise progressively increases it, whereas learning progressively decreases it.

More generally, this paper contributes to a large literature on the use of performance measures for evaluation and compensation purposes (Holmstrom (1979), Holmstrom and Tirole (1993)). To our knowledge, it is the first to explicitly model the stock price process in continuous time in a principal-agent model *and* to identify some tradeoffs associated with the use of any given point in this process, and with the number of points of this process to be included in the compensation contract.

This paper is also related to a more recent literature which discusses the advantages of short-term versus long-term compensation. Long-term compensation may prevent managers from overinvesting in general skills, at the expense of firm-specific skills (Giannetti (2009)). In Bizjak, Brickley, and Coles (1993), higher (and persistent) information asymmetries translate into greater managerial opportunities for short-termism, which can be addressed by postponing compensation. Two more recent papers counter that stronger information asymmetries may on the contrary result in more short-term compensation. In Bolton, Scheinkman and Xiong (2006) and Gopalan, Milbourn, Song, and Thakor (2011), mispricing on the stock market enables a better-informed manager to benefit from overvaluation by selling his stocks early. Bhattacharyya and Cohn (2008) emphasize that short-term equity-based compensation also allows a risk averse manager to quickly reduce his exposure to the firm's risk, which enhances his willingness to undertake risky but valuable projects. Nevertheless, Aghion and Stein (2008) warn that short-term stock-based compensation may induce the manager to inefficiently cater to the stock market.

Since we assume that the manager takes only one action and is paid only once, we do not consider the balance between short-term and long-term incentives (Holmstrom and Tirole (1993)), consumption smoothing (Wang (1997)), borrowing and saving, or conditioning the manager's compensation on his past performances. These last three issues have been extensively studied in the dynamic moral hazard literature, especially in Fama (1980) and Holmstrom (1999). In these papers with repeated moral hazard, the passage of time enables a more precise assessment of the agent's talent, a result also obtained in Radner (1981). In a related paper, Edmans, Gabaix, Sadzik, and Sannikov (2011) also focus on the temporal dimension of a principal-agent model. The effects that they emphasize, however, are different, notably because the passage of time is not accompanied by learning in their model.

This paper makes three contributions. The first is methodological: we model progressive learning in continuous time with a Brownian motion process (learning can take the form of a progressive elimination of information asymmetries). This new method has many potential applications in the economics and finance literatures. Second, we identify the point in time when the stock price process is "most informative" about the manager's action. We argue that it results from a tradeoff between the accumulation of noise in the stock price process and progressive learning. The model thus generates a new theory which relates the optimal timing of managerial compensation to firm characteristics and the degree of stock market efficiency. Third, we present both a puzzle – why isn't the whole stock price process used for compensation purposes? – and a potential explanation – this would induce costly signal jamming on the part of managers.

The paper proceeds as follows. The first section presents the model. The second section derives the stock price process. The third section describes the optimal contract supposing that the manager is paid at one point in time only. The fourth section determines under which conditions this timing of compensation is optimal. The fifth section revisits the model under an alternative information structure. The sixth section discusses the model's predictions, and the seventh section concludes. All proofs are in the Appendix.

1 The model

Basic setup

We consider a firm initially controlled and at least partly owned by a risk-neutral "founder", whose objective is to maximize firm value net of compensation costs, and run by a risk averse manager.[5] More generally, the "founder" can be thought of as a large blockholder with a controlling stake, or as a private equity fund which will take the firm public at time 0. Contracting occurs between the founder and the manager, whom we refer to as firm insiders. This part of the model is in the spirit of Holmstrom (1979) and Dittmann and Maug (2007).

Then we extend the model by explicitly modelling the stock price process, which is set continuously by a multitude of risk-neutral investors according to their information. An important although *not crucial* assumption (as discussed in section 5) is that the manager and the founder have inside information about the firm. This type of information structure with asymmetrically informed investors has already been used in a large literature – it is notably a cornerstone of the pecking order theory of the capital structure and of the dividend signaling theory. Furthermore, the assumption that the founder (or blockholder) is better informed than outside investors is consistent with a recent literature which finds that blockholders have strong incentives to acquire information (Edmans (2009)), and that larger shareholders are indeed more informed (Parrino, Sias, and Starks (2003) and Bushee and Goodman (2007)). In addition, we make the *crucial* assumption that investors progressively learn about firm value. We will propose a new method to model this learning process of the "marginal investor" in continuous time, and we will show that it has important consequences for the timing of equity-based managerial compensation.

[5]It is crucial that the CEO be risk averse, otherwise the first-best outcome could be implemented and the timing of compensation would not matter. The assumed risk neutrality of the founder could be microfounded by viewing him as a well-diversified shareholder or private equity firm, say. These different risk preferences for the principal (founder) and the agent (manager) with respect to firm risk is often used in the literature to obtain a tradeoff between incentives and risk sharing and motivate why the structure of the compensation contract matters.

Production technology, contracting, and trading on the stock market

The contracting stage involves a standard principal-agent relationship. At $t = -2$, the founder offers a compensation contract to the manager. We assume that the manager has CRRA utility and coefficient of relative risk aversion γ (so that $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$ for $\gamma \neq 1$ and $u(x) = \ln(x)$ for $\gamma = 1$)[6] and outside employment opportunities which give him a reservation utility of $\bar{U}$. He will therefore accept the contract if and only if his expected utility is at least equal to $\bar{U}$ in equilibrium. Define his reservation wage as $\bar{W}$, where $u(\bar{W}) \equiv \bar{U}$. For simplicity, the pre-existing wealth of the manager is normalized at zero.[7] If he accepts, the manager then chooses and exerts the level of effort $e \in [0, \infty)$ at $t = -1$. The manager incurs a personal cost of effort additive in the utility of wealth of $C(e)$, where C is increasing and convex, and satisfies the Inada conditions that $\lim_{e \to 0} C'(e) = 0$ and $\lim_{e \to \infty} C'(e) = \infty$ (this guarantees an interior solution to the effort choice problem). Effort represents any action which increases firm value but is personally costly to the manager.

Since the level of effort does not have any effect on the optimal timing of compensation, as we make clear below, we focus on the first step of optimal contracting in Grossman and Hart (1983), which consists of deriving the contract which induces a given level of effort $e^\star$ from the manager at the minimum expected cost. This approach is standard in papers which focus on the structure of compensation (e.g. Hall and Murphy (2002), Dittmann and Maug (2007)).[8] Besides, for any level of $e^\star$, any contract which does not satisfy this first step cannot be optimal. We also assume that $e^\star$ is common knowledge, although our results would be unchanged if the investors had an improper prior about $e^\star$ instead.

[6] The assumption of CRRA utility is common in the managerial compensation literature (e.g., Hall and Murphy (2002), Dittmann and Maug (2007), Edmans, Gabaix, Sadzik, and Sannikov (2011)). We use it for tractability: as explained further in section 3, it enables us to characterize the form of the optimal contract at a point in time. However, our results related to the timing of compensation are robust to an alternative risk averse utility function.

[7] This does not matter for our results related to the timing of compensation. It only matters insofar as the functional form relating the stock price to the pay of the manager is concerned.

[8] Another approach is to rely on the "maximum effort principle", whereby possible levels of effort are bounded from above and it is optimal for the principal to induce the maximum possible effort from the manager (Edmans and Gabaix (2011) and Edmans, Gabaix, Sadzik, and Sannikov (2011)).

To simplify the exposition, we assume that the initial shareholders hedge at $t = -2$ the cost of paying the manager. That is, at $t = -2$ they pay to a third party a premium equal to the expected cost of managerial compensation given the contract offered to the manager. If the manager were paid ex-post instead, a higher stock price would imply a higher compensation, and therefore a higher cost to the firm which would diminish firm value correspondingly. With this hedging assumption, we do not have to consider this feedback mechanism, whose influence on the stock price is second-order in big companies in any case (since firm value is very large relative to the cost of CEO compensation). Note that the risk neutral founder is indifferent between hedging the cost of compensation and paying the manager ex-post.

At $t = T$, shareholders earn the noncontractible payoff produced by the firm,[9] which is equal to

$$a + e + \sigma^N B_T^N \tag{1}$$

where a is a constant which captures the productivity of the firm's technology, and B_t^N is an unobserved Brownian motion defined on $[0, T]$.[10] The assumption that the variability of the stock price due to noise is increasing over time from an ex-ante perspective is standard in the literature (e.g. Dittmann and Maug (2007)), and it could also be microfounded by assuming that at every point in time firm value is affected by an exogenous shock, with i.i.d. shocks. We also assume that a liquidation of the firm at any time is infinitely costly, so that investment is irreversible, and no profitable renegotiation is possible in any case. For simplicity, the discount

[9] The assumption that a contract cannot be written on the payoff per se enables us to focus exclusively on the stock price process as a performance measure. This is consistent with the fact that performance-based pay in big US companies is primarily equity-based, as bonuses are largely independent of performance (Murphy (1999)). The "payoff" of a project may not be contractible for several reasons. First, the payoff may not be purely monetary − think about a project which provides private benefits of some kind to shareholders, say because it is a "socially responsible investment". Second, the project may benefit shareholders in ways which are not captured by the profit and loss statement − think about a project which can generate synergies with other firms that shareholders also own. Third, the payoff may not be realized at time T − in our model, time T is essentially the time when the learning process stops, as we make clear below; the learning process obviously stops if the payoff is realized at time T, but the payoff might also be realized after time T, in which case the expression in (1) is merely the expectation of the payoff as of time T.

[10] It is therefore characterized by the standard properties of Brownian motions. In particular, it is equal to 0 at $t = 0$, and it writes as $B_t^N = \int_0^t dB_s^N$, for any $t \geq 0$. For any $s > t$, the term $B_s^N - B_t^N$ is normally distributed with mean 0 and variance $s - t$.

factor is zero.

From time $t = 0$ to $t = T$, the stock price is set by a multitude of risk-neutral, unconstrained, and competitive investors. As is standard in the literature (e.g. Edmans, Gabaix, Sadzik, and Sannikov (2011)), the manager cannot trade the firm stock at any time, because of the prohibition of insider trading, and because this would enable him to undo the contract. Likewise, the founder cannot trade the firm stock at any time $t \in [0, T]$, unless he places a market order prior to $t = 0$. For example, he may at $t = -2$ commit himself to selling part or all of his stake at some time $t \in [0, T]$. This assumption, which is relaxed in section 5, can be microfounded by invoking the prohibition of insider trading, or by considering that no trade can occur ex-post in equilibrium between a risk neutral founder with inside information and asymmetrically informed risk neutral investors. This being said, trade can occur if the insider places a *market order ex-ante*, before the investors receive a signal on firm value (as described below) and set the stock price accordingly.

