WORKING PAPERS

N° IDEI-834

June 2014

"Risk-sharing or risk-taking? An incentive theory of counterparty risk, clearing and margins"

Bruno Biais, Florian Heider and Marie Hoerova

INSTITUT D'ECONOMIE INDUSTRIELLE

Toulouse School of Economics

Risk-sharing or risk-taking? An incentive theory of counterparty risk, clearing and margins*

Bruno Biais† Florian Heider‡ Marie Hoerova§

June 19, 2014

Abstract

Derivatives activity, motivated by risk–sharing, can breed risk–taking. Bad news about the risk of the asset underlying the derivative increases the expected liability of a protection seller and undermines her risk–prevention incentives. This limits risk–sharing, and may create endogenous counterparty risk and contagion from news about the hedged risk to the balance sheet of protection sellers. Margin calls after bad news can improve protection sellers' incentives and enhance the ability to share risk. Central clearing can provide insurance against counterparty risk but must be designed to preserve risk–prevention incentives.

JEL classification: G21, G22, D82

Keywords: Hedging; Insurance; Derivatives; Moral hazard; Risk management; Counterparty risk; Contagion; Central clearing; Margin requirements

June 19, 2014

*We would like to thank the Editor (Cam Harvey), an Associate Editor and two anonymous referees, our discussants Ulf Axelson, Jonathan Berk, Sugato Bhattacharyya, Bruce Carlin, Simon Gervais, Artashes Karapetyan, Christine Parlour, Alp Simsek, Lauri Vilmi as well as numerous seminar and conference participants for their comments and suggestions. A previous version of the paper was circulated under the title "Risk-sharing or risk-taking? Counterparty risk, incentives and margins". The views expressed do not necessarily reflect those of the European Central Bank or the Eurosystem. Biais gratefully acknowledges the support of the European Research Council.

†Toulouse School of Economics (CNRS–CRM, Fédération des Banques Françaises Chair on the Investment Banking and Financial Markets Value Chain at IDEI), email: bruno.biais@univ-tlse1.fr

‡European Central Bank, Financial Research Division, email: florian.heider@ecb.int

§European Central Bank, Financial Research Division, email: marie.hoerova@ecb.int

1 Introduction

Derivatives activity has grown strongly over the past fifteen years. For example, credit default swaps (CDS), bilateral over-the-counter contracts used to insure credit risk, alone saw total notional amounts outstanding increase from around \$180 billion in 1998 to a peak of over \$60 trillion by mid-2008 (Acharya et al., 2012). But the insurance provided by derivatives is effective only if counterparties can honor their contractual obligations and do not default. When Lehman Brothers filed for bankruptcy protection in September 2008, it froze the positions of more than 900,000 derivative contracts (about 5% of all derivative transactions globally) in which Lehman Brothers was a party (Fleming and Sarkar, 2014). The sudden awareness of the possibility of counterparty risk in derivatives and of the associated loss of protection marked the beginning of the global financial crisis.

What are the interactions between counterparty–risk and derivatives activity? Can risk-sharing via derivatives perversely lead to risk–taking by financial institutions? How can derivatives activity be made more resilient to risk? In this paper, we explain how derivatives positions affect risk–taking incentives. We show how margin deposits and clearing arrangements can be designed to mitigate counterparty risk. We provide new empirical predictions about the extent of derivatives activity in a given financial environment and the default risk of institutions selling protection through derivatives.

Our model features risk-averse protection buyers who want to insure against a common exposure to risk (any idiosyncratic component of risk can be diversified among protection buyers themselves). To insure the common risk, they contact risk-neutral protection sellers whose assets are risky, but who are not directly exposed to the risk buyers want to insure. Because of limited liability, protection sellers can make insurance payments only if their assets are sufficiently valuable. The value of a protection seller's assets is affected by her actions. Specifically, we assume protection sellers can prevent downside risk, and hence maintain a sufficient value for their assets, by exerting costly effort. For example, when choosing their investments they can carefully scrutinize their quality. Instead of careful and costly own scrutiny, protection sellers can "shirk" and avoid the cost by relying on external, ready-made credit ratings or simple backward–looking measures of risk, as pointed out by Ellul and Yerramilli (2013). A failure of protection sellers to exert the risk-prevention effort (which we call "risk-taking") leads to counterparty risk for protection buyers. Since financial institutions' balance-sheets and activities are opaque and complex, lack of risk-prevention effort is difficult to observe and to detect for outsiders. This creates a moral hazard problem for protection sellers, the key friction in our model.

Our model builds on two important characteristics of derivatives activity. First, during the life of a derivative contract, new information about the value of the underlying asset becomes available. Such news affect the expected pay-offs of the contracting parties: it makes the derivative position an asset for one party and a liability for the other. Second, derivative exposures, and hence the associated potential liabilities, can be large. According to the Quarterly Report on Bank Derivatives Activities by the Office of the Comptroller of the Currency, total credit exposure from derivatives reached more than \$1.5 trillion in 2008.[1] The total credit exposure of the top five financial institutions was two to ten times larger than their risk-based capital.

A key insight of our analysis is that a large derivative exposure undermines a protection seller's incentives to exert the risk-prevention effort when news makes the derivative position an expected liability for her. In that case, she bears the full cost of the risk-prevention effort while the benefit of this effort partly accrues to her counterparty in the form of payments from the derivative contract. This is reminiscent of debt-overhang (Myers, 1977) but there is an important difference. In our analysis, the liability arises endogenously in the context of an optimal contract and it only materializes when negative news occur.

The optimal contract takes one of two forms, depending on the severity of the moral-hazard problem. Either the contract maintains protection sellers' risk–prevention incentives, but this comes at the cost of less ex–ante risk-sharing for protection buyers. Or it promises more risk–sharing but gives up on risk–prevention incentives, which creates counterparty risk for protection buyers. Thus, our first contribution is to show how the risk-sharing potential from derivatives contracts is limited either by just the potential or the actual presence of endogenous counterparty risk.

[1] Total credit exposure is the sum total of net current credit exposure (NCCE) and potential future exposure (PFE). NCCE is the gross positive fair value of derivatives contracts less the dollar amount of netting benefits. PFE is an estimate of what the current credit exposure could be over time, based upon a supervisory formula in the risk-based capital rules.

Our second contribution is to identify a channel through which derivatives activity can propagate risk. Without moral hazard, we assume for simplicity that the pay-offs from protection seller assets and from protection buyer assets are independent. In contrast, with moral hazard, bad news about protection buyer assets can increase the likelihood of low pay-offs from protection seller assets, because bad news undermine the protection seller's risk–prevention incentives. Moral-hazard in derivatives activity can therefore generate contagion (endogenous correlation) between two, otherwise unrelated, asset classes.

For example, prior to the recent crisis commercial banks frequently reduced their capital requirements by purchasing derivatives.[2] A bank exposed to sub-prime mortgages could purchase CDS on those mortgages and save on regulatory capital. Conditional on the drop in real estate prices (which started well in advance of the crisis), those CDS contracts became expected liabilities for those institutions that sold them, typically investment banks. Our model predicts that financial institutions with larger short CDS positions exposed their balance sheets more to downside risks as bad news about the housing market emerged. This creates correlation between the mortgage values and the values of financial institutions' assets without direct exposure to mortgage default. By contrast, those same institutions would not have increased their risk exposure after good news about the housing market. Importantly, in our model the exposure to downside risk is not the consequence of mistakes or incompetence. It is a calculated *choice* of trading-off ex–ante risk-sharing and downside risk exposure after bad news.

The third contribution of our paper is to characterize the optimal design of margin calls and central clearing, two institutional arrangements that aim to mitigate counterparty risk in derivatives activity. Both margins and central clearing received much focus in the regulatory overhaul of financial markets in the aftermath of the financial crisis. The Dodd-Frank Wall Street Reform Act in the U.S. and the European Market Infrastructure Regulation in Europe require certain derivative trades to occur via central clearing platforms (CCPs). There is, however, still considerable debate about the optimal design of CCPs for derivatives (see, e.g., Dudley, 2014, and Economist, 2014).

To examine the effects of central clearing, our model features a CCP that interposes

[2] 72% of the CDS AIG had sold by December 2007 were used by banks for capital relief (European Central Bank, 2009).

between protection buyers and sellers. The benefit of the CCP is that it mutualizes the idiosyncratic part of counterparty risk. In a bilateral contract, each protection buyer is exposed to the counterparty risk of his own protection seller. The CCP instead pools the resources from all protection sellers. Any losses from the default of individual sellers are therefore shared across all protection buyers.

The CCP is also in charge of implementing margin calls. We emphasize the incentive role of margins. The party subject to a margin call has to deposit assets with the CCP. She no longer has control over the deposited assets, which are therefore "ring-fenced" from moral hazard. Risk-prevention effort only concerns the remaining, now smaller fraction of assets over which she still has control. The cost of risk-prevention effort is therefore lower, which improves risk-prevention incentives. While ring–fencing is the benefit of margins, it comes at a cost. The loss of control goes hand-in-hand with a loss of income. Safe assets on a margin account earn a lower return than risky assets left on financial institutions' balance-sheets. Margins will therefore be used only when the ring-fencing benefit outweighs their cost, e.g., when the moral hazard problem is severe, or when the opportunity cost of depositing assets in the margin account is not too large.

Our analysis implies margins can be an attractive substitute to equity capital. Margins improve incentives by making the asset side of the balance sheet less susceptible to moral-hazard. With less moral-hazard, the assets can support larger liabilities. Consequently, margins allow protection sellers to engage in incentive-compatible derivative trading with less equity. An advantage of margins is their contingent nature. They are called only when individual derivative positions deteriorate.

Our mechanism design approach clarifies how two important reform proposals to make derivative markets more resilient, namely margins and central clearing, interact and need to be designed together. While central clearing allows mutualizing counterparty risk, margins provide incentives to avoid counterparty risk. Without margins, CCPs would bear too much risk and without a CCP, contracting parties would have to put up higher margins. And it is the CCP who must design and mandate the margin calls. Otherwise, there would be free-riding on the insurance it offers.

Our model also generates new, testable implications. First, we predict that derivatives contracts that offer ample insurance but increase exposure to downside risk (of protection sellers) are likely to be underwritten in a “benign” macroeconomic and financial environment. Second, the relation between derivatives exposures and the pledgeability of a financial institution’s assets (measured, e.g., by the efficiency of its risk-management practices) is U-shaped. Financial institutions with an intermediate level of risk-management efficiency choose small derivatives exposures while financial institutions on the other two sides of efficiency spectrum choose large exposures. Third, optimal margins are higher when i) risk-free rates are high compared to the return on productive investment opportunities, and ii) risk-management costs increase strongly with the amount of assets under management.

While the financial insurance literature typically focuses on moral hazard on the part of the buyer of protection,[3] Thompson (2010) assumes moral hazard on the part of the seller of protection. Our analysis shares this feature with that of Thompson (2010), but the two papers consider very different information–asymmetry problems. In particular, in our analysis, in contrast with Thompson (2010), moral hazard impedes the provision of insurance.

Bolton and Oehmke (2013) rely on a modeling framework similar to ours but consider different issues. They show that effective seniority for derivatives transfers credit risk to the firm’s debtholders that could be borne more efficiently by the derivative market.

Acharya and Bisin (2011) analyze the externalities arising between several protection buyers contracting with the same protection seller. They show how centralized clearing can internalize externalities among protection buyers, via optimally designed pricing schedules. This differs from our moral hazard setting where externalities are not a key issue, and quantities as well as prices must be controlled to restore incentives.[4]

Our paper explains how derivatives activity, through its effect on incentives, can generate contagion between asset classes whose risk is independent in the absence of incentive problems. This novel form of contagion channel adds to the literature on shock propagation, which emphasized interregional financial connections (Allen and Gale, 2000, Freixas,

[3]See, e.g., Parlour and Plantin (2008) in the context of credit risk transfer in banking.

[4]In the context of a model with dividend externalities among interconnected banks, Acharya et al. (2013) show how margin and capital requirements imposed by clearing-houses make banks internalize costs of their default on each other.

Parigi and Rochet, 2000), information contagion (Acharya and Yorulmazer, 2007, King and Wadhwani, 1990) and fire sales (Allen and Carletti, 2006, Cifuentes, Ferrucci and Shin, 2005).

Margins can be interpreted as a form of collateral. Collateral is usually analyzed in models in which agents borrow to finance investments (see, e.g, Bolton and Scharfstein (1990), Holmstrom and Tirole (1997), Acharya and Viswanathan (2011)). Our paper offers the first analysis of the incentives role of collateral in derivatives trading. This new context brings about new features that set margins apart from standard collateral. Standard collateral, say a house that backs up a mortgage, is transferred from the borrower to the creditor after decisions have been taken and pay-offs are realized, e.g., when the borrower defaults. By contrast, margin calls in our analysis, as in derivatives markets, occur before contracts mature, i.e., final pay-offs are realized, and, importantly, before effort and risk–taking decisions are made.

Our modelling of moral hazard, where the agent chooses between effort and shirking is in line with Holmstrom and Tirole (1997, 1998), and we borrow from their analysis the terminology "pledgeable income," to refer to the future output that can be promised by the agent without jeopardizing her incentives. In our setting, however, incentives can be undermined by the arrival of information about the risk underlying the derivative contract before effort decisions are made, and this problem can be mitigated with margin calls. These key features of our model are absent from the standard moral–hazard model studied in Holmstrom and Tirole (1997, 1998).

The remainder of the paper is organized as follows. The model is presented in Section 2, which also analyzes the benchmark case in which there is no moral–hazard problem. Section 3 analyzes optimal contracting under moral hazard. Section 4 presents extensions and discusses robustness. Section 5 contains empirical implications and Section 6 policy implications of our analysis. Section 7 concludes. Proofs are in the Appendix.

2 Model and First–Best Benchmark

2.1 The model

There are three dates, $t = 0, 1, 2$, a mass–one continuum of protection buyers, a mass–one continuum of protection sellers and a Central Clearing Platform, hereafter referred to as the CCP. At $t = 0$, the parties design and enter the contract. At $t = 1$, investment decisions are made. At $t = 2$, payoffs are received.

Players and assets. Protection buyers are identical, with twice differentiable concave utility function u, and are endowed with one unit of an asset with random return $\tilde{\theta}$ at $t = 2$.[5] For simplicity, we assume $\tilde{\theta}$ can only take on two values: $\bar{\theta}$ with probability π and $\underline{\theta}$ with probability $1 - \pi$, and we denote $\Delta\theta = \bar{\theta} - \underline{\theta}$. The risk $\tilde{\theta}$ is the same for all protection buyers.[6]

Protection buyers seek insurance against the risk $\tilde{\theta}$ from protection sellers who are risk-neutral and have limited liability. Each protection seller j has an initial amount of cash A. At time $t = 1$, this initial balance sheet can be split between two types of assets: i) low risk, low return assets such as Treasuries (with return normalized to 1), and ii) risky assets returning $\tilde{R}_j$ per unit at $t = 2$. The protection seller has unique skills (unavailable to the protection buyer or the CCP) to manage the risky assets and earn excess return. After this initial investment allocation decision, the protection seller makes a risk-management decision at $t = 1$. To model risk-management in the simplest possible way, we assume that each seller j can undertake a costly effort to make her assets safer. If she undertakes the effort, the per unit return $\tilde{R}_j$ is R with probability one. If she does not exert the effort, then the return is R with probability $p < 1$ and zero with probability $1 - p$. The risk-management process reflects the unique skills of the protection seller and is therefore difficult to observe and monitor by outside parties. Combined with limited liability, effort unobservability generates moral–hazard.

[5] The concavity of the objective function of the protection buyer can reflect institutional, financial or regulatory constraints, such as leverage constraints or risk–weighted capital requirements. For an explicit modeling of hedging motives see Froot, Scharfstein and Stein (1993). Rampini and Viswanatan (2010) examine how a firm's hedging policy interacts with its financing policy in a dynamic context.

[6] At the cost of unnecessarily complicating the analysis, we could also assume that the risk has an idiosyncratic component. This component would not be important as protection buyers could hedge this risk among themselves, without seeking insurance from protection sellers.

Exerting the effort costs C per unit of assets under management at $t = 1$.[7] Because protection seller assets are riskier without costly effort, we also refer to the decision not to exert effort as "risk-taking".[8] Undertaking risk-management effort is efficient,

$$R - C > pR, \tag{1}$$

i.e., the expected net return is larger with risk-management effort than without it. We also assume that when protection seller exerts risk management effort, return on his assets is higher than the return on the safe asset,

$$R - C > 1, \tag{2}$$

For simplicity, conditional on effort, $\tilde{R}_j$ is independent across sellers and independent of protection buyers' risk $\tilde{\theta}$. To allow protection sellers that exert effort to fully insure buyers, we assume $AR \geq \pi\Delta\theta$.

