JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 56, No. 6, Sept. 2021, pp. 2039–2071
© THE AUTHOR(S), 2020. PUBLISHED BY CAMBRIDGE UNIVERSITY PRESS ON BEHALF OF THE MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON
doi:10.1017/S0022109020000642

Dynamic Compensation Under Uncertainty Shocks and Limited Commitment

Felix Zhiyu Feng
University of Washington Foster School of Business
ffeng@uw.edu (corresponding author)

Abstract

This article studies dynamic compensation and risk management under cash-flow volatility shocks. The optimal contract depends critically on firms' ability to make good on promised future payments to managers. When volatility is low, firms with full commitment ability implement high pay–performance sensitivity to motivate effort from managers, and they impose large penalties on the arrival of volatility shocks to incentivize prudent risk management. In contrast, firms with limited commitment may allow excessive risk taking in exchange for low pay–performance sensitivity. When volatility becomes high, firms with full commitment defer compensation more, whereas firms with limited commitment must expedite payments.

I. Introduction

Firms today rely heavily on incentive pay, which constitutes the majority of overall firm cost as well as total employee compensation. Economically, the prevalence of incentive pay stems from the uncertainty over how managers' performance reflects their true effort. Moreover, uncertainty shocks have been argued as an important contributing factor to business cycles. In particular, many studies (e.g., Bloom (2009), He and Krishnamurthy (2013), Atkeson, Eisfeldt, and Weill (2017), and Di Tella (2017)) document the sudden and drastic escalation of uncertainty as a prominent feature of financial crises, during which times corporate compensation practices usually come under intense public scrutiny.

Despite the importance of understanding the relationship between compensation and uncertainty, theoretical models that study this relationship have been mostly static (e.g., Diamond (1998), Biais and Casamatta (1999), and Palomino

This article is the main chapter of my PhD dissertation at Duke University. I am indebted to Barney Hartman-Glaser, Adriano Rampini, Curtis Taylor, and S. Viswanathan for their invaluable guidance. I thank an anonymous referee, Jennifer Conrad (the editor), Attila Ambrus, Bruno Biais, Philip Bond, Micheal Brolley, Brendan Daley, Mike Fishman, George Georgiadis, Brett Green, Dirk Jenter, Rui Li, Andrey Malenko, Philipp Sadowski, Ajay Subramanian, Alexei Tchistyi, Felipe Varas, Paolo Volpin, Mark Westerfield, and Ming Yang and seminar and conference participants at Duke University, Georgia State University, the University of Notre Dame, the University of Oxford, the University of Washington, the 2013 EFA Annual Meeting, the 2015 American Finance Association (AFA) Annual Meeting, the 2017 Northwestern Relational Contracts Conference, the 2018 Society for Economic Dynamics (SED) Annual Meeting, and the 2018 Econometric Society North American Summer Meeting (NASM) for helpful comments and suggestions. All errors are my own.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

and Prat (2003)). Recently, the continuous-time contracting techniques pioneered by DeMarzo and Sannikov (2006), Biais, Mariotti, Plantin, and Rochet (2007), and Sannikov (2008) allow the study of dynamic agency and risk management while maintaining analytic tractability. However, among existing models, the uncertainty shock either terminates the contract immediately (DeMarzo, Livdan, and Tchistyi (2013)) or is assumed to be transitory (Li and Williams (2018), Wong (2019)), leaving the contracting environment unchanged before and after the shock. Arguably, neither approach is ideal when uncertainty shocks represent persistent changes to economic regimes, as in the case of financial crises. Moreover, all of the aforementioned studies assume that firms have perfect commitment power, which is generally hard to come by in practice.[1] Importantly, the concern over firm commitment varies with the degree of uncertainty and is likely the greatest when uncertainty is high and firm value is low.

The innovation of this article is thus twofold: First, it offers a theoretical framework that studies the *dynamic* interactions between compensation and *permanent* uncertainty shocks. The basic structure involves a principal–agent model with multidimensional moral hazard and regime switching in uncertainty. Second, the article highlights the importance of commitment power on the principal's side when reconciling the theoretical predictions with empirical observations. In particular, the predictions of the model when the principal has only limited commitment power may shed new light on managerial compensation and risk choices in practice, especially during crisis times.

In the baseline model, a firm hires a manager to run its operations. The manager privately controls the productivity and riskiness of the firm's operations. Riskiness is modeled as the arrival intensity of a volatility shock, after which the firm's cash flows become permanently more volatile. The firm must offer sufficient incentives to the manager to not only work hard to achieve high productivity but also to exercise prudence to achieve low risk because the manager would prefer to do neither. The contract uses promises of future compensation, that is, the manager's continuation utility, to provide the necessary incentives. Continuation utility increases whenever the firm realizes high cash flows *and* is not hit by the volatility shock, and it decreases otherwise. Once it clears a "payment" hurdle, the manager gets to cash out the difference (in the forms of cash bonuses or stock dividends); once it falls below a "termination" hurdle, the manager is fired.

Absent any other contracting friction, the optimal contract of the baseline model implements low risk from managers by imposing a large penalty for the arrival of the volatility shock and uses high pay–performance sensitivity (PPS) as an incentive for high productivity. When the volatility shock arrives, the contract *raises* the payment hurdle. This means that the manager must achieve stronger cumulative performance before receiving actual payments, effectively making the manager's compensation *more back-loaded* during high-volatility times.

[1] For example, at-will employment is ubiquitous in many countries: Firms can terminate employment contracts without having to establish just cause or give warning. They may default on labor contracts because of limited or costly access to financial markets (e.g., Palacios and Stomper (2014), Ellul, Pagano, and Schivardi (2017)). Finally, firms can also liquidate at any time, after which they are no longer liable for any future compensation promised to employees.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

This benchmark theoretical prediction, however, does not reconcile with empirical findings on how managerial compensation varies with firm risk. For example, Peters and Wagner (2014) show that at the industry level, higher volatility of equity value is strongly associated with *higher* current compensation, whereas Edmans, Gabaix, and Jenter (2017) find a similar relationship based on firm-level volatility.[2] The financial sector's behavior during the Great Recession is also at odds with the benchmark result. Despite being in a period of heightened uncertainty, many firms actually expedited their employees' compensation.[3] Overall, huge losses of company wealth notwithstanding, many bankers and executives accused of being responsible for leading their firms into financial turmoil still received substantial compensation. Public questioning of the fairness of the current compensation structure and scholarly debate over its efficiency, unsurprisingly, ensued.

In light of these observations, I reexamine the implications of uncertainty for compensation and find that the discrepancy between the benchmark predictions and actual observations stems from an overlooked but critical market friction: firms' ability to commit to long-term incentive contracts. I show that the theoretical predictions of the model are drastically different when firms have only *limited commitment*, that is, when they can unilaterally terminate contracts at any time. Limited commitment restricts firms' ability to use promises of future payments as incentives, exacerbating the moral-hazard problem. Consequently, firms with limited commitment may forgo prudent risk management in order to reduce the incentive necessary to maintain high productivity. In other words, they may allow excessive risk taking in exchange for the low PPS needed. When the volatility shock arrives, the total surplus of the contractual relationship decreases, and firms with limited commitment are forced to *lower* the payment hurdle and award managers cash payments after small positive performance. That is, managers' compensation becomes *less back-loaded* during high-volatility times.[4]

The results of this article have potential policy implications. High compensation during financial turmoil is often regarded as a sign of managerial entrenchment and suboptimal contracts. In the aftermath of the Great Recession, many blame the managerial compensation scheme for not aligning managers' interests with those of investors, and the widely touted solution is delaying and capping executive compensation.[5] However, the effectiveness of such measures hinges on the credibility of firms' promises to pay the full value in the future. If managers believe that firms'

[2]See Edmans and Gabaix (2016) for a summary of existing theories and empirical findings on volatility and managerial compensation.

[3]For example, Bank of America announced in 2009 that it would accelerate the vesting period of its employees' stocks from 2012 to 2011, stating that such action is "critical for retention of talent" during volatile times. Deloitte reported instances of financial firms lowering performance hurdles during the crisis so that employees would still be entitled to payments such as bonuses. See *Forbes Insights* (2009) and *Management Today* (2010).

[4]These managers are, however, simultaneously subject to a higher probability of contract termination. The overall effect of the volatility shock on their continuation utility is still negative. As such, those payments should not be construed as a "reward" for excessive risk taking.

[5]For instance, the Troubled Asset Relief Program (TARP) limits the ability of the executives of TARP firms to cash out their restricted stocks until the government is repaid in full. See TARP Standards for Compensation and Corporate Governance, 74 Fed. Reg. 28,394, 28,410 (June 15, 2009).

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

ability to deliver on promised future payments will be compromised during times of stress when investors withdraw investment, they will demand even higher and more immediate compensation during such times. Moreover, the lack of commitment can limit the extent to which risk-taking behaviors will be penalized and may in turn lead managers to expose firms to even greater risks ex ante.

This article bridges several strands of literature. It follows DeMarzo and Sannikov (2006), Biais et al. (2007), Sannikov (2008), and Zhu (2013) in modeling productivity management. It belongs to the broad research on risk management in static settings (e.g., Diamond (1998), Biais and Casamatta (1999), Palomino and Prat (2003), Hellwig (2009), and Ray and Robson (2012)) and in dynamic settings (Makarov and Plantin (2015), Barron, Georgiadis, and Swinkels (2020)). These studies typically have a 1-dimensional agency problem: The agent either chooses effort or risk. In contrast, this article allows the agent to take both actions privately. The agent's risk choice is modeled as the arrival intensity of a regime switch of cash-flow volatility, adopting the technique from Hoffmann and Pfeil (2010), Piskorski and Tchistyi (2010), Narita (2011), and DeMarzo, Fishman, He, and Wang (2012). These studies, however, assume exogenous probability of the regime switch, whereas this article features endogenous probability and hence endogenous risk-taking incentives.

The interaction of dynamic productivity and risk management in this article shares the most similarities with DeMarzo et al. (2013), Li and Williams (2018), and Wong (2019) but with critical differences. First, a central message of DeMarzo et al. (2013) and Wong (2019) is that incentives for high productivity increase the agent's desire for excessive risk taking. Consequently, high effort and efficient risk management may not be simultaneously feasible. In my model, incentives to prevent excessive risk taking lead to stronger incentives being required to maintain high productivity. The optimal contract may forgo efficient risk management even when it is jointly feasible with implementing high productivity. Put differently, whereas existing studies consider two mechanisms that determine the optimal risk choice, incentive compatibility (IC) and the agent's limited liability, I consider an additional mechanism: the principal's optimality. Second, the "risky" events in Li and Williams (2018) and Wong (2019) incur an exogenous amount of capital/cash-flow losses. This implies that the effect of the risky event is *transitory*: Each event results in some loss of the agent's continuation utility (the state variable), but the contracting environment is the same before and after the occurrence of such an event. This article, in contrast, considers a *persistent* shock from which the loss of firm value is endogenous because of the difficulty of maintaining high productivity in the high-uncertainty state. The contracting relationship can continue after the volatility shock arrives, but the environment is different. This also differs from DeMarzo et al. (2013), in which the risky event exogenously terminates the contract immediately. Consequently, the model of this article is arguably more suitable for studying the compensation and risk-management practices following market-wide events such as financial crises.

Finally, all of the aforementioned studies assume full commitment on the part of the principal, whereas my article highlights that the implications of uncertainty for managerial compensation are distinctly different under limited commitment.

To the best of my knowledge, this is the first article to jointly consider 2-dimensional dynamic moral hazard (productivity and risk management) in the presence of volatility regime switching and limited commitment.

The article evolves as follows: Section II describes the main model and the optimal contract. Section III derives the model implications and discusses their empirical relevance. Section IV explores several extensions. Section V concludes. All proofs are in Appendix A.

II. Model

The model is introduced in 2 steps. First, I describe a principal–agent environment with volatility regime switching and characterize the optimal contract, assuming that the principal has full commitment power. Then, I relax the assumption of full commitment power and derive the new resulting optimal contract.