The information structure

Both the manager and the founder have inside information about the firm. More specifically, at $t = -2$, only the manager and the founder know the noncontractible value of a in (1) while the investors have an improper prior on a. This reflects information asymmetries on the firm's productivity. In addition, we assume that the founder cannot credibly communicate the value of a to the investors, and that neither the founder nor the investors observe the manager's choice of effort e. We also assume that the investors have improper priors on $\bar{U}$ and $C(e)$ (for any e), so that they do not draw any inferences about a after observing the compensation contract. Without these assumptions, there would not be any meaningful information asymmetries, since investors could potentially infer the value of a by observing the compensation contract. In an extension of the model without information asymmetries presented in section 5, all the assumptions described in this paragraph are relaxed. It should be emphasized that the results of this paper are not so much driven by the information asymmetries on a per se, but rather by the learning process of the investors that we describe in the next paragraphs. Once again, we refer to section 5 for a more thorough discussion of the implications of these assumptions.

The investors progressively learn the value of the conditional expected payoff $a + e + \sigma^N B_t^N$: for all $t \in [0, T]$, they all observe a noncontractible signal v_t at time t. Their initial information is therefore crystallized in the signal v_0. (Since neither the manager nor the founder do anything after $t = -1$, it is inconsequential whether or not they observe the process v_t at any time $t \in [0, T]$.) We assume that the time 0 signal, v_0, and the time t signal, v_t, are respectively

$$v_0 = a + e - \int_0^T \sigma_s^U dB_s^U \tag{2}$$

$$v_t = a + e + \sigma^N B_t^N - \int_t^T \sigma_s^U dB_s^U \tag{3}$$

where B_t^U is a Brownian motion independent from B_t^N, with time-varying, deterministic, and common knowledge diffusion process σ_t^U, which is continuous. The Brownian motion B_t^U is fully realized before $t = 0$, and never observed by any agent at any time (however, the signal v_t is observed by the investors at time t). As will be clear later, the assumption that σ^N is constant (over time) is without loss of generality and can be viewed as a normalization, since what matters at any time is the *relative* level of σ^N and σ_t^U.

The dynamics of the σ_t^U process determine whether most learning occurs at the beginning or at the end of the time interval $[0, T]$: it mostly occurs at the beginning if σ_t^U is decreasing in t, and mainly towards the end if σ_t^U is increasing in t. The overall magnitude of σ_t^U captures the extent of learning, i.e., the extent of information asymmetries: the less private information that firm insiders possess relative to the investors, the smaller the overall magnitude of σ_t^U. If σ_t^U were identically zero, the investors would not learn anything over time, and the passage of time would only add noise to the signal process v_t, where noise is represented by the Brownian motion B_t^N.

This representation of the signal process has the advantage of being very tractable. It also has the following properties. First, at any time t the signal v_t is an unbiased estimate of the conditional expected payoff, i.e., $a + e + \sigma^N B_t^N$, since its probability distribution is centered around this conditional expected payoff. Second, the variance between the signal v_t and the conditional expected payoff at time t is decreasing in t, which represents the progressive attenuation of the mispricing due to learning.

2 The stock price

As already mentioned, the discount factor is zero, and the investors are competitive and risk-neutral. For any $t \in [0, T]$ and $\tau \in [t, T]$, we therefore have in equilibrium:

$$S_t = E_t[S_\tau | \{v_s : s \in [0, t]\}, e^\star] \tag{4}$$

At $t = T$, the payoff $a + e + \sigma^N B_T^N$ of the stock is known, so

$$S_T = a + e + \sigma^N B_T^N \tag{5}$$

Setting $\tau = T$ in (4) and substituting,

$$S_t = E_t[a + e + \sigma^N B_T^N | \{v_s : s \in [0, t]\}, e^\star] \tag{6}$$

At any time $t \in [0, T]$, the market clearing price S_t, at which the net demand from the investors equals the net supply of the stock, is equal to the conditional expectation of the final payoff.[11] Moreover, v_t is a sufficient statistic for $\{v_s : s \in [0, t]\}$. Indeed, for any $\tau \in [0, t)$ we can write

$$v_\tau = S_T + \sigma^N (B_\tau^N - B_T^N) - \int_\tau^T \sigma_s^U dB_s^U \tag{7}$$

[11] Since all investors have the same beliefs, there would admittedly not be any trading in the model, for any $t \in (0, T]$. This being said, we could postulate a mechanism in which a market-maker receives orders from investors, which he must balance against a net supply of the stock of zero. The stock price announced by the market-maker will therefore be the market clearing price, which ensures that the net demand from investors is zero.

$$v_t = S_T + \sigma^N (B_t^N - B_T^N) - \int_t^T \sigma_s^U dB_s^U \quad (8)$$

This implies that, conditional on S_T, we can write $v_\tau = v_t + u_{\tau,t}$ for any $\tau \in [0,t)$, where $u_{\tau,t}$ is a (normally distributed) random variable independent from v_t. Standard arguments (cf. the Proof of Proposition 2) then show that v_t is a sufficient statistic for $\{v_s : s \in [0,t]\}$.

Using this fact, we substitute for v_t (as defined in (3)) in (6) to rewrite it as

$$S_t = E_t\Big[v_t + \int_t^T \sigma_s^U dB_s^U + \sigma^N (B_T^N - B_t^N)|v_t, e^\star\Big] = v_t \quad (9)$$

where the second equality follows from investors' improper prior on a, which implies an improper prior on firm value. This assumption ensures that the updating of beliefs at time t is very simple. We consider a different information structure in section 5, in which the stock price at any time t is a weighted average of the prior belief on firm value and the signal v_t. Applying the equality in (9) to (2) and (3) respectively yields the time 0 and time t stock prices:

$$S_0 = a + e - \int_0^T \sigma_s^U dB_s^U \quad (10)$$

$$S_t = a + e + \sigma^N B_t^N - \int_t^T \sigma_s^U dB_s^U \quad (11)$$

Note that, with a performance measure S_t additively separable into effort and a normally distributed random variable, the monotone likelihood ratio property is satisfied at any given time t. Combining (10) and (11) gives another expression for S_t:

$$S_t = S_0 + \sigma^N B_t^N + \int_0^t \sigma_s^U dB_s^U \quad (12)$$

Finally, the dynamics of S_t may be derived from either (11) or (12):

$$dS_t = \sigma^N dB_t^N + \sigma_t^U dB_t^U \quad (13)$$

This deserves clarification. The investors have an improper prior on firm value, and at time t they do not observe any of the different terms on the right-hand-side of (11). Instead, at any time t they observe v_t, and set the stock price S_t according to this information. Equation (11) neatly captures the two forces at play. On the one hand, the value of the conditional expected payoff, $a + e + \sigma^N B_t^N$, progressively diverges from its ex-ante value of $a + e$ following the arrival of additive shocks, so that noise accumulates at a constant rate in the stock price as time passes. On the other hand, the investors progressively learn the value of the conditional expected payoff, by continuously receiving signals whose precision is increasing over time. Loosely speaking, our model uses the Brownian motion B_t^U to *reduce* uncertainty as time passes – from the investors' perspective.

It is noteworthy that our specification of the signal process v_t in (3) ensures that the stock price at any given time is normally distributed. This is not crucial for our main results. As discussed in the next section, the form of the probability distribution affects the functional form relating managerial compensation to the level of the stock price at a given point in time. But our results related to the timing of managerial compensation are robust to the assumption of a lognormally distributed stock price, which is more common in the literature (e.g. Hall and Murphy (2002), Dittmann and Maug (2007)) and also follows from the assumption that stock returns are i.i.d. and continuously compounded. To see why, for any $t \in [0, T]$, consider the random variable $P_t \equiv \exp\{S_t\}$. Since S_t is normally distributed, P_t is lognormally distributed. In addition, since there is a one-to-one mapping between S_t and P_t, and denoting by $W(S_t)$ the optimal contract conditional on S_t, the optimal contract conditional on P_t is simply $w(P_t) = W(\ln(P_t))$. It is therefore equivalent to work with the lognormally distributed variable P_t or with the normally distributed variable S_t (in addition, the limited liability constraint that $w(P_t) \geq 0$ will be satisfied if $W(\ln(P_t)) = W(S_t) \geq 0$). The choice of a lognormal probability distribution would only affect the results on the functional form relating managerial compensation to the level of the stock price at a given point in time, which is not the issue that we focus on in this paper.

3 Stock price informativeness and the optimal timing of compensation

Supposing that only one point in the stock price process is used in a compensation contract, this section establishes which point in time is optimal, as well as the form of the contract conditional on this stock price. Accordingly, we denote by $\mathbf{1}_t$ the dummy variable equal to one if the stock price at time $t \in [0, T]$ is used in the compensation contract, and equal to zero otherwise.

Among the set of contracts which induce participation and effort $e^\star$ from the manager, we look for the contract which minimizes the cost of compensation.

At the first-best, the manager exerts effort $e^\star$. Since the manager is risk averse and the founder is risk-neutral, the manager is paid a fixed wage which exactly compensates him for his reservation utility $\bar{U}$ and the cost of effort $C(e^\star)$. The timing of compensation therefore does not matter in this case.