Advance information. At the beginning of $t = 1$, before investment and risk management decisions are made, a public signal $\tilde{s}$ about protection buyers' risk $\tilde{\theta}$ is observed. For example, when $\tilde{\theta}$ is the credit risk of real–estate portfolios, $\tilde{s}$ can be the real–estate price index. Denote the conditional probability of a correct signal by

$$\lambda = \text{prob}[\bar{s}|\bar{\theta}] = \text{prob}[\underline{s}|\underline{\theta}].$$

The probability π of a good outcome $\bar{\theta}$ for the protection buyer's risk is updated to $\bar{\pi}$ upon observing a good signal $\bar{s}$ and to $\underline{\pi}$ upon observing a bad signal $\underline{s}$, where, by Bayes' law,

$$\bar{\pi} = \text{prob}[\bar{\theta}|\bar{s}] = \frac{\lambda\pi}{\lambda\pi + (1-\lambda)(1-\pi)} \text{ and } \underline{\pi} = \text{prob}[\bar{\theta}|\underline{s}] = \frac{(1-\lambda)\pi}{(1-\lambda)\pi + \lambda(1-\pi)}.$$

We assume that $\lambda \geq \frac{1}{2}$. If $\lambda = \frac{1}{2}$, then $\bar{\pi} = \pi = \underline{\pi}$ and the signal is completely uninformative. If $\lambda > \frac{1}{2}$, then $\bar{\pi} > \pi > \underline{\pi}$, i.e., observing a good signal $\bar{s}$ increases the probability of a good outcome $\bar{\theta}$ whereas observing a bad signal $\underline{s}$ decreases the probability of a good outcome $\bar{\theta}$. If $\lambda = 1$ the signal is perfectly informative.

[7] We show later that our results are unchanged when we allow the unit cost C to increase (linearly) with assets under management, which makes the overall cost of risk-management effort convex.

[8] Here risk-management effort improves returns in the sense of first-order stochastic dominance. In an earlier version of the paper we show that our results are robust when effort improves returns in the sense of second–order stochastic dominance, so that lack of effort corresponds to risk–shifting.

Central counterparty, contracts and margins.

In practice, protection buyers and protections sellers contract bilaterally, and the CCP then interposes between contracting parties. Thus, the contract between the protection buyer and protection seller is transformed into two contracts, one between the seller and the CCP and another one between the buyer and the CCP (a process called novation). In our model, for simplicity, we by-pass the first step (bilateral contracting), and analyze directly the contracts between the CCP and protection buyers and sellers. This enables us to approach the problem from a mechanism design viewpoint in which the CCP designs an optimal mechanism for buyers and sellers.

Correspondingly, the CCP is modeled as a public utility designed to maximize the welfare of its members (i.e., it acts as the social planner). For simplicity, we assume the CCP maximizes expected utility of protection buyers subject to the participation constraint of the protection sellers.[9]

At $t = 0$, the CCP specifies transfers τ^S between protection sellers and the CCP at $t = 2$, and transfers τ^B between protection buyers and the CCP at $t = 2$. Positive transfers $\tau^S, \tau^B > 0$ represent payments from the CCP to sellers and buyers, while negative transfers represent payments from sellers and buyers to the CCP. The transfers τ^S and τ^B at $t = 2$ are contingent on all available information at that time. This information consists of the buyers' risk $\tilde{\theta}$, the signal $\tilde{s}$ and the set of all the protections sellers' asset returns $\tilde{R}$. Hence, we write $\tau^S(\tilde{\theta}, \tilde{s}, \tilde{R})$ and $\tau^S(\tilde{\theta}, \tilde{s}, \tilde{R})$. Since the transfers are contingent on final asset values as well as advance public information about those values (that could be conveyed, e.g., by asset prices), we can think of them as transfers specified by derivative contracts.

The transfers between the CCP and its members reflect the initial underlying bilateral contract, which is novated, and mutualization across all bilateral contracts. Hence, the transfers depend not only on a protection seller individual asset return $\tilde{R}_j$, as would be the case in a bilateral contract without the CCP, but depend on all sellers' asset returns $\tilde{R}$. This is because the latter affect the amount of resources available to the CCP to insure its members against counterparty risk.

[9] While this is only one point on the Pareto frontier, in the first-best all other Pareto optima would entail the same real decisions, i.e., the same risk–sharing and productive efficiency. In the second-best, changing the bargaining does not alter our qualitative results.

Figure 1 illustrates how the CCP sits in between protection buyers and sellers.

Insert Figure 1 here

The contract between the CCP and its members not only specifies transfers, it can also request margin deposits. Because the CCP has no ability to manage risky, opaque assets, it only accepts as margin deposits safe, transparent ones, such as cash or Treasuries that are not subject to information asymmetry problems.[10] One can therefore interpret margins as an institutional arrangement that affects the time–1 split of the seller's balance sheet between transparent assets and opaque investments. Margins "ring-fence" a fraction of the protection sellers' assets from moral-hazard. However, margins incur the opportunity cost of foregoing the excess return of the risky asset, $R - C - 1$. The margin can be contingent on all information available at time 1, i.e., the signal $\tilde{s}$. We denote the fraction of the protection seller's balance sheet deposited on the margin account by $\alpha(\tilde{s})$.

The CCP is subject to budget-balance, or feasibility, constraints at $t = 2$. For each joint realization of buyers' risk $\tilde{\theta}$, the signal $\tilde{s}$ and sellers' asset returns $\tilde{R}$, aggregate transfers to protection buyers cannot exceed aggregate transfers from protection sellers (the CCP has no resources of its own):

$$\tau^B(\theta, s, R) \leq -\tau^S(\theta, s, R), \quad \forall(\theta, s, R). \tag{3}$$

Transfers from protection sellers are constrained by limited liability,

$$-\tau^S(\theta, s, R) \leq \alpha(s)A + (1 - \alpha(s))AR, \quad \forall(\theta, s, R). \tag{4}$$

A protection seller cannot make transfers larger than what is returned by the fraction $(1 - \alpha(s))$ of assets under her management and by the fraction $\alpha(s)$ of assets she deposited on the margin account. Finally, the fraction of assets deposited must be between zero and one,

$$\alpha(s) \in [0, 1] \quad \forall s. \tag{5}$$

The sequence of events is summarized in Figure 2.

Insert Figure 2 here

[10] That assets with low information sensitivity are used as collateral is in line with Gorton and Pennacchi (1990).

2.2 First-best: observable effort

In this subsection we consider the case in which protection sellers' risk-management effort is observable, so that there is no moral hazard and the first-best is achieved. While implausible, this case offers a benchmark against which we will identify the inefficiencies arising when protection seller's risk-management effort is not observable.

Protection sellers are requested to exert risk-management effort when offering protection since doing so increases the resources available for risk-sharing (see (1)). Margins are not used since they are costly (see (2)) and offer no benefit when risk-management effort is observable. The CCP chooses transfers to buyers and sellers, $\tau^B(\tilde{\theta}, \tilde{s}, \tilde{R})$ and $\tau^S(\tilde{\theta}, \tilde{s}, \tilde{R})$, to maximize buyers' utility

$$E[u(\tilde{\theta} + \tau^B(\tilde{\theta}, \tilde{s}, \tilde{R})] \tag{6}$$

subject to the feasibility (3) and limited liability (4) constraints, as well as the constraint that protection sellers participate and join the CCP. By joining (and exerting effort), sellers obtain $E[\tau^S(\tilde{\theta}, \tilde{s}, \tilde{R})] + A(R - C)$. If they do not join and thus do not sell protection, they obtain $A(R-C)$.[11] The protection sellers' participation constraint in the first-best therefore is

$$E[\tau^S(\tilde{\theta}, \tilde{s}, \tilde{R})] \geq 0. \tag{7}$$

Proposition 1 states the first-best outcome. Since protection sellers exert risk-management effort, the return $\tilde{R}$ is always equal to R and we drop the reference to the return in the transfers τ^B and τ^S for ease of notation.

Proposition 1 *When effort is observable, the optimal contract entails effort, provides full insurance, is actuarially fair and does not react to the signal. Margins are not used. The transfers are given by*

$$\begin{aligned}
\tau^B(\bar{\theta}, \bar{s}) &= \tau^B(\bar{\theta}, \underline{s}) = E[\tilde{\theta}] - \bar{\theta} = -(1-\pi)\,\Delta\theta < 0 \\
\tau^B(\underline{\theta}, \bar{s}) &= \tau^B(\underline{\theta}, \underline{s}) = E[\tilde{\theta}] - \underline{\theta} = \pi\Delta\theta > 0 \\
\tau^B(\theta, s) &= -\tau^S(\theta, s), \forall(\theta, s)
\end{aligned}$$

[11] Without derivative trading, protection sellers always exert effort since it is efficient to do so (see condition (2)).

The first-best contract fully insures protection buyers. Their marginal utility, and hence their consumption, is the same across all realizations of their risky asset θ and the signal s. The transfers are independent of the signal and ensure a consumption level equal to the expected value of the risky asset, $E[\tilde{\theta}]$. The first-best insurance contract is actuarially fair since the expected transfer from protection sellers to protection buyers is zero, $E[\tau^B(\tilde{\theta}, \tilde{s})] = -E[\tau^S(\tilde{\theta}, \tilde{s})] = 0$. We assume

$$AR > \pi\Delta\theta, \tag{8}$$

so that, in the first–best, the aggregate resources of the protection sellers are large enough to fully insure the protection buyers.

In our simple model, when effort is observable, each transfer to a protection buyer τ^B is matched by an opposite transfer from a protection seller and margins are not needed. Thus the contract can be implemented bilaterally and the CCP is not needed. Of course, this reflects our simplifying assumption that, under effort, R is obtained for sure. If protection sellers could default, even with high effort, the CCP would be useful, in the first best, to mutualize default risk. As shown in the next sections, even in the simple case where effort precludes default, with moral hazard, the CCP plays a useful role.

The first best transfers, $\tau^B(\theta, s)$ and $\tau^S(\theta, s)$, can be implemented with forward contracts. The protection buyer sells the underlying asset forward, at price $F = E[\tilde{\theta}]$. When the final value of the asset is worth $\bar{\theta}$, the protection buyer must deliver at the relatively low forward price F. But, when the final value of the asset is low $\underline{\theta}$, the forward price is relatively high. This provides insurance to the protection buyer.

While we only consider transfers at $t = 2$, and not explicitly at $t = 1$, this is without loss of generality, because any other trading arrangement can be replicated with transfers at $t = 2$ and margins. Consider for example spot trading in which at $t = 1$, before the realization of the signal, the protection seller uses some of his initial assets A to acquire the protection buyers' asset at price S. Because there is no discounting, this is equivalent for the protection buyer to a constant transfer S at time 2. This can be achieved within the mechanism we analyze, by depositing S on the margin account at $t = 1$ and letting $\tau^B(\theta, s) = S$, irrespective of the realization of θ and s. Proposition 1 shows, however, that this is dominated by forward trading. Forward trading is more efficient, because it makes it possible to keep the assets under the management of the protection seller until $t = 2$ and earn a larger return $(R - C)$ than when investing in the risk free asset.

3 Protection-seller moral-hazard

In the previous section, we examined the hypothetical case in which protection sellers' risk-management effort is observable and can therefore be requested by protection buyers. We now move on to the more realistic situation in which risk-management effort is not observable and there is moral-hazard on the side of protection sellers.

If protection buyers want protection sellers to exert risk-management effort, then it must be in sellers' own interest to do so after observing the signal s about buyers' risk $\tilde{\theta}$. The incentive compatibility constraint under which a protection seller exerts effort after observing s is:

$$\begin{aligned} E[\tau^S(\tilde{\theta}, \tilde{s}, \tilde{R}) + \alpha(\tilde{s})A + &(1 - \alpha(\tilde{s}))A(\tilde{R} - C)|e = 1, \tilde{s} = s] \\ &\geq E[\tau^S(\tilde{\theta}, \tilde{s}, \tilde{R}) + \alpha(\tilde{s})A + (1 - \alpha(\tilde{s}))A\tilde{R}|e = 0, \tilde{s} = s]. \end{aligned}$$

The left-hand side is a protection seller's expected payoff if she exerts risk-management effort. The effort costs C per unit of assets she still controls, $(1 - \alpha(s))A$. The right-hand side is her (out-of-equilibrium) expected payoff if she does not exert effort and therefore does not incur the cost C.

Without effort, her assets under management return R with probability p and zero with probability $1 - p$. In order to relax the incentive constraint, the CCP requests the largest possible transfer from a protection seller when $\tilde{R} = 0$: $-\tau^S(\tilde{\theta}, \tilde{s}, 0) = \alpha(\tilde{s})A$. This rationalizes the stylized fact that, in case of default of the protection seller, the CCP seizes her deposits and uses them to pay protection buyers.

With effort, the protection seller's assets under management are safe, with $\tilde{R} = R$. For brevity, we write $\tau^S(\tilde{\theta}, \tilde{s}, R)$ as $\tau^S(\tilde{\theta}, \tilde{s})$. The incentive constraint after observing s is then

$$\begin{aligned} E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = s] + \alpha(s)A + &(1 - \alpha(s))A(R - C) \\ &\geq p\left(E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = s] + \alpha(s)A + (1 - \alpha(s))AR\right), \end{aligned}$$

or, using the notion of "pledgeable return" $\mathcal{P}$ (see Tirole, 2006),

$$\mathcal{P} \equiv R - \frac{C}{1-p}, \tag{9}$$

the incentive compatibility constraint rewrites as

$$\alpha(s)A + (1-\alpha(s))A\mathcal{P} \geq E[-\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = s]. \tag{10}$$

The right–hand side is what protection sellers expect to pay to the CCP after seeing the signal about buyers' risk.[12] The left-hand side is the amount that protection sellers' can pay (or pledge) to the CCP without undermining their incentive to exert risk-management effort. The left-hand side is positive since the assumption that effort is efficient, condition (1), ensures positive pledgeable return, $\mathcal{P} > 0$. The right–hand side is positive when, conditional on the signal, a protection seller expects, on average, to make transfers to the CCP. If after seeing the signal she expects, on average, to receive transfers from the CCP, then the right-hand side is negative and the incentive constraint does not bind. This is an important observation to which we return later.

When the pledgeable return $\mathcal{P}$ is sufficiently high, then protection sellers' incentive problem does not matter because the first-best allocation (stated in Proposition 1) satisfies the incentive-compatibility constraint (10) after any signal. The exact condition is given in the following lemma.

Lemma 1 *When risk–management effort is not observable, the first-best can be achieved if and only if the pledgeable return on assets is high enough:*

$$A\mathcal{P} \geq (\pi - \underline{\pi})\Delta\theta = E[\tilde{\theta}] - E[\tilde{\theta}|\tilde{s} = \underline{s}]. \tag{11}$$

The threshold for the pledgeable return on assets, beyond which full risk-sharing is possible despite protection seller moral-hazard, is given by the difference between the unconditional expectation of buyers' risk $\tilde{\theta}$ and the conditional expectation of this risk after a low signal (indicating a bad outcome is more likely). The threshold increases, making it more

[12] In our simple model this promised payment reflects a single trade. With multiple trades, the relevant expected payment would reflect the net exposure of the protection seller. In addition, when several trades are conducted with several counterparties, contractual externalities may arise. In this context a potential benefit of centralized clearing is to internalize externalities (see Acharya and Bisin, 2013).

difficult to attain the first-best, when buyers' assets are riskier (larger $\Delta\theta$) and, interestingly, when there is *better* information about this risk (larger λ leading to a lower $\underline{\pi}$). Thus, Lemma 1 has the following corollary.

Corollary 1 *When the signal is uninformative, $\lambda = \frac{1}{2}$, the first-best is always reached since $(\pi - \underline{\pi})\Delta\theta = 0$.*

In what follows, we focus on the case in which protection seller moral-hazard matters and full insurance is not feasible, as (11) does not hold.

3.1 Effort after both signals

In this section, we study the contract providing the protection seller the incentives to exert risk-management effort both after positive and after negative signals. While margins were not useful without moral-hazard (as discussed in Subsection 2.2), they may be useful now. When a protection seller exerts risk-management effort after both signals, her participation constraint is

$$E[\alpha(\tilde{s})A + (1 - \alpha(\tilde{s}))A(\tilde{R} - C) + \tau^S(\tilde{\theta}, \tilde{s}, \tilde{R})|e = 1] \geq A(R - C).$$

Since, on the equilibrium path, the protection sellers exert effort, we have $\tilde{R} = R$ and again, for brevity, we write the transfer to a protection seller as $\tau^S(\tilde{\theta}, \tilde{s})$. Collecting terms, the participation constraint is

$$E[\tau^S(\tilde{\theta}, \tilde{s})] \geq E[\alpha(\tilde{s})]A(R - C - 1), \tag{12}$$

The expected transfers from the CCP to a protection seller (left-hand-side) must be high enough to compensate her for the opportunity cost of the expected use of margins (right-hand-side). Thus, if margins are used, the contract is not actuarially fair.

The CCP chooses transfers to protection buyers $\tau^B(\tilde{\theta}, \tilde{s})$ and protection sellers, $\tau^S(\tilde{\theta}, \tilde{s})$, as well as margins $\alpha(\tilde{s})$, to maximize buyers' utility (6) subject to the feasibility constraints (3), the constraint that the fraction α be in $[0, 1]$ (5), and the incentive (10), limited liability (4), and participation (12) constraints.

The next proposition collects first results on how resources are optimally transferred between protection sellers and protection buyers.

Proposition 2 *In the optimal contract with risk-management effort, the feasibility constraints (3) bind for all* (θ, s)*, the limited liability constraints (4) are slack in state* $(\bar{\theta}, s)$ *for each* s*, and the participation constraint (12) binds.*

Protection sellers earn no rents and all resources available for insurance are passed on to protection buyers. Protection sellers' limited liability is not an issue when the value of the protection buyers' asset is $\bar{\theta}$, since in that state risk-sharing implies positive transfers to protection sellers.