A. Basic Environment

Time is continuous. A principal (she), representing a firm, must hire an agent, representing a manager (he), to run a project. Both the principal and the agent are risk neutral. The cash flow Y_t of the project is as follows:

$$dY_t = (\mu - s_t)dt + \sigma_t dZ_t, \tag{1}$$

where Z_t is a standard Brownian motion. The agent manages the cash-flow growth rate, or the productivity of the project, via a binary variable $s_t \in \{0, S\}$ $(S > 0)$, representing *active* productivity management (i.e., high cash-flow growth μ) and *passive* productivity management (i.e., low cash-flow growth $\mu - S$). In standard moral-hazard problems, these are often referred to as "working" and "shirking." Shirking results in lower productivity but also private benefit (specified later) for the agent. The principal observes Y_t but not s_t. She discounts future cash flows by r and the agent by $\gamma > r$, so the agent is more impatient. The agent is protected by limited liability via an outside option with a value normalized to 0. The principal has an outside option $\underline{V} > 0$, which can be interpreted as the firm's liquidation value or the net profit from firing the old manager and hiring a new one.

The volatility of the cash flow, σ_t, is stochastic, taking one of two values, σ_l or σ_h, where $\sigma_l < \sigma_h$, representing, respectively, "normal" and "crisis" times. For simplicity, I assume that the economy starts with σ_l and transitions to σ_h, the arrival of which is characterized by a Poisson (jump) process N_t. The σ_h state is absorbing (no jumping back to σ_l), but the main results of this article remain under recurring states (details available from the author). Because there can be at most one state transition in equilibrium, I refer to it as the *volatility shock*. This volatility shock can be contracted on because it is publicly observable.

The arrival intensity of the volatility shock is an endogenous choice of the agent. I assume that the shock arrives with Poisson rate $\pi + a$, where $\pi > 0$, and $a \in \{0, A\}$ is a binary choice between a *prudent* action (low risk) and a *risky* action (high risk), respectively. The agent chooses a privately and derives personal benefit

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

by opting for the risky action. Here, a can be understood as the agent's risk management, and $a = A$ represents *excessive* risk taking that is undesirable to the principal but can only be eliminated through effort that is costly for the agent.[6] Because there is a positive probability ($\pi > 0$) that a volatility shock will arrive even if the agent acts prudently, the agent's actual risk-management choice remains unobservable to the principal. Finally, there is no risk-management choice after the volatility shock arrives because the σ_h state is absorbing.

In sum, the agency problem (in the low-volatility state) is *2-dimensional*: It can manifest in the agent's productivity-management choice s and in his risk-management choice a. I assume the agent's personal benefit, denoted as $B(s,a)$, has the following properties:

Assumption 1. $B(0,0) = 0$; $B(S,0) = \lambda S$, $\lambda \in (0,1)$; $B(0,A) = \delta A$, $\delta \in (0,1)$; and

$$\frac{\partial^2 B(s,a)}{\partial s \partial a} < 0.$$

In words, the agent prefers not to "work hard" in general, whether the job is to maintain a high level of productivity ($s = 0$) or a low level of risk ($a = 0$). He derives no personal benefit from the $(s = 0, a = 0)$ action profile and enjoys maximum personal benefit by avoiding both management duties. However, the personal benefit of avoiding one duty and that of avoiding the other are (imperfect) substitutes, captured by the negative cross-derivative. If the action space is continuous, this is equivalent to the personal-benefit function being jointly concave in the two choice variables. This is economically reasonable: Concave personal benefit is necessary for a well-defined, socially optimal choice of actions. In the case of binary actions, a jointly concave private benefit $B(s,a)$ implies that $\max\{\lambda S, \delta A\} < B(S,A) < \lambda S + \delta A$.[7]

An important result to be derived in this article is that although both low productivity and excessive risk are socially suboptimal, it may not be optimal for the principal to prevent both. To reduce the complexity of the analysis, for most of the article, I impose the following assumption:

Assumption 2. S is sufficiently large such that the principal always prefers to implement high productivity ($s = 0$).

[6] One can also interpret π as systemic risk and a as idiosyncratic risk. Although the elimination of the latter is possible, it requires due diligence on the part of the manager, which is personally costly to him.

[7] An example of $B(s,a)$ that satisfies Assumption 1 is the constant elasticity of substitution (CES) function $B(s,a) = [(\lambda s)^\rho + (\delta a)^\rho]^{\frac{1}{\rho}}$ with $\rho > 1$. Alternatively, one can write the model in terms of costly effort, and the assumption would be that the cost function is jointly convex: The marginal cost of effort in maintaining high productivity is increasing with the level of effort in sustaining low risk. The model is written based on personal benefits mainly because of its simpler exposition: The term $B(s_t, a_t)dt$ in equation (3) and the term δa in equation (8) are both 0 under IC. If the model is written based on costly effort, there would be extra terms in equations (3), (8), and (9) and a number of other places. Appendix B shows that the assumption of joint concavity or convexity is equivalent to a risk-averse agent with a nonadditive utility function and a continuous-action space.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

This is a standard simplification, particularly in studies that also involve risk management (e.g., DeMarzo et al. (2013), Wong (2019)). Detailed discussion on the exact requirement for high productivity to be always optimal is provided in Section IV.

B. Incentive Compatibility

To find the optimal contract, it is necessary to establish the IC conditions first. Formally, let $(\Omega,\mathcal{F},P)$ denote the probability space and $\{\mathcal{F}_t\}_{t\geq 0}$ the filtration generated by the cash-flow history $\{Y_t\}_{t\geq 0}$ and the state-transition history $\{N_t\}_{t\geq 0}$. A contract specifies a nondecreasing payment process $\{C_t\}_{t\geq 0}$ to the agent, a stopping time τ when the contract is terminated, and a sequence of recommended actions $\{s_t\}_{t\geq 0},\{a_t\}_{t\geq 0}$. Given the contract, the agent chooses a given set of actions $\{\hat{s}_t,\hat{a}_t\}_{t\geq 0}$. The agent's objective function is the expected discounted value of consumption plus private benefits:

$$W_t^{\hat{s},\hat{a}} = E_{\hat{s},\hat{a}}\left[\int_t^{\tau} e^{-\gamma(u-t)}(dC_u + B(\hat{s}_u,\hat{a}_u)du)|\mathcal{F}_t\right]. \tag{2}$$

As in the previous literature, W_t is referred to as the agent's *continuation utility*. A contract is defined as being *incentive compatible* if $(s_t,a_t)=(\hat{s}_t,\hat{a}_t)|_{t\geq 0}$. The following proposition characterizes the dynamics of W_t and the IC condition:

Proposition 1. Given any contract and any sequence of the agent's choices, there exist predictable, finite processes $\beta_t(0\leq t\leq\tau)$ and $\phi_t(0\leq t\leq\tau)$, such that in the low-volatility state, W_t evolves according to

$$\begin{aligned} dW_t = {} & \gamma W_t dt - B(s_t,a_t)dt - dC_t + \beta_t(dY_t - (\mu - s_t)dt) \\ & - \phi_t(dN_t - (\pi + a_t)dt). \end{aligned} \tag{3}$$

The IC condition is

$$\{s_t,a_t\} = \arg\max_{\hat{s}_t,\hat{a}_t} B(\hat{s}_t,\hat{a}_t) - \beta_t\hat{s}_t - \phi_t\hat{a}_t. \tag{4}$$

If $(s=0,a=0)$ is implemented, then

$$dW_t = \gamma W_t dt - dC_t + \beta_t\sigma_l dZ_t - \phi(dN_t - \pi dt). \tag{5}$$

If $(s=0,a=A)$ is implemented, then

$$dW_t = \gamma W_t dt - \delta A dt - dC_t + \beta_t\sigma_l dZ_t - \phi(dN_t - (\pi + A)dt). \tag{6}$$

dW_t and the IC condition in the high-volatility state follow equations (3) and (4) with $\phi_t = a_t = 0$.

Equation (3) can be derived using the standard martingale method in Hoffmann and Pfeil (2010), Piskorski and Tchistyi (2010), and DeMarzo et al. (2012). The first three terms on its right-hand side reflect dynamic promise

keeping, and the last two terms reflect the agent's incentives. β_t is the agent's PPS: For every dollar of extra cash flow $(dY_t - (\mu - s_t)dt)$, the agent's continuation utility changes by β_t dollars. ϕ_t is the adjustment of the agent's continuation utility when the volatility shock occurs. If $\phi > 0$, it acts as a punishment to the agent. Before the volatility shock hits, the agent is compensated with $(\pi + a_t)\phi_t$ to keep W a martingale.

The IC condition in equation (4) is the solution to a maximization over the agent's flow payoff. For the contract to implement a particular (s,a) combination, it must be that the agent does not derive a higher payoff from any other combination. This allows me to derive the IC conditions, summarized as the following:

Proposition 2. $\{s_t = 0, a_t = 0\}$ are implemented if and only if

$$(IC : \alpha = 0) \qquad \beta_t \geq \lambda, \quad \phi_t \geq \delta,$$

and $\{s_t = 0, a_t = A\}$ are implemented if and only if

$$(IC : \alpha = A) \qquad \beta_t \geq b \equiv \max\left\{\lambda - (\delta - \phi)\frac{A}{S}, \frac{B(S,A) - \delta A}{S}\right\}, \quad \phi_t < \delta.$$

In particular, $\phi < \delta$ implies $b < \lambda$.

Intuitively, because the agent prefers to perform neither active productivity management nor prudent risk management, in order to implement $s = a = 0$, the contract must provide the agent with sufficient "skin in the game" $(\beta \geq \lambda)$ and punish the agent with a sufficiently large penalty $(\phi \geq \delta)$ whenever the volatility shock occurs. In contrast, to implement $s = 0$ and $a = A$, the principal cannot punish the agent too much $(\phi < \delta)$.

Importantly, if the contract allows excessive risk taking, the minimal degree of PPS (β) needed to maintain high productivity is lower than that needed when the contract also incentivizes low risk. That is because the agent's marginal private benefit from adopting passive productivity management is λ when $a = 0$ and $(B(S,A) - \delta A)/S < \lambda$ when $a = A$. In other words, when free from the duty of risk management, the agent requires less incentive to focus on the duty of productivity management. This result plays a critical role in the later analysis of which level of risk the optimal contract implements.

C. Optimal Contract with Full Commitment

I now proceed to solve the optimal contract, assuming first that the principal has full power of commitment. The principal's objective function is the expected discounted value of the cash flow minus payments to the agent, with the expectation taken under the probability measure associated with the agent's choices $(\hat{s}, \hat{a})$:

$$V_t^{\hat{s},\hat{a}} = E_{\hat{s},\hat{a}}\left[\int_t^\tau e^{-r(u-t)}(dY_u - dC_u) + e^{-r\tau}\underline{V}\middle|\mathcal{F}_t\right]. \tag{7}$$

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

The *optimal contract* is defined as a contract that i) maximizes the principal's objective function over the set of contracts, ii) satisfies the IC conditions, and iii) grants the agent nonnegative initial utility W_0. The optimal contract can be summarized with two endogenously linked value functions of the principal, one for each volatility state, given as the following:

Proposition 3. The optimal contract under volatility regime switching with full commitment on the principal's side is characterized by a pair of concave and twice-differentiable value functions $V_j(W)$, $j \in \{l,h\}$ defined on $W \in [0 W_j^C]$, respectively. $V_l(W)$ solves the following Hamilton–Jacobi–Bellman (HJB) equation in the low-volatility state:

$$rV_l(W) = \max_{\beta,\phi,a} \; \mu + [\gamma W - \delta a + (\pi + a)\phi] V_l'(W) + \frac{1}{2}\beta^2 \sigma_l^2 V_l''(W) + (\pi + a)[V_h(W - \phi) - V_l(W)], \tag{8}$$

subject to the IC condition (IC: $a = 0$) if $a = 0$ or (IC: $a = A$) if $a = A$. W_l^C is chosen such that $V_l'(W_l^C) = -1$ and $V_l''(W_l^C) = 0$. Payments to the agent $dC_t = \max\{W_t - W_l^C, 0\}$. The contract terminates when $W = 0$, at which point $V_l(0) = \underline{V}$.

$V_h(W)$ solves the following HJB equation in the high-volatility state:

$$rV_h(W) = \max_{\beta \geq \lambda} \left[\mu + \gamma W V_h'(W) + \frac{1}{2}\beta^2 \sigma_h^2 V_h''(W) \right]. \tag{9}$$

subject to $V_h'(W_h^C) = -1$ and $V_h''(W_h^C) = 0$. Payments to the agent $dC_t = \max\{W_t - W_h^C, 0\}$. The contract terminates when $W_t = 0$, at which point $V_h(0) = \underline{V}$.

In the high-volatility state, there is no future state transition and thus no risk-management choice. The HJB equation and the associated boundary conditions for V_h are identical to those in DeMarzo and Sannikov (2006). Nevertheless, two features of V_h are worth mentioning because they are related to the case of limited commitment, to be discussed later.