At the second-best, the provision of incentives via a compensation contract contingent upon the manager's performance, as reflected in the stock price, will necessitate a deviation from the first-best risk sharing rule. Using standard arguments, it can be shown that this deviation will be costly. The optimal contract associated with any given level of effort will minimize this cost.[12]

Supposing that managerial compensation is contingent upon the time t stock price only ($\mathbf{1}_t = 1$ and $\mathbf{1}_s = 0$ for any $s \neq t$), the incentive constraint which guarantees that the manager optimally exerts effort $e^\star$ under the contract $W(S_t)$ writes as

$$e^\star = \arg \max_{e \in [0,\infty)} E_{-1}[u(W(S_t))] \tag{14}$$

To solve the model, we rely on the validity of the first-order approach. As in Dittmann, Maug,

[12]More precisely, given a level of effort, we can decompose the expected payment to the manager at the second-best into two components: the first-best payment, and the expected payment in excess of the first-best payment. The latter is often referred to as the agency cost of the contract. Clearly, minimizing the expected payment to the manager at the second-best is equivalent to minimizing the agency cost.

and Spalt (2010), we assume that $C''(e)$ is sufficiently high – i.e., the cost function $C(e)$ is sufficiently convex – for the second-order condition to the manager's effort choice problem to be satisfied, which guarantees the validity of the first-order approach. The optimal level of effort $e^\star$ given a contract $W(S_t)$ then solves

$$\frac{d}{de}E_{-1}[u(W(S_t))|e^\star] = C'(e^\star) \tag{15}$$

Given (11), this rewrites as

$$\frac{d}{dS_t}E_{-1}[u(W(S_t))|e^\star] = C'(e^\star) \tag{16}$$

Given a contract $W(S_t)$ and an equilibrium effort $e^\star$, the manager accepts the contract if and only if the following participation constraint is satisfied:

$$E_{-1}[u(W(S_t))|e^\star] \geq \bar{U} \tag{17}$$

Given the constraints in (16) and in (17) and the objective function of the founder, we can determine the form of this optimal contract at a given point in time:

Proposition 1: If the first-order approach is valid, then the optimal contract which is such that $\mathbf{1}_t = 1$ for a given $t \in [0, T]$ and $\mathbf{1}_s = 0$ for $s \neq t$ has the form $W(S_t) = \max\{(\alpha_0 + \alpha_1 S_t)^{\frac{1}{\gamma}}, 0\}$ for two constants α_0 and α_1.

This result is adapted from Dittmann and Maug (2007), who derive it with the conditions described in Holmstrom (1979). Since the monotone likelihood ratio property is satisfied at any given time t, as already noted, we know that the optimal compensation contract described in Proposition 1 is increasing in S_t.

The curvature of the optimal contract depends on the assumed degree of relative risk aversion. With CRRA utility and a level of relative risk aversion lower (respectively higher) than one, the contract is convex (resp. concave) on the region where the limited liability constraint is not binding. With log utility ($\gamma = 1$), the optimal contract takes the form of a call option on the stock price at time t. As already emphasized in Hemmer, Kim, and Verrecchia (2000) and Dittmann and Maug (2007) for example, the functional form $W(S_t)$ relating compensation to the stock price at a given point in time is sensitive to the assumed managerial preferences. Likewise, assuming another probability distribution would alter the form of the optimal contract (see e.g. Dittmann and Maug (2007) in the case of the lognormal distribution, and Hemmer, Kim, and Verrecchia (2000) in the case of the gamma distribution). We do not lay too much emphasis on this result on the form of the optimal contract, because this issue has already been addressed in the literature, and because this particular result is not required for our subsequent analysis.

Given the form of the optimal compensation contract at a given point in time, we now address the optimal timing of compensation. We first establish this important preliminary result:

Proposition 2: If $\mathbf{1}_t = 1$ for a given $t \in [0, T]$ and $\mathbf{1}_s = 0$ for $s \neq t$, then the optimal $t^\star$ is such that $t^\star = \arg\min_t var(S_t)$.

If the manager is only paid based on the stock price at one date, then Proposition 2 says that it is optimal to select the date at which the variance of the stock price from an ex-ante ($t = -2$) perspective is minimal. With the normal distribution, this means that the stock price to select for contingent compensation is the one with the highest informativeness with respect to e, which is intuitive. This is because, with the normal distribution, the variance fully captures the dispersion of the stock price distribution around its mean.[13]

In turn, the date which minimizes the variance of the stock price depends on the *relative* magnitudes of the accumulation of noise over time and of the learning process. From an ex-ante

[13]With other distributions which are not fully described by their mean and variance, this problem is potentially more complex, since there may be tradeoffs between higher-order moments. For example, a given time t may have a lower variance than another time s but a higher kurtosis. In this case, it is a priori not clear which one is "more informative."

perspective, the variance of the stock price at any time $t \in [0,T]$ conditional on a and e writes as

$$var(S_t) = var\Big(\sigma^N B_t^N - \int_t^T \sigma_s^U dB_s^U\Big) = {\sigma^N}^2 t + \int_t^T {\sigma_s^U}^2 ds \tag{18}$$

where we used (11). The optimal timing of compensation is the time t which minimizes this expression. Note that the effect of an increase in σ^N on the variance of the time t stock price in (18) is increasing in t, whereas the effect of a uniform increase in σ_s^U on the variance of the time t stock price in (18) is decreasing in t. This suggests that the optimal timing of compensation is negatively related to the rate σ^N at which the noise accumulates, but positively related to the overall magnitude of information asymmetries.

The economic intuition is the following: on the one hand, a manager is only exposed to the noise which accumulates in the stock price until the time of compensation; on the other hand, a manager is only exposed to the mispricing due to information asymmetries which have not yet been resolved at this time. The optimal timing of compensation trades off these two sources of risk, so that the point in time used in the contract is the one which minimizes the variability of the stock price, and therefore the variability of the manager's equity-based pay, from an ex-ante ($t = -2$) perspective.

In the following Proposition, we distinguish between four especially interesting cases.

Proposition 3: If $\mathbf{1}_t = 1$ for a given $t \in [0,T]$ and $\mathbf{1}_s = 0$ for $s \neq t$, then the optimal timing of compensation $t^\star \equiv \arg\min_t var(S_t)$ is as follows:

- If $\sigma_t^U \geq \sigma^N$ for all $t \in [0,T]$, then $t^\star = T$.
- If $\sigma_t^U \leq \sigma^N$ for all $t \in [0,T]$, then $t^\star = 0$.
- If none of the above applies and σ_t^U is increasing in t on $[0,T]$, then $t^\star = 0$ if $\int_0^T {\sigma_t^U}^2 dt < {\sigma^N}^2 T$, and $t^\star = T$ if $\int_0^T {\sigma_t^U}^2 dt > {\sigma^N}^2 T$.
- If none of the above applies and σ_t^U is decreasing in t on $[0,T]$, then $t^\star$ is implicitly defined

by $\sigma_{t^\star}^U = \sigma^N$.

First, if information asymmetries outweigh noise, compensation should be contingent on the end-of-period stock price, in order for learning to be complete. Second, if noise outweighs information asymmetries, compensation should be contingent on the initial date's stock price, so that the manager is thoroughly protected from noise. Third, if the rate of learning increases over time, the manager should be compensated either based on the initial or the end-of-period stock price. Fourth, if the rate of learning decreases over time, the manager should be compensated at the time when the rate in question is equal to the rate at which noise accumulates. As a rule, the stronger the noise is relative to information asymmetries, the earlier the manager should be paid.

At the second-best optimum, it is noteworthy that the first-best outcome of $e^\star$ and a constant managerial compensation is reached in two special cases. If there are no information asymmetries ($\sigma_t^U \equiv 0$), then compensation is contingent on S_0 ($t^\star = 0$), so that the manager is not exposed to the accumulation of noise, and does not bear any risk. In the absence of noise ($\sigma^N = 0$), it is costless to wait for learning to be complete: compensation is then contingent on S_T ($t^\star = T$), and the manager does not bear any risk either.

4 The tradeoff between manipulation and diversification

This section establishes the conditions under which it is optimal to use the stock price process at one point in time only for compensation purposes. That is, it establishes the conditions under which the results of the previous section apply.

To this end, we extend our stylized model by allowing for stock price "manipulation". Our objective is to capture a situation where the manager can divert firm resources to temporarily increase the stock price. For the problem to be nontrivial, a few conditions need to be satisfied. First, it must be in the interests of a manager with stock-based compensation to manipulate the stock price at some point, otherwise manipulation would not occur. Second, manipulation needs to be costly for the firm, otherwise it would be inconsequential. Third, stock price manipulation needs to be unobservable and undetectable, otherwise the firm could easily either prevent it or prevent the manager from benefiting from it. We make a series of assumptions that guarantee that these conditions are satisfied in our setting, and which are further discussed at the end of this section.

First, for a given arbitrarily small $\Delta > 0$, we assume that the manager can at any time $t \in [0, T - 2\Delta]$ "manipulate" the stock price until time s, where $s \in [t + \Delta, T - \Delta]$. In this case, the dynamics of the signal process become (where we used (9) and (13)):

$$\begin{aligned} dv_\tau &= \mu d\tau + \sigma^N dB_\tau^N + \sigma_\tau^U dB_\tau^U && \text{for } \tau \in [t, t + \Delta] \\ dv_\tau &= \sigma^N dB_\tau^N + \sigma_\tau^U dB_\tau^U && \text{for } \tau \in [t + \Delta, s] \\ dv_\tau &= -\mu d\tau + \sigma^N dB_\tau^N + \sigma_\tau^U dB_\tau^U && \text{for } \tau \in [s, s + \Delta] \end{aligned}$$

For reasons that will become clear below, we assume that $\mu \to 0$. Notice that stock price manipulation will not result in any jump in the stock price process. This assumption that manipulation results in a temporary increase in the stock price is consistent with the empirical evidence in Bhojraj, Hribar, Picconi, and McInnis (2009). Second, we assume that stock price manipulation is costly, with an instantaneous cost of c, i.e., it reduces profits by c for each unit interval where it is undertaken. For example, if the manager manipulates the stock price from time t to time s, then the total cost is $c(s - t)$. More generally, if stock price manipulation is undertaken on intervals whose union has measure Λ, then the total cost is $c\Lambda$, so that the payoff at time T becomes $a + e - c\Lambda + \sigma^N B_T^N$. Third, we assume that stock price manipulation is unobservable per se, and that the manager cannot commit himself ex-ante not to manipulate the stock price. Lastly, we assume that a manager who is indifferent between manipulating and not manipulating does not manipulate.

With $\mu \to 0$, stock price manipulation cannot be detected. Indeed, if we denote by I_t the dummy variable equal to one if stock price manipulation is occurring at a given time t, then, for any k,

$$Pr(dv_t \leq k | I_t = 1) \longrightarrow_{\mu \to 0} Pr(dv_t \leq k | I_t = 0) \tag{19}$$

As $\mu \to 0$, the distribution of dv_t is independent from the value of I_t. There are two implications. First, $Pr(I_t = 1 | dv_t) \to Pr(I_t = 1)$, so that stock price changes are not informative about stock price manipulation.[14] Second, since stock price manipulation does not affect the distribution of dv_t in the limit as $\mu \to 0$, the same arguments as in section 2 yield $dS_t = dv_t$.