Using the binding feasibility constraints, we can rewrite the incentive constraint (10) as

$$\alpha(s) A + (1 - \alpha(s)) A\mathcal{P} \geq E[\tau^B(\tilde{\theta}, \tilde{s}) | \tilde{s} = s] \tag{13}$$

Incentive compatibility implies that the expected transfers to the protection buyer be no larger than the sum of the returns on i) the assets deposited on the margin account and on ii) those left under the protetction seller's management. The pledgeable return on assets under management is smaller than the physical net return, $\mathcal{P} < R - C$, because there is moral hazard when exerting effort to manage the risk of those assets. The pledgeable return on assets deposited on the margin account is equal to their physical return of one since they are "ring-fenced" from moral-hazard in risk-management. When the moral hazard is severe, $\mathcal{P} < 1$, then depositing assets on the margin account relaxes the incentive constraint and thus allows for higher transfers to protection buyers. This is the benefit of margins. But assets deposited on the margin account incur an opportunity cost $R - C - 1$ to protection sellers. This basic tradeoff leads to the following proposition:

Proposition 3 *In the optimal contract with risk-management effort, margins are not used after* s *if the incentive constraint given* s *is slack or if the moral-hazard is not severe, i.e.,* $\mathcal{P} \geq 1$.

When the incentive constraint after s is slack, then depositing assets on the margin account offers no incentive benefit and only incurs the opportunity cost. When the pledgeable return of assets under management (weakly) exceeds the pledgeable return of assets deposited on the margin account, then margins also do not offer any incentive benefit since they actually tighten the incentive constraint.

To keep the next steps of the analysis tractable, we make the following simplifying assumption:

$$AR > \bar{\pi}\Delta\theta - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} A\mathcal{P}, \tag{14}$$

The assumption guarantees, as we will show, a slack limited liability constraint for transfers from a protection seller to the CCP when there is a good signal, $\bar{s}$, but buyers' asset return is low, $\underline{\theta}$. We discuss this assumption in more detail once we have solved for the optimal transfers, $\tau^B(\tilde{\theta}, \tilde{s})$ and $\tau^S(\tilde{\theta}, \tilde{s})$. Given (14) and Proposition 2, we only need to consider the limited liability constraint in state $(\underline{\theta}, \underline{s})$.

The next proposition states that moral-hazard problem matters only after a bad signal.

Proposition 4 *In the optimal contract with risk-management effort, the incentive constraint (13) binds after a bad signal, but is slack after a good signal. Hence there is no margin call after a good signal, i.e.,* $\alpha(\bar{s}) = 0$.

After observing a bad signal about the underlying risk, a protection seller's position is a liability to her, $E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] < 0$. This undermines her incentives to exert risk-management effort. She has to bear the full cost of effort while the benefit of staying solvent accrues in part to protection buyers in the form of the (likely) transfer to the CCP. This is in line with the debt-overhang effect (Myers, 1977).

In contrast, there is no moral-hazard problem for a protection seller after observing a good signal. A good signal indicates that her position is an asset at this point of time, $E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}] > 0$. This strengthens her incentives to exert risk-management effort. In a sense, after a good signal, since the protection seller's position has become an asset for her, it increases the income she can pledge. In contrast, the loss she expects after a bad signal reduces her pledgeable income.

We are now ready to characterize the optimal contract between the CCP, protection buyers and protections seller that exert risk management effort. It is convenient to first characterize optimal transfers as a function of the margin after a bad signal, $\alpha(\underline{s})$, and later examine the optimal margin call after a bad signal. Expected transfers conditional on the signal (as a function of $\alpha(\underline{s})$) are given by the binding participation constraint (Proposition

2) and the incentive constraint after a bad signal (Proposition 4),

$$E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}] = A\left[\alpha(\underline{s}) + (1-\alpha(\underline{s}))\mathcal{P}\right] \tag{15}$$

$$E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}] = -\frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}A\left[\alpha(\underline{s})\,(R-C) + (1-\alpha(\underline{s}))\mathcal{P}\right] \tag{16}$$

The next proposition characterizes the transfers in each possible state:

Proposition 5 *The transfers to protection buyers are*

$$\tau^B(\bar{\theta},\bar{s}) = (E[\tilde{\theta}|\bar{s}]-\bar{\theta}) - \frac{prob[\underline{s}]}{prob[\bar{s}]}A\left[\alpha(\underline{s})\,(R-C) + (1-\alpha(\underline{s}))\mathcal{P}\right] < 0, \tag{17}$$

$$\tau^B(\underline{\theta},\bar{s}) = (E[\tilde{\theta}|\bar{s}]-\underline{\theta}) - \frac{prob[\underline{s}]}{prob[\bar{s}]}A\left[\alpha(\underline{s})\,(R-C) + (1-\alpha(\underline{s}))\mathcal{P}\right] > 0,$$

so that (14) implies the limited liability constraint does not bind in state $(\underline{\theta},\bar{s})$. Furthermore, if the limited liability constraint is slack in state $(\underline{\theta},\underline{s})$, the transfers to protection buyers after a bad signal are

$$\tau^B(\bar{\theta},\underline{s}) = (E[\tilde{\theta}|\underline{s}]-\bar{\theta}) + A\left[\alpha(\underline{s}) + (1-\alpha(\underline{s}))\mathcal{P}\right] < 0 \tag{18}$$

$$\tau^B(\underline{\theta},\underline{s}) = (E[\tilde{\theta}|\underline{s}]-\underline{\theta}) + A\left[\alpha(\underline{s}) + (1-\alpha(\underline{s}))\mathcal{P}\right] > 0.$$

Otherwise, the transfers after a bad signal are

$$\tau^B(\bar{\theta},\underline{s}) = \alpha(\underline{s})A - (1-\alpha(\underline{s}))A\frac{(1-\underline{\pi})R-\mathcal{P}}{\underline{\pi}} \tag{19}$$

$$\tau^B(\underline{\theta},\underline{s}) = \alpha(\underline{s})A + (1-\alpha(\underline{s}))AR > 0.$$

In the optimal contract, if the limited liability constraint is slack in state $(\underline{\theta},\underline{s})$, then there is full risk sharing given the signal. That is, for a given signal s, the consumption of the protection buyer is the same irrespective of whether $\bar{\theta}$ or $\underline{\theta}$ realizes. On the other hand, in contrast with the first best, transfers vary with the signal. This is because, after a bad signal, it is difficult to provide incentives to the agent. Thus, incentive compatibility reduces the transfers that can be requested from the protection seller. Correspondingly, due to incentive problems, the protection buyer is exposed to signal risk, as her consumption is larger after a good signal than after a bad signal. Cross–subsidization across signals mitigates that effect, but only imperfectly, due to incentive constraints. Cross–subsidization across realizations of the signal is possible because the parties commit to the contract at time 0, before advance information is observed. If the contract was written after that information had been observed, such cross–subsidization would be not be possible. This would reduce the scope for insurance, in the line with the Hirshleifer (1971) effect.

To further analyze these effects consider the structure of the transfers in Proposition (5). Each of the transfers in (17) has two components. The first one is the transfer implementing full risk–sharing conditional on a good signal. The second one reflects cross–subsidization across signals. Transfers in (18) have the same structure except that the first component now reflects full risk–sharing conditional on a bad signal.

The expectation of the first component of these transfers, taken over signals and final realizations of θ is 0. This is what would arise with actuarially fair insurance. But the insurance offered by the protection seller is not actuarially fair. It involves a premium, to compensate the protection seller for the efficiency loss induced by margins: $prob[\underline{s}]\alpha(\underline{s})(R - C - 1)$. This premium is equal to the expectation of the second component of the transfers in (17) and (18).

The structure of the transfers in (19) is different. When limited liability binds in state $(\underline{\theta}, \underline{s})$, full risk–sharing conditional on the signal is no longer possible, as protection sellers' resources in state $(\underline{\theta}, \underline{s})$ are insufficient. Conditional on a bad signal, the transfers in (19) implement whatever risk–sharing is still possible given the binding limited liability constraint.

Now, turn to the determination of the optimal margin call after a bad signal. We first note that putting all the assets of the protection seler in the margin account cannot be optimal.

Proposition 6

$$\alpha^*(\underline{s}) < 1. \tag{20}$$

The logic underlying Proposition 6 is the following. When assets are put in the margin account, they earn lower return than under the management of the protection seller exerting effort. This reduces the resources available to pay insurance to the protection buyer. To cope with this dearth of resources, when $\alpha^*(\underline{s}) = 1$ all the assets in the margin account must be transferred to the protection buyer when $\underline{\theta}$ realizes. In this case, as can be seen by inspecting (19) for $\alpha^*(\underline{s}) = 1$, the structure of transfers is highly constrained. In fact, it is so constrained that very little risk sharing can be achieved. Hence, a contract requesting $\alpha^*(\underline{s}) = 1$ is suboptimal.

To analyze the precise amount of margin the calls, it is useful to consider the ratio of the marginal utility of a protection buyer after a bad and a good signal. Denoting this ratio by φ, we have

$$\varphi = \frac{u'(\bar{\theta} + \tau^B(\bar{\theta}, \underline{s}))}{u'(\bar{\theta} + \tau^B(\bar{\theta}, \bar{s}))} \tag{21}$$

In the first-best, there is full insurance and φ is equal to 1. With moral hazard, the protection buyer is exposed to signal risk. This makes insurance imperfect and drives φ above one.

Given the transfers in Proposition 5, φ is a known function of exogenous variables and $\alpha(\underline{s})$. (17) implies that $\tau^B(\bar{\theta}, \bar{s})$ is decreasing in $\alpha(\underline{s})$. Hence the denominator of ϕ is increasing in $\alpha(\underline{s})$. On the other hand, the numerator of is decreasing in $\alpha(\underline{s})$ (irrespective of whether the limited liability condition in state $(\underline{\theta}, \bar{s})$ binds or not). Hence, φ is decreasing in $\alpha(\underline{s})$. Higher margins reduce φ, as they reduce the wedge between consumption after a good signal and after a bad one, i.e., they improve insurance against signal risk. Optimal margins tradeoff this benefit with their cost: assets in the margin account are less profitable than under the management of the protection seller exerting effort. This tradeoff gives rise to the following proposition.

Proposition 7 *If $\mathcal{P} > 1$, margins are not used. Otherwise, we have the following: If $\varphi(0) < 1 + \frac{R-C-1}{1-\mathcal{P}}$, then it is optimal not to use margins. Otherwise, there are two cases. If*

$$\varphi\left(1 - \frac{\underline{\pi}\Delta\theta}{A(R-\mathcal{P})}\right) < 1 + \frac{R-C-1}{1-\mathcal{P}}, \tag{22}$$

the limited liability constraint is slack in state $(\underline{\theta}, \underline{s})$ and the optimal margin solves

$$\varphi(\alpha^*(\underline{s})) = 1 + \frac{R-C-1}{1-\mathcal{P}}, \tag{23}$$

while, if (22) does not hold, the optimal margin solves

$$\varphi(\alpha^*(\underline{s})) = 1 + \frac{R-C-1}{1-\mathcal{P}} + \frac{1-\underline{\pi}}{1-\mathcal{P}} \frac{u'(\underline{\theta} + \tau^B(\underline{\theta}, \underline{s})) - u'(\bar{\theta} + \tau^B(\bar{\theta}, \underline{s}))}{u'(\bar{\theta} + \tau^B(\bar{\theta}, \bar{s}))}. \tag{24}$$

The right-hand side of (23) reflects the tradeoff between the costs and benefits of margins. The numerator, $R-C-1$, is the opportunity cost of depositing a margin. The denominator goes up as $\mathcal{P}$ decreases, i.e., as the incentive problem gets more severe.

When margins are as in (23), consistency requires that there be enough resources to provide full insurance conditional on the signal. This is the case if (22) holds. Consistent with intuition, this is the case if R is large enough. When there is full risk sharing conditional on the signal, the last term on the right hand–side of (24) is 0. In that case, (24) simplifies to (23). This case is illustrated in Figure 3. The figure is useful to examine graphically the effect an increase in p, reducing pledgeable income $\mathcal{P}$. The decrease in $\mathcal{P}$ shifts curve φ upwards while shifting $1+\frac{R-C-1}{1-\mathcal{P}}$ downwards. This raises the optimal margin in (23). When incentive problems become more severe, margins are needed more, to relax the incentive constraint.

Insert Figure 3 here

On the other hand, when the limited liability constraint binds in state $(\underline{\theta}, \underline{s})$, full risk–sharing conditional on the signal is not achievable, so that $u'(\underline{\theta}+\tau^B(\underline{\theta}, \underline{s})) > u'(\bar{\theta}+\tau^B(\bar{\theta}, \underline{s}))$. The last term on the right hand–side of (24) is strictly positive, and, correspondingly, margins are lower than when the limited liability condition is slack. Again, this is because (taking as given that there is effort) margins reduce the amount of resources eventually available to pay insurance. When limited liability binds, these resources are sorely needed. So it is perferable to reduce margins, in order to increase the amount of resources available. The following corollary gives a sufficient condition for (22) to hold.

Corollary 2 *A sufficient condition for the limited liability condition to be slack in state* $(\underline{\theta}, \underline{s})$ *is*

$$1-\frac{\underline{\pi}\Delta\theta}{A\left(R-\mathcal{P}\right)} > \frac{(1-\underline{\pi})R-\mathcal{P}}{\underline{\pi}+(1-\underline{\pi})R-\mathcal{P}}. \tag{25}$$

Condition (25) holds if $\underline{\pi}\Delta\theta$ is not too large. In that case, full risk–sharing after a bad signal does not request too large resources, and can thus be implemented.

In the first-best the transfers depend only on the realization of θ and the optimal contract can be implemented with a simple forward contract. In contrast, with moral-hazard and risk-management effort after both signals, the transfers depend on the realizations of θ *and* s.

The optimal contract can be implemented by the sale of a forward contract on the underlying asset θ by protection buyers (as in the first-best) together with the purchase of a forward contract on the signal s. The forward contract on s generates a gain for protection sellers in state $\underline{s}$. This gain increases their pledgeable income after a bad signal and thus restores incentive compatibility in the light of the liability from the forward contract on θ.[13]

3.2 No effort after a bad signal (risk-taking)

Incentive compatibility after a bad signal reduces risk–sharing. Protection buyers may find this reduction in insurance too costly. They may instead choose to accept shirking on risk prevention effort (risk-taking) by protection sellers in exchange for a better sharing of the risk associated with $\tilde{\theta}$. In this subsection, we characterize the optimal contract with risk-taking after a bad signal.

After a good signal, protection sellers exert risk-management effort so that $\tilde{R}_j = R$ for all j. After a bad signal, protection sellers do not not exert risk-management effort so that $\tilde{R}_j = R$ for a proportion p of sellers and $\tilde{R}_j = 0$ for a proportion $1-p$ of sellers. Hence, the transfer τ^S from the CCP to a protection seller must now be contingent on the realization of $\tilde{R}_j$. By contrast, the transfer τ^B from the CCP to a protection buyer does not have to be contingent on the realization of a particular $\tilde{R}_j$. The CCP can mutualize counterparty risk and provide insurance to risk-averse protection buyers. However, the aggregate amount of resources protection sellers generate differs after a good signal and after a bad signal. After a bad signal, only proportion p of protection sellers generate return R while proportion $1-p$ of sellers generate a zero return and cannot make any payments to the CCP as they are protected by limited liability.

The CCP chooses transfers to buyers and sellers, $\tau^B(\tilde{\theta}, \tilde{s}, \tilde{R})$ and $\tau^S(\tilde{\theta}, \tilde{s}, \tilde{R})$, to maximize buyers' utility

$$\begin{aligned} &\pi\lambda u(\bar{\theta} + \tau^B(\bar{\theta}, \bar{s}, R)) + (1-\pi)(1-\lambda)u(\underline{\theta} + \tau^B(\underline{\theta}, \bar{s}, R)) \\ &+ \pi(1-\lambda)u(\bar{\theta} + \tau^B(\bar{\theta}, \underline{s}, pR)) + (1-\pi)\lambda u(\underline{\theta} + \tau^B(\underline{\theta}, \underline{s}, pR)) \end{aligned} \tag{26}$$

[13] While this implementation is plausible, it is not unique. Other financial contracts with gains for protection sellers after $\underline{s}$ such as options can be used.

where, after a bad signal, τ^B is written as a function of pR to indicate mutualization of counterparty risk by the CCP.

The feasibility constraints of the CCP after good and a bad signal, respectively, are given by

$$\tau^B(\theta,\bar{s}) \leq -\tau^S(\theta,\bar{s},R) \quad \forall(\theta,\bar{s}) \tag{27}$$

$$\tau^B(\theta,\underline{s}) \leq -p\tau^S(\theta,\underline{s},R)-(1-p)\tau^S(\theta,\underline{s},0) \quad \forall(\theta,\underline{s}) \tag{28}$$

The limited liability constraints for sellers whose assets generate R and for those whose assets generate 0, respectively, are given by:

$$-\tau^S(\theta,s,R) \leq \alpha(s)A+(1-\alpha(s))AR \quad \text{for } R_j=R \tag{29}$$

$$-\tau^S(\theta,s,0) \leq \alpha(s)A \quad \text{for } R_j=0 \tag{30}$$

The seller's incentive constraint after a good signal is, as before,

$$\alpha(\bar{s})A+(1-\alpha(\bar{s}))A\mathcal{P} \geq -E[\tau^S(\theta,\bar{s},R)], \tag{31}$$

whereas after a bad signal, the seller must prefer not to exert effort

$$E[\tau^S(\theta,\underline{s},R)]+\alpha(\underline{s})A+(1-\alpha(\underline{s}))A(R-C) \leq$$
$$pE[\tau^S(\theta,\underline{s},R)]+(1-p)E[\tau^S(\theta,\underline{s},0)]+\alpha(\underline{s})A+(1-\alpha(\underline{s}))pAR,$$

or, equivalently,

$$(1-\alpha(\underline{s}))A\mathcal{P} \leq -E[\tau^S(\theta,\underline{s},R)]+E[\tau^S(\theta,\underline{s},0)]. \tag{32}$$

Finally, the seller's participation constraint with risk-taking is

$$\begin{aligned}&\text{prob}[\bar{s}]\alpha(\bar{s})A(R-C-1)+\text{prob}[\underline{s}]\alpha(\underline{s})A(pR-1)+\text{prob}[\underline{s}](1-p)A\mathcal{P}\\ &\leq \text{prob}[\bar{s}]E[\tau^S(\theta,\bar{s},R)]+\text{prob}[\underline{s}](pE[\tau^S(\theta,\underline{s},R)]+(1-p)E[\tau^S(\theta,\underline{s},0)])\end{aligned} \tag{33}$$

The expected transfer from the CCP to a protection seller (right-hand side) is positive. If a seller enters the position, she must be compensated for the potential efficiency loss (left-hand side). The loss is due to two factors: 1) costly margins after good and a bad signal (where $R-C-1$ is the opportunity cost of margins when a seller exerts effort and $pR-1$ is the opportunity cost of margins when she does not) and 2) the loss of pledgeable income in the event of default, which occurs with probability $\text{prob}[\underline{s}](1-p)$. Thus, the contract with no effort after a bad signal is actuarially unfair. The higher the pledgeable income, the greater the efficiency loss generated by risk-taking after a bad signal, the more actuarially unfair the contract.