First, V_h is concave, and because β only appears on the $V_h''(W) < 0$ term, it is always optimal to choose $\beta = \lambda$. Intuitively, β is set at the minimal level allowed by the IC condition because the cost of incentivizing productivity management is captured by the likelihood of hitting the termination boundary $W = 0$ (i.e., the likelihood of a costly separation as a result of bad luck). Conditional on the contract being incentive compatible, the principal prefers the lowest possible variability in dW, which comes from the lowest possible level of PPS. This is an important mechanism in play throughout the model.

Second, the payment boundary W_h^C is a reflecting boundary. Once crossed, an instant payment $dC_t = W_t - W_h^C$ is made to bring W_t back to W_h^C.[8] W_h^C is optimally

[8] W_h^C can be interpreted as a bonus hurdle or, as shown by DeMarzo and Sannikov (2006), the dividend payment boundary if the contract is implemented with inside equity.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

set based on two conditions: a "smooth-pasting" condition $V_h'(W_h^C) = -1$ and a "super-contact" condition $V_h''(W_h^C) = 0$.[9] Substituting those boundary conditions back to the HJB equation yields

$$rV_h(W_h^C) + \gamma W_h^C = \mu. \tag{10}$$

Intuitively, it is beneficial to postpone payment to the agent by making W_h^C larger because doing so reduces the risk of early contract termination. Postponing payment is sensible until the condition in equation (10) is met, when "the principal and agent's required expected returns exhaust the available expected cash flows" (DeMarzo and Sannikov (2006)). I refer to equation (10) as the *optimal payment frontier*. This frontier is "second best" compared with the "first-best" frontier $r(V + W) = \mu$. The efficiency loss from the second-best frontier stems from the fact that the agent is impatient and discounts his promised utility by $\gamma > r$, in addition to the possibility of contract termination. The position of W_h^C in relation to this frontier relies crucially on the principal's having full commitment power and will be reexamined later under limited commitment.

The value function in the low-volatility state, $V_l(W)$, has one major difference from that in the high-volatility state: Besides β, the principal now has two additional controls: ϕ and a. Compared with equation (9), the HJB equation for $V_h(W)$, the additional term $[-\delta a + (\pi + a)\phi]V_l'(W)$ in equation (8) comes from the promise-keeping terms in dW (equation (3)), and the additional term $(\pi + a)[V_h(W - \phi) - V_l(W)]$ reflects the expected change in the principal's valuation due to the possibility of the volatility shock.

The optimal choices of β, ϕ, and a naturally depend on the state variable W. Therefore, they should always be understood as concise representations of $\beta(W)$, $\phi(W)$, and $a(W)$. The following corollary summarizes how they are determined:

Corollary 1. For any $W \in [0, W_l^C]$, the principal's optimal controls $\{\beta(W), \phi(W), a(W)\}$ have the following properties:

- β is chosen such that the IC condition (IC: $a = 0$ or IC: $a = A$) always binds. That is, $\beta(W) = \lambda$ if $a(W) = 0$, and $\beta(W) = b < \lambda$ if $a(W) = A$.
- ϕ is chosen based on three conditions:
  - The IC condition (IC: $a = 0$ or IC: $a = A$) is met; that is, $\phi(W) \geq \delta$ *if* $a(W) = 0$, and $\phi(W) < \delta$ *if* $a(W) = A$.
  - The agent's limited liability is not violated; that is, $\phi(W) \leq W$.
  - If neither the IC constraint nor the limited-liability constraint is binding, then $\phi(W)$ solves $V_h'(W - \phi(W)) = V_l'(W)$.

[9] Specifically, smooth pasting exists because the principal can always make a lump-sum payment of dC to the agent, moving the agent from W to $W - dC$ and changing the principal's value by $V_h(W - dC) - dC$. Super contact exists because W_C is optimally chosen and the principal has linear utility, so the marginal value of continuation utility $V_h'(W)$ must also vary smoothly. See Dumas (1991) for a general discussion of the smooth-pasting and super-contact conditions.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

- a is chosen by comparing the total payoff to high and low risk; that is, $a(W) = 0$ if

$$(11)\quad \frac{1}{2}\lambda^2\sigma_l^2 V_l''(W) \geq (\phi - \delta)AV_l'(W) + \frac{1}{2}b^2\sigma_l^2 V_l''(W) + A[V_h(W-\phi) - V_l(W)].$$

These properties are intuitive. First, as in the case of the high-volatility state, β only appears on the $V_l''(W) < 0$ term. Therefore, it is optimal to set β to its minimal value allowed by the IC condition. Second, ϕ is bounded by the IC condition and the agent's limited-liability constraint. If neither is binding, ϕ satisfies a first-order condition, which implies that the marginal value of the agent's continuation utility to the principal before the volatility shock equals that after the shock. Finally, the choice of a depends on the total payoff to high and low risk. The left-hand side of equation (11) is the cost of implementing low risk ($V_l''(W) < 0$). As stated in Proposition 2, incentivizing lower risk requires stronger incentives ($\beta = \lambda$) to maintain high productivity. The right-hand side is the cost of allowing high risk, which is composed of three parts: the $(\phi - \delta)A$ term coming from the drift of dW in equation (6); the cost of the productivity-management incentive, which is now lower than that on the left-hand side ($\beta = b < \lambda$); and the increasing probability of the volatility shock and the loss of firm value $V_h(W-\phi) - V_l(W)$.

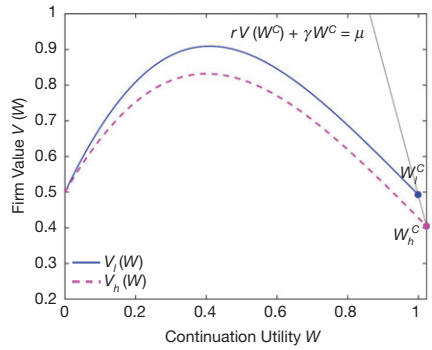
In sum, the optimal contract when the principal has full commitment power is a pair of value functions $V_l(W)$ and $V_h(W)$, both defined between a termination boundary and a payment boundary. Meanwhile, β is set to the minimal value required to maintain high productivity, and ϕ is set based on a slope-matching (optimality) condition, subject to the IC constraint and the agent's limited-liability constraint.

An example of the optimal contract is illustrated in Figure 1. Note that the value function in the high-volatility state is lower for any given W, which supports the interpretation of the high-volatility state as the "crisis" state. High volatility implies lower firm value because more volatile cash flows increase the likelihood of (unfortunate) termination of the manager, which is necessary in order to provide the manager with incentives for high effort. Another important observation is that the

FIGURE 1

Optimal Contract: Full Commitment

Figure 1 plots the principal's value functions under different cash-flow volatility σ_j and with full commitment on the principal's side. The parameter values are $\underline{V} = 0.5, \gamma = 0.04, r = 0.02, \mu = 0.5, \lambda = 0.2, \sigma_l = 0.28, \sigma_h = 0.306, \pi = 0.001$ and $\delta = 0.1$. W_l^C and W_h^C indicate the payment boundaries, and $rV\left(W^C\right) + \gamma W^C = \mu$ indicates the "optimal payment frontier."



payment boundary is higher in the high-volatility state ($W_h^C > W_l^C$). This result is examined in more depth in Section III and is subject to significant changes when limited commitment is introduced.

D. Optimal Contract with Limited Commitment

So far, the structure of the optimal contract relies on the principal's commitment to all future payments once the contract is signed. However, before the agent's continuation utility W hits the payment boundary W^C, the agent is not actually paid. His continuation utility measures the present value of the total amount of payments he expects to receive in the future, only if the principal honors the contract. In practice, the principal will likely be tempted to exercise her outside option, which is liquidating the project and receiving $\underline{V}$, if firm value V drops below her liquidation value before W reaches the payment boundary. If payment enforcement is not perfect, the optimal contract will consequently be different.

In this article, I consider a simple yet standard form of limited commitment: The principal can terminate the contract anytime. Once the contract is terminated, both parties receive the value of their outside option: $\underline{V}$ for the principal and 0 for the agent. Conditional on the continuation of the contract, the principal can still commit to all payments once the payment boundary is reached.[10] This assumption of limited commitment implies a participation constraint for the principal:

$$V_j(W) \geq \underline{V} \quad \forall j \in \{l,h\}. \tag{12}$$

Considering only the high-volatility state (and thus dropping the state subscript j) for a moment, the principal's value function must lie in the area bounded by $W \geq 0$, $V \geq \underline{V}$ and the "optimal payment frontier" $rV(W) + \gamma W = \mu$ (i.e., equation 10). Consequently, the boundary condition for the payment boundary W^C depends on whether the value function crosses the optimal payment frontier first or the principal's participation constraint $V(W) = \underline{V}$ first. If the value function meets the former first, the super-contact condition holds, $V''(W^C) = 0$ is still true, and the commitment constraint is not binding. If the value function reaches $V(W) = \underline{V}$ first, the principal cannot delay cash payment to the agent as much as she wants to without violating her participation constraint. Thus, it must be that $rV(W) + \gamma W < \mu$ for all W. This inequality, combined with the principal's HJB equation, implies $V''(W) < 0$ for all W, including at the payment boundary. That is, the super-contact condition $V''(W^C) = 0$ is replaced with a physical boundary condition $V(W^C) = \underline{V}$.[11]

[10] More precisely, the principal can commit to a *continuous* path of dC_t but not a discrete one. This is mainly a technical assumption to ensure that a reflecting payment boundary is still well defined. One might think that a limited-commitment principal should also not be able to "commit" to terminating the contract immediately when the agent's participation constraint binds. For example, she may "reset" the contract through renegotiation. I address this concern in Appendix C and show that the main results remain intact under different versions of limited commitment.

[11] Meanwhile, the smooth-pasting condition $V'(W^C) = -1$ is still true under limited commitment because the principal can still make a lump-sum payment of dC to the agent, moving the agent from W to $W - dC$ and changing the principal's value by $V(W - dC) - dC$, as long as $V(W) \geq \underline{V}$ is satisfied.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

Let $\widetilde{X}$ represent any variable X in the limited-commitment environment. The following proposition summarizes the optimal contract under limited commitment:

Proposition 4. The optimal contract under volatility regime switching with limited commitment on the principal's side defines a pair of value functions $\widetilde{V}_j(W)$ and payment boundaries $\widetilde{W}^C_j, j\in\{l,h\}$, where $\widetilde{V}_l(W)$ satisfies the same ordinary differential equation (ODE) as equation (8) and $\widetilde{V}_h(W)$ satisfies the same ODE as (9), subject to boundary conditions $\widetilde{V}_j(0)=\underline{V}$ and $\widetilde{V}'_j\left(\widetilde{W}^C_j\right)=-1$, and

$$\widetilde{V}''_j\left(\widetilde{W}^C_j\right)=0 \text{ if } \widetilde{V}_j\left(\widetilde{W}^C_j\right)\geq\underline{V}, \text{ and} \tag{13}$$

$$\widetilde{V}_j\left(\widetilde{W}^C_j\right)=\underline{V}, \text{ otherwise }. \tag{14}$$

The optimal choice of β, ϕ, and a follows the same rules described in Corollary 1.

The boundary conditions specified in this proposition imply that the optimal contract can be categorized into three types, depending on when the limited-commitment constraint binds. One: It is *not* binding in either volatility state. Two: It binds in one state but not the other, and because higher volatility implies lower firm value, the state in which it binds must be the high-volatility state. Three: It binds in both states. Clearly, type 1 is equivalent to the contract with full commitment, whereas type 3 offers the sharpest contrast. For the remainder of the article, I restrict discussion to the third type. That is, I make the following assumption:

Assumption 3. The model parameters are such that under full commitment, $V_l\left(W^C_l\right)<\underline{V}$.

Holding other parameters constant, DeMarzo and Sannikov (2006) show that both a larger σ_l and smaller $\gamma-r$ lead to a lower $V_l\left(W^C_l\right)$. Either would suffice in this model.[12]

A numerical example of the optimal contract under limited commitment is shown in Figure 2, using the same parameter values as in Figure 1. Note that in Figure 1, firm value in both states is below the liquidation value $\underline{V}$. Therefore, the payment boundaries in either state cannot be sustained without full commitment by the principal. That is, Figure 2 illustrates the type of contract in which limited commitment binds in both states. Comparing to Figure 1, both value functions are lower than their previous counterparts because of the additional binding-commitment constraint. Moreover, there is a significant change to the position of the payment boundaries: The boundary is lower in the high-volatility state, the implication of which is discussed in detail next.