First of all, we study the case in which manipulation does not occur because it is impossible ($\mu = 0$). In this case, we show that it is generically suboptimal to leave any point of the stock price process out of the contract.

PROPOSITION 4: If $\mu = 0$ and if σ_t^U is only constant on sets of measure zero, then any optimal contract is such that $\mathbf{1}_t = 1$ almost everywhere.[15]

This is a strong result, which implies that the *whole* stock price process should be used for compensation purposes – as opposed to merely using the stock price at a few dates or on some subintervals. The stock price at any given date contains some unique information because of the coexistence of two processes – one which increases uncertainty, and one which reduces uncertainty (with respect to $a + e$). Therefore, any subset of stock prices does *not* generate a sufficient statistic for the whole stock price process. The informativeness principle (Holmstrom (1979)) then dictates that the whole stock price process be used for compensation purposes.

[14] Without this assumption of $\mu \to 0$, changes in the stock price would be significantly (but not fully) informative about stock price manipulation, which shareholders want to discourage. In this case, it might become optimal for shareholders to punish the manager following a temporarily large increase in the stock price. This may however not be feasible in practice. First, any compensation scheme in which the manager's pay is strictly decreasing in the stock price (or profits) could induce the manager to either burn profits (in some way) at any time t when dS_t reaches a threshold. A similar argument was used more formally in Innes (1990) to rule out compensation schemes in which the manager's pay is a discontinuous function of profits. Second, any compensation scheme which penalizes the manager for temporary bumps in the stock price process could induce him to inefficiently spend resources to smooth out the stock price process (in our model, these "bumps" could indeed occur in the absence of manipulation).

[15] Almost everywhere means on $[0, T]$, with the possible exception of subsets of measure zero. Note that the condition that σ_t^U is only constant on sets of measure zero is sufficient but not necessary.

We now study the case in which manipulation is possible ($\mu > 0$). We show that, if manipulation is sufficiently costly for the firm (c is sufficiently high), then it is optimal to base managerial compensation on the stock price at one point in time only.

Proposition 5: If $\mu > 0$ and

$$c \geq \frac{1}{\Delta}(E[\alpha_0 + \alpha_1(S_{t^\star})] - \bar{W}) \tag{20}$$

then it is optimal to have $\mathbf{1}_t > 0$ for $t = t^\star$ and $\mathbf{1}_t = 0$ for $t \neq t^\star$.

The intuition for this result is the following. If anticipated but impossible to detect (because $\mu \to 0$), stock price manipulations will significantly affect the level of the stock price, but they will not significantly affect its dynamics. Given that manipulation nevertheless temporarily increases the stock price slightly (because $\mu > 0$), it is ex-post optimal for the manager to manipulate the stock price at any time when his compensation depends on it. The assumption that μ is positive but arbitrarily small is therefore crucial in our setting, because it makes manipulation ex-post optimal for the manager *and* impossible to detect. Consequently, when setting the optimal compensation contact, the firm needs to trade off the informativeness benefits of using the whole stock price process (Proposition 4), or at least some points in this process, against the cost of stock price manipulation. For a sufficiently high cost of manipulation c, the latter outweighs the former so much that it is optimal to use only one point in the stock price process for compensation purposes.

Proposition 5 proposes an explanation for the fact that managerial compensation is only contingent on a small subset of stock prices – even though this set is not a sufficient statistic for the whole stock price process, and current technology makes it feasible to write a contract based on the whole stock price process. Indeed, this practice limits the extent of the diversion of firm resources away from long-term profit maximization and toward the manipulation of the information at the disposal of investors. Put differently, it limits inefficient signal-jamming.

The first-best outcome with respect to stock price manipulation would be attained if stock price manipulation were not costly to the firm in equilibrium. This would be the case if (i) the manager could credibly commit not to manipulate the stock price, if (ii) stock price manipulation or its manifestations were contractible, if (iii) manipulation were impossible ($\mu = 0$), or costless for the firm ($c = 0$), or if (iv) the manager could be monitored. In practice, none of these conditions is likely to be easily satisfied. First, given the unobservability and the ex-post optimality of stock-price manipulation, it is a priori not clear which mechanism could be put in place to make such a commitment on the part of the manager possible. Second, it seems unlikely that a contract could specify managers' jobs to such an extent that it would make all instances of manipulation impossible, or that manipulation could always be identified when it occurs. Third, it seems reasonable to assume that, in most cases, stock price manipulation is possible to some extent, and costly for the firm. Legislation such as Sarbanes-Oxley surely imposes some limits on this practice, but it would be a stretch to argue that it is possible to effectively eliminate all instances of stock price manipulation. Fourth, since the manager has a lot of leeway in that regard, it seems implausible that monitoring could completely solve this problem at a reasonable cost. This being said, if a monitoring technology could detect manipulation with a positive probability, we would have a tradeoff between the intensity of (costly) monitoring and the informativeness benefits of using more stock prices for managerial compensation purposes. Even in this case, Proposition 5 would still hold as long as monitoring is sufficiently costly.

In the remainder of this section, we discuss whether several stock prices could be used in a contract. So far, we established that two polar cases are possible: when manipulation is impossible, it is optimal to use the whole stock price process for compensation purposes (Proposition 4); when manipulation is possible and sufficiently costly to the firm, it is optimal to condition managerial compensation on the stock price at one point in time only (Proposition 5). Now suppose that manipulation is possible but less costly than in Proposition 5. Even though positive costs of manipulation might make it too expensive to use the whole stock price process as a basis for compensation, the optimal contract may nevertheless use stock prices at several different points in time (i.e., $\mathbf{1}_t = 1$ for several t). In this case, for the benefits of incremental informativeness to be maximal, we expect that the times t such that $\mathbf{1}_t = 1$ will in general *not* be clustered on a small subset of $[0, T]$, but rather spread out on the whole time interval. This statement needs to be qualified, though. For example, consider the case of strong information asymmetries at the beginning of the project's life (relative to the noise), and a fast learning process which results in information asymmetries being fully resolved at time $\frac{T}{2}$, say. In this case, there would be no benefits to setting $\mathbf{1}_t = 1$ for any $t > \frac{T}{2}$, since the stock price at time $\frac{T}{2}$ would be a sufficient statistic for stock prices at times $s > \frac{T}{2}$. Consequently, all times t such that $\mathbf{1}_t = 1$ would lie in $[0, \frac{T}{2}]$.

Another technical difficulty associated with the case in which several stock prices are used in the contract is the fact that optimal compensation is not necessarily additively separable in these stock prices. For example, suppose that $\mathbf{1}_t = 1$ for $t = \{t_1, t_2\}$ and for these dates only, and denote the unconditional mean of the stock price by $A \equiv a + e$. Then, if the first-order approach applies, the optimal contract takes the form:

$$\frac{1}{u'(W(S_{t_1}, S_{t_2}))} = \lambda + \mu \frac{\varphi_e(S_{t_1}, S_{t_2}; A)}{\varphi(S_{t_1}, S_{t_2}; A)} \tag{21}$$

on the region where the limited liability constraint is not binding, where $\frac{1}{u'(x)} = x$ with log utility. Since the variables S_{t_1} and S_{t_2} are jointly Gaussian (cf. the proof of Proposition 4 with $\mathcal{T} = t_1$ and $\tau = t_2$ for a formal demonstration) with an unconditional mean of A, their joint density can be written as

$$\varphi(S_{t_1}, S_{t_2}; A) = \frac{\exp\left\{ - \frac{\frac{(S_{t_1} - A)^2}{\sigma_{t_1}^2} + \frac{(S_{t_2} - A)^2}{\sigma_{t_2}^2} - 2\rho \frac{S_{t_1} - A}{\sigma_{t_1}} \frac{S_{t_2} - A}{\sigma_{t_2}}}{2(1 - \rho^2)} \right\}}{2\pi \sigma_{t_1} \sigma_{t_2} \sqrt{1 - \rho^2}} \tag{22}$$

where ρ is the correlation coefficient between S_{t_1} and S_{t_2}, and $\sigma_{t_1}^2$ and $\sigma_{t_2}^2$ are their respective variances. In this case, if $\rho \neq 0$, then substituting from (22) in (21), we *cannot* write

$W(S_{t_1}, S_{t_2}) = W(S_{t_1}) + W(S_{t_2})$ (even with log utility). In our model, $\rho = 0$ if and only if $t_1 = 0$ and $t_2 = T$.

5 Initial uncertainty about firm value

In this section, we alter the model by assuming that all agents have the same information about a at the contracting stage. That is, we consider the case where there are no information asymmetries between firm insiders and investors at any time. This also enables us to relax the assumption that the investors have improper priors on $\bar{U}$ and $C(e)$, as well as the assumption that firm insiders cannot trade the stock. Crucially, we will show that this alternative information structure does not affect most of results.

We now assume that all agents (the founder, the manager, and the investors) observe the following public signal at $t = -2$

$$v_{-2} = a + \sigma\epsilon \tag{23}$$

where ϵ is normally distributed with a mean of zero and a variance of one. We explain below why it is important to assume initial uncertainty about a, which is here parameterized by $\sigma > 0$, even if it does not take the form of information asymmetries.

Following the same arguments used in section 2, at any time $t \in [0, T]$ the stock price S_t is determined according to

$$S_t = E[a + e + \sigma^N B_t^N | v_{-2}, v_t, e^\star] \tag{24}$$

In addition,

$$var[v_{-2} + e^\star | a + e + \sigma^N B_t^N] = {\sigma^N}^2 t + \sigma^2 \equiv \Sigma_{-2}(t) \tag{25}$$

$$var[v_t | a + e + \sigma^N B_t^N] = \int_t^T {\sigma_s^U}^2 ds \equiv \Sigma(t) \tag{26}$$

Given that $e = e^\star$ in equilibrium, conditional on $a + e + \sigma^N B_t^N$ (the conditional expected payoff at time t), the variables v_{-2} and v_t are normally distributed with mean $a + e + \sigma^N B_t^N$ and variance $\Sigma_{-2}(t)$ and $\Sigma(t)$, respectively. The standard formula for the Bayesian updating of beliefs with normally distributed variables gives

$$S_t = \frac{\Sigma(t)(v_{-2} + e^\star) + \Sigma_{-2}(t)v_t}{\Sigma(t) + \Sigma_{-2}(t)} \tag{27}$$

At any time t, the stock price is a weighted average of $v_{-2} + e^\star$, the prior belief of investors on firm value, and of the signal v_t that they receive at time t. Note that S_t is a linear combination of normally distributed variables, so that it is also normally distributed. Intuitively, the signal v_t is incrementally informative relative to v_{-2} at any time t if $\sigma > 0$. In addition, the signal v_t is fully revealing about the value of the conditional expected payoff $a + e + \sigma^N B_t^N$ at any time t if $\sigma_t^U \equiv 0$ for any t.