We can re-write the seller's participation constraint with risk-taking as

$$\begin{aligned} &A\text{prob}[\bar{s}]\alpha(\bar{s})(R-C-1) + A\text{prob}[\underline{s}]\left[R-C-(\alpha(\underline{s})+(1-\alpha(\underline{s}))pR)\right] \\ &\leq \text{prob}[\bar{s}]E[\tau^S(\theta,\bar{s},R)] + \text{prob}[\underline{s}](pE[\tau^S(\theta,\underline{s},R)] + (1-p)E[\tau^S(\theta,\underline{s},0)]) \end{aligned} \tag{34}$$

On the left-hand side, there is again the efficiency loss from entering the contract with risk-taking. After a good signal, the seller exerts effort but there is an opportunity cost of margins, given by $R-C-1$. After a bad signal, the seller does not exert effort and the efficiency loss is given by the difference between $R-C$, the return on assets when not entering the contract and doing effort, and $\alpha(\underline{s})+(1-\alpha(\underline{s}))pR$, the expected return under the contract with risk-taking.

We first show that in the optimal contract with risk-taking, the feasibility constraints and the participation constraint must bind, i.e., protection sellers earn no rents and all resources available for insurance are passed on to protection buyers.

Proposition 8 *In the optimal contract with risk-taking after a bad signal, the feasibility constraints bind for all* (θ, s) *and the participation constraint binds.*

The next proposition characterizes the use of margins in the contract with risk-taking and narrows down the parameter space for which risk-taking after a bad signal can be optimal:

Proposition 9 *In the optimal contract with risk-taking after a bad signal, margins are not used after signal* $\bar{s}$ *if the incentive constraint given* $\bar{s}$ *is slack or if the moral-hazard is not severe, i.e.,* $\mathcal{P} \geq 1$. *After signal* $\underline{s}$, *margins are not used if* $pR \geq 1$. *If* $pR < 1$, *then* $\alpha^*(\underline{s}) = 1$. *Such contract is, however, dominated by the one with effort after a bad signal.*

Without effort after a bad signal, the expected per-unit return on the seller's balance sheet is pR. If $pR < 1$, this is lower than what assets return on the margin account. Hence, it is more profitable to deposit all of the protection seller's assets in the margin account, $\alpha = 1$, where they earn a greater return and are ring-fenced from moral hazard. But protection buyers can do at least as well by requesting effort after a bad signal since, there too, $\alpha = 1$ can be selected (but, as we know from Proposition 6, it is never optimal). It follows that the contract with margins and no effort after a bad signal can only be strictly optimal if $pR \geq 1$.

The next proposition characterizes the optimal transfers in the contract with risk-taking after a bad signal.

Proposition 10 *If $pR < 1$, then risk–taking is suboptimal. Otherwise, the optimal contract with risk-taking after a bad signal provides full insurance to protection buyers if and only if*

$$pAR \geq \pi\Delta\theta - (1-p)\, prob[\underline{s}]A\mathcal{P}. \tag{35}$$

The transfers are given by

$$\begin{aligned}
\tau^B(\bar{\theta}, \bar{s}) &= \tau^B(\bar{\theta}, \underline{s}) = -(1-\pi)\Delta\theta - prob[\underline{s}](1-p)A\mathcal{P} < 0, \\
\tau^B(\underline{\theta}, \bar{s}) &= \tau^B(\underline{\theta}, \underline{s}) = \pi\Delta\theta - prob[\underline{s}](1-p)A\mathcal{P} > 0.
\end{aligned}$$

In contrast to the contract with effort after a bad signal, the contract with risk-taking does not react to the signal, i.e., $\tau^B(\tilde{\theta}, \bar{s}) = \tau^B(\tilde{\theta}, \underline{s})$. The consumption of the buyer is equalized across states (i.e., there is full insurance, as in the first-best) as long as the amount of resources generated under risk-taking (by the protection sellers who succeed), equal to pAR, is sufficiently high. However, since protection buyers must compensate protection sellers for the efficiency loss due to risk-taking (given by the loss of pledgeable income in the event of default after a bad signal, prob$[\underline{s}](1-p)A\mathcal{P}$), the consumption of protection buyers falls short of the first-best consumption levels. Condition (35) ensures that the limited liability constraints are slack under full insurance. On the left-hand side are the aggregate resources generated by protection sellers. On the right–hand side is the transfer that would be paid in the first-best, minus the payment requested by protection–sellers to offset the efficiency loss they incur due to risk–taking.

Risk–taking can be optimal only if it is not too inefficient, i.e., if $pR \geq 1$. In that case, margins are not used. Since protection sellers engage in risk–taking after a bad signal, margins do not help with incentives. Margins are also not needed to insure buyers against counterparty risk since it is mutualized by the CCP. Thus, mutualization tackles ex–post counterparty risk in the contract with risk-taking, while margins tackle ex–ante incentives in the contract with effort.

Condition (35) can be re-written as

$$A\left[(1-\text{prob}[\underline{s}])\, pR + \text{prob}[\underline{s}]\,(R-C)\right] \geq \pi\Delta\theta.$$

Since $R - C > 1$ and $pR \geq 1$ (the latter condition is necessary for the contract with risk-taking to be optimal), it follows that

$$A\left[(1-\text{prob}[\underline{s}])\, pR + \text{prob}[\underline{s}]\,(R-C)\right] > A.$$

Hence, a sufficient condition for (35) to hold is

$$A \geq \pi\Delta\theta. \tag{36}$$

In the optimal contract with risk-taking after bad news, thanks to the mutualization of counterparty risk by the CCP, transfers are not contingent on signals or on individual protection seller's returns. Hence the optimal contract can be implemented with a single forward contract (as in the first-best) provided it is insured by the CCP (unlike in the first–best). The forward contract, however, is sold at a discount relative to the expected value of the underlying risk, in order to compensate the protection sellers for the loss of pledgeable income in default.

3.3 Risk-sharing and risk-taking

The contract under which protection sellers exert effort after both signals entails limited risk-sharing for buyers but entails no risk-taking by sellers (Subsection 3.1), while the contract with no effort after a bad signal entails full risk-sharing for protection buyers but is actuarially unfair and falls short of the first-best due to the loss of resources in default (Subsection 3.2). The next proposition characterizes the optimal choice between the two contracts as a function of the probability of success under risk-taking, p.

Proposition 11 *Assume (36) holds. There exists a threshold value of the success probability under no effort $\hat{p}$ such that risk–prevention effort after bad news is optimal if and only if $p \leq \hat{p}$.*

The logic of the proposition is illustrated in Figure 4. Consider the expected utility of the protection buyer when effort is requested after bad news. It decreases when p increases. For this contract, indeed, the only effect of an increase in p is to tighten the incentive constraint, and thus reduce risk–sharing. Now turn to the expected utility of the protection buyer when effort is not requested after bad news. In contrast with the previous case, it increases when p increases. Indeed, for this contract, the only effect of an increase in p is to increase the amount of resources available after bad news. Hence the result, stated in the proposition, that risk–prevention effort after bad news is optimal if and only if p is lower than a threshold.

Insert Figure 4 here

4 Extensions and Robustness

4.1 Renegotiation

The optimal contract inducing effort after both good and bad news is contingent on the signal s. One may wonder whether the optimal outcome could also be achieved by renegotiating - at time 1 after s is observed - an initial contract, $\tau^B(\theta)$, independent of the signal.

For brevity and simplicity, rather than offering a general treatment of this question, we discuss the underlying economic forces in the context of an example. Suppose we start from an initial contracttual transfer $\tau^B(\theta)$ independent of the signal. For example, suppose we take it to be the transfer prevailing in the optimal contingent contract after good news $\tau^B(\theta, \bar{s})$. Would both parties (protection buyer and protection seller) agree to switch from $\tau^B(\theta) = \tau^B(\theta, \bar{s})$ to $\tau^B(\theta, \underline{s})$ after observing bad news?

First consider the protection seller. Sticking to $\tau^B(\theta, \bar{s})$ after a bad signal violates her incentive compatibility constraint. Thus, she does not exert risk-management effort and fails with probability $1 - p$. Her expected gain is then:[14]

$$\underline{\pi}p\left(AR - \tau^B(\bar{\theta}, \bar{s})\right) + (1 - \underline{\pi})p\left(AR - \tau^B(\underline{\theta}, \bar{s})\right). \tag{37}$$

If instead she switches to $\tau^B(\theta, \underline{s})$, and thus exerts risk-management effort, she expects to obtain

[14] Recall that with probability $1 - p$, $\tilde{R}_j = 0$ and that $\tau^B(\theta, \tilde{s}, 0) = 0$.

$$\underline{\pi}\left(AR - \tau^B(\bar{\theta}, \underline{s})\right) + (1 - \underline{\pi})\left(AR - \tau^B(\underline{\theta}, \underline{s})\right) - AC. \tag{38}$$

By switching, the protection seller increases the expected payoff on her assets. She also reduces the payment to the protection buyers as $\tau^B(\underline{\theta}, \underline{s}) < \tau^B(\underline{\theta}, \bar{s})$. Thus, switching is quite attractive for her, as we now establish more formally. Substituting for the transfers and re-arranging, (38) is larger than (37) if and only if

$$A\mathcal{P} < E[\theta] - \text{prob}[\bar{s}]E[\tilde{\theta}|\underline{s}] \tag{39}$$

which is satisfied under our assumption that (11) does not hold.

Now turn to protection buyers. Sticking to $\tau^B(\theta, \bar{s})$ after a bad signal implies higher transfers from the CCP, but undermines the incentives of the protection seller. When the CCP insures against counterparty risk, the protection buyer does not internalize the cost of default of his counterparty. Consequently, the protection buyer does not accept to switch from the initial contract to $\tau^B(\theta, \underline{s})$ after bad news. Thus, the simple renegotiation game we proposed does not implement the optimal contract. This negative result extends to a larger class of renegotiation games. To the extent that they are insured against counterparty risk, investors are not willing to downscale initially generous insurance promises to preserve incentives. This suggests that, with centralized clearing, the adjustment of transfers, contingent on the arrival of information, should be factored in the initial contract.

What if, instead, trading occurs bilaterally over-the-counter and there is no centralized clearing? Then, in the simple renegotiation game proposed above, after observing bad news the protection buyer knows he will be exposed to counterparty risk if he sticks to $\tau^B(\theta, \bar{s})$. In this case his expected utility is

$$\underline{\pi}u(\bar{\theta} + \tau^B(\bar{\theta}, \bar{s})) + (1 - \underline{\pi})pu(\underline{\theta} + \tau^B(\underline{\theta}, \bar{s})). \tag{40}$$

If instead he switches to $\tau^B(\theta, \underline{s})$, the protection buyer's expected utility is

$$\underline{\pi}u(\bar{\theta} + \tau^B(\bar{\theta}, \underline{s})) + (1 - \underline{\pi})u(\underline{\theta} + \tau^B(\underline{\theta}, \underline{s})). \tag{41}$$

Substituting for the transfers, (41) is larger than (40) if and only if

$$p < \frac{\frac{u(E(\theta|\underline{s}) + A\mathcal{P})}{u\left(E(\theta|\bar{s}) - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}A\mathcal{P}\right)} - \underline{\pi}}{1 - \underline{\pi}}. \tag{42}$$

If (42) holds, i.e., if effort strongly improves productive efficiency, then although they started from an initial noncontingent contract $\tau^B(\theta)$, both parties are happy to switch to $\tau^B(\theta, \underline{s})$ after bad news, in order to preserve incentives. This suggests that, with bilateral trading and non–centralized clearing, initially non–contingent contracts could, in some cases, be successfully renegotiated to the optimal contract.

On the other hand, if p is relatively large and (42) does not hold, protection buyers don't find it very attractive to renegotiate to lower insurance payments after bad news. In that case, renegotiation is unlikely to implement the optimal contract.

4.2 Derivative's payoffs

The payoff from an interest rate swap is symmetric, while the payoff from a credit-default swap is highly skewed: most of the time, protection sellers collect a small insurance premium but in the rare case of default, they have to make large payments to protection buyers. Does this skewness in the payoff have an effect on incentives?

To analyze the effect of an increase in the skewness of the hedged risk on incentives formally, we increase the probability π of a good outcome for the protection buyer's risk θ while keeping its mean and the standard deviation constant.[15] An increase of π increases the amount of risk to be hedged, $\Delta\theta$. Consequently, protection buyers demand more insurance, which increases the incentive problem for protection sellers. There is, however, a counterveiling effect when the skewness π is already large. In that case, the good outcome of the hedged risk is quite likely and the information content of a bad signal $\underline{s}$ is low. Thus, at high levels of π, a further increase of skewness mutes the negative effect of bad news on incentives. But, as long as $\pi < \lambda$ (the precision of the signal s), the negative effect on incentives from larger amounts of risk dominates and more skewness leads to more severe incentive problems. In this case, it is more difficult to maintain risk-management incentives when the underlying risk is skewed.

[15] The mean μ and the standard deviation σ of $\tilde{\theta}$ are $\mu = \pi\bar{\theta} + (1-\pi)\,\underline{\theta}$ and $\sigma = \sqrt{\pi\,(1-\pi)}\Delta\theta$, respectively. We can therefore write $\bar{\theta}$ and $\underline{\theta}$ as a function of π as follows: $\bar{\theta} = \mu + \sigma\sqrt{\frac{1-\pi}{\pi}}$ and $\underline{\theta} = \mu - \sigma\sqrt{\frac{\pi}{1-\pi}}$. Holding the mean and standard deviation constant, an increase in π leads to more skewness (when $\pi > \frac{1}{2}$).

4.3 Non-linear cost of risk-management effort

Up to now, we assumed the cost of risk–management effort increased linearly in the assets under management. We now relax this assumption and allow the cost of effort to be convex in assets under management.[16] This reflects the notion that, while controlling and preventing risk is relatively easy when the amount of assets under management is low, it gets more complex and costly when this amount is large.[17] Thus, we assume the cost of risk–prevention effort, when assets under management are $(1-\alpha)A$, is equal to

$$c(1-\alpha)A + \gamma(1-\alpha)^2 A^2. \tag{43}$$

In the analysis above we had $\gamma = 0$. $\gamma > 0$ gives rise to a new effect: as margins increase, assets under management decrease, and so does the marginal cost of risk–management. We hereafter analyze the optimal contract arising in this case. Since margins do not play any role in the contract without risk-management effort after bad news, we need only consider the contract with effort. As in Section 3.1, the feasibility and participation constraints bind: there is no reason to have idle resources or to leave rents to protection sellers. Moreover, the incentive constraint is slack after a good signal,and there is no margin call, while it binds after a bad signal, in which case there may be a margin call. As in Section 3, the incentive compatibility condition after bad news simplifies to

$$\alpha(\underline{s})A + (1-\alpha(\underline{s}))A\mathcal{P}(\alpha(\underline{s})) \geq E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s} = \underline{s}]. \tag{44}$$

The difference with Section 3 is that now the pledgeable return now depends on the size of the margin call after a bad signal:

$$\mathcal{P}(\alpha(\underline{s})) \equiv R - \frac{c + \gamma(1-\alpha(\underline{s}))A}{1-p}. \tag{45}$$

Margins improve risk-sharing when they relax the incentive constraint after a bad signal, i.e., when the left–hand–side of (44) is increasing in α, i.e.,

$$\mathcal{P}(\alpha(\underline{s})) - (1-\alpha(\underline{s}))\,\mathcal{P}' < 1. \tag{46}$$

[16] Cost convexity is a classical assumption in micro–economics. It leads to well behaved problems, in contrast with cost concavity for which optimality is more difficult to characterize.

[17] This is in line with Berk and Green (2004)'s assumption that funds' marginal returns decrease with assets under management.

It is easier to satisfy condition (46) when the cost of effort is convex (so that $\mathcal{P}' > 0$) than when it is linear (and $\mathcal{P}' = 0$). This reflects the above mentioned effect that, as margins increase, the marginal cost of risk–management decreases.

To determine the optimal margin with convex costs, we proceed as in Section 3.1. The transfers τ^B and τ^S have the same structure as in Proposition 5, except that $\mathcal{P}$ is now given by (45). We obtain the following proposition (where, as before, φ denotes the ratio of the marginal utility of a protection buyer after a bad and a good signal).