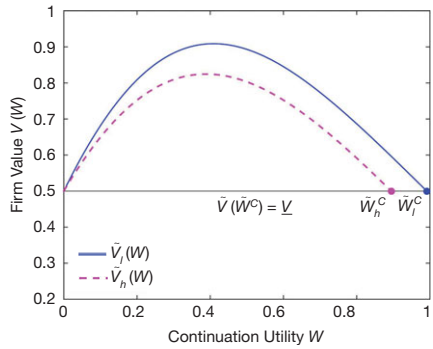
[12]Moreover, Assumption 3 is mainly for simplifying the analysis and not necessary for most of the ensuing results. A case in which equation (14) binds only in the high-volatility state (a type 2 contract) can produce very similar results.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

FIGURE 2

Optimal Contract: Limited Commitment

Figure 2 plots the principal's value functions under different cash-flow volatility σ_j and with limited commitment on the principal's side. Parameter values are the same as in Figure 1. $\widetilde{W}_l^C$ and $\widetilde{W}_h^C$ indicate the payment boundaries in the low- and high-volatility state, respectively. $\bar{V}\left(\widetilde{W}^C\right) = \underline{V}$ indicates the principal's participation constraint due to her limited commitment.



III. Contract Implications

In this section, I derive implications of the optimal contract under different levels of commitment and discuss their empirical relevance. Three implications are presented: The first pertains to the payment boundary, the second pertains to the optimal level of risk management and PPS, and the third pertains to the agent's compensation in the presence of the volatility shock.

A. Theoretical Predictions

The contracting implications for the payment boundary can be summarized as follows:

Prediction 1. When volatility becomes higher, the optimal contract under full commitment raises the payment hurdle, whereas the optimal contract under limited commitment lowers the payment hurdle.

As shown in Figure 1, under full commitment, the payment boundary is *higher* when volatility is higher. That is, $W_h^C > W_l^C$. Intuitively, because the cost of providing incentives to the agent is the possibility of early termination after sufficiently poor performance, it is higher when volatility is higher because higher uncertainty of cash flows increases the likelihood of sufficiently poor performance and the subsequent early termination. The principal adjusts the contract optimally by extending the payment boundary W_h^C further, which allows the agent to build up a larger amount of continuation utility and thus lower the likelihood of early termination.

In contrast, as shown in Figure 2, under limited commitment, the payment boundary is *lower* when volatility is higher. That is, $\widetilde{W}_h^C < \widetilde{W}_l^C$. The reason is that limited commitment implies that the contract must guarantee the principal at least

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

$\underline{V}$, which restricts the amount of future cash flow generated by continuing the project that can be credibly pledged as compensation to the agent. Higher volatility increases the likelihood of early contract termination and thus reduces the total value generated by the contract. A principal lacking the ability to commit to future payments is forced to lower the performance hurdle because she can credibly promise less compensation.

The contracting implications for optimal risk management and PPS can be summarized as follows:

Prediction 2. High risk is always implemented near the termination boundary. Away from that boundary, the optimal contract under full commitment implements low risk and thus a high PPS, whereas the optimal contract under limited commitment may implement high risk and a lower PPS.

The fact that the contract cannot prevent high risk near the termination boundary is straightforward: When W is too low, in particular if $W < \delta$, then the agent's limited liability implies $\phi \leq W < \delta$, and high risk taking cannot be prevented. This intuition is similar to that featured in previous studies (e.g., DeMarzo et al. (2013) and Wong (2019)) regarding why socially suboptimal risk taking is inevitable even under the optimal contract. Put differently, the principal is "constrained" in her risk-management policy because incentivizing low risk is simply not feasible when W is too low.

Away from the termination boundary, the principal becomes "unconstrained" in her risk-management choice: It is possible to either incentivize low risk or allow high risk. Her exact choice depends on the trade-off summarized in Corollary 1. Most importantly, there is a benefit for allowing high risk in exchange for a lower PPS (β) required. Take the risk choice at the payment boundary as an example. Under full commitment, the cost of providing incentives for maintaining high productivity, captured by the $V_l''(W)$ term in the HJB equation, is 0 at the payment boundary ($V_l''\left(W_l^C\right) = 0$, as in Proposition 3. In contrast, under limited commitment when the principal's participation constraint is binding, the contract cannot completely eliminate the cost of productivity management ($\widetilde{V}_l''\left(\widetilde{W}_l^C\right) < 0$, as in Proposition 4. Because the agent's private benefit from low productivity and from high risk are substitutes, the principal may find it optimal to forgo active risk management, allowing the agent to take high risk in exchange for the lower PPS needed to still maintain high productivity.

Figure 3 provides an illustration of this intuition. Each graph plots three value functions: $\{V_l, a = 0\}$, $\{V_l, a = A\}$, and V_h. Here, $\{V_l, a = 0\}$ should be understood as implementing $a = 0$ whenever it is "feasible."[13] The value function corresponding to implementing low risk is higher than that corresponding to high risk under full commitment. In contrast, under limited commitment, although the commitment constraint is binding for both level of risk choice, the one corresponding to high risk is less severely affected by the commitment constraint and thus is higher, thanks to the smaller β (the variation in dW) when high risk is allowed.

Prediction 2 relies on the assumption that the agent's private benefit from low productivity and his private benefit from high risk are substitutes. In other words,

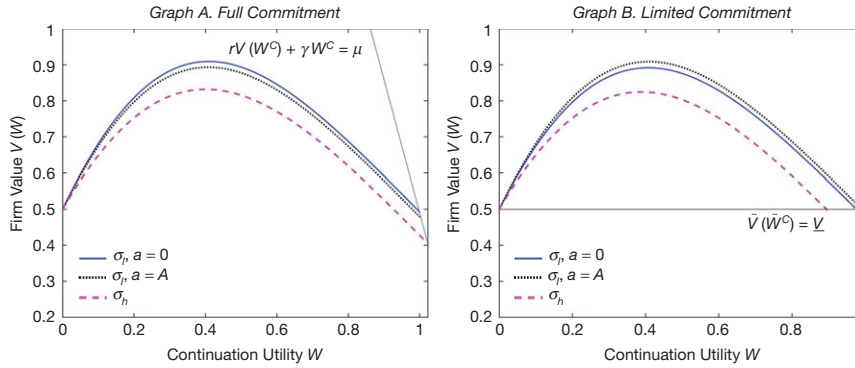
[13]That is, $W \geq \phi$, where ϕ follows the slope-matching procedure specified in Corollary 1.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

FIGURE 3

Value Functions under Different Risk Choices

Figure 3 presents the value functions under different risk choices. The solid line represents the value function of the contract that implements low risk ($a=0$) whenever it is feasible (i.e., $W \geq \phi$, where ϕ follows the slope-matching procedure specified in Corollary 1. The dotted line represents the value function of the contract that always implements high risk ($a=A$). The dashed line represents the value function in the high-volatility state. Parameter values are the same as those in Figures 1 and 2.



the cross-derivative of $B(s,a)$, as stated in Assumption 1, is negative. Alternatively, Prediction 2 also holds if the agent bears a private cost of taking the socially optimal actions ($s=0, a=0$) and the cost function is jointly convex in (s,a). This version of the model is presented in Appendix B, where I also assume the agent is risk averse with a continuous-action space. The resulting IC conditions show that the degree of PPS needed to maintain any level of productivity is still higher if the principal also wishes to implement lower risk, and Prediction 2 would still apply if the limited commitment constraint is binding and the cost of incentives for high productivity cannot be completely eliminated, even at the payment boundary.[14]

The intuition behind Prediction 2 also sheds light on which parameters are important for this prediction to hold. First, it naturally depends on the cross-derivative of $B(s,a)$. The more negative the cross-derivative is, the more likely Prediction 2 holds. Meanwhile, it also depends on the parameters that affect the cost of incentives for productivity management. Those are the parameters appearing in the $V_l''(W)$ term in the HJB equation, in particular, λ (which determines the level of PPS needed for high productivity) and σ_l (which determines the volatility of the agent's continuation utility under the incentive contract). The higher they are, the more likely Prediction 2 applies.

[14]The assumptions of a risk-neutral agent and binary actions in the main text are made primarily for simplicity of exposition. As a result of those two assumptions, $B(S,A)$ being jointly concave is necessary for a meaningful interaction of the two private actions. If $B(S,A)$ is linear in S and A, the two private actions are completely independent of each other. However, this should be regarded as a special case only because if the agent is risk averse and actions are continuous, assuming that $B(S,A)=\lambda S+\delta A$ is equivalent to assuming that the utility function is additive (i.e., $u(c,s,a)=u(c)+\lambda s+\delta a$), which is arguably very restrictive. In short, one can assume either linear utility with a jointly concave private benefit function or concave utility with linear private benefits. Instead, having linearity everywhere would eliminate any interesting interaction between the two private actions.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

Prediction 1 and 2 together imply the next result:

Prediction 3. With full commitment, entering the high-volatility state causes the agent's continuation value to be adjusted away from the payment boundary, and the agent will only be paid in the future after a long series of strong cash flows. With limited commitment, entering the high-volatility state causes the agent's continuation value to be adjusted toward the payment boundary; the agent will be paid after only a short spell of good performance.

This result is a combination of the position of the payment boundary and the level of the agent's continuation utility after the volatility shock hits, which is captured by $W-\phi$. Under full commitment, the payment boundary is higher in the high-volatility state (Prediction 1). Moreover, because the contract implements low risk near the payment boundary (Prediction 2), the punishment ϕ when the volatility shock occurs is large. Together, these imply that the agent's continuation utility immediately after the volatility shock ($W-\phi$) is far away from the payment boundary. As a result, the agent must produce strong cumulative performance that raises W sufficiently high in order to clear the hurdle for cash payments. In comparison, under limited commitment, the payment boundary is lower in the high-volatility state. Because the contract implements high risk near the payment boundary, the punishment ϕ cannot be too large. Consequently, $W-\phi$ is close to the payment boundary, and a small success that raises W marginally is enough for the agent to receive monetary payments.[15]

Figure 4 illustrates this intuition by demonstrating how ϕ varies with the agent's cumulative performance. In both graphs, $W \leq \hat{W}$ represents the region in which the contract is terminated once the volatility shock arrives. Conditional on survival (i.e., $W > \hat{W}$), each dotted segment links the agent's continuation utility before the volatility shock (the circles on $V_l(W)$, the solid line) with that after the shock (the circles on $V_h(W)$, the dashed line). The projection of those segments on the W-axis measures the size of ϕ. Under full commitment, the optimal contract implements low risk; thus, ϕ must be large enough, and the agent's continuation utility after the volatility shock is substantially lower than the payment boundary W_h^C.[16] Meanwhile, the optimal contract implements high risk under limited commitment; agents are more likely to receive cash payments with a small amount of good performance after the shock because ϕ is small and $\widetilde{W}_h^C$ is low.[17]

B. Empirical Relevance

The theoretical predictions of the model can help reconcile empirical observations regarding risk management and compensation in practice. Because high risk in this model is undesirable to the principal but generates private benefit to the

[15]Whereas high risk is always implemented for low W, regardless of the commitment level, contracts with low W are more likely to be terminated if the volatility shock occurs.

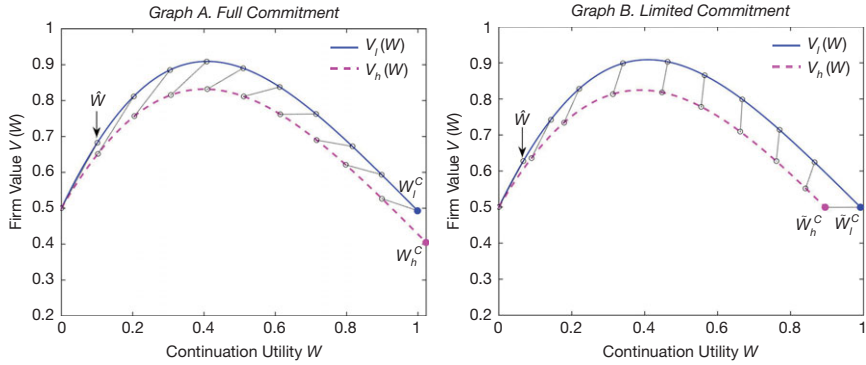
[16]For example, if the volatility shock occurs at $W = W_l^C$, then $\phi \geq \delta$ implies that the agent's continuation utility after the volatility shock is $W_l^C - \delta$, and its distance to the high-volatility-state payment boundary is $\left(W_h^C - W_l^C\right) + \delta$.