From an ex-ante ($t = -2$) perspective, we show in the Appendix (in "Optimal timing of compensation in section 5") that, as in section 3, the optimal timing of compensation $t^\star$ in this setting is the time which minimizes $var(v_t)$, where

$$var(v_t) = \sigma^{N2} t + \int_t^T \sigma_s^{U2} ds \tag{28}$$

The tradeoff between learning and the accumulation of noise is the same as in the baseline model, and so are the comparative statics of $t^\star$ with respect to a change in σ^N or to a uniform change in σ_s^U. The predictions related to the optimal timing of compensation therefore remain the same as in the baseline model, which suggests that they are not primarily driven by the presence of information asymmetries per se, but rather by the coexistence of the accumulation of noise and of learning.

This is because two new effects cancel out. On the one hand, the relative weight of the signal process v_t in the determination of the stock price process increases over time, which increases the variability of the stock price from an ex-ante perspective, all else equal. This is in turn due to two factors. First, v_t is informative about the noise accumulated from time 0 to time t, while v_{-2} is not. Second, progressive learning means that the precision of v_t with respect to the conditional expected payoff $a + e + \sigma^N B_t^N$ increases over time. Thus, the stock price tends to deviate more from its expected value as of $t = -2$, which notably depends on v_{-2}, so that the manager is more exposed to risk as time passes and t increases, ceteris paribus. On the other hand, the expected firm value at $t = -2$, which is $v_{-2} + e^{\star}$, is not affected by the actual effort e of the manager (it is only affected by the expected, equilibrium level of effort $e^{\star}$), whereas v_t is. To provide adequate effort incentives, the sensitivity of the compensation to the stock price must therefore be increasing in the weight of v_{-2} in the determination of the stock price in (27). For example, in the extreme case where the stock price at a point in time is fully determined by v_{-2} (this would be the case at $t = 0$ if $\sigma = 0$), it would be impossible to provide any effort incentives by conditioning managerial compensation on the stock price at this point in time. Indeed, paying the manager based on the stock price at $t = 0$ would then necessarily violate incentive compatibility.[16] All in all, the two effects discussed in this paragraph cancel out.

Three special cases are especially interesting. First, as $\sigma \to \infty$, the signal v_{-2} is not informative about a, and the stock price is only driven by the signal process $\{v_t : t \in [0, T]\}$. Second, even as $\sigma \to 0$, the stock price at any time $t > 0$ is *not* only driven by the prior v_{-2}, as can be seen in (27). This is because, for any $t > 0$, the signal process is informative about the accumulated noise – even though it is useless for learning about a, since a is approximately known as $\sigma \to 0$. Third, if $\sigma_t^U \equiv 0$ for all t (i.e., there is no additional learning after time 0), then S_t is only driven by v_t for any t, as can be seen in (27). Intuitively, v_0 reveals the value of $a + e$, and v_t reveals the accumulated noise at time t. In this case, it is optimal to pay the manager conditional on the $t = 0$ stock price (indeed, setting $t = 0$ minimizes (28)).

It is noteworthy that most of our previous results are unchanged under this alternative information structure. Proposition 1 relies on the normality of the stock price, which still holds. Proposition 2 relies on the informativeness principle, which still applies. In these two cases, the proof would follow exactly the same lines. Proposition 3 still holds, since the optimal timing of compensation minimizes the same function as in section 3. Proposition 4 relies on the normal

[16]More precisely, supposing that $e = e^{\star}$, then $S_0 = a + e^{\star}$. But then it is impossible to satisfy the incentive constraint (16) by paying the manager based on S_0, since S_0 does not depend on e, which contradicts $e = e^{\star}$.

distribution of stock prices and the informativeness principle, so that it still holds. The proof would follow the same steps, although the notations would change. Proposition 5 relies on the high cost of manipulation, so that it still holds too – once again, the proof would follow the exact same lines.

6 Predictions: firms characteristics and the timing of managerial compensation

The model suggests that the optimal timing of compensation is determined by the relative rates at which noise accumulate, and at which learning occurs or information asymmetries are resolved. More precisely, our results in section 3 suggest the following predictions:

PREDICTION 1: Firms in which information asymmetries between firm insiders and shareholders are stronger (respectively weaker) relative to the variability of profits will tend to have longer (resp. shorter) vesting periods for equity-based managerial compensation.

PREDICTION 2: Firms in which information asymmetries are mostly resolved early will tend to have short vesting periods for equity-based managerial compensation. Firms in which information asymmetries are mostly resolved late will tend to have either a very short or a long vesting period for equity-based managerial compensation.

The corresponding predictions suggested by section 5 are the following:

PREDICTION 3: To the extent that there is some initial uncertainty about the firm's technology, firms in which learning is more important (respectively less important) relative to the variability of profits will tend to have longer (resp. shorter) vesting periods for equity-based managerial compensation.

PREDICTION 4: To the extent that there is some initial uncertainty about the firm's technology, firms in which most learning occurs early will tend to have short vesting periods for equity-based managerial compensation. Firms in which most learning occurs late will tend to have either a very short or a long vesting period for equity-based managerial compensation.

These predictions have a number of implications.

First, firms which manage to produce an almost noise-free performance measure, whether because their profits are not significantly affected by noise, or because they can filter it out, will delay compensation more, ceteris paribus. This will mainly depend on the industrial sector of the firm.

Second, for a given level of noise, firms in which there is not a lot of scope for learning, for example because their insiders do not have a lot of private information (think about firms with recurring investments, or firms in traditional industries that investors easily comprehend) will pay managers relatively early. On the contrary, firms in which there is a lot of scope for learning, for example because insiders have a strong informational advantage relative to shareholders (think about firms with new projects, investment banks with opaque strategies,[17] and firms in ascending industries that investors do not fully understand yet) will use compensation schemes that are more long-term.

Third, all else equal, firms whose stock prices are under more scrutiny should use more short-term stock-price-based compensation. Thus, we expect more (informationally) efficient markets to be associated with more short-term managerial compensation.

The existing empirical evidence, although not conclusive, tends to be consistent with these predictions. Kole (1997) finds that chemicals, machinery and producer firms have higher mean minimum and average waits to exercise stock-options award. Furthermore, this group of firms imposes a longer vesting period on restricted stock grants (50 months) than metals, foods, and

[17] Gopalan, Milbourn, Song, and Thakor (2011) find that executive pay in the financial industry is long-term relative to other industries in 2006-2008.

consumer firms (20 months). Kole describes the former group of firms as belonging to highly innovative industries which require specialized knowledge. It seems that there is more scope for learning in these industries. More recently, Gopalan, Milbourn, Song, and Thakor (2011) rank industries according to the pay duration of their CEO. The top 10 (longest duration) includes finance - trading, defense, utilities, petroleum and natural gas, and pharmaceutical products, arguably industries where inside knowledge is important. The bottom 10 (shortest duration) includes food, retail, printing and publishing, restaurants and hotels, automobiles, apparel, entertainment, agriculture, and precious metals, arguably industries where inside knowledge is less important but which are strongly affected by noise (i.e., random shocks).

Gopalan, Milbourn, Song, and Thakor (2011) also find that pay duration is longer in firms with more R&D expenditures (which may be viewed as a proxy for the scope of learning about the firm's technology), more debt (the pecking order theory of the capital structure predicts that equity financing is more costly when information asymmetries are stronger), and firms whose stock price is less mispriced, and has a lower volatility. It is unclear whether we can relate these latest two findings to our model. First, in their paper, mispricing is measured by stock liquidity and the dispersion of analysts' earnings forecasts. While these two measures are good proxies for mispricing and uncertainty, it is unclear whether they reflect more the presence of noise or the existence of information asymmetries. Second, in our model stock price volatility is driven both by the learning process and by noise. Hence, neither of these two findings can be interpreted in light of our model.

Finally, the growth rate of the firm provides a good measure of information asymmetries. For example, Smith and Watts (1992) hypothesize that information asymmetries are greater in high-growth firms, whereas managerial effort is more easily observable in low-growth firms. In line with our model's predictions, Bizjak, Brickley, and Coles (1993), Cadman, Rusticus, and Sunder (2010), and Gopalan, Milbourn, Song, and Thakor (2011) present evidence that executive compensation in high-growth firms is relatively more long-term.

7 Conclusion

This paper proposes a new model of CEO pay and offers a new explanation for deferred compensation. In our setting, the optimal timing of compensation results from a tradeoff between the progressive accumulation of noise in the stock price process over time on the one hand, and progressive learning by outside investors on the other hand. This tradeoff invalidates the view that the timing of compensation should generally coincide with the horizon of investment projects or with the realization of uncertainty. It also suggests a number of predictions that could explain the variability of vesting periods across firms. While the existing empirical evidence seems broadly consistent with these predictions, future empirical work could more rigorously test this aspect of the model, for example by using the measure of pay duration proposed in Gopalan, Milbourn, Song, and Thakor (2011).

We have also shown that a subset of stock prices is generally not a sufficient statistic for the whole stock price process, so that a model of deferred CEO pay must explain why CEO compensation is typically contingent on the stock price at a few points in time only. A possible response that we have emphasized is that managers can take inefficient actions to temporarily lift the stock price at a cost to the firm. To discourage this behavior, it may be optimal to use only a restricted subset of stock prices for managerial compensation purposes.