Proposition 12 *With a convex cost of risk-management effort, $\gamma > 0$, an optimal interior margin after a bad signal $\alpha^*(\underline{s})$ is given by*

$$\varphi(\alpha^*(\underline{s})) = 1 + \frac{R - C - 1}{1 - [\mathcal{P}(\alpha^*(\underline{s})) - (1 - \alpha^*(\underline{s}))\,\mathcal{P}']}. \tag{47}$$

As in Proposition 7, the optimal interior margin reflects the tradeoff between improved risk-sharing across signals and the opportunity cost of margin deposits. But now $\mathcal{P}' > 0$, which lowers the right–hand–side of the inequality. Holding $\mathcal{P}$ fixed (to reason other things equal) this increases the value of $\alpha^*(\underline{s})$ (the solution of (47)). Thus, we obtain the following comparative static result:

Proposition 13 *Other things equal, the greater the convexity of the cost of risk prevention, the larger the optimal margin.*

5 Empirical implications

According to our theory, a strong and pledgeable asset base ($A\mathcal{P}$) helps maintaining protection–sellers' risk–prevention incentives.[18] Asset pledgeability decreases with the cost of risk–prevention, the ineficiency of risk-management practices,[19] and the opacity and complexity of financial institutions and their activities. Our model (in particular Propositions 5 and 10) predicts a non-linear, U-shaped, relation between derivatives exposures and the pledgeability of assets:

[18] While for simplicity protection sellers have no initial debt in our model, to gauge this implication empirically one should consider assets net of liabilities.

[19] Ellul and Yerramili (2013) propose a Risk Management Index measuring the organizational strength and independence of the risk management function within financial insititutions.

Empirical implication 1. *Financial institutions with efficient risk-management and transparent activities optimally choose large derivatives exposures; financial institutions with less efficient risk-management and more opaque activities choose small derivatives exposures; financial institutions with very inefficient risk management and opaque activities choose large exposures (associated with significant counterparty risk).*

Derivatives exposure and protection sellers' incentives also depend on the macroeconomic and financial environment in which financial institutions operate. For example, an environment characterized by a low probability of failure even when there is no risk-management effort (high p) can be viewed as a "benign"/low-risk economic situation. Derivatives contracts that offer ample insurance but undermine risk-management incentives will be traded in such a benign environment (see Proposition 11). This resonates with the notion that risk builds up in "good" times (see, e.g., Borio, 2011).

In this context, consider the effect of bad news. For example, when the underlying risk insured is that of mortgage defaluts, declining house prices convey bad news. After bad news, protection sellers give up on risk–prevention. Hence they become more likely to default. This creates correlation between the mortgage values and the values of financial institutions' assets without direct exposure to mortgage default.

An increase in the precision of the public information signal (λ) increases this endogenous correlation. Information about the performance of mortgage-backed securities and CDS contracts written on them was unavailable before 2006.[20] The ABX.HE indices providing this information were introduced only in January 2006. As of early 2007, the prices for the index on AAA securitizations and those on BBB securitizations, which were virtually identical until then, started to diverge. Our theoretical analysis implies that the information then conveyed by the ABX.HE undermined the incentives of protection sellers. To the extent that ample insurance kept being written, it came at the expense of risk-taking. We summarize this discussion in our next empirical implication:

Empirical implication 2. *Derivatives contracts with large exposures are more likely to be underwritten when the economic environment seems benign. In this context, after bad news about the hedged risk, the expected value of the other assets of protection sellers decreases.*

[20] Although the issuance of mortgage-backed securites was around \$ 2 trillion in every year from 2002 until 2006 (see, e.g., Fender and Scheicher, 2008).

The more accurate the information about the hedged risk, the stronger this contagion.

The use of margins depends on their opportunity cost and the degree to which they alleviate protection sellers' incentive problem. The opportunity cost of margins depends on the risk-free rate (normalized to one in our analysis) since this is the rate assets on the margin account earn. When risk-free rates are low compared to the return on productive investment opportunities, the opportunity cost of margins increases and the optimal margin is lower.

Empirical implication 3. *When risk-free rates are low compared to the return on productive investment opportunities, the optimal margin deposit is lower.*

In terms of alleviating the incentive problem, margins are particularly beneficial when the cost of risk-management effort is convex, and the optimal margin is higher the more convex risk-management costs are (see Proposition 13). Convexity in risk-management costs implies that the risk of each additional unit of assets is more costly to manage. This could be a feature of complex and opaque (information-sensitive) assets which require intense monitoring and information collection, which becomes more expensive as the size of assets under management increases. Convexity in risk-management cost could also be related to liquidity of assets under management, with larger positions being more illiquid (e.g., due to a larger price impact and higher execution costs in case the position needs to be closed). The above discussion is summarized in our next implication:

Empirical implication 4. *The more risk-management costs increase with assets under management, the higher the optimal margin.*

6 Policy implications

6.1 Margins and equity capital

We showed that margins allow for more incentive-compatible insurance as they ring-fence assets from protection seller moral-hazard. Would capital requirements offer alternative mechanisms to reduce moral-hazard? What are the similarities and the differences between margins and equity capital in the context of our analysis? These questions are particularly relevant since the regulatory overhaul in the aftermath of the 2007-2009 financial crisis includes both margins and capital requirements. As argued below, our theoretical analysis implies margins can be an attractive substitute to capital.

Margins reduce the need for equity capital: In our model, at $t = 0$, protection sellers have assets A and no liabilities. Hence, the book value of their equity capital (the difference between assets and liabilities) is A. Its market value, reflecting rationally anticipated future cash flows, is AR. At $t = 1$, after a good signal, the derivative position is an expected asset for a protection seller, and the value of her equity increases. After a bad signal, however, the derivative position is an expected liability for a protection seller. The optimal contract with effort limits this liability to

$$A[\alpha(\underline{s}) + (1 - \alpha(\underline{s}))\mathcal{P})], \tag{48}$$

to preserve protection seller's incentives to exert risk-management effort (see (15)). Thus, the value of a protection seller's equity capital after a bad signal at $t = 1$ is

$$(1 - \alpha(\underline{s}))\,(R - \mathcal{P})\,A > 0, \tag{49}$$

which is the difference between the value of protection seller's assets, $A\,[\alpha(\underline{s}) + (1 - \alpha(\underline{s}))R]$, and the value of her liability, (48). The interpretation is that the optimal contract with effort requires protection sellers to hold a minimum amount of equity (i.e., keep enough skin in the game) to make sure the incentive compatibility constraint holds.

Without margin calls (e.g., if there was no enforcement mechanism for margins), the incentive compatibility condition would be more demanding. Hence protection sellers would need to have a higher amount of equity (more skin in the game) to ensure that effort remains incentive compatible. In that sense, margins are a substitute to equity capital. Margins improve incentives by making the asset side of the balance sheet less susceptible to moral-hazard. With less moral-hazard, the assets can support larger liabilities. Consequently, margins allow protection sellers to engage in incentive-compatible derivative trading with less equity.

Higher capital is an alternative to margins, but can be infeasible. Another way to relax the incentive-compatibility constraint after a bad signal would be to increase the protection seller's initial equity capital. This could be difficult to implement, however. In our simple agency-theoretic framework, raising capital from dispersed outside investors doesn't improve the incentives of the manager. Quite to the contrary, it dilutes her ownership of the firm and reduces her incentives to exert effort. Thus, increasing capital relaxes incentive-compatibility only if the additional capital belongs to the agent (increasing her skin in the game) or to investors closely monitoring the agent (which reduces the severity of the moral hazard problem.) When these conditions cannot be met, margin requirements are more effective than capital requirements.

Moreover, margins, unlike equity, are linked to derivative positions. A margin call only occurs when the derivative position turns into a liability (which depends on information about the underlying asset). Capital requirements, independent of the development of derivative positions, could be wasteful, as they would require equity capital even when derivative positions are profitable.

6.2 CCP design

The key advantage of the CCP over bilateral contracting is the mutualization of counterparty default risk. By insuring protection buyers, it makes them more eager to contract with protection sellers. At the same time, it makes each of them less eager to take costly action to reduce protection sellers' default risk. Margin calls, in our analysis, are one of the key instruments to reduce that risk. Thus, to implement the optimal contract characterized in this paper, one cannot delegate to the trading parties the task of designing their own individual margin calls. Such decentralization would lead to insufficient margining and excessive counterparty default. To see this, consider the case where the optimal contract calls for high effort even after bad news. Suppose the CCP offers the optimal transfers $\tau^S(\theta, \bar{s})$ and $\tau^B(\theta, \underline{s})$ described above, while letting each protection–seller/protection–buyer pair choose their own margin call. A limited–liability protection seller and a protection buyer insured by the CCP against counterparty risk, both prefer to set $\alpha(s) = 0, \forall s$. This implies the incentive-compatibility condition of the protection seller does not hold, and results in excessive counterparty default risk. This is a form of free-riding, since the cost of that default is borne by all the other members of the CCP. To avoid such free-riding, margin calls must be mandated by the CCP.

7 Conclusion

We analyze optimal contracts in the context of hedging with derivatives. We show how contracts designed to engineer risk-sharing can generate incentives for risk-taking. When the position of the protection seller becomes a liability for her, it undermines her incentives to exert risk prevention effort. The failure to exert such effort may lead to the default of the protection seller. Thus, a bad signal about derivative positions can propagate to other lines of business of financial institutions and, when doing so, create endogenous counterparty risk.

When the seller's moral hazard is moderate, margins enhance the scope for risk-sharing. Our emphasis on the positive consequence of margins contrasts with the result that margins can be destabilizing, as shown by Gromb and Vayanos (2002) and Brunnermeier and Pedersen (2009). The contrast stems from differences in assumptions. Gromb and Vayanos (2002) and Brunnermeier and Pedersen (2009) take margin constraints as given and, for these margins, derive equilibrium prices. Greater margins force intermediaries to sell more after bad shocks, which pushes prices down and can generate spirals. In contrast, we endogenize margins, but take as given the value of assets a protection seller deposits on a margin account. It would be interesting in future research to combine the two approaches and study how endogenous margins could destabilize equilibrium prices. This would be in the line of Acharya and Viswanathan (2011)'s analysis of the equilibrium price at which borrowers liquidate assets and the corresponding fire-sales negative externality.

References

Acharya, V. V. and A. Bisin, 2011, “Counterparty Risk Externality: Centralized versus Over-the-counter Markets,” NBER working paper No. 17000.

Acharya, V. V., R. F. Engle, S. Figlewski, A. W. Lynch and M. G. Subrahmanyam (2012), “Centralized Clearing for Credit Derivatives,” Chapter 11 in Restoring Financial Stability: How to Repair a Failed System Viral V. Acharya and Matthew Richardson eds.

Acharya, V.V., H. Le and H.S. Shin, 2013, “Bank Capital and Dividend Externalities,” NBER Working Paper No. 19707.

Acharya, V. V. and S. Viswanathan, 2011, “Leverage, moral hazard and liquidity,” *Journal of Finance*, 66 (1), 99-138.

Acharya, V.V. and T. Yorulmazer, 2008, “Information Contagion and Bank Herding,” *Journal of Money, Credit and Banking* 40 (1), 215–231.

Allen, F. and D. Gale, 2000, “Financial Contagion,” *Journal of Political Economy* 108(1), 1-33.

Berk, J. and R. Green, 2004, “Mutual Fund Flows and Performance in Rational Markets,” *Journal of Political Economy*, 112 (6), 1269-1295.

Biais, B. and C. Casamatta, 1999, “Optimal Leverage and Aggregate Investment,” *Journal of Finance*, 54 (4), 1291-1323.

Billio, M., M. Getmansky, A. Lo and L. Pelizzon, 2012,“Econometric measures of connectedness and systemic risk in the finance and insurance sectors,” *Journal of Financial Economics*, 104 (3), 535-559.

Bolton, P. and M. Oehmke, 2011, “Credit Default Swaps and the Empty Creditor Problem,” *Review of Financial Studies*, forthcoming.

Bolton, P. and M. Oehmke, 2013, “Should Derivatives be Senior?,” *Journal of Finance*, forthcoming.

Bolton, P. and D. Scharfstein, 1990, “A Theory of Predation Based on Agency Problems in Financial Contracting,” *American Economic Review*, 80, 93–106.

Borio, C., 2011. Rediscovering the macroeconomic roots of ï¬ï¿½nancial stability policy: journey, challenges and a way forward. Annual Review of Financial Economics 3, 87â€“117.

Brunnermeier, M. and L. Pedersen, 2009, "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22 (6), 2201-2238.

Cifuentes, R., G. Ferrucci and H. S. Shin, 2005, "Liquidity Risk and Contagion," *Journal of the European Economic Association* 3 (2-3), 556-566.

DeMarzo, P., M. Fishman, Z. He, and N. Wang, 2011, "Dynamic Agency and q Theory of Investment," *Journal of Finance*, forthcoming.

Dudley, William C, 2014, "Global financial stability - the road ahead," remarks by Mr William C Dudley, President and Chief Executive Officer of the Federal Reserve Bank of New York, at the Tenth Asia-Pacific High Level Meeting on Banking Supervision, Auckland, New Zealand, 26 February 2014.

Duffie, D., A. Li, and T. Luebke, 2010, "Policy Perspectives on OTC Derivatives Market Infrastructure," MFI Working Paper 2010-002.

Economist, 2014, "Centrally Cleared Derivatives: Clear and Present Danger," http://www.economist.co

Ellul, A. and V. Yerramilli, 2013, "Stronger Risk Controls, Lower Risk: Evidence from US Bank Holding Companies," *Journal of Finance*, 68 (5), 1757-1803.

European Central Bank, 2009, "Credit Default Swaps and Counterparty Risk," August.

Fender, I. and M. Scheicher, 2008,"The ABX: how doe the markets price subprime mortage risk?," *BIS Quarterly Review*, September, 67-81.

Fleming, M. and A. Sarkar, 2014, "The Failure Resolution of Lehman Brothers," *Economic Policy Review* Vol. 20, No. 2.

Freixas, X., B. Parigi and J.-C. Rochet, 2000, "Systemic Risk, Interbank Relations, and Liquidity Provision by the Central Bank," *Journal of Money, Credit and Banking* 32 (3), 611-638.

Froot, K., Scharfstein, D. and J. Stein, 1993, "Risk Management: Coordinating Corporate Investment and Financing Policies," *Journal of Finance*, 48 (5), 1629-1658.

Gorton, G., and G. Pennacchi, 1990, "Financial Intermediaries and Liquidity Creation," *Journal of Finance*, 45 (1), 49–71.

Gromb D, and D. Vayanos, 2002, "Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs," *Journal of Financial Economics*, 66 (2-3), 361-407.

Holmström, B. and J. Tirole, 1998, "Private and Public Supply of Liquidity," *Journal of*

Political Economy, 106 (1), 1-40.

Jensen, M., and Meckling, 1976, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, 3 (4), 305-360.

King, M. and S. Wadhwani, 1990, "Transmission of volatility between stock markets," *Review of Financial Studies* 3 (1), 5-33.

Myers, S., 1977, "Determinants of Corporate Borrowing," *Journal of Financial Economics*, 5 (2), 147-175.

Parlour, C. and G. Plantin, 2008, "Loan Sales and Relationship Banking," *Journal of Finance*, 63 (3), 1291-1314.

Rampini, A. and S. Viswanathan, 2010, "Collateral, Risk Management, and the Distribution of Debt Capacity," *Journal of Finance*, 65 (6), 2293-2322.

Thompson, J.R., 2010, "Counterparty Risk in Financial Contracts: Should the Insured Worry about the Insurer?," *Quarterly Journal of Economics*, 125 (3), 1195-1252.

Tirole, J., 2006, *The Theory of Corporate Finance* (Princeton University Press, Princeton).

8 Appendix

Proof of Proposition 1 Form the Lagrangian using the objective (6), the feasibility constraints (3) with multiplier μ_{FC} and the participation constraint (7) with multiplier μ. For the moment we ignore the limited liability constraints (4) in the first-best. We then show that first-best transfers do not violate limited liability given our assumption $AR > \pi\Delta\theta$. Since $\tilde{R} = R$ under effort, we do not explicitly write the dependence of the transfers on $\tilde{R}$.

The first-order conditions of the Lagrangian with respect to $\tau^B(\theta, s)$ and $\tau^S(\theta, s)$ are, respectively,

$$\text{prob}[\theta, s]u'(\theta + \tau^B(\theta, s)) - \mu_{FC}(\theta, s) = 0 \qquad \forall(\theta, s) \tag{50}$$

$$\mu\text{prob}[\theta, s] - \mu_{FC}(\theta, s) = 0 \qquad \forall(\theta, s). \tag{51}$$

Since marginal utility is strictly positive, it follows from (50) that $\mu_{FC}(\theta, s) > 0$ for all (θ, s) and hence the feasibility constraints bind. Since $\mu_{FC}(\theta, s) > 0$, it follows from (51) that the participation constraint binds. After substituting (50) into (51), it follows that buyers' marginal utility is the same across all states. That is, there is full risk-sharing.

From equal marginal utility across all states, we obtain, first, that $\theta + \tau^B(\theta, \bar{s}) = \theta + \tau^B(\theta, \underline{s})$ and hence $\tau^B(\theta, \bar{s}) = \tau^B(\theta, \underline{s})$ for $\theta = \bar{\theta}, \underline{\theta}$. Second, we obtain that $\bar{\theta} + \tau^B(\bar{\theta}, s) = \underline{\theta} + \tau^B(\underline{\theta}, s)$ and hence $\tau^B(\underline{\theta}, s) - \tau^B(\bar{\theta}, s) = \Delta\theta$ for $s = \bar{s}, \underline{s}$.