[17]For example, if the volatility shock occurs at $W = \widetilde{W}_l^C$, as long as $\widetilde{W}_h^C - \widetilde{W}_h^C < \delta$, his continuation utility after the volatility shock is very close to $\widetilde{W}_h^C$ because $\widetilde{V}_l'\left(\widetilde{W}_l^C\right) = \widetilde{V}_h'\left(\widetilde{W}_h^C\right) = -1$.

FIGURE 4

Changes in W upon the Arrival of the Volatility Shock

Figure 4 displays ϕ, the changes in the agent's continuation utility at the time of the volatility shock, determined according to Corollary 1. Each dotted segment links the agent's continuation utility before the volatility shock (the circles on $V_l(W)$, the solid line) with that after the shock (the circles on $V_h(W)$, the dashed line). $\hat{W}$ represents the cutoff point where contracts with $W < \hat{W}$ are terminated when the volatility shock arrives. Parameter values are the same as those in Figures 1 and 2.



agent, one can interpret the agent's tendency to pick $a = A$ as a *risk shifting* behavior. Although both canonical agency theory (e.g., Jensen and Meckling (1976)) and modern dynamic agency models (e.g., DeMarzo et al. (2013), Wong (2019)) predict risk shifting among distressed firms, the empirical evidence on when risk shifting takes place has been largely mixed. On the one hand, Eisdorfer (2008) and Landier, Sraer, and Thesmar (2015) show that firms increase risk taking following poor performance. On the other hand, Andrade and Kaplan (1998), Gilje (2016), and Aretz, Banerjee, and Pryshchepa (2019) find no evidence that risk shifting occurs only among distressed firms. The articles by Rauh (2008) and Huang, Sialm, and Zhang (2011) paint a similarly mixed picture in the asset-management industry.

Prediction 2 offers an explanation for this discrepancy by revealing a relationship between risk shifting and firm performance different from existing models. Among firms with strong commitment power, the model generates a time-series prediction whereby firms engage in risk shifting only after a series of negative shocks. However, among firms with different levels of commitment, the model makes a cross-sectional prediction whereby even firms with strong histories of performance may optimally choose to engage in risk shifting if they lack commitment power. Given that most empirical studies utilize a sample of firms within certain industries or over certain time periods,[18] the mixed empirical evidence regarding the relation between risk shifting and a firm's performance history may be attributed to the average commitment power of the specific industry or time period studied.

In terms of managerial compensation and turnover, Predictions 1 and 3 offer results consistent with empirical findings (e.g., Kaplan and Minton (2012), Eisfeldt

[18]For example, Landier et al. (2015) focus on the financial industry, whereas Gilje (2016) focuses on energy firms.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

and Kuhnen (2013), and Jenter and Kanaan (2015)) and anecdotal evidence regarding the compensation practice during market downturns. In particular, observations from the Great Recession suggest that the high level of cash payments can be attributed to firms setting lower performance hurdles. Importantly, the model distinguish PPS (measured by β_t) from the actual payments (measured by dC_t).[19] In addition to allowing risk shifting in exchange for lower PPS, firms without strong commitment power are also more likely to make larger immediate payments to their managers conditional on negative uncertainty shocks. These results provide theoretical support for empirical studies based on observed compensation practices, for example, Peters and Wagner (2014), who find higher industry-level equity volatility associated with *higher* managerial compensation, and Edmans et al. (2017), who find a similar relationship based on firm-level volatility.

It should be emphasized that although the majority of existing studies focus on executive compensation, the increasing adoption of incentive contracts for rank-and-file employees suggests that the predictions of this article may be applicable to a broader range of nonexecutive employees as well. Moreover, existing studies typically do not explicitly control for firms' ability to commit to long-term labor contracts. Empirical work that examines executive compensation or general labor costs while taking into account firms' commitment power should help better test the theoretical predictions of this article.[20]

IV. Discussion and Extensions

This section provides additional discussion regarding three crucial elements of the model: first, the advantage of modeling uncertainty shocks through endogenous regime switching; second, the interpretation of the high-volatility state as the "crisis" state; and third, the possibility of state-varying outside options.

A. Regime Switching Versus Comparative Statics

This subsection demonstrates the importance of modeling uncertainty shocks as the endogenous regime switching of cash-flow volatility. To do so, I compare the predictions introduced in Section III with those made from two simpler benchmarks: first, removing regime switching entirely and replacing it with comparative statics of volatility from standard 1-dimensional moral-hazard models (e.g., DeMarzo and Sannikov (2006) or Zhu (2013)); second, maintaining regime switching of volatility but removing risk management (i.e., assuming the arrival intensity of the volatility shock is exogenous). I refer to the first benchmark as *comparative statics* and the second as *exogenous regime switching.*

[19]Thus, the model's main predictions are complements to findings of studies on risk and PPS (e.g., Edmans, Gabaix, and Landier (2008), Cheng, Hong, and Scheinkman (2015)).

[20]One measure that potential empirical studies could plausibly use as a proxy for commitment is legal protection at the regional or country level. In the state of New York, judges on average favor employers during labor disputes involving cash payments. The burden is on the employees to provide evidence that they are entitled to their bonuses and other incentive pay. In contrast, California is more lenient to employees in similar disputes. More broadly, European countries generally have stronger protection for employees relative to the United States.

Of the three predictions introduced in Section III, Prediction 1, a statement regarding the positions of the payment boundary, can be obtained under the comparative-statics benchmark. Indeed, the shape of the value functions plotted in Figures 1 and 2 can be replicated using DeMarzo and Sannikov (2006) or Zhu (2013) with two different levels of volatility. The implication of limited commitment remains the same: Under full commitment, an increase in cash-flow volatility σ in DeMarzo and Sannikov (2006) or Zhu (2013) would result in an increase of the payment boundary W^C. Under limited commitment, an increase in cash-flow volatility decreases the payment boundary for the same reason as given in the analysis following Prediction 1.

Meanwhile, Prediction 2, which is a statement involving the agent's risk choice, is naturally missing in both the comparative-statics and the exogenous regime-switching benchmarks. Besides the direct implication for risk management, the contracting implication for the PPS of managerial contracts, which stems from the interaction between the incentives for productivity management and for risk management, also no longer exists without agency frictions in risk management.

Interestingly, Prediction 3, which is a statement on the dynamics of compensation after the volatility shock, is actually a joint result of both limited commitment and private risk management. This is because Prediction 3 hinges critically on the distance between the agent's continuation utility and the payment boundary after the volatility shock. In the comparative-statics benchmark, such a distance does not exist because there is no volatility shock and, therefore, no discontinuous jump in the agent's continuation utility (i.e., ϕ is not defined). In the exogenous regime-switching benchmark, when the volatility shock arrives, although the payment boundary W_h^C moves up under full commitment, the agent's continuation utility also jumps *upward* (i.e., $\phi(W) > 0$ for sufficiently high W). Intuitively, this is because higher volatility makes incentives for high productivity more costly by increasing the likelihood of contract termination. Because the volatility shock contains no information about the agent's private action, the optimal contract raises the agent's continuation utility to reduce the likelihood of contract termination. Consequently, it is possible that agents entering the high-volatility state can still receive cash payments just after a short series of high cash flows, despite the fact that the payment boundary has also been raised. In contrast, when risk management is also part of the agency frictions, Prediction 2 shows that the optimal contract will implement low risk under full commitment for sufficiently high continuation utility. The implementation of low risk requires that continuation utility jumps *downward* when the volatility shock arrives as a form of punishment. As a result, the combination of the upward movement of the payment boundary and the downward jump of the continuation utility implies that the agent receives cash payments only after a long positive series of outcomes under full commitment contracts.

In short, although a simpler model may be able to predict similar movements of the payment boundaries under different levels of volatility, the main predictions of this article for risk management, contract PPS, and the dynamics of compensation following volatility shocks rely critically on modeling volatility shocks through regime switching with endogenous intensity.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

B. Real Impact of Volatility Shocks

Throughout the article, the high-volatility state is interpreted as the "crisis" period. Although it is true that in this model, firm value is lower in this state, it is also reasonable to imagine a crisis being the state in which average productivity is low: that is, a lower *drift* (μ) of the cash-flow process. In this subsection, I compare the differences of a drift-shock model with the current volatility-shock model. I argue that modeling the crisis as a volatility shock provides more generality than modeling it as a drift/productivity shock.

The first and foremost difference is that a negative drift shock may produce Prediction 1 of the article but does not necessarily imply the other two predictions. It does not imply Prediction 2 because while the agent's payment boundary W^C decreases with μ, the value function at the payment boundary $V\left(W^C\right)$ does not necessarily also decrease.[21] Put differently, the drift shock when W is sufficiently close to the payment boundary does not necessarily imply a negative event for the principal. Therefore, the principal's optimal risk-management policy for sufficiently high W is theoretically ambiguous. Similarly, because Prediction 3 relies critically on knowing the principal's risk-management policy, the ambiguity in such a policy for sufficiently high W will lead to ambiguity in the prediction regarding the dynamics of compensation after the drift shock arrives. In sum, modeling the crisis as a volatility shock allows a richer set of results and clearer predictions than does modeling it as a drift shock.

Moreover, a volatility shock in this model can actually induce an endogenous decrease in cash-flow drift. To see this, it is necessary to first examine the condition under which high productivity is preferred by the principal. Consider the high-volatility state under full commitment. A contract that involves no payment but simply allows the agent to shirk forever defines a pair of payoffs (W_S, V_S) for the agent and the principal, respectively, where

$$W_S = \frac{\lambda S}{\gamma}; \tag{15}$$

$$V_S = \frac{1}{r}\left(\mu - \frac{\lambda S}{\gamma}\right). \tag{16}$$

The following proposition characterizes the condition for always maintaining high productivity under the optimal contract:

Proposition 5. Implementing $s_t = 0$ is always optimal if and only if

$$V_S \leq \min_{W} \; V_j(W) + \frac{\gamma}{r}(W_S - W)V_j'(W)\,, \tag{17}$$

where $j \in \{l,h\}$, and W_S and V_S are given by equations (15) and (16), respectively.

[21] Recall that the payment boundary under full commitment is set at $rV\left(W^C\right) + \gamma W^C = \mu$. When μ is lower, W^C is also lower, but the change in $V\left(W^C\right)$ is unclear.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

Proposition 5, simply put, says that whether the optimal contract allows shirking or not depends on the level of V_S. High productivity is preferred by the principal only if S is sufficiently low (thus Assumption 2). If equation (17) does not hold, the principal may be better off stopping incentive provision *sometimes*, and the agent will choose to shirk, receiving no payment while being compensated by his private benefit from shirking. The optimal contract is static, unrelated to the agent's performance, and therefore involves no termination.

In particular, it is possible that equation (17) is satisfied in one state only. The optimal contract may prevent shirking only in the low-volatility state and switch to the static contract in the high-volatility state. In that case, the dynamics of the optimal contract follow the system described in Propositions 3 and 4, with one difference: The value function V_h is replaced by the static payoff given by equation (16).[22]

The change in the agent's equilibrium effort is a feature of increasing volatility but not necessarily of decreasing profitability. Although lower average cash flow μ does bring down firm value under an incentive-compatible contract for any given W, it does the same to V_S. As a result, $s = 0$ can still be the optimal effort to induce. In contrast, V_S does *not* depend on σ. Therefore, when volatility becomes higher, firm value becomes lower until eventually falling below V_S, at which point the incentive-compatible contract becomes dominated by the static contract, a unique feature of stochastic volatility.

C. Outside Options

I have so far assumed that the principal and the agent receive constant outside value whenever the contract is terminated. Although this simplifies the analysis, it is not necessary to establish the findings of this article. In this subsection, I discuss one straightforward variation of making the outside options state contingent and show that the main results still hold.

Suppose firms can replace the incumbent manager with a new manager anytime by paying a search cost C^R (as in, e.g., Krishna, Lopomo, and Taylor (2013)). The contract with the new manager will specify an initial W, which maximizes firm value in that state. That is, $\underline{V}_j = V_j(W^*) - C^R$, where $W^* = \arg\max V_j(W)$. Figure 5 illustrates an example of the value functions and payment boundaries with state-contingent outside (liquidation) value for the principal.