More generally, the new framework presented, especially the modelization of the progressive resolution of information asymmetries, could be used and extended to address many issues in the economics and finance literatures. This notably includes a number of issues related to the timing of compensation that we did not specifically address in this paper. For example, while we focused on equity-based compensation, the model could be extended to also study the optimality of profits-based compensation. Next, in practice managers can typically choose the exercise date of their options within some time interval, which differs across firms (Kole (1997)). Likewise, managers typically do not cash in when they receive their options, but when they sell the associated stocks, at a time which may or may not coincide.[18] To further our

[18]This being said, anecdotal evidence suggests that managers typically exercise their options *and* sell the

understanding of deferred compensation and vesting, future research could notably attempt to merge the model of Axelson and Baliga (2009), which deals with these issues and also relies on information asymmetries between managers and shareholders, with the one presented in this paper. It could also be interesting to model the stock price process of another firm, or of a market index, to determine the characteristics of an optimal contract which allows for relative performance evaluation. Lastly, the analytical framework of this paper could be extended to study the positive abnormal returns that tend to follow grants of stock-options to CEOs (Yermack (1997)).

8 References

Aghion, P. and Stein, J.C., 2008, Growth versus margins: destabilizing consequences of giving the stock market what it wants, *Journal of Finance* 63, 1025-1058.

Axelson, Ulf, and Sandeep Baliga, 2009, Liquidity and manipulation of executive compensation schemes, *Review of Financial Studies* 22 (10) 3907-3939.

Bhattacharyya, Sugato, and Jonathan B. Cohn, 2008, The temporal structure of equity compensation, working paper.

Bhojraj, S., Hribar, P., Picconi, M., and McInnis, J., 2009, Making sense of cents: an examination of firms that marginally miss or beat analyst forecasts, *Journal of Finance*, 64, 2361-2388.

Bizjak, John M., James A. Brickley, and Jeffrey L. Coles, 1993, Stock-based incentive compensation and investment behaviour, *Journal of Accounting and Economics* 16, 349-372.

Bolton, Patrick, Jose Scheinkman, Wei Xiong, 2006, Executive compensation and shorttermist behavior in speculative markets, *Review of Economic Studies* 73, 577-610.

Bushee, Brian and Theodore Goodman, 2007, Which institutional investors trade based on private information about earnings and returns? *Journal of Accounting Research* 45, 289-321.

associated stocks at the same date, unless restrictions apply.

Cadman, Brian D., Sunder, Jayanthi, and Rusticus, Tjomme O., 2010, Stock option grant vesting terms: economic and financial reporting determinants, working paper.

Dittmann, Ingolf, and Ernst Maug, 2007, Lower salaries and no options? On the optimal structure of executive pay, *Journal of Finance* 62 (1), 303-343.

Dittmann, Ingolf, Ernst Maug, and Oliver Spalt, 2010, Sticks or carrots? Optimal CEO compensation when managers are risk averse, *Journal of Finance* 65 (6), 2015-2050.

Dittmann, Ingolf, and Ko-Chia Yu, 2009, How important are risk-taking incentives in executive compensation?, working paper.

Edmans, Alex, 2009, Blockholder trading, market efficiency, and managerial myopia, *Journal of Finance* 64 (6), 2481-2513.

Edmans, Alex, Xavier Gabaix, 2011, The effect of risk on the CEO market, *Review of Financial Studies* 24 (8), 2822-2863.

Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, Yuliy Sannikov, 2011, Dynamic CEO compensation, forthcoming, *Journal of Finance.*

Fama, E., 1980, Agency problems and the theory of the firm, *Journal of Political Economy*, 88 (2), 288-307.

Giannetti, Mariassunta, 2009, Serial CEO incentives and the structure of managerial contracts, working paper.

Goldman, Eitan, and Steve K. Slezak, 2006, An equilibrium model of incentive contracts in the presence of information manipulation, *Journal of Financial Economics* 80, 603-626.

Grossman, Sanford J., and Oliver D. Hart, 1983, An analysis of the principal-agent problem, *Econometrica* 51 (1), 7-45.

Hall, Brian J., Kevin J. Murphy, 2002, Stock options for undiversified executives, *Journal of Accounting and Economics* 33, 3-42.

Hemmer, T., O. Kim, and R.E. Verrecchia, 2000, Introducing convexity into optimal compensation contracts, *Journal of Accounting and Economics* 28, 307-327.

Holmstrom, Bengt, 1979, Moral hazard and observability, *Bell Journal of Economics* 10 (1), 74-91.

Holmstrom, Bengt, and Jean Tirole, 1993, Market liquidity and performance monitoring, *Journal of Political Economy* 101 (4) 678-709.

Holmstrom, Bengt, 1999, Managerial incentive problems: a dynamic perspective, *Review of Economic Studies* 66 (1), 169-182.

Ibarrola, P. and Perez-Palomares, A., 2003, Linear sufficiency and linear admissibility in a continuous time Gauss-Markov model, *Journal of Multivariate Analysis* 87, 315-327.

Innes, Robert D., 1990, Limited liability and incentive contracting with ex-ante action choices, *Journal of Economic Theory* 52, 45-67.

Kole, Stacey R., 1997, The complexity of compensation contracts, *Journal of Financial Economics* 43 (1), 79-104.

Murphy, Kevin J., 1999, Executive compensation, in Orley Ashenfelter and David E. Card, eds.: *Handbook of Labor Economics*, North Holland.

Narayanan, M.P., 1985, Managerial incentives for short-term results, *Journal of Finance* 40, 1469-1484.

Parrino, Robert, Richard Sias, and Laura Starks, 2003, Voting with their feet: institutional ownership changes around forced CEO turnover, *Journal of Financial Economics* 68, 3-46.

Peng, Lin, and Ailsa Roell, 2008, Managerial Incentives and Stock Price Manipulation, working paper.

Radner, Roy, 1981, Monitoring cooperative agreements in a repeated principal-agent relationship, *Econometrica* 49 (5), 1127-1148.

Smith, Clifford W., Jr., and Ross L. Watts, 1992, The investment opportunity set and corporate financing, dividend, and compensation policies, *Journal of Financial Economics* 32 (3), 263-292.

Stein, Jeremy C., 1989, Efficient capital markets, inefficient firms: a model of myopic corporate behavior, *Quarterly Journal of Economics* 104, 655-669.

Tirole, Jean, 2006, *The Theory of Corporate Finance* Princeton University Press.

von Thadden, Ernst-Ludwig, 1995, Long-term contracts, short-term investment and monitoring, *Review of Economic Studies* 62, 557-575.

Wang, Cheng, 1997, Incentives, CEO compensation, and shareholder wealth in a dynamic agency model, *Journal of Economic Theory* 76, 72-105.

Yermack, D., 1997, Good timing: CEO stock option awards and company news announcements, *Journal of Finance* 52 (2), 449-476

9 Appendix

Proof of Proposition 1:

To derive the form of the optimal contract, we follow Dittmann and Maug (2007). The Holmstrom (1979) condition below describes the optimal contract $W(S_t)$ when the principal is risk-neutral and the first-order approach applies:

$$\frac{1}{u'(W(S_t))} = \lambda + \mu \frac{\varphi_e(S_t)}{\varphi(S_t)} \tag{29}$$

where λ and μ are the Lagrange multipliers on the participation constraint and the incentive constraint, respectively. With a normal distribution, the likelihood ratio is linear in performance. With CRRA utility, the marginal utility of x is equal to $x^{-\gamma}$. Taking into account the limited liability constraint that $W(S_t) \geq 0$ for any S_t, the optimal contract takes the form:

$$W(S_t) = \begin{cases} (\alpha_0 + \alpha_1 S_t)^{\frac{1}{\gamma}} & \text{if } S_t \geq \bar{S}_t \\ 0 & \text{if } S_t < \bar{S}_t \end{cases}$$

where $\bar{S}_t \equiv -\frac{\alpha_0}{\alpha_1}$, and α_0 and α_1 are two constants which are determined to satisfy the participation constraint and the incentive constraint.

Proof of Proposition 2:

Using (11), the variance of the stock price conditional on a and e at any time $t \in [0, T]$ writes as

$$var(S_t) = var\Big(\sigma^N B_t^N - \int_t^T \sigma_s^U dB_s^U\Big) \equiv \sigma_t^2 \tag{30}$$

Furthermore, S_t is normally distributed, for any t, and we let $t^\star$ be defined by $t^\star \equiv \arg\min_t var(S_t)$.

For any $t \in [0, T]$, we can then write

$$S_t = a + e + u_t + y_t \tag{31}$$

where u_t is normally distributed with mean 0 and variance $\sigma_{t^\star}^2$, and y_t is normally distributed with mean 0 and variance $\sigma_t^2 - \sigma_{t^\star}^2$, and these two variables are independent, for any given t. This notably implies that

$$S_{t^\star} = a + e + u_{t^\star} \tag{32}$$

From an ex-ante perspective, any contract contingent on $S_{t^\star}$ is equivalent to a contract contingent on

$$U_t \equiv a + e + u_t \tag{33}$$

for any t. This is because the unconditional probability distribution of $a+e+u_{t^\star}$ is identical to the unconditional probability distribution of $a+e+u_t$, for any t. Accordingly, we will compare the contract contingent on S_t for a given t to the contract contingent on U_t – which is equivalent to compare the contract contingent on S_t to the contract contingent on $S_{t^\star}$.

Let $A \equiv a+e$ denote the unconditional mean of S_t for any $t \in [0, T]$. Let $\psi(U_t, y_t; A)$ denote the density of the joint distribution of U_t and y_t given A, $\phi(U_t; A)$ denote the density of U_t given A, and $\varphi(y_t)$ denote the density of y_t. Given that a and e are constants, and the random variables u_t and y_t are independent, the joint density of U_t and y_t is simply equal to the product of densities:

$$\psi(U_t, y_t; A) = \phi(U_t; A)\varphi(y_t) \tag{34}$$

That is, y_t is uninformative about A.

The next part of the proof largely follows the proof of the first part of Proposition 3 in Holmstrom (1979).