Using $\tau^S(\theta, s) = -\tau^B(\theta, s)$ (from the binding feasibility constraints) and $\tau^B(\theta, \bar{s}) = \tau^B(\theta, \underline{s})$, we can write the binding participation constraint as

$$-(\text{prob}[\bar{\theta}, \bar{s}] + \text{prob}[\bar{\theta}, \underline{s}])\tau^B(\bar{\theta}, \bar{s}) - (\text{prob}[\underline{\theta}, \bar{s}] + \text{prob}[\underline{\theta}, \underline{s}])\tau^B(\underline{\theta}, \bar{s}) = 0 \tag{52}$$

Using $\tau^B(\underline{\theta}, \bar{s}) - \tau^B(\bar{\theta}, \bar{s}) = \Delta\theta$ to substitute for $\tau^B(\bar{\theta}, \bar{s})$ and since $\text{prob}[\bar{\theta}, \bar{s}] + \text{prob}[\bar{\theta}, \underline{s}] = \text{prob}[\bar{\theta}] = \pi$ (and similarly for $1 - \pi$), the binding participation constraint yields $\tau^B(\underline{\theta}, \bar{s}) = \pi\Delta\theta$, from which the remaining transfers in the proposition follow immediately. QED

Proof of Lemma 1 Plugging the first-best transfers from Proposition 1 into the incentive conditions (10) and using $\alpha(\tilde{s}) = 0$ yields $A\mathcal{P} \geq (\pi - \bar{\pi})\Delta\theta$ and $A\mathcal{P} \geq (\pi - \underline{\pi})\Delta\theta$. When the signal is informative, $\lambda > \frac{1}{2}$, we have $\bar{\pi} > \pi > \underline{\pi}$. The result in the lemma follows. QED

Proof of Proposition 2 Form the Lagrangian using the objective (6), the feasibility constraints (3) with multiplier $\mu_{FC}(\theta, s)$, the limited liability constraints (4) with multipliers $\mu_{LL}(\theta, s)$, the feasibility constraints on margins (5) with $\mu_0(s)$ for $\alpha(s) \geq 0$ and $\mu_1(s)$ for $\alpha(s) \leq 1$, the incentive compatibility constraints (10) with multipliers $\mu_{IC}(s)$ and the participation constraint (12) with multiplier μ.

The first-order conditions of the Lagrangian with respect to $\tau^B(\theta, s)$ and $\tau^S(\theta, s)$ are

$$\begin{aligned} \text{prob}[\theta, s]u'(\theta + \tau^B(\theta, s)) - \mu_{FC}(\theta, s) &= 0 \qquad \forall(\theta, s) &(53)\\ \mu\text{prob}[\theta, s] + \mu_{LL}(\theta, s) + \text{prob}[\theta|s]\mu_{IC}(s) - \mu_{FC}(\theta, s) &= 0 \qquad \forall(\theta, s). &(54) \end{aligned}$$

Since marginal utilities are positive, it follows from (53) that $\mu_{FC}(\theta, s) > 0$ and hence all feasibility constraints bind:

$$\tau^B(\theta, s) = -\tau^S(\theta, s), \forall(\theta, s). \tag{55}$$

Using (53) to substitute for $\mu_{FC}(\theta, s)$ in (54) and rearranging, we obtain

$$u'(\theta + \tau^B(\theta, s)) = \mu + \frac{\mu_{LL}(\theta, s)}{\text{prob}[\theta, s]} + \frac{\mu_{IC}(s)}{\text{prob}[s]} \qquad \forall(\theta, s) \tag{56}$$

where we used that $\text{prob}[\theta|s]\text{prob}[s] = \text{prob}[\theta, s]$.

We next show that the limited liability constraint in state $(\bar{\theta}, s)$ is slack for each s. The proof proceeds in two steps. First, we show that the limited liability constraints cannot bind for both the state $(\bar{\theta}, s)$ and the state $(\underline{\theta}, s)$. Suppose not. Since both limited liability constraints after the signal s bind, we have $-\tau^S(\bar{\theta}, s) = \alpha(s)A + (1-\alpha(s))AR$ and $-\tau^S(\underline{\theta}, s) = \alpha(s)A + (1 - \alpha(s))AR$. Hence,

$$E[-\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = s] = \alpha(s)A + (1 - \alpha(s))AR \quad \forall s$$

But since $R > \mathcal{P}$, this violates the incentive compatibility constraint (10) after the signal s. Hence, at least one limited liability constraint after the signal s must be slack.

Second, we show that the limited liability constraint in state $(\bar{\theta}, s)$ is always slack for each s. Suppose not, so that $-\tau^S(\bar{\theta}, s) = \alpha(s)A + (1 - \alpha(s))AR$. We have just shown that at least one limited liability constraint after the signal s must be slack. Hence, we must have that $-\tau^S(\underline{\theta}, s) < \alpha(s)A + (1 - \alpha(s))AR$ and $\mu_{LL}(\underline{\theta}, s) = 0$. Using the binding feasibility constraints (55), we therefore have $\tau^B(\bar{\theta}, s) > \tau^B(\underline{\theta}, s) \quad \forall s$, which implies $u'(\bar{\theta} + \tau^B(\bar{\theta}, s)) < u'(\underline{\theta} + \tau^B(\underline{\theta}, s)), \quad \forall s$, since $\bar{\theta} > \underline{\theta}$. However, using $\mu_{LL}(\underline{\theta}, s) = 0$ in (56) implies $u'(\bar{\theta} + \tau^B(\bar{\theta}, s)) \geq u'(\underline{\theta} + \tau^B(\underline{\theta}, s)), \quad \forall s$. A contradiction. Hence, the limited liability constraint is slack in state $(\bar{\theta}, s)$ and $\mu_{LL}(\bar{\theta}, s) = 0$ for all s.

Finally, we show by contradiction that the participation constraint (12) binds. Suppose not. Plugging $\mu = 0$ and $\mu_{LL}(\bar{\theta}, s) = 0$ (just shown above) into (56) implies that $\mu_{IC}(s) > 0$ for all s. Hence, both incentive constraints bind, $-E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = s] = \alpha(s)A + (1 - \alpha(s))A\mathcal{P}$ for $s = \bar{s}, \underline{s}$. Therefore,

$$E[\tau^S(\tilde{\theta}, \tilde{s})] = E[E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s}]] = -E[\alpha(\tilde{s})A + (1 - \alpha(\tilde{s}))A\mathcal{P}] \tag{57}$$

From the participation constraint, we have

$$\begin{aligned} 0 &\leq E[\tau^S(\tilde{\theta}, \tilde{s})] - E[\alpha(\tilde{s})]A(R - C - 1) \\ &= -E[\alpha(\tilde{s})A + (1 - \alpha(\tilde{s}))A\mathcal{P}] - E[\alpha(\tilde{s})]A(R - C - 1) \quad \text{[using (57)]} \\ &= -E[(1 - \alpha(\tilde{s}))A\mathcal{P} + \alpha(\tilde{s})A(R - C)]. \end{aligned}$$

The last expression is strictly negative since $R - C > \mathcal{P} > 0$ and $0 \leq \alpha(\tilde{s}) \leq 1$. A contradiction. Hence, the participation constraint binds and also $\mu > 0$. QED

Proof of Proposition 3

The first-order conditions of the Lagrangian from the proof of Proposition 2 with respect to $\alpha(s)$ are

$$\frac{\mu_0(s) - \mu_1(s)}{A} + \mu_{IC}(s)(1 - \mathcal{P}) = \mu\text{prob}[s](R - C - 1) + (R - 1)\mu_{LL}(\underline{\theta}, s) \qquad \forall s, \tag{58}$$

where we have used $\mu_{LL}(\bar{\theta}, s) = 0$ for all s (Proposition 2).

The right-hand side of (58) is strictly positive since $R - C > 1$ and $\mu > 0$ (see the end of the proof of Proposition 2). If the incentive constraint is slack for a signal s, then $\mu_s = 0$, implying that $\mu_0(s) > 0$ must hold and $\alpha(s) = 0$. Similarly, if $\mathcal{P} \geq 1$, then $\mu_0(s) > 0$ for each s must hold and $\alpha(s) = 0$ for all s. QED

Proof of Proposition 4

It cannot be that both incentive constraints are slack since we assume that the first-best is not attainable, $A\mathcal{P} < (\pi - \underline{\pi})\Delta\theta$. It also cannot be that both incentive constraints bind (see the argument that the participation constraint binds in the proof of Proposition 2).

We now show by contradiction that the incentive constraint following a bad signal binds. Suppose not and hence $\mu_{IC}(\underline{s}) = 0$. After the good signal, the limited liability constraints are slack, $\mu_{LL}(\bar{\theta}, \bar{s}) = 0$ by Proposition 2 and $\mu_{LL}(\underline{\theta}, \bar{s}) = 0$ since we are considering a relaxed problem - see condition (14)). Equations (56) for $s = \bar{s}$ then imply that $u'(\bar{\theta} + \tau^B(\bar{\theta}, \bar{s})) = u'(\underline{\theta} + \tau^B(\underline{\theta}, \bar{s}))$. There is full risk-sharing conditional on the good signal. For transfers after a good signal we thus have

$$\tau^B(\underline{\theta}, \bar{s}) - \tau^B(\bar{\theta}, \bar{s}) = \Delta\theta > 0 \tag{59}$$

After the bad signal, limited liability constraint in state $(\bar{\theta}, \underline{s})$ is slack, $\mu_{LL}(\bar{\theta}, \underline{s}) = 0$ by Proposition 2. In state $(\underline{\theta}, \underline{s})$, we have two cases to consider, depending on whether the limited liability constraint is slack or whether it binds.

Consider first the case when the limited liability constraint in state $(\underline{\theta}, \underline{s})$ is slack, $\mu_{LL}(\underline{\theta}, \underline{s}) = 0$. Equations (56) for $s = \underline{s}$ then imply that there is also full risk-sharing conditional on the bad signal, $u'(\bar{\theta} + \tau^B(\bar{\theta}, \underline{s})) = u'(\underline{\theta} + \tau^B(\underline{\theta}, \underline{s}))$, and thus

$$\tau^B(\underline{\theta}, \underline{s}) - \tau^B(\bar{\theta}, \underline{s}) = \Delta\theta > 0 \tag{60}$$

Since $\mu_{IC}(\underline{s}) = 0$ and $\mu_{LL}(\underline{\theta}, \underline{s}) = \mu_{LL}(\underline{\theta}, \bar{s}) = 0$, it follows from equations in (56) that $u'(\underline{\theta} + \tau^B(\underline{\theta}, \underline{s})) \leq u'(\underline{\theta} + \tau^B(\underline{\theta}, \bar{s}))$, and thus

$$\tau^B(\underline{\theta}, \underline{s}) \geq \tau^B(\underline{\theta}, \bar{s}). \tag{61}$$

From the binding participation constraint

$$\text{prob}[\bar{s}]E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}] + \text{prob}[\underline{s}]E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] = E[\alpha(\tilde{s})]A(R - C - 1) \geq 0$$

and $E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}] < 0$ (binding incentive constraint after a good signal), we know that

$$E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] > 0 \tag{62}$$

Using full risk-sharing conditional on the signal (equations (59) and (60)) we can write

$$\begin{aligned} E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] &= \underline{\pi}\tau^S(\bar{\theta}, \underline{s}) + (1 - \underline{\pi})\,\tau^S(\underline{\theta}, \underline{s}) \\ &= \tau^S(\underline{\theta}, \underline{s}) + \underline{\pi}\left[\tau^S(\bar{\theta}, \underline{s}) - \tau^S(\underline{\theta}, \underline{s})\right] \\ &= \tau^S(\underline{\theta}, \underline{s}) + \underline{\pi}\left[\tau^S(\bar{\theta}, \bar{s}) - \tau^S(\underline{\theta}, \bar{s})\right] \end{aligned}$$

Using (61) and the binding feasibility conditions (55), we have $\tau^S(\underline{\theta}, \underline{s}) \leq \tau^S(\underline{\theta}, \bar{s})$. And since $\underline{\pi} < \bar{\pi}$ (the signal is informative), we have

$$\begin{aligned} E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] &\leq \tau^S(\underline{\theta}, \bar{s}) + \underline{\pi}\left[\tau^S(\bar{\theta}, \bar{s}) - \tau^S(\underline{\theta}, \bar{s})\right] \\ &< \tau^S(\underline{\theta}, \bar{s}) + \bar{\pi}\left[\tau^S(\bar{\theta}, \bar{s}) - \tau^S(\underline{\theta}, \bar{s})\right] \end{aligned}$$

and thus $E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] < E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}]$. But since $E[\tau^S(\tilde{\theta}, \tilde{s})|\bar{s}] < 0$ (by the binding incentive constraint after a good signal), we have a contradiction with (62).

Now, consider the case when the limited liability constraint in state $(\underline{\theta}, \underline{s})$ binds. Since $\mu_{LL}(\bar{\theta}, \underline{s}) = 0$ (by Proposition 2) and $\mu_{IC}(\underline{s}) = 0$, equations (56) for $s = \underline{s}$ imply that $u'(\underline{\theta} + \tau^B(\underline{\theta}, \underline{s})) \geq u'(\bar{\theta} + \tau^B(\bar{\theta}, \underline{s}))$, and thus

$$\tau^B(\underline{\theta}, \underline{s}) - \tau^B(\bar{\theta}, \underline{s}) \leq \Delta\theta. \tag{63}$$

Since $\alpha(\underline{s}) = 0$ (incentive constraint after a bad signal is slack in contradiction), the binding limited liability constraint is $AR = -\tau^S(\underline{\theta}, \underline{s})$. Together with (63) in conjunction with the binding feasibility constraints (55), we then have

$$\begin{aligned} -E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] &= -\left[\underline{\pi}\tau^S(\bar{\theta}, \underline{s}) + (1 - \underline{\pi})\,\tau^S(\underline{\theta}, \underline{s})\right] \\ &= -\tau^S(\underline{\theta}, \underline{s}) - \underline{\pi}\left[\tau^S(\bar{\theta}, \underline{s}) - \tau^S(\underline{\theta}, \underline{s})\right] \\ &\geq AR - \underline{\pi}\Delta\theta \end{aligned}$$

Since $\underline{\pi} < \pi$ (informative signal) and $AR > \pi\Delta\theta$ (limited liability constraints are slack in the first-best), we have $-E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] > (\pi - \underline{\pi})\Delta\theta$. But since the incentive constraint after a bad signal is slack, $A\mathcal{P} > -E[\tau^S(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}]$, this would mean that $A\mathcal{P} > (\pi - \underline{\pi})\Delta\theta$ and the first-best can be reached, which is a contradiction.

Consequently, the incentive constraint after a bad signal binds and the incentive constraint after a good signal must be slack. QED

Proof of Proposition 5

After a good signal, we have full risk-sharing (see the derivation of equation (59) in the proof of Proposition 4). Using (59) and (16), we obtain the transfers $\tau^B(\bar{\theta}, \bar{s})$ and $\tau^B(\underline{\theta}, \bar{s})$.

After a bad signal, we have to distinguish two cases, depending on whether the limited liability constraint in state $(\underline{\theta}, \underline{s})$ is slack or not. If it is slack, then we have full risk-sharing (see the derivation of equation (60) in the proof of Proposition 4). Using (60) and (15), we obtain the transfers $\tau^B(\bar{\theta},\underline{s})$ and $\tau^B(\underline{\theta},\underline{s})$. If the limited liability constraint binds, we have $\alpha(\underline{s})A+(1-\alpha(\underline{s}))AR=-\tau^S(\underline{\theta},\underline{s})$, which we plug into (15) to obtain $\tau^B(\bar{\theta},\underline{s})$.

Finally, we check that, under (14), the limited liability constraint in $(\underline{\theta},\bar{s})$ is slack. Since $\alpha(\bar{s})=0$, the limited liability constraint (4) writes as $\tau^B(\underline{\theta},\bar{s})<AR$. Now, Proposition 5 implies that $\tau^B(\underline{\theta},\bar{s})$ decreases in $\alpha(\underline{s})$. So $\tau^B(\underline{\theta},\bar{s})<AR$ for all $\alpha(\underline{s})$ if and only if it is for $\alpha(\underline{s})=0$. After simplifications, $\tau^B(\underline{\theta},\bar{s})<AR$ for $\alpha(\underline{s})=0$ is equivalent to (14).

QED

Proof of Proposition 6

We claim that $\alpha^*(\underline{s})<1$. Suppose not and $\alpha^*(\underline{s})=1$. First, note that $\mu_{LL}(\underline{\theta},\underline{s})>0$ must hold in this case. Suppose not, and $\mu_{LL}(\underline{\theta},\underline{s})=0$. Then, equations (56) for $s=\underline{s}$ imply that that there is full risk-sharing conditional on the bad signal. Hence, the individual transfers after the bad signal are given by (18) so that $\tau^B(\underline{\theta},\underline{s})=-\tau^S(\underline{\theta},\underline{s})=\underline{\pi}\Delta\theta+A>A$. But the limited liability constraint requires $-\tau^S(\underline{\theta},\underline{s})\leq A$, a contradiction. Since $\mu_{LL}(\underline{\theta},\underline{s})>0$, the limited liability constraint binds and the individual transfers after a bad signal are as in (19). In particular, $\tau^B(\bar{\theta},\underline{s})=A>0$. Equations (56) and binding incentive constraint after a bad signal imply that $\tau^B(\bar{\theta},\bar{s})\geq\tau^B(\bar{\theta},\underline{s})=A>0$. However, by equation (17), $\tau^B(\bar{\theta},\bar{s})<0$. A contradiction.