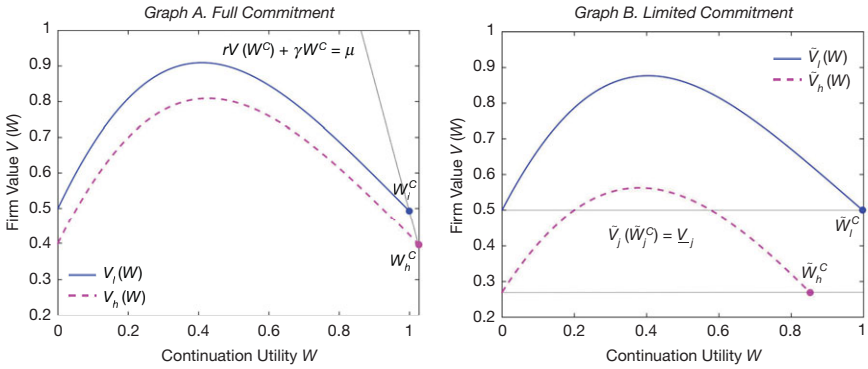
Under both full commitment and limited commitment, the liquidation value is lower in the high-volatility state. However, the positions of the payment boundaries still resemble the baseline case shown in Figures 1 and 2. That is, the payment boundary is higher in the high-volatility state under full commitment and lower in the high-volatility state under limited commitment. Intuitively, higher volatility implies a lower value of the principal's outside option, which reduces the tightness of the commitment constraint. Therefore, whether the payment boundary in the high-volatility state is higher or lower simply depends on whether the decrease

[22]Consequently, $\phi(W)$ is also different. Once the volatility shock occurs, the agent's continuation utility may jump to W_S instead of the state-contingent jumps described in Corollary 1.

FIGURE 5

Endogenous State-Contingent Liquidation Value

Figure 5 plots the value functions with endogenous state-contingent $\underline{V}_j, j \in \{l, h\}$. The full-commitment case is shown in Graph A. The limited-commitment case is shown in Graph B. Parameter values are the same as in Figures 1 and 2.



in $\underline{V}_j$ or the increase in σ_j dominates, which is an empirical question outside the scope of this study.[23] In short, large variations in volatility relative to small changes in the outside option would likely produce results consistent with those of the present model.

D. Alternative Commitment Mechanism

Economic research has long recognized that firms do not possess full commitment power over labor contracts.[24] Recent research (e.g., Palacios and Stomper (2014), Ellul et al. (2017)) shows that firms can also default on labor contracts because of limited or costly access to financial markets. Furthermore, most of these studies also assume that firms, in addition to not being able to commit to intertemporal compensation, also cannot commit to terminal payments, such as severance pay.[25] In practice, severance pay, especially for incentive purposes, is rarely paid in the event of firm liquidation, and its execution is also subject to changes in the external environment.[26]

[23]If contract termination is interpreted as replacing the incumbent manager, it is also unclear whether the value of the principal's outside options decreases unambiguously during high-volatility, crisis times. Because more managers are laid off during those times, the search cost C^R could be either lower, due to a larger managerial talent pool, or higher, due to greater information asymmetry regarding the true ability of the potential replacement manager.

[24]Many studies of risk sharing in labor contracts emphasize the lack of commitment on the firm's side (e.g., Thomas and Worrall (1988), Abreu, Pearce, and Stacchetti (1990), Ray (2002), Grochulski and Zhang (2011), and Miao and Zhang (2014)). More generally, the relational contract literature (e.g., Atkeson (1991), Levin (2003), and Opp and Zhu (2015)) studies the lack of commitment for both contracting parties.

[25]For instance, Berk, Stanton, and Zechnar (2010), Ai and Li (2015), and Bolton, Wang and Yang (2019). Even when severance pay is included in managerial contracts, as long as it is not fully guaranteed, limited commitment obtains, and the argument of this article still applies.

[26]For example, during the bankruptcies of Hostess Brands Inc. and Hawker Beechcraft Corp., the U.S. Justice Department blocked proposals to grant extra bonuses to the executives of those companies.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

Although mechanisms that provide managers with payments at the time of contract termination exist, it is unlikely that their implementation can avoid the limited-commitment problem. Imagine that a firm sets up an escrow account with balance W and allows the agent to withdraw the balance in the event of contract termination. This mechanism for implementing the terminal payment, however, must necessarily give the principal access to the account in order to update W according to the agent's performance while at the same time forbid the agent access to the account to prevent balance manipulation. This is essentially giving firms the control right to the account and asking them to "commit" to paying the agent out of the balance of that account, which is effectively no different from giving firms the control right to cash flows, as this article assumes. In other words, the same limit on commitment remains because it is intrinsically difficult to prevent the principal from reducing the balance of any account to which she has the exclusive control right before the agent can withdraw from it.

V. Conclusion

It has been argued that the buildup and sudden release of risk within the financial industry is the preeminent feature of financial crises. Moreover, the bankers and executives seemingly responsible for the aggressive risk taking still received substantial compensation in the wake of the recent crisis. To make sense of these observations, this article derives the optimal contract in a principal–agent model where the agent has private control over both firm productivity and risk (the probability of a volatility shock). Uncertainty and agency frictions require the provision of incentives in the form of back-loaded compensation, which increases with good performance in the absence of the volatility shock. The agent receives payments only if his cumulative performance exceeds a performance hurdle. Thus, specific features of the compensation contract depend crucially on the degree to which firms can credibly pledge future payments: Firms that lack sufficient commitment power may allow managers to take on large risks in exchange for the lower PPS needed to maintain high productivity. When uncertainty is high, the value of continuing operations is low, and commitment-constrained firms must provide managers with more immediate compensation. They do so by lowering the performance hurdle, which takes on the appearance of rewarding managers for excessive risk taking by means of less back-loaded compensation.

There are several directions in which this model might be fruitfully extended. Consider, for example, liquidity management and financing policies in response to volatility shocks. Recent empirical articles (e.g., Campello, Giambona, Graham, and Harvey (2011)) compare liquidity management and financing policies under different market conditions. A slightly modified dynamic model á la Bolton, Chen and Wang (2011) or Decamps, Gryglewicz, Morellec, and Villeneuve (2016) would be neatly equipped to provide theoretical insights into the empirical observations documented. Finally, as the global economy dives into another crisis triggered by the 2019–2020 coronavirus pandemic, it would be interesting to test the theoretical implications of this article in this new environment.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

Appendix A. Proofs

Proof of Proposition 1. Consider dW_t in the low-volatility state. Define the agent's total expected utility received under a contract conditional on his information at time t as:

$$G_t = E^{\hat{s},\hat{a}}\left[\int_0^\tau e^{-\gamma u} dC_u + \int_0^\tau e^{-\gamma u} B(s_u, a_u) du | \mathcal{F}_t\right],$$

where the process $G = \{G_t, \mathcal{F}_t; 0 \leq t < \tau\}$ is an $\mathcal{F}_t$ martingale. The expectation is taken with respect to the probability measure induced by $\{\hat{s},\hat{a}\}$. Then, by the martingale representation theorem, there exist $\mathcal{F}_t$ predictable, integrable processes β and ϕ, such that

$$G_t = G_0 + \int_0^t e^{-\gamma u} \beta_u \sigma_l dZ_u^{\hat{s},\hat{a}} - \int_0^t e^{-\gamma u} \phi \left(dN_u^{\hat{a}} - (\pi + \hat{a}_u) du\right). \tag{A-1}$$

Differentiating equation (A-1) and substituting in the definition of $W_t^{\hat{s},\hat{a}}$ yields

$$dG_t = e^{-\gamma t} dC_t + e^{-\gamma t} B(\hat{s},\hat{a}) dt - \gamma e^{-\gamma t} W_t^{\hat{s},\hat{a}} dt + e^{-\gamma t} dW_t^{\hat{s},\hat{a}};$$

therefore,

$$dW_t^{\hat{s},\hat{a}} = \gamma W_t^{\hat{s},\hat{a}} dt - dC_t - B(\hat{s},\hat{a}) dt + \beta_t \sigma_l dZ_t^{\hat{s},\hat{a}} + \phi\left(dN_t^{\hat{a}} - (\pi + \hat{a}_t) dt\right),$$

which implies the evolution of the agent's continuation utility given in equation (3), the evolution given in equation (5) for $\{\hat{s},\hat{a}\} = \{0,0\}$, and the evolution given in equation (6) for $\{\hat{s},\hat{a}\} = \{0,A\}$. Proof of dW_t in the high-volatility state follows the previous argument without the terms associated with dN_t and ϕ.

To prove the IC condition, let $\hat{G}_t$ denote the payoff to a strategy consisting of $\{\hat{s},\hat{a}\}$ until time $t < \tau$ and $\{s,a\}$ thereafter; then

$$\hat{G}_t = \int_0^t e^{-\gamma u} dC_u + \int_0^t e^{-\gamma u} B(\hat{s},\hat{a})_u du + e^{-\gamma t} W_t^{s,a}.$$

Differentiating $\hat{G}_t$ yields

$$\begin{aligned} e^{\gamma t} d\hat{G}_t = {} & B(\hat{s},\hat{a}) dt - B(s,a) dt + \beta_t(\hat{s}_t - s_t) dt + \beta_t \sigma_l dZ_t^{\hat{s},\hat{a}} \\ & - \phi_t(\hat{a}_t - a_t) dt + \phi_t\left(dN_t^{\hat{a}} - (\pi + \hat{a}_t - a_t) dt\right). \end{aligned}$$

If equation (4) does not hold on a set of positive measure, then the agent could choose $\{\hat{s},\hat{a}\}$, such that

$$B(\hat{s},\hat{a}) + \beta_t \hat{s}_t - \phi_t \hat{a}_t > B(s,a) + \beta_t s_t - \phi_t a_t.$$

Therefore, implementing $\{\hat{s},\hat{a}\} = \{s,a\}$ requires equation (4). □

Proof of Proposition 2. Given equation (4), implementing $\{s,a\}=\{0,0\}$ requires that

$$\lambda S-\beta S\le 0, \tag{A-2}$$

$$\delta A-\phi A\le 0, \tag{A-3}$$

$$B(S,A)-\beta S-\phi A\le 0. \tag{A-4}$$

The first two lines imply $\beta\ge\lambda$ and $\phi\ge\delta$. Together, they imply the third line because $B(S,A)<\lambda S+\delta A\le\beta S+\phi A$.

Similarly, implementing $\{s,a\}=\{0,A\}$ requires that

$$B(S,A)-\delta A-\beta S\le 0, \tag{A-5}$$

$$\delta A-\phi A>0, \tag{A-6}$$

$$\lambda S+\phi A-\beta S-\delta A\le 0. \tag{A-7}$$

The first line implies $\beta\ge(B(S,A)-\delta A)/S$. The second line implies $\phi<\delta$. The third line implies $\beta\ge(\lambda S-(\delta-\phi)A)/S$. Because $B(S,A)<\delta A+\lambda S$, these conditions together imply

$$\beta\ge b\equiv\max\left\{\frac{\lambda S-(\delta-\phi)A}{S},\frac{B(S,A)-\delta A}{S}\right\}. \qquad \square$$

Proof of Propositions 3 and 4. Full commitment. In Proposition 3, the value function $V_h(W)$ is identical to the value function in Proposition 1 in DeMarzo and Sannikov (2006), with a mere change of notations. The proof is thus omitted, and the existence and concavity of $V_h(W)$ are taken as given in the following argument. This reduces the complexity of the analysis of $V_l(W)$ in Propositions 3 and 4 because the term $V_h(W-\phi)$ in the HJB equation (8) can be regarded as a known function of W, therefore reducing equation (8) to a standard second-order ODE for $V_l(W)$. Consequently, the proof associated with the value function $V_l(W)$ largely follows the procedure in Hoffmann and Pfeil (2010), Piskorski and Tchistyi (2010), and Li (2017), with one technically minor difference in the argument for concavity, which is shown as follows: Define the *social surplus* $F_l(W)\equiv V_l(W)+W$. $F_l(W)$ and $F_l'(W)$ satisfy the following:

$$\begin{aligned}(r+(\pi+a))F_l(W)=\mu+(r-\gamma)+(\gamma W-(\pi+a)\phi(W))F_l''(W)\\ +\frac{\beta^2\sigma_l^2}{2}F_l'''(W)+(\pi+a)[(W-\phi)F_h(W-\phi)-V_l(W)],\end{aligned} \tag{A-8}$$

$$(r-\gamma)F_l'(W)=(r-\gamma)+(\gamma W-(\pi+a)\phi(W))F_l''(W)+\frac{\beta^2\sigma_l^2}{2}F_l'''(W). \tag{A-9}$$

From the boundary conditions, there exists $\varepsilon>0$, such that $F_l''(W^C-\varepsilon)<0$ and $F_l'''(W^C-\varepsilon)>0$. Assume that there exists a point $\bar{W}\equiv\sup\{W<W_l^C: F_l''(W)\ge 0\}$ (i.e., $\bar{W}$ is the largest point below which F is not concave). Then,

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

by continuity, $F_l''(\breve{W})=0$ and $F_l'''(\breve{W})<0$, which, from equation (A-9), implies that $F_l'(\breve{W})>1$. Now consider two points $W^1<\breve{W}<W^2$ close to $\breve{W}$, such that $F_l''(W^1)>0>F_l''(W^2)$ and $W^1F_l'(W^1)=W^2F_l'(W^2)$. Evaluating equation (A-8) at W^1 and W^2 yields $F'(W^1)>F'(W^2)$, in contradiction to $F_l'(\breve{W})>1$. Therefore, F_l (and V_l) must be strictly concave.