Consider any contract $W(U_t, y_t)$. For every U_t, define $W(U_t)$ such that

$$\int u(W(U_t, y_t))\varphi(y_t)dy_t = \int u(W(U_t))\varphi(y_t)dy_t \tag{35}$$

In addition,

$$\int u(W(U_t))\varphi(y_t)dy_t = u(W(U_t))\int \varphi(y_t)dy_t = u(W(U_t)) \tag{36}$$

Using successively (34), (35), and (36),

$$\int u(W(U_t, y_t))\psi(U_t, y_t; A)dU_t dy_t = \int u(W(U_t, y_t))\phi(U_t; A)\varphi(y_t)dU_t dy_t$$

$$= \int u(W(U_t))\phi(U_t; A)\varphi(y_t)dU_t dy_t = \int u(W(U_t))\phi(U_t; A)dU_t \tag{37}$$

For any A, the expected utility of the manager conditional on A under the contract $W(U_t, y_t)$ is the same as under the contract $W(U_t)$. It follows that the manager chooses the same effort e and enjoys the same expected utility in equilibrium with both contracts. Moreover, using the strict concavity of u, Jensen inequality, and (35), we get:

$$\int W(U_t, y_t)\varphi(y_t)dy_t > \int W(U_t)\varphi(y_t)dy_t \tag{38}$$

Since this holds for every U_t, integrating yields

$$\int W(U_t, y_t)\psi(U_t, y_t; A)dU_t dy_t > \int W(U_t)\psi(U_t, y_t; A)dU_t dy_t \tag{39}$$

For a given t, this shows that the expected cost of compensation is strictly lower with the contract contingent on U_t, or equivalently on $S_{t^\star}$, than with a contract contingent on both U_t and

y_t, which includes a contract contingent on S_t.

PROOF OF PROPOSITION 3:

Given Proposition 2, we need only prove in each case that $t^\star = \arg\min_t var(S_t)$. The variance of the stock price at a given time t conditional on a and e is

$$var(S_t|e) = \sigma^{N^2} t + \int_t^T \sigma_s^{U^2} ds \tag{40}$$

First, if $\sigma_t^U \geq \sigma^N$ for all t, then it immediately follows from (40) that $t^\star = T$.

Second, if $\sigma_t^U \leq \sigma^N$ for all t, then it immediately follows from (40) that $t^\star = 0$.

Third, suppose that σ_t^U is increasing in t, and none of the former two cases apply. Since we assumed that σ_t^U is continuous, we can define r by $\sigma_r^U \equiv \sigma^N$. Setting $t > r$ results in a lower variance than with $t = r$, since $\sigma_t^U > \sigma^N$ for any $t > r$ (as in the first case above). Setting $t < r$ results in a lower variance than with $t = r$, since $\sigma_t^U < \sigma^N$ for any $t < r$ (as in the second case above). The optimum is therefore given by a corner solution: either $t^\star = 0$, or $t^\star = T$. The variance is then respectively:

$$var(S_0|e) = \int_0^T \sigma_t^{U^2} dt \tag{41}$$

$$var(S_T|e) = \sigma^{N^2} T \tag{42}$$

It follows that $t^\star = 0$ if $\sigma^{N^2} T > \int_0^T \sigma_t^{U^2} dt$, and $t^\star = T$ if $\sigma^{N^2} T < \int_0^T \sigma_t^{U^2} dt$.

Fourth, suppose that σ_t^U is decreasing in t, and none of the first two cases apply. Then, since we assumed that σ_t^U is continuous, there exists s, $0 \leq s \leq T$ such that $\sigma_s^U = \sigma^N$. This implicitly defines s. Setting $t > s$ results in a higher variance than with $t = s$, since $\sigma_t^U < \sigma^N$ for any $t > s$ (as in the second case above). Setting $t < s$ results in a higher variance than with $t = s$, since $\sigma_t^U > \sigma^N$ for any $t < s$ (as in the first case above). The optimum is therefore characterized by an interior solution:

$$\sigma_{t^\star}^U = \sigma^N \tag{43}$$

Proof of Proposition 4:

We denote the set of times t such that $\mathbf{1}_t = 1$ by $\mathcal{T}$. Any contract which satisfies the incentive constraint is such that $\mathcal{T}$ is nonempty. Given any nonempty $\mathcal{T}$, we show below that a contract such that $\mathcal{T} \neq [0, T]$ is suboptimal.

For a given contract, we denote by $\mathcal{I} \equiv [t_1, t^1] \bigcup [t_2, t^2] \bigcup \ldots \bigcup [t_m, t^m]$ the set of intervals of measure nonzero such that $\mathbf{1}_t = 1$ if and only if $t \in [t_i, t^i]$ for $i \in \{1, 2, \ldots, m\}$. In addition, we denote by $\mathcal{J} \equiv \{\bar{t}_1, \bar{t}_2, \ldots, \bar{t}_n\}$ the set of times $t \in [0, T]$ such that, for any $t \in \mathcal{J}$, we have $\mathbf{1}_t = 1$, $\lim_{t \to t_j, t < t_j} \mathbf{1}_t = 0$, and $\lim_{t \to t^j, t > t^j} \mathbf{1}_t = 0$. By definition of $\mathcal{T}$, we have $\mathcal{T} = \mathcal{I} \bigcup \mathcal{J}$. Since $\mathcal{T}$ is nonempty, we know that $m + n \geq 1$.

Suppose that a contract is such that there exists at least one nonempty subset of $[0, T]$, which we denote by $(\underline{\tau}, \bar{\tau})$, such that $\mathbf{1}_t = 0$ for any $t \in (\underline{\tau}, \bar{\tau})$.

Consider the Gaussian process S_t on any given interval $[t_i, t^i] \subset \mathcal{I}$. The covariance $cov(S_t, S_s)$ exists and is known for any $\{t, s\} \in [t_i, t^i]^2$, and the process S_t is square mean continuous on any such interval because of (13). We then know from Theorem 4 in Ibarrola and Perez-Palomares (2003) that linear sufficiency and sufficiency (i.e., ordinary sufficiency) are equivalent, i.e., there exists a function $\lambda(t)$ such that $\theta_i \equiv \int_{t_i}^{t^i} \lambda(t) S_t dt$ is a sufficient statistic for the stochastic process $\{S_t : t \in [t_i, t^i]\}$.

Given a contract and the associated $\mathcal{I}$ and $\mathcal{J}$, it is equivalent to use the set of sufficient statistics $\{\theta_1, \theta_2, \ldots, \theta_m\}$ and $\{S_{\bar{t}_1}, S_{\bar{t}_2}, \ldots, S_{\bar{t}_n}\}$ or to use the set of stock prices in $\mathcal{I}$ and $\{S_{\bar{t}_1}, S_{\bar{t}_2}, \ldots, S_{\bar{t}_n}\}$. In addition, for any t, S_t is normally distributed. Moreover, since θ_i is equal to a sum of normally distributed random variables, θ_i is normally distributed for any i.

We also know that, given a set of normally distributed variables, linear sufficiency and sufficiency are equivalent (e.g., Drygas (1983), Mueller (1987)). That is, there are functions $\eta(i)$ and $\mu(j)$ such that

$$\theta \equiv \sum_{i=1}^{m} \eta(i) \theta_i + \sum_{j=1}^{n} \mu(j) S_{\bar{t}_j} \tag{44}$$

is a sufficient statistic for the set of variables $\{\theta_1, \theta_2, \ldots, \theta_m; S_{\bar{t}_1}, S_{\bar{t}_2}, \ldots, S_{\bar{t}_n}\}$. That is, it is a sufficient statistic for $\{S_t : t \in \mathcal{T}\}$.

Consider one subinterval $(\underline{\tau}, \bar{\tau})$ such that $\mathbf{1}_t = 0$ for any $t \in (\underline{\tau}, \bar{\tau})$. We know from Holmstrom (1979) that, given any $\tau \in (\underline{\tau}, \bar{\tau})$, θ as defined in (44) is a sufficient statistic for $\{S_t : t \in [0, T]\}$ only if it is possible to write

$$\psi(\theta, S_\tau; A) = \phi(\theta; A)\varphi(\theta, S_\tau) \tag{45}$$

where $A \equiv a + e$ is the unconditional mean of S_t for any $t \in [0, T]$. If this equality held, then the process $\{S_t : t \in \mathcal{T}\}$ would generate a sufficient statistic for $\{S_t : t \in \mathcal{T}\} \bigcup S_\tau$. We will show that this is not the case, which implies a fortiori that $\{S_t : t \in \mathcal{T}\}$ does not generate a sufficient statistic for $\{S_t : t \in [0, T]\}$.

A linear combination of θ and S_τ writes as

$$\delta_1\Big(\sum_{i=1}^{m}\eta(i)\int_{t\in[t_i,t^i]}\lambda(t)\Big(A+\sigma^N B_t^N-\int_t^T\sigma_s^U dB_s^U\Big)dt+\sum_{j=1}^{n}\mu(j)\Big(A+\sigma^N B_{\bar{t}_j}^N-\int_{\bar{t}_j}^T\sigma_s^U dB_s^U\Big)\Big)$$

$$+\delta_2\Big(A+\sigma^N B_\tau^N-\int_\tau^T\sigma_s^U dB_s^U\Big) \tag{46}$$

for two constants δ_1 and δ_2. This expression involves sums of normally distributed random variables, so that it is normally distributed, for any δ_1 and δ_2. We conclude that any linear combination of θ and S_τ is normally distributed, so that θ and S_τ are jointly normally distributed.

The density of the joint distribution of θ and S_τ can be written as

$$\psi(\theta, S_\tau; A) = \frac{\exp\Big\{-\frac{\frac{(\theta-A)^2}{\sigma_\theta^2}+\frac{(S_\tau-A)^2}{\sigma_\tau^2}-2\rho\frac{\theta-A}{\sigma_\theta}\frac{S_\tau-A}{\sigma_\tau}}{2(1-\rho^2)}\Big\}}{2\pi\sigma_\theta\sigma_\tau\sqrt{1-\rho^2}} \tag{47}$$

where σ_τ^2 is the variance of S_τ, σ_θ^2 is the variance of θ, and ρ is the correlation coefficient between θ and S_τ.

Condition (45) holds if and only if we can rewrite the expression inside the exponential function in (47) as

$$h_1(\theta, S_\tau) + h_2(\theta; A) \tag{48}$$

for two functions h_1 and h_2. However, some calculations show that, factoring out the term $S_\tau A$ within the exponential function in (47), the associated coefficient is

$$-\frac{1}{2(1-\rho^2)}\Big[-\frac{2}{\sigma_\tau^2} + 2\rho\frac{1}{\sigma_\tau\sigma_\theta}\Big] \tag{49}$$

which is different from zero if and only if $\frac{1}{\sigma_\tau} = \frac{\rho}{\sigma_\theta}$, which, given the definition of ρ ($\rho \equiv \frac{cov(\theta,S_\tau)}{\sigma_\theta\sigma_\tau}$), is satisfied if and only if $\frac{cov(\theta,S_\tau)}{\sigma_\theta^2} = 1$.