QED

Proof of Proposition 7

Since the incentive constraint after a good signal is slack (see Proposition 4), it follows from Proposition 3 that $\alpha^*(\bar{s})=0$. It remains to characterize the optimal margin after a bad signal.

We now derive the optimal margin after a bad signal, $\alpha^*(\underline{s})$. Using equations (56) to subsitute for μ, $\mu_{IC}(\underline{s})$ and $\mu_{LL}(\underline{\theta},\underline{s})$ in equation (58), we get

$$\begin{aligned}\frac{u'(\bar{\theta}+\tau^B(\bar{\theta},\underline{s}))}{u'(\bar{\theta}+\tau^B(\bar{\theta},\bar{s}))} &= 1+\frac{R-C-1}{1-\mathcal{P}}+\frac{\mu_1(\underline{s})-\mu_0(\underline{s})}{u'(\bar{\theta}+\tau^B(\bar{\theta},\bar{s}))\text{prob}[\underline{s}]\,(1-\mathcal{P})\,A} \\ &\quad +\frac{1-\underline{\pi}}{1-\mathcal{P}}\frac{u'(\underline{\theta}+\tau^B(\underline{\theta},\underline{s}))-u'(\bar{\theta}+\tau^B(\bar{\theta},\underline{s}))}{u'(\bar{\theta}+\tau^B(\bar{\theta},\bar{s}))}\end{aligned} \tag{64}$$

where we used $\mu_{IC}(\bar{s})=0$ (Proposition 4).

Denote the RHS of (64) by φ. Note that $\frac{\partial\tau^B(\bar{\theta},\bar{s})}{\partial\alpha} = -\frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}A\left(R-C-\mathcal{P}\right) < 0$. For $\mathcal{P}<1$, $\frac{\partial\tau^B(\bar{\theta},\underline{s})}{\partial\alpha} > 0$. (When the limited liability constraint is slack, we have $\frac{\partial\tau^B(\bar{\theta},\underline{s})}{\partial\alpha} = A\left(1-\mathcal{P}\right) > 0$ and when the limited liability constraint binds, we have $\frac{\partial\tau^B(\bar{\theta},\underline{s})}{\partial\alpha} = A\left[1+\frac{(1-\underline{\pi})R-\mathcal{P}}{\underline{\pi}}\right] > 0$ since $R-\mathcal{P} > R-1 > \underline{\pi}(R-1)$). Hence, φ is decreasing in α. If $\varphi\left(0\right) < 1+\frac{R-C-1}{1-\mathcal{P}}$, then

$$\varphi\left(0\right) < 1+\frac{R-C-1}{1-\mathcal{P}}+\frac{1-\underline{\pi}}{1-\mathcal{P}}\frac{u'(\underline{\theta}+\tau^B(\underline{\theta},\underline{s}))-u'(\bar{\theta}+\tau^B(\bar{\theta},\underline{s}))}{u'(\bar{\theta}+\tau^B(\bar{\theta},\bar{s}))}$$

for any $\alpha\in[0,1]$ (since the last term is non-negative). By equation (64) we have $\mu_0>0$ and hence $\alpha^*(\underline{s})=0$.

Otherwise, there are two cases depending on whether or not the limited liability constraint in state $(\underline{\theta},\underline{s})$ is slack. If it is slack, then marginal utilities after the bad signal are equalized (equation (60)), and the last term in equation (64) vanishes. The optimal margin $\alpha^*(\underline{s})\in(0,1)$ is given by $\varphi\left(\alpha^*(\underline{s})\right) = 1+\frac{R-C-1}{1-\mathcal{P}}$ in this case. If the limited liability constraint binds, then the optimal margin $\alpha^*(\underline{s})\in(0,1)$ solves

$$\frac{u'(\bar{\theta}+\tau^B(\bar{\theta},\underline{s}))}{u'(\bar{\theta}+\tau^B(\bar{\theta},\bar{s}))}-\frac{1-\underline{\pi}}{1-\mathcal{P}}\frac{u'(\underline{\theta}+\tau^B(\underline{\theta},\underline{s}))-u'(\bar{\theta}+\tau^B(\bar{\theta},\underline{s}))}{u'(\bar{\theta}+\tau^B(\bar{\theta},\bar{s}))} = 1+\frac{R-C-1}{1-\mathcal{P}}$$

We now check under what conditions the limited liability constraints are slack. By Proposition 2, we only need to check limited liability constraints in states $(\underline{\theta},\bar{s})$ and $(\underline{\theta},\underline{s})$. First, consider the case when $\mathcal{P}\geq 1$ and margins are not used. The limited liability constraints are slack if and only if: $AR > -\tau^S(\theta,s,R) = \tau^B(\theta,s,R)$, $\forall(\theta,s,R)$. Since $\tau^B(\underline{\theta},\bar{s})\geq\tau^B(\underline{\theta},\underline{s})$, we only need to check when the limited liability constraint is slack in state $(\underline{\theta},\bar{s})$. It is slack if and only if:

$$\bar{\pi}\Delta\theta-\frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}A\mathcal{P} < AR \tag{65}$$

or, equivalently,

$$AR-\pi\Delta\theta > \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}\left[\left(\pi-\underline{\pi}\right)\Delta\theta-A\mathcal{P}\right] > 0.$$

Now consider the case when $\mathcal{P}<1$. The limited liability constraints in this case are slack if and only if: $\alpha(s)A+(1-\alpha(s))AR > -\tau^S(\theta,s,R)$, $\forall(\theta,s,R)$, with $\alpha^*\left(\bar{s}\right)=0$ and $\alpha^*(\underline{s})\geq 0$. The limited liability constraint in state $(\underline{\theta},\bar{s})$ is slack if and only if:

$$\bar{\pi}\Delta\theta-\frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}A\left[\alpha^*(\underline{s})\left(R-C\right)+(1-\alpha^*(\underline{s}))\mathcal{P}\right] < AR$$

Since $R - C > \mathcal{P} > 0$, we have:

$$\bar{\pi}\Delta\theta - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} A\left[\alpha^*(\underline{s})\left(R - C\right) + (1 - \alpha^*(\underline{s}))\mathcal{P}\right] < \bar{\pi}\Delta\theta - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} A\mathcal{P}$$

Hence, condition (65) is sufficient for the limited liability constraint to be slack in state $(\underline{\theta}, \bar{s})$.

The limited liability constraint in state $(\underline{\theta}, \underline{s})$ is slack if and only if:

$$\alpha^*(\underline{s}) < 1 - \frac{\underline{\pi}\Delta\theta}{A\left(R - \mathcal{P}\right)}$$

Since the optimal interior margin when the limited liability constraint is slack is given by

$$\alpha^*(\underline{s}) = \varphi^{-1}\left(1 + \frac{R - C - 1}{1 - \mathcal{P}}\right),$$

the constraint in state $(\underline{\theta}, \underline{s})$ is slack if and only if

$$\varphi^{-1}\left(1 + \frac{R - C - 1}{1 - \mathcal{P}}\right) < 1 - \frac{\underline{\pi}\Delta\theta}{A\left(R - \mathcal{P}\right)}.$$

Note that if the limited liability constraint in state $(\underline{\theta}, \underline{s})$ is slack, it must be that $\tau^B(\bar{\theta}, \underline{s}) < 0$ (equation (18)) implying that

$$\alpha^*(\underline{s}) < \frac{(1 - \underline{\pi})\Delta\theta - A\mathcal{P}}{A\left(1 - \mathcal{P}\right)}$$

must hold if the limited liability constraint in state $(\underline{\theta}, \underline{s})$ is slack.

In case the limited liability constraint binds, it also must be that $\tau^B(\bar{\theta}, \underline{s}) < 0$. This is because equations (19) imply that

$$\begin{aligned}
\tau^B(\underline{\theta}, \underline{s}) &= \alpha(\underline{s})A + (1 - \alpha(\underline{s}))AR > \\
E[\tau^B(\tilde{\theta}, \tilde{s})|\tilde{s} &= \underline{s}] = \alpha(\underline{s})A + (1 - \alpha(\underline{s}))A\mathcal{P} \qquad \text{[since } R > \mathcal{P} \text{ and } \alpha^*(\underline{s}) < 1\text{]} \\
&> 0 > \tau^B(\bar{\theta}, \underline{s}) \qquad \text{[since } E[\tau^B(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] = \underline{\pi}\tau^B(\bar{\theta}, \underline{s}) + (1 - \underline{\pi})\,\tau^B(\underline{\theta}, \underline{s})\text{]}
\end{aligned}$$

For $\tau^B(\bar{\theta}, \underline{s})$ to be negative if the limited liability constraint in state $(\underline{\theta}, \underline{s})$ binds, it must be that

$$\alpha^*(\underline{s})\left[1 + \frac{(1 - \underline{\pi})R - \mathcal{P}}{\underline{\pi}}\right] < \frac{(1 - \underline{\pi})R - \mathcal{P}}{\underline{\pi}}$$

or, equivalently,

$$\alpha^*(\underline{s}) < \frac{(1 - \underline{\pi})R - \mathcal{P}}{\underline{\pi} + (1 - \underline{\pi})R - \mathcal{P}} < 1$$

It follows that a sufficient condition for the limited liability constraint in state $(\underline{\theta}, \underline{s})$ to be slack is

$$1-\frac{\underline{\pi}\Delta\theta}{A\left(R-\mathcal{P}\right)}>\frac{(1-\underline{\pi})R-\mathcal{P}}{\underline{\pi}+(1-\underline{\pi})R-\mathcal{P}}.$$

QED

Proof of Proposition 8

Form the Lagrangian using the objective (26), the feasibility constraints (27) and (28) with multipliers $\mu_{FC}\left(\theta,s\right)$, the limited liability constraints (29) and (30) with multipliers $\mu_{LL}\left(\theta,s,R\right)$, the feasibility constraints on margins (5) with $\mu_0\left(s\right)$ for $\alpha\left(s\right)\geq 0$ and $\mu_1\left(s\right)$ for $\alpha\left(s\right)\leq 1$, the incentive compatibility constraints (31) and (32) with multipliers $\mu_{IC}(s)$ and the participation constraint (33) with multiplier μ.

The first-order conditions of the Lagrangian with respect to $\tau^B(\theta,s)$ are

$$\text{prob}[\theta,s]u'(\theta+\tau^B(\theta,s))-\mu_{FC}\left(\theta,s\right)=0 \qquad \forall(\theta,s) \tag{66}$$

The first-order conditions of the Lagrangian with respect to $\tau^S(\theta,\bar{s},R)$, $\tau^S(\theta,\underline{s},R)$ and $\tau^S(\theta,\underline{s},0)$ are

$$\begin{aligned}
\mu\text{prob}[\theta,\bar{s}]+\mu_{LL}(\theta,\bar{s},R)+\text{prob}[\theta|\bar{s}]\mu_{IC}(\bar{s})-\mu_{FC}\left(\theta,\bar{s}\right) &= 0 \qquad \forall(\theta,\bar{s},R) &(67)\\
\mu\text{prob}[\theta,\underline{s}]+\frac{\mu_{LL}(\theta,\underline{s},R)}{p}-\text{prob}[\theta|\underline{s}]\frac{\mu_{IC}(\underline{s})}{p}-\mu_{FC}\left(\theta,\underline{s}\right) &= 0 \qquad \forall(\theta,\underline{s},R) &(68)\\
\mu\text{prob}[\theta,\underline{s}]+\frac{\mu_{LL}(\theta,\underline{s},0)}{1-p}+\text{prob}[\theta|\underline{s}]\frac{\mu_{IC}(\underline{s})}{1-p}-\mu_{FC}\left(\theta,\underline{s}\right) &= 0 \qquad \forall(\theta,\underline{s},0) &(69)
\end{aligned}$$

Since marginal utilities are positive, it follows from (66) that $\mu_{FC}\left(\theta,s\right)>0$ and hence the feasibility constraints (27) and (28) bind.

Using (66) to substitute for $\mu_{FC}\left(\theta,s\right)$ in (67)-(69) and rearranging, we obtain

$$\begin{aligned}
u'(\theta+\tau^B(\theta,\bar{s})) &= \mu+\frac{\mu_{LL}(\theta,\bar{s},R)}{\text{prob}[\theta,\bar{s}]}+\frac{\mu_{IC}(\bar{s})}{\text{prob}[\bar{s}]} \qquad \forall(\theta,\bar{s},R) &(70)\\
u'(\theta+\tau^B(\theta,\underline{s})) &= \mu+\frac{\mu_{LL}(\theta,\underline{s},R)}{p\text{prob}[\theta,\underline{s}]}-\frac{\mu_{IC}(\underline{s})}{p\text{prob}[\underline{s}]} \qquad \forall(\theta,\underline{s},R) &(71)\\
u'(\theta+\tau^B(\theta,\underline{s})) &= \mu+\frac{\mu_{LL}(\theta,\underline{s},0)}{(1-p)\,\text{prob}[\theta,\underline{s}]}+\frac{\mu_{IC}(\underline{s})}{(1-p)\,\text{prob}[\underline{s}]} \qquad \forall(\theta,\underline{s},0) &(72)
\end{aligned}$$

where we used that $\text{prob}[\theta|s]\text{prob}[s]=\text{prob}[\theta,s]$.

Combining (71) and (72) yields

$$(1-p)\,\mu_{LL}(\theta,\underline{s},R)-p\mu_{LL}(\theta,\underline{s},0)=\text{prob}[\theta|\underline{s}]\mu_{IC}(\underline{s}) \qquad \forall(\theta,\underline{s}) \tag{73}$$

We next show that the limited liability constraint in state $(\bar{\theta},\bar{s},R)$ is slack. The proof proceeds in two steps. First, we show that the limited liability constraints cannot bind for both the state $(\bar{\theta},\bar{s},R)$ and the state $(\underline{\theta},\bar{s},R)$. Suppose not. Since both limited liability constraints after the signal $\bar{s}$ bind, we have $-\tau^S(\bar{\theta},\bar{s},R)=\alpha(\bar{s})A+(1-\alpha(\bar{s}))AR$ and $-\tau^S(\underline{\theta},\bar{s},R)=\alpha(\bar{s})A+(1-\alpha(\bar{s}))AR$. Hence, $E[-\tau^S(\theta,\bar{s},R)]=\alpha(\bar{s})A+(1-\alpha(\bar{s}))AR$. But since $R>\mathcal{P}$, this violates the incentive compatibility constraint (31) after the good signal. Hence, at least one limited liability constraint after the signal $\bar{s}$ must be slack.

Second, we show that the limited liability constraint in state $(\bar{\theta},\bar{s},R)$ is always slack. Suppose not, so that $-\tau^S(\bar{\theta},\bar{s},R)=\alpha(\bar{s})A+(1-\alpha(\bar{s}))AR$. We have just shown that at least one limited liability constraint after the signal $\bar{s}$ must be slack. Hence, we must have that $-\tau^S(\underline{\theta},\bar{s},R)<\alpha(\bar{s})A+(1-\alpha(\bar{s}))AR$ and $\mu_{LL}(\underline{\theta},\bar{s},R)=0$. Using the binding feasibility constraints (27), we have $\tau^B(\bar{\theta},\bar{s},R)>\tau^B(\underline{\theta},\bar{s},R)$, which implies

$$u'(\bar{\theta}+\tau^B(\bar{\theta},\bar{s},R))<u'(\underline{\theta}+\tau^B(\underline{\theta},\bar{s},R))$$

since $\bar{\theta}>\underline{\theta}$. However, using $\mu_{LL}(\underline{\theta},\bar{s},R)=0$ in (70) implies

$$u'(\bar{\theta}+\tau^B(\bar{\theta},\bar{s},R))\geq u'(\underline{\theta}+\tau^B(\underline{\theta},\bar{s},R)).$$

A contradiction. Hence, the limited liability constraint is slack in state $(\bar{\theta},\bar{s},R)$ and $\mu_{LL}(\bar{\theta},\bar{s},R)=0$.

Third, we show by contradiction that $\mu>0$ and the participation constraint (33) binds. Suppose not, i.e. $\mu=0$. Using $\mu=0$ in (71), it follows that $\mu_{LL}(\theta,\underline{s},R)>0$ must hold for $\theta=\bar{\theta},\underline{\theta}$. Using $\mu=0$ and $\mu_{LL}(\bar{\theta},\bar{s},R)=0$ (just shown above) in (70), it follows that $\mu_{IC}(\bar{s})>0$ and the incentive constraint in state $\bar{s}$ binds. Now, there are two possibilities in state $\underline{s}$: either the incentive constraint binds or it is slack.