Verification. Let τ_N denote the time of the volatility shock and W_{t^-} denote the agent's continuation utility before the shock. Let $V(W_t;\sigma_t)$ denote $V_h(W_t)$ ($V_l(W_t)$) if $\sigma_t=\sigma_h$ ($\sigma_t=\sigma_l$) at t. Given equation (3) and by Ito's lemma, for any incentive-compatible contract at time $T\in[0,\tau\wedge\tau_N]$, it must be that

$$V_l(W_{0^-})=e^{-rT}V(W_T,\sigma_T)-G_1-G_2-G_3, \tag{A-10}$$

where

$$G_1=\int_0^T\Big[e^{-rT}(\gamma W_{t^-}-(\pi+a_t)\phi_t)V_l'(W_{t^-})+\frac{\beta_t^2}{\sigma_t^2}V_l''(W_{t^-}) +(\pi+a_t)[V_h(W_{t^-}-\phi_t)-V_l(W_{t^-})]\Big]\le-\int_0^T e^{-rT}\mu dt, \tag{A-11}$$

and

$$G_2=-\int_0^T e^{-rT}V_l'(W_{t^-})dC_t\le-\int_0^T e^{-rT}dC_t, \tag{A-12}$$

$$G_3=\int_0^T e^{-rT}V_l'(W_{t^-})\beta_t\sigma_t dZ_t+\int_0^T e^{-rT}[V_h(W_{t^-}-\phi_t)-V_l(W_{t^-})]d\widetilde{N}_t=0, \tag{A-13}$$

where $d\widetilde{N}_t=dN_t-(\pi+a_t)dt$ is the compensated (martingale) jump process. The inequality for G_1 follows equation (8), and the inequality for G_2 follows $V_l'(W)\ge-1$ and $dC_t\ge0$. The equality holds in both cases only under the contract characterized in Proposition 3. The equality $G_3=0$ comes from the fact that Z_t and $\widetilde{N}_t$ are both martingales. Therefore, we have

$$V_l(W_{0^-})\ge\mathrm{E}\left[\int_0^{(\tau\wedge\tau_N)}e^{-rt}(\mu dt-dC_t)+e^{-(\tau\wedge\tau_N)t}V(W_{(\tau\wedge\tau_N)};\sigma_{(\tau\wedge\tau_N)})\right], \tag{A-14}$$

which shows that $V_l(W_{0^-})$ is an upper bound for the principal's expected discounted payoff at time 0, and equation (A-14) holds with equality only when under the contract described in Proposition 3.

Limited commitment. The proof of Proposition 4 follows the same procedure but with one difference: One of the boundary conditions, the super-contact condition, is replaced with the principal's participation constraint. To see that this alternative boundary condition does not change the concavity of the value function, recall that in the high-volatility state, there is no risk-management choice, and $\beta=\lambda$. The social benefit function is $F_h(W)=V_h(W)+W$ at the boundary W_h^C satisfies:

$$F_h''(W_h^C)=\frac{rF_h(W_h^C)+(\gamma-r)W_h^C-\mu}{\frac{1}{2}\lambda^2\sigma_h^2}.$$

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

Recall that $\widetilde{X}$ denotes a variable X under limited commitment. When the principal's participation constraint is binding, $\widetilde{F}_h\left(W_h^C\right)=\underline{V}+\widetilde{W}_h^C$, which implies

$$\widetilde{F}_h''\left(\widetilde{W}_h^C\right)=\frac{r\underline{V}+\gamma\widetilde{W}_h^C-\mu}{\frac{1}{2}\lambda^2\sigma_h^2}.$$

Suppose $\widetilde{F}_h''(\widetilde{W}_h^C)>0$; that is, $r\underline{V}+\gamma\widetilde{W}_h^C>\mu$. This implies that $\widetilde{V}_h''(\widetilde{W}_h^C)>0$ because $\widetilde{V}_h(\widetilde{W}_h^C)=\underline{V}$ and $r\widetilde{V}_h(W_h^C)+\gamma\widetilde{W}_h^C>\mu$. Compare this result to the case of full commitment, where $rV_h(W_h^C)+\gamma W_h^C=\mu$. If $\widetilde{W}_h^C<W_h^C$, because $rV_h(W)+\gamma W<\mu$ for all $W<W^C$, it must be that $rV(\widetilde{W}^C)+\gamma\widetilde{W}_h^C<\mu$, which implies $V_h(\widetilde{W}_h^C)<\widetilde{V}_h(\widetilde{W}_h^C)$. However, this is a contradiction to $V_h(W)\geq\widetilde{V}_h(W)$ for every W because imposing limited commitment cannot increase the principal's payoff above that without such a constraint. If $\widetilde{W}_h^C>W_h^C$, but $\widetilde{V}_h(\widetilde{W}_h^C)=\underline{V}$ and $V_h(W_h^C)<\underline{V}$, then $\widetilde{V}_h(\widetilde{W}_h^C)>V_h(\widetilde{W}_h^C)$, which is again a contradiction. The argument in the low-volatility state and the subsequent verification are similar to those used previously and are therefore omitted. □

Finally, Propositions 3 and 4 imply the following result, which is useful for the discussion of contract implications in Section III:

Corollary A.1. Under Assumptions 1–3, $V_l(W)>V_h(W)$ and $\widetilde{V}_l(W)>\widetilde{V}_h(W)$ for all $W>0$. In particular, $W_h^C>W_l^C$ and $\widetilde{W}_h^C<\widetilde{W}_l^C$.

Proof. $W_h^C>W_l^C$ comes directly from (DeMarzo and Sannikov (2006), Appendix B). To show $\widetilde{W}_h^C<\widetilde{W}_l^C$, first notice that from the HJB equation (8), if $V_l(W)<V_h(W)$ for some $W>0$, then the principal could set $\beta(W)=\lambda\sigma_h/\sigma_l>\lambda$, which would satisfy the IC condition but increase her valuation. Thus, by the single-crossing property in Zhu (2013), $V_l(W)>V_h(W),\forall W$. The same applies to the limited-commitment case; therefore, $\widetilde{V}_l(W)>\widetilde{V}_h(W),\forall W$. Now suppose $\widetilde{W}_h^C>\widetilde{W}_l^C$; then $\widetilde{V}_h(\widetilde{W}_h^C)=\widetilde{V}_l(\widetilde{W}_l^C)=\underline{V}$, $\widetilde{V}_h'(\widetilde{W})=-1$ and the concavity of $\widetilde{V}_h(W)$ imply that $\widetilde{V}_h(\widetilde{W}_h^C)>\widetilde{V}_l(\widetilde{W}_h^C)$, a contradiction to $\widetilde{V}_l(W)>\widetilde{V}_h(W),\forall W$. Therefore, it must be that $\widetilde{W}_h^C<\widetilde{W}_l^C$. □

Proof of Proposition 5. The proof under full commitment follows the same argument as in DeMarzo and Sannikov (2006), Hoffmann and Pfeil (2010), and Zhu (2013) and is thus omitted. Under limited commitment, this condition can be further simplified to an explicitly lower bound for S using the model primitives, which is summarized in the following corollary:

Corollary A.2. Under Assumption 3 and limited commitment, $s=0$ is always optimal if

$$S>\frac{r}{\lambda(1-\gamma)}\left(\frac{\mu}{r}-\underline{V}\right). \tag{A-15}$$

Proof. Substitute $V_j(W)$ and $V_j'(W)$ from equation (17) using the fact that $V_j(W)\geq\underline{V}$ and $V_j'(W)\geq-1$ for all W under limited commitment. □

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

Appendix B. Risk-Averse Agent and Continuous-Action Space

This article assumes that the agent is risk neutral and that the agent's actions $\{a_t, s_t\}$ are binary. This is mainly for tractability, allowing the application of known techniques from existing studies (e.g., Hoffmann and Pfeil (2010), Piskorski and Tchistyi (2010), and Li (2017)) in deriving the optimal contract. However, the underlying mechanism and the intuition of this article do not hinge on those assumptions. Here I present an alternative version of the model in which the agent is risk averse with continuous-action space. Specifically, the agent has the utility function

$$u(c_t, s_t, a_t), \quad s_t \in [0, S], \ a_t \in [0, A],$$

with the usual concavity restrictions $u_c > 0, u_s > 0, u_a > 0, u_{cc} < 0, u_{ss} < 0, u_{aa} < 0$. The following proposition summarizes the agent's continuation utility and the IC conditions:

Proposition 6. The agent's continuation utility in the low-volatility state evolves according to

$$dW_t = \gamma W_t - u(c_t, s_t, a_t)dt + \beta_t(dY_t - (\mu - s_t)dt) - \phi_t(dN_t - (\pi + a_t)dt). \tag{A-16}$$

The IC condition is

$$\{s_t, a_t\} = \arg\max_{\hat{s}_t, \hat{a}_t} \ u(c_t, \hat{s}_t, \hat{a}_t) - \beta_t \hat{s}_t - \phi_t \hat{a}_t. \tag{A-17}$$

That is, s_t is implemented if and only if

$$\beta_t = u_s(c_t, s_t, a_t), \tag{IC: s}$$

and a_t is implemented if and only if

$$\phi_t = u_a(c_t, s_t, a_t). \tag{IC: a}$$

The proof is similar to that of Propositions 1 and 2 and thus is omitted. The IC conditions (IC: *s*) and (IC: *a*) imply a connection between incentives for productivity management and incentives for risk management similar to that stated in Proposition 2. To see this more transparently, assuming, for example, a constant absolute risk aversion (CARA) utility function:

$$u(c, s, a) = -\frac{1}{\eta} e^{-\eta(c + \theta^s s + \theta^a a)},$$

where θ^s and θ^a capture the marginal benefit from low productivity and excessive risk taking. It is straightforward to see that (IC: *s*) and (IC: *a*) imply

$$\beta_t = -\theta^s \eta u_t, \quad \phi_t = -\theta^a \eta u_t,$$

and therefore

$$\beta_t = \left(\frac{\theta^s}{\theta^a}\right)\phi_t. \tag{IC: \{s,a\}}$$

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

Comparing this result with that from Proposition 2, the most critical implication of Proposition 2, that the level of incentive required for high productivity depends on the level of incentive for low risk, remains intact. The lower the risk the principal wishes to implement, the higher the PPS (β) she needs to provide in order to maintain her desired level of productivity, which is exactly the takeaway of Proposition 2.

Appendix C. Renegotiation-Proof Contracts

In this article, the principal terminates the agent's contract whenever $W = 0$ because of the agent's limited liability. However, a principal with only limited commitment power may find it tempting to renegotiate the contract when W approaches 0. Although the explicit analysis of renegotiation requires a slight modification of the model, the main predictions are qualitatively intact, as explained in the following paragraphs.