We will show that $\frac{cov(\theta,S_\tau)}{\sigma_\theta^2} = 1$ can only hold on sets of measure zero. Using the same expression for θ and S_τ as in (46), we get

$$\begin{aligned}
cov(\theta, S_\tau) = cov\Big(&\sum_{i=1}^{m}\eta(i)\int_{t\in[t_i,t^i]}\lambda(t)\Big(A + \sigma^N B_t^N - \int_t^T \sigma_s^U dB_s^U\Big)dt \\
&+\sum_{j=1}^{n}\mu(j)\Big(A + \sigma^N B_{\bar{t}_j}^N - \int_{\bar{t}_j}^T \sigma_s^U dB_s^U\Big),\ A + \sigma^N B_\tau^N - \int_\tau^T \sigma_s^U dB_s^U\Big) \\
&= \sum_{i=1}^{m}\eta(i)\int_{t\in[t_i,t^i]}\Big(\lambda(t)\min\{t,\tau\}{\sigma^N}^2\Big)dt + \sum_{j=1}^{n}\mu(j){\sigma^N}^2\min\{\bar{t}_j,\tau\} \\
&+\sum_{i=1}^{m}\eta(i)\int_{t\in[t_i,t^i]}\Big(\lambda(t)\int_{\max\{t,\tau\}}^T {\sigma_s^U}^2 ds\Big)dt + \sum_{j=1}^{n}\mu(j)\int_{\max\{\bar{t}_j,\tau\}}^T {\sigma_s^U}^2 ds
\end{aligned} \tag{50}$$

To alleviate notations, define

$$\hat{\lambda}(t) = \begin{cases} \eta_i\lambda(t) & \text{if } \eta_i\lambda(t) > 0 \\ 0 & \text{if } \eta_i\lambda(t) = 0 \end{cases}$$

Using these notations, then splitting the sums and integrals into two parts, we can rewrite (50) as

$$cov(\theta, S_\tau) = \int_0^T \hat{\lambda}(t)\min\{t,\tau\}{\sigma^N}^2 dt + \sum_{j=1}^{n}\mu(j){\sigma^N}^2\min\{\bar{t}_j,\tau\}$$

$$+\int_0^T \hat{\lambda}(t) \int_{\max\{t,\tau\}}^T \sigma_s^{U^2} ds dt + \sum_{j=1}^n \mu(j) \int_{\max\{\bar{t}_j,\tau\}}^T \sigma_s^{U^2} ds$$

$$= \int_0^\tau \hat{\lambda}(t) t \sigma^{N^2} dt + \sum_{j:\bar{t}_j<\tau} \mu(j)\sigma^{N^2}\bar{t}_j + \int_\tau^T \hat{\lambda}(t)\tau\sigma^{N^2} dt + \sum_{j:\bar{t}_j>\tau} \mu(j)\sigma^{N^2}\tau$$

$$+\int_0^\tau \hat{\lambda}(t) \int_\tau^T \sigma_s^{U^2} ds dt + \sum_{j:\bar{t}_j<\tau} \mu(j) \int_\tau^T \sigma_s^{U^2} ds$$

$$+\int_\tau^T \hat{\lambda}(t) \int_t^T \sigma_s^{U^2} ds dt + \sum_{j:\bar{t}_j>\tau} \mu(j) \int_{\bar{t}_j}^T \sigma_s^{U^2} ds \tag{51}$$

The first derivative of $cov(\theta, S_\tau)$ with respect to τ writes as

$$\frac{d}{d\tau}\{cov(\theta, S_\tau)\} = \hat{\lambda}(\tau)\tau\sigma^{N^2}dt - \hat{\lambda}(\tau)\tau\sigma^{N^2}dt + \int_\tau^T \hat{\lambda}(t)\sigma^{N^2}dt + \sum_{j:\bar{t}_j>\tau} \mu(j)\sigma^{N^2}$$

$$+\hat{\lambda}(\tau)\int_\tau^T \sigma_s^{U^2}ds + \int_0^\tau \hat{\lambda}(t)(-\sigma_\tau^{U^2})dt - \hat{\lambda}(\tau)\int_\tau^T \sigma_s^{U^2}ds + \sum_{j:\bar{t}_j<\tau}\mu(j)(-\sigma_\tau^{U^2}) \tag{52}$$

Eliminating the terms that cancel out, we get

$$\frac{d}{d\tau}\{cov(\theta, S_\tau)\} = \int_\tau^T \hat{\lambda}(t)\sigma^{N^2}dt + \sum_{j:\bar{t}_j>\tau} \mu(j)\sigma^{N^2}$$

$$-\int_0^\tau \hat{\lambda}(t)\sigma_\tau^{U^2}dt - \sum_{j:\bar{t}_j<\tau}\mu(j)\sigma_\tau^{U^2} \tag{53}$$

The second derivative of $cov(\theta, S_\tau)$ with respect to τ writes as

$$\frac{d^2}{d\tau^2}\{cov(\theta, S_\tau)\} = -\hat{\lambda}(\tau)\sigma^{N^2} - \hat{\lambda}(\tau)\sigma_\tau^{U^2} - \int_0^\tau \hat{\lambda}(t)\frac{d\sigma_\tau^{U^2}}{d\tau}dt - \sum_{j:\bar{t}_j<\tau}\mu(j)\frac{d\sigma_\tau^{U^2}}{d\tau} \tag{54}$$

Since by construction $\mathbf{1}_\tau = 0$, we have $\hat{\lambda}(\tau) = 0$, and the second-derivative of $cov(\theta, S_\tau)$ with respect to τ has the sign of $-\frac{d\sigma_\tau^{U^2}}{d\tau}$. But we assumed that σ_t^U is only constant on sets of measure zero, so that $-\frac{d\sigma_\tau^{U^2}}{d\tau}$ is either strictly positive or strictly negative almost everywhere. This implies that the second-derivative of $cov(\theta, S_\tau)$ with respect to τ is strictly convex or strictly concave almost everywhere in a neighborhood of any τ such that $\mathbf{1}_\tau = 0$. This in turn implies that, for any such τ, $cov(\theta, S_\tau) = 1$ can only hold on a set of measure zero, i.e., $\frac{cov(\theta,S_\tau)}{\sigma_\theta^2} \neq 1$ almost everywhere.

Proof of Proposition 5:

First, we show that the manager manipulates the stock price at time $t - \Delta$ if and only if $\mathbf{1}_t = 1$. Denote the set of times t such that $\mathbf{1}_t = 1$ by $\mathcal{T}$. Given that $dS_t = dv_t$ for any t and that managerial compensation is increasing in S_t, it immediately follows from the effect of manipulation on v_t as described in section 4 that a manager will manipulate the stock price on all intervals $[t - \Delta, t]$, for any $t \in \mathcal{T}$. Denoting by Λ the measure of the union of all such intervals, the total cost of manipulation is $c\Lambda$.

Second, given a compensation contract W, the participation constraint in (17) imposes that $E[u(W)] \geq \bar{U} \equiv \frac{\bar{W}^{1-\gamma}}{1-\gamma}$. Jensen's inequality and the concavity of the utility function then imply that $E[W] > \bar{W}$. The reservation wage $\bar{W}$ is therefore a lower bound for the expected cost of compensation.

Third, we compare the relative cost of the optimal contract with $\mathbf{1}_t = 1$ for only one $t \in [0, T]$ considered above (call it a contract of type A) and any contract with $\mathbf{1}_t = 1$ for more than one time $t \in [0, T]$ (call it a contract of type B). On the one hand, the expected cost of compensation with a contract of type A is equal to $E[\alpha_0 + \alpha_1(S_t)]$, while the expected cost of compensation with a contract of type B is bounded below by $\bar{W}$. On the other hand, the cost of a contract of type A in terms of manipulation is $c\Delta$, while the cost of a contract of type B is at least equal to $2c\Delta$. It follows that, for the parameters α_0, α_1, and $t^\star$ of the optimal type A contract, this contract dominates any contract of type B if

$$c\Delta \geq E[\alpha_0 + \alpha_1(S_{t^\star})] - \bar{W} \tag{55}$$

Optimal timing of compensation in section 5

Define

$$A_t \equiv \frac{\Sigma(t)}{\Sigma(t) + \Sigma_{-2}(t)} \tag{56}$$

$$B_t \equiv \frac{\Sigma_{-2}(t)}{\Sigma(t) + \Sigma_{-2}(t)} \tag{57}$$

$$\sigma_t^2 \equiv \int_t^T {\sigma_s^U}^2 ds + {\sigma^N}^2 t \tag{58}$$

$$t^\star \equiv \arg\min_t var(v_t) \tag{59}$$

As of $t = -2$, for any $t \in [0, T]$ we can decompose any v_t into the following additive components:

$$v_t = a + e + u_t + y_t \tag{60}$$

where u_t is normally distributed with mean 0 and variance $\sigma_{t^\star}^2$, and y_t is normally distributed with mean 0 and variance $\sigma_t^2 - \sigma_{t^\star}^2$, and these two variables are independent, for any given t. This notably implies that, for any t, v_t has the same unconditional distribution as $v_{t^\star} + y_t$.

Note that, because of (27), S_t can be written as

$$S_t = A_t v_{-2} + B_t v_t \tag{61}$$

and $v_{t^\star}$ can be written as

$$v_{t^\star} = B_{t^\star}^{-1}(S_{t^\star} - A_{t^\star} v_{-2}) \tag{62}$$

Hence, from a $t = -2$ perspective, it is equivalent to make compensation contingent on S_t or on

$$Z_t \equiv A_t v_{-2} + B_t \Big(B_{t^\star}^{-1}(S_{t^\star} - A_{t^\star} v_{-2}) + y_t \Big) \tag{63}$$

because Z_t and S_t have the same unconditional probability distribution, by construction of Z_t.

But

$$Z_t = S_{t^\star} + (A_t - B_t B_{t^\star}^{-1} A_{t^\star}) v_{-2} + B_t y_t \tag{64}$$

where $(A_t - B_t B_{t^\star}^{-1} A_{t^\star}) v_{-2}$ and B_t are known constants at $t = -2$, and y_t is a white noise.

We have shown that it is equivalent to use either S_t, or a linear transformation of $S_{t^\star}$ to which is added a white noise, as a measure of performance. Standard arguments (cf. the Proof of Proposition 2) then show that a compensation contract contingent on $S_{t^\star}$ dominates a contract contingent on S_t.