If renegotiation is allowed, that is, if the principal cannot commit to any termination policy in addition to any discrete path of dC_t, then a feasible optimal contract must also be renegotiation-proof. This requires the slope of the principal's value function to be nonpositive, which is mechanically ruled out in the main context of this article because $\widetilde{V}(0) = \widetilde{V}\left(\widetilde{W}^C\right) = \underline{V}$ when the limited-commitment constraint binds. Thus, a nontrivial contract must have a region where the principal's valuation is increasing in the agent's continuation value W. To allow a more meaningful discussion of renegotiation-proof contracts and in light of existing studies (e.g., Ai and Li (2015) and Bolton et al. (2019)), I now assume that the principal will only withdraw the investment when the firm value is below a certain threshold (e.g., 0). The corresponding constraint on the payment boundary is $V\left(W^C\right) \geq 0$, and the dynamics of the agent's continuation value under renegotiation-proof contracts follow:

$$\begin{aligned} dW_t = {} & \gamma W_t dt - B(s_t, a_t) - dC_t + \beta_t (dY_t - (\mu - s_t)dt) \\ & - \phi_t (dN_t - (\pi + a_t)dt) + dX_t. \end{aligned} \tag{A-18}$$

The new term dX_t defines a reflecting termination boundary $\underline{W}$ with boundary condition $V(\underline{W}) = \underline{V}$ and $V'(\underline{W}) = 0$. Termination is stochastic at this boundary with probability $dX_t/\underline{W}$.

Figure 6 illustrates an example of the value functions for the renegotiation-proof contracts, where both the endogenous renegotiation boundaries and the payment boundaries are displayed. Critically, the relative positions of the payment boundaries for both the full- and limited-commitment contracts are still the same as those summarized in Section III. Therefore, the renegotiation-proof contracts lead to the same predictions.

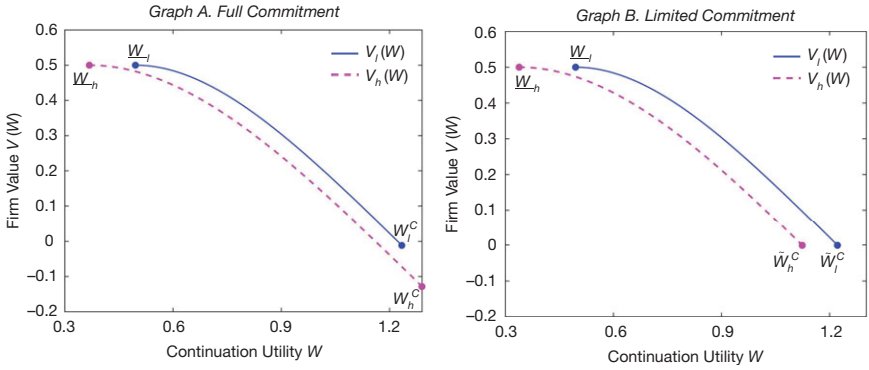
As a final remark, renegotiation-proofness is not a necessary feature for the contract to be optimal despite limited commitment. The principal is still able to rule out further renegotiation because the only action she cannot commit to is not to withdraw when firm value is negative. In particular, the principal can commit to the random termination schedule described previously, which is fundamental in maintaining proper managerial incentives (see Anderson, Bustamante, Guibaud and Zervos (2018) for discussions regarding the optimality of committing to random termination.) Further, the assumption that investors withdraw their investment when firm value drops below 0 replaces the earlier assumption of liquidation at any time. Firm value at the termination boundary is still the liquidation value, and the principal is able to credibly commit to termination once that boundary is reached.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

FIGURE 6

Renegotiation-Proof Contracts

Figure 6 plots firm value functions of renegotiation-proof optimal contracts. The parameter values are the same as those in Figures 1 and 2, except $\sigma_l = 0.36$ and $\sigma_h = 0.42$. $\underline{W}$ represents the renegotiation boundary given by $V(W) = \underline{V}$ and $V'(\underline{W}) = 0$.



References

Abreu, D.; D. Pearce; and E. Stacchetti. "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica*, 58 (1990), 1041–1063.

Ai, H., and R. Li. "Investment and CEO Compensation under Limited Commitment." *Journal of Financial Economics*, 116 (2015), 452–472.

Anderson, R. W.; M. C. Bustamante; S. Guibaud; and M. Zervos. "Agency, Firm Growth, and Managerial Turnover." *Journal of Finance*, 73 (2018), 419–464.

Andrade, G., and S. N. Kaplan. "How Costly Is Financial (Not Economic) Distress? Evidence from Highly Leveraged Transactions That Became Distressed." *Journal of Finance*, 53 (1998), 1443–1493.

Aretz, K.; S. Banerjee; and O. Pryshchepa. "In the Path of the Storm: Does Distress Risk Cause Industrial Firms to Risk-Shift?" *Review of Finance*, 23 (2019), 1115–1154.

Atkeson, A. "International Lending with Moral Hazard and Risk of Repudiation." *Econometrica*, 59 (1991), 1069–1089.

Atkeson, A. G.; A. L. Eisfeldt; and P. O. Weill. "Measuring the Financial Soundness of US Firms, 1926–2012." *Research in Economics*, 71 (2017), 613–635.

Barron, D.; G. Georgiadis; and J. Swinkels. "Optimal Contracts with a Risk-Taking Agent." *Theoretical Economics*, 15 (2020), 715–761.

Berk, J.; R. Stanton; and J. Zechnar. "Human Capital, Bankruptcy, and Capital Structure." *Journal of Finance*, 65 (2010), 891–926.

Biais, B., and C. Casamatta. "Optimal Leverage and Aggregate Investment." *Journal of Finance*, 54 (1999), 1291–1323.

Biais, B.; T. Mariotti; G. Plantin; and J.-C. Rochet. "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications." *Review of Economic Studies*, 74 (2007), 345–390.

Bloom, N. "The Impact of Uncertainty Shocks." *Econometrica*, 77 (2009), 623–685.

Bolton, P.; H. Chen; and N. Wang. "A Unified Theory of Tobin's q, Corporate Investment, Financing, and Risk Management." *Journal of Finance*, 66 (2011), 1545–1578.

Bolton, P.; N. Wang; and J. Yang. "Optimal Contracting, Corporate Finance, and Valuation with Inalienable Human Capital." *Journal of Finance*, 74 (2019), 1363–1429.

Campello, M.; E. Giambona; J. Graham; and C. Harvey. "Liquidity Management and Corporate Investment during a Financial Crisis." *Review of Financial Studies*, 24 (2011), 1944–1979.

Cheng, I.-H.; H. Hong; and J. A. Scheinkman. "Yesterday's Heroes: Compensation and Risk at Financial Firms." *Journal of Finance*, 70 (2015), 839–879.

Decamps, J.-P.; S. Gryglewicz; E. Morellec; and S. Villeneuve. "Corporate Policies with Permanent and Transitory Shocks." *Review of Financial Studies*, 30 (2016), 162–210.

DeMarzo, P.; M. Fishman; Z. He; and N. Wang. "Dynamic Agency and the Q Theory of Investment." *Journal of Finance*, 67 (2012).

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

DeMarzo, P., and Y. Sannikov. "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model." *Journal of Finance*, 61 (2006), 2681–2724.
DeMarzo, P. M.; D. Livdan; and A. Tchistyi. "Risking Other People's Money: Gambling, Limited Liability, and Optimal Incentives." Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2324879 (2013).
Di Tella, S. "Uncertainty Shocks and Balance Sheet Recessions." *Journal of Political Economy*, 125 (2017), 2038–2081.
Diamond, P. "Managerial Incentives: On the Near Linearity of Optimal Compensation." *Journal of Political Economy*, 106 (1998), 931–957.
Dumas, B. "Super Contact and Related Optimality Conditions." *Journal of Economic Dynamics and Control*, 15 (1991), 675–685.
Edmans, A., and X. Gabaix. "Executive Compensation: A Modern Primer." *Journal of Economic Literature*, 54 (2016), 1232–1287.
Edmans, A.; X. Gabaix; and D. Jenter, "Executive Compensation: A Survey of Theory and Evidence." "The Handbook of the Economics of Corporate Governance," vol. 1 Elsevier (2017), 383–539.
Edmans, A.; X. Gabaix; and A. Landier. "A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium." *Review of Financial Studies*, 22 (2008), 4881–4917.
Eisdorfer, A. "Empirical Evidence of Risk Shifting in Financially Distressed Firms." *Journal of Finance*, 63 (2008), 609–637.
Eisfeldt, A., and C. Kuhnen. "CEO Turnover in a Competitive Assignment Framework." *Journal of Financial Economics*, 109 (2013), 351–372.
Ellul, A.; M. Pagano; and F. Schivardi. "Employment and Wage Insurance within Firms: Worldwide Evidence." *Review of Financial Studies*, 31 (2017), 1298–1340.
Gilje, E. P. "Do Firms Engage in Risk-Shifting? Empirical Evidence." *Review of Financial Studies*, 29 (2016), 2925–2954.
Grochulski, B., and Y. Zhang. "Optimal Risk Sharing and Borrowing Constraints in a Continuous-Time Model with Limited Commitment." *Journal of Economic Theory*, 146 (2011), 2356–2388.
He, Z., and A. Krishnamurthy. "Intermediary Asset Pricing." *American Economic Review*, 103 (2013), 732–770.
Hellwig, M. F. "A Reconsideration of the Jensen–Meckling Model of Outside Finance." *Journal of Financial Intermediation*, 18 (2009), 495–525.
Hoffmann, F., and S. Pfeil. "Reward for Luck in a Dynamic Agency Model." *Review of Financial Studies*, 23 (2010), 3329–3345.
Huang, J.; C. Sialm; and H. Zhang. "Risk Shifting and Mutual Fund Performance." *Review of Financial Studies*, 24 (2011), 2575–2616.
Jensen, M. C., and W. H. Meckling. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics*, 3 (1976), 305–360.
Jenter, D., and F. Kanaan. "CEO Turnover and Relative Performance Evaluation." *Journal of Finance*, 70 (2015), 2155–2184.
Kaplan, S., and B. Minton. "How Has CEO Turnover Changed?" *International Review of Finance*, 12 (2012), 57–87.
Krishna, R. V.; G. Lopomo; and C. R. Taylor. "Stairway to Heaven or Highway to Hell: Liquidity, Sweat Equity, and the Uncertain Path to Ownership." *RAND Journal of Economics*, 44 (2013), 104–127.
Landier, A.; D. A. Sraer; and D. Thesmar. "The Risk-Shifting Hypothesis: Evidence from Subprime Originations." Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1786542 (2015).
Levin, J. "Relational Incentive Contracts." *American Economic Review*, 93 (2003), 835–857.
Li, R. "Dynamic Agency with Persistent Observable Shocks." *Journal of Mathematical Economics*, 71 (2017), 74–91.
Li, R., and N. Williams. "Optimal Contracts with Hidden Risk." Working Paper, University of Wisconsin (2018).
Makarov, I., and G. Plantin. "Rewarding Trading Skills without Inducing Gambling." *Journal of Finance*, 70 (2015), 925–962.
Miao, J., and Y. Zhang. "A Duality Approach to Continuous-Time Contracting Problems with Limited Commitment." *Journal of Economic Theory*, 159 (2014), 929–988.
Narita, F. "Hidden Actions, Risk-Taking, and Uncertainty Shocks." Working Paper, University of Minnesota (2011).
Opp, M. M., and J. Y. Zhu. "Impatience Versus Incentives." *Econometrica*, 83 (2015), 1601–1617.
Palacios, M., and A. Stomper. "Training the Doubtful and Timid." Working Paper, Vanderbilt University and Humboldt University (2014).
Palomino, F., and A. Prat. "Risk Taking and Optimal Contracts for Money Managers." *RAND Journal of Economics*, (2003), 113–137.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press

Peters, F., and A. Wagner. "The Executive Turnover Risk Premium." *Journal of Finance*, 69 (2014), 1529–1563.
Piskorski, T., and A. Tchistyi. "Optimal Mortgage Design." *Review of Financial Studies*, 23 (2010), 3098–3140.
Rauh, J. D. "Risk Shifting Versus Risk Management: Investment Policy in Corporate Pension Plans." *Review of Financial Studies*, 22 (2008), 2687–2733.
Ray, D. "The Time Structure of Self-Enforcing Agreements." *Econometrica*, 70 (2002), 547–582.
Ray, D., and A. Robson. "Status, Intertemporal Choice, and Risk-Taking." *Econometrica*, 80 (2012), 1505–1531.
Sannikov, Y. "A Continuous-Time Version of the Principal-Agent Problem." *Review of Economic Studies*, 75 (2008), 957–984.
Thomas, J., and T. Worrall. "Self-Enforcing Wage Contracts." *Review of Economic Studies*, 61 (1988), 541–554.
Wong, T.-Y. "Dynamic Agency and Endogenous Risk-Taking." *Management Science*, 65 (2019), 4032–4048.
Zhu, J. Y. "Optimal Contracts with Shirking." *Review of Economic Studies*, 80 (2013), 812–839.

https://doi.org/10.1017/S0022109020000642 Published online by Cambridge University Press