Consider first the case when the incentive constraint in state $\underline{s}$ binds. Using the binding limited liability constraints in states $(\bar{\theta},\underline{s},R)$ and $(\underline{\theta},\underline{s},R)$ in the incentive constraint in state $\underline{s}$, we get

$$(1-\alpha\,(\underline{s}))A\mathcal{P}=\alpha(\underline{s})A+(1-\alpha(\underline{s}))AR+\underline{\pi}\tau^S(\bar{\theta},\underline{s},0)+(1-\underline{\pi})\tau^S(\underline{\theta},\underline{s},0)$$

or, equivalently,

$$\alpha(\underline{s})A + (1 - \alpha\,(\underline{s}))A\,(R - \mathcal{P}) = -\underline{\pi}\tau^S(\bar{\theta}, \underline{s}, 0) - (1 - \underline{\pi})\tau^S(\underline{\theta}, \underline{s}, 0) \tag{74}$$

If the limited liability constraints (30) are slack, we have $-\tau^S(\bar{\theta}, \underline{s}, 0) < \alpha(\underline{s})A$ and $-\tau^S(\underline{\theta}, \underline{s}, 0) < \alpha(\underline{s})A$ so that the right-hand side of (74) is strictly smaller than $\alpha(\underline{s})A$. Since $(1-\alpha\,(\underline{s}))A\,(R - \mathcal{P}) \geq 0$, the left-hand side of (74) is greater or equal to $\alpha(\underline{s})A$. A contradiction. If the limited liability constraints (30) are binding, then all limited liability constraints in state $\underline{s}$ bind. Using the binding limited liability constraints in state $\underline{s}$ and the binding incentive constraint in state $\bar{s}$ in the (weakly slack) participation constraint (33), we get

$$\begin{aligned}
&\text{prob}[\bar{s}]\alpha\,(\bar{s})\,A\,(R - C - 1) + \text{prob}[\underline{s}]\alpha\,(\underline{s})\,A(pR - 1) + \text{prob}[\underline{s}](1 - p)A\mathcal{P} \\
&\leq -\text{prob}[\bar{s}]\,(\alpha(\bar{s})A + (1 - \alpha(\bar{s}))A\mathcal{P}) - \text{prob}[\underline{s}](p\,(\alpha(\underline{s})A + (1 - \alpha(\underline{s}))AR) + (1 - p)\alpha(\underline{s})A)
\end{aligned}$$

Simplifying yields

$$\text{prob}[\bar{s}]\,[\alpha\,(\bar{s})\,A\,(R - C) + (1 - \alpha(\bar{s}))A\mathcal{P}] + \text{prob}[\underline{s}]A\,[(1 - p)\mathcal{P} + pR] \leq 0 \tag{75}$$

Since both terms on the right-hand side of (75) are strictly positive, we have a contradiction.

Now consider the case when the incentive constraint in state $\underline{s}$ is slack so that $\mu_{IC}(\underline{s}) = 0$. Since $\mu_{LL}(\bar{\theta}, \underline{s}, R) > 0$ and $\mu_{LL}(\underline{\theta}, \underline{s}, R) > 0$, using $\mu_{IC}(\underline{s}) = 0$ in (73) implies that $\mu_{LL}(\bar{\theta}, \underline{s}, 0) > 0$ and $\mu_{LL}(\underline{\theta}, \underline{s}, 0) > 0$ must hold. Hence, all limited liability constraints in state $\underline{s}$ bind. But we have just shown in the previous step that this is incompatible with the weakly slack participation constraint. A contradiction.

We conclude that $\mu > 0$ and the participation constraint must bind.

Fourth, we show that $\mu_{LL}(\bar{\theta}, \underline{s}, R) = 0$ and $-\tau^S(\bar{\theta}, \underline{s}, R) \leq \alpha(\underline{s})A + (1 - \alpha(\underline{s}))AR$. The proof proceeds in two steps. First, we show that it cannot be that both $\mu_{LL}(\bar{\theta}, \underline{s}, R) > 0$ and $\mu_{LL}(\underline{\theta}, \underline{s}, R) > 0$. Suppose not. When both $\mu_{LL}(\bar{\theta}, \underline{s}, R) > 0$ and $\mu_{LL}(\underline{\theta}, \underline{s}, R) > 0$, then

$$-\tau^S(\bar{\theta}, \underline{s}, R) = -\tau^S(\underline{\theta}, \underline{s}, R) = \alpha(\underline{s})A + (1 - \alpha(\underline{s}))AR \tag{76}$$

Using (76) in the incentive constraint after a bad signal (32) yields

$$-E[\tau^S(\theta, \underline{s}, 0)] + (1 - \alpha\,(\underline{s}))A\mathcal{P} < \alpha(\underline{s})A + (1 - \alpha(\underline{s}))AR$$

since $-E[\tau^S(\theta,\underline{s},0)] \leq \alpha(\underline{s})A$ and $\mathcal{P} < R$. Hence, the incentive constraint after a bad signal is slack and $\mu_{IC}(\underline{s}) = 0$. Since $\mu_{LL}(\bar{\theta},\underline{s},R) > 0$ and $\mu_{LL}(\underline{\theta},\underline{s},R) > 0$, using $\mu_{IC}(\underline{s}) = 0$ in (73) implies that $\mu_{LL}(\bar{\theta},\underline{s},0) > 0$ and $\mu_{LL}(\underline{\theta},\underline{s},0) > 0$ must hold. Hence, all limited liability constraints in state $\underline{s}$ bind. Using the binding limited liability constraints in state $\underline{s}$ in the binding participation constraint (33), we get

$$\begin{aligned}&\text{prob}[\bar{s}]\alpha\left(\bar{s}\right)A\left(R-C-1\right)+\text{prob}[\underline{s}]\alpha\left(\underline{s}\right)A(pR-1)+\text{prob}[\underline{s}](1-p)A\mathcal{P}\\ &\qquad=\text{prob}[\bar{s}]E[\tau^S(\theta,\bar{s},R)]-\text{prob}[\underline{s}](p\left(\alpha(\underline{s})A+(1-\alpha(\underline{s}))AR\right)+(1-p)\alpha(\underline{s})A)\end{aligned}$$

Simplifying yields

$$\text{prob}[\bar{s}]\alpha\left(\bar{s}\right)A\left(R-C-1\right)+\text{prob}[\underline{s}]ApR+\text{prob}[\underline{s}](1-p)A\mathcal{P}=\text{prob}[\bar{s}]E[\tau^S(\theta,\bar{s},R)] \quad (77)$$

For equation (77) to hold, it must be that $E[\tau^S(\theta,\bar{s},R)] > 0$. By the binding feasibility constraint (27), this is equivalent to $E[\tau^B(\theta,\bar{s},R)] < 0$. There can be two cases: either the incentive constraint after a good signal binds or it is slack. First, consider the case when the incentive constraint after a good signal binds. Then, $E[\tau^S(\theta,\bar{s},R)] = -\left(\alpha(\bar{s})A+(1-\alpha(\bar{s}))A\mathcal{P}\right) < 0$. A contradiction with (77). Second, consider the case when the incentive constraint after a good signal is slack. Then, $\mu_{IC}(\bar{s}) = 0$. Using $\mu_{LL}(\theta,\bar{s},R) = 0$ and $\mu_{IC}(\bar{s}) = 0$ in (70) and $\mu_{LL}(\bar{\theta},\underline{s},R) > 0$ and $\mu_{IC}(\underline{s}) = 0$ in (71), we have

$$u'(\bar{\theta}+\tau^B(\bar{\theta},\bar{s})) < u'(\bar{\theta}+\tau^B(\bar{\theta},\underline{s}))$$

implying that $\tau^B(\bar{\theta},\bar{s}) > \tau^B(\bar{\theta},\underline{s})$. So, we have:

$$\begin{aligned}\tau^B(\bar{\theta},\bar{s}) &> \tau^B(\bar{\theta},\underline{s}) = -p\tau^S(\bar{\theta},\underline{s},R)-(1-p)\tau^S(\bar{\theta},\underline{s},0) && \text{[using binding feasibility constraint]}\\ &= p\left[\alpha(\underline{s})A+(1-\alpha(\underline{s}))AR\right]+(1-p)\alpha(\underline{s})A && \text{[using binding LL constraints in state } \underline{s}]\\ &= \alpha(\underline{s})A+p(1-\alpha(\underline{s}))AR > 0 && (78)\end{aligned}$$

Now, there are two cases to consider: either the limited liability constraint in state $(\underline{\theta},\bar{s},R)$ binds or it is slack. If it binds, then $\tau^B(\underline{\theta},\bar{s}) = \alpha(\bar{s})A+(1-\alpha(\bar{s}))AR > 0$. Together with (78), this implies that $E[\tau^B(\theta,\bar{s},R)] > 0$, a contradiction with (77). If the limited liability constraint in state $(\underline{\theta},\bar{s},R)$ is slack, then $\mu(\underline{\theta},\bar{s},R) = 0$. Then, there is full risk-sharing after a good signal, $\bar{\theta}+\tau^B(\bar{\theta},\bar{s},R) = \underline{\theta}+\tau^B(\underline{\theta},\bar{s},R)$, and $\tau^B(\underline{\theta},\bar{s},R) = \tau^B(\bar{\theta},\bar{s},R)+\Delta\theta > 0$.

Together with (78), this implies that $E[\tau^B(\theta,\bar{s},R)] = -E[\tau^S(\theta,\bar{s},R)] > 0$, a contradiction with (77).

Hence, we showed that at least one of the $\mu_{LL}(\theta,\underline{s},R)$'s must be zero. We now show that it is μ_{LL} in state $(\bar{\theta},\underline{s},R)$. Suppose not, i.e., $\mu_{LL}(\bar{\theta},\underline{s},R) > 0$ and $\mu_{LL}(\underline{\theta},\underline{s},R) = 0$. Using $\mu_{LL}(\underline{\theta},\underline{s},R) = 0$ in (73), it follows that $\mu_{LL}(\underline{\theta},\underline{s},0) = 0$ and $\mu_{IC}(\underline{s}) = 0$. Using $\mu_{LL}(\bar{\theta},\underline{s},R) > 0$ and $\mu_{IC}(\underline{s}) = 0$ in (73), it follows that $\mu_{LL}(\bar{\theta},\underline{s},0) > 0$. Hence,

$$\tau^B(\bar{\theta},\underline{s}) = p\left(\alpha(\underline{s})A + (1-\alpha(\underline{s}))AR\right) + (1-p)\alpha(\underline{s})A \tag{79}$$

Using $\mu_{LL}(\bar{\theta},\underline{s},R) > 0$ and $\mu_{LL}(\underline{\theta},\underline{s},R) = 0$ in (71), we have $u'(\bar{\theta}+\tau^B(\bar{\theta},\underline{s})) > u'(\underline{\theta}+\tau^B(\underline{\theta},\underline{s}))$, implying that $\bar{\theta}+\tau^B(\bar{\theta},\underline{s}) < \underline{\theta}+\tau^B(\underline{\theta},\underline{s})$. Since $\bar{\theta} > \underline{\theta}$, this means that

$$\tau^B(\bar{\theta},\underline{s}) < \tau^B(\underline{\theta},\underline{s}) \tag{80}$$

must hold. However, we also have that

$$\begin{aligned}
\tau^B(\underline{\theta},\underline{s}) &= -p\tau^S(\underline{\theta},\underline{s},R) - (1-p)\tau^S(\underline{\theta},\underline{s},0) \qquad \text{[using binding feasibility constraint]} \\
&\leq p\left(\alpha(\underline{s})A + (1-\alpha(\underline{s}))AR\right) + (1-p)\alpha(\underline{s})A \qquad \text{[using limited liability constraints]} \\
&= \tau^B(\bar{\theta},\underline{s}) \qquad \text{[using (79)]}
\end{aligned}$$

which contradicts (80). Hence, we must have that $\mu_{LL}(\bar{\theta},\underline{s},R) = 0$.

Fifth, we claim that $\mu_{LL}(\bar{\theta},\underline{s},0) = 0$ and $\mu_{IC}(\underline{s}) = 0$. This claim follows immediately from substituting $\mu_{LL}(\bar{\theta},\underline{s},R) = 0$ in (73).

QED

Proof of Proposition 9

The first-order conditions of the Lagrangian from the proof of Proposition 8 with respect to $\alpha(\bar{s})$ and $\alpha(\underline{s})$ are

$$\begin{aligned}
\frac{\mu_0(\bar{s}) - \mu_1(\bar{s})}{A} + \mu_{IC}(\bar{s})(1-\mathcal{P}) &= \mu\text{prob}[\bar{s}](R-C-1) + (R-1)\mu_{LL}(\underline{\theta},\bar{s},R) \qquad (81) \\
\mu_{LL}(\underline{\theta},\underline{s},0) + \frac{\mu_0(\underline{s}) - \mu_1(\underline{s})}{A} &= \mu\text{prob}[\underline{s}](pR-1) + (R-1)\mu_{LL}(\underline{\theta},\underline{s},R) \qquad (82)
\end{aligned}$$

where we have used $\mu_{LL}(\bar{\theta},\bar{s},R) = 0$, $\mu_{LL}(\bar{\theta},\underline{s},R) = 0$, $\mu_{LL}(\bar{\theta},\underline{s},0) = 0$ and $\mu_{IC}(\underline{s}) = 0$ (all shown in the previous Proposition).

Consider first state $\bar{s}$. The right-hand side of (81) is strictly positive since $R - C > 1$ and $\mu > 0$ (see Proposition 8). If the incentive constraint is slack after a good signal, then $\mu_{IC}(\bar{s}) = 0$, implying that $\mu_0(\bar{s}) > 0$ must hold and $\alpha^*(\bar{s}) = 0$. Similarly, if $\mathcal{P} \geq 1$, then $\mu_0(\bar{s}) > 0$ must hold and $\alpha^*(\bar{s}) = 0$.

Consider now state $\underline{s}$. Using $\mu_{IC}(\underline{s}) = 0$ (as shown in the previous Proposition) in (73) yields

$$\mu_{LL}(\underline{\theta}, \underline{s}, 0) = \frac{1-p}{p}\mu_{LL}(\underline{\theta}, \underline{s}, R)$$

Substituting for $\mu_{LL}(\underline{\theta}, \underline{s}, 0)$ in (82) yields

$$\frac{\mu_0(\underline{s}) - \mu_1(\underline{s})}{A} = (pR - 1)\left[\mu\text{prob}[\underline{s}] + \frac{\mu_{LL}(\underline{\theta}, \underline{s}, R)}{p}\right] \tag{83}$$

If $pR \geq 1$, then the right-hand side of (83) is non-negative, implying that $\mu_0(\underline{s}) \geq 0$ and $\alpha^*(\underline{s}) = 0$. If $pR < 1$, then the right-hand side of (83) is negative, implying that $\mu_1(\underline{s})) > 0$ and $\alpha^*(\underline{s}) = 1$. We now claim that the contract with risk-taking and $\alpha^*(\underline{s}) = 1$ is dominated by the contract with effort after a bad signal. Note that $\alpha(\underline{s}) = 1$ is also feasible under the contract with effort. However, it is never chosen (Proposition 6), implying that the optimal contract with effort is strictly preferred to the contract with risk-taking and $\alpha(\underline{s}) = 1$.

QED

Proof of Proposition 10

The optimal transfers follow from asserting full risk-sharing across all states and using the binding participation constraint. Condition (35) follows from checking that all limited liability constraints are satisfied for these transfers. It remains to check that, in the proposed contract, the incentive constraint after a good signal is slack and margins are not used. Using $\alpha(\bar{s}) = 0$ and the transfers in state $\bar{s}$ in the incentive constraint (31) we have:

$$A\mathcal{P} > 0 > -(\bar{\pi} - \pi)\Delta\theta - \text{prob}[\underline{s}](1-p)A\mathcal{P} = E[\tau^B(\theta, \bar{s}, R)] = -E[\tau^S(\theta, \bar{s}, R)]$$

so that the incentive constraint after $\bar{s}$ is indeed slack at $\alpha(\bar{s}) = 0$. Since $pR \geq 1$, it is not optimal to use margins after a bad signal either (Proposition **9**).

QED

Proof of Proposition 11

We first show that for $p < \max\left\{\frac{R-C-1}{R-1}, \frac{1}{R}\right\}$ the contract with effort is optimal. First, consider $p \leq \frac{R-C-1}{R-1}$. In this case, we have that $\mathcal{P} \geq 1$. Combining with condition (36) yields

$A\mathcal{P} \geq A \geq \pi\Delta\theta > (\pi - \underline{\pi})\Delta\theta$. By Lemma 1, the first-best (which entails effort) is reached. Second, consider $p < \frac{1}{R}$. By Proposition 9, the contract with effort strictly dominates the contract with risk-taking in this case.

We now consider the case when $p \geq \max\left\{\frac{R-C-1}{R-1}, \frac{1}{R}\right\}$. Note that p must always be lower than $\frac{R-C}{R}$ since we require that $\mathcal{P} > 0$.

We now show that the expected utility of the contract with effort is decreasing in p. Consider first the case when the limited liability constraint in state $(\underline{\theta}, \underline{s})$ is slack. Then, there is full risk-sharing conditional on the signal and, using Proposition 5, the expected utility of the protection buyer under effort is given by

$$\text{prob}[\bar{s}]u\left(E[\tilde{\theta}|\bar{s}] - \frac{\text{prob}[\underline{s}]A\left[\alpha^*(\underline{s})\left(R-C\right) + (1-\alpha^*(\underline{s}))\mathcal{P}\right]}{\text{prob}[\bar{s}]}\right) + \text{prob}[\underline{s}]u\left(E[\tilde{\theta}|\underline{s}] + A\left[\alpha^*(\underline{s}) + (1-\alpha^*(\underline{s}))\mathcal{P}\right]\right)$$

The derivative of the expected utility with respect to p is given by

$$-\text{prob}[\bar{s}]u'\left(E[\tilde{\theta}|\bar{s}] - \frac{\text{prob}[\underline{s}]A\left[\alpha^*(\underline{s})\left(R-C\right) + (1-\alpha^*(\underline{s}))\mathcal{P}\right]}{\text{prob}[\bar{s}]}\right)\frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}A(1-\alpha^*(\underline{s}))\frac{\partial\mathcal{P}}{\partial p} + \text{prob}[\underline{s}]u'\left(E[\tilde{\theta}|\underline{s}] + A\left[\alpha^*(\underline{s}) + (1-\alpha^*(\underline{s}))\mathcal{P}\right]\right)A(1-\alpha^*(\underline{s}))\frac{\partial\mathcal{P}}{\partial p} = \text{prob}[\underline{s}]A(1-\alpha^*(\underline{s}))\frac{\partial\mathcal{P}}{\partial p}\times \left[u'\left(E[\tilde{\theta}|\underline{s}] + A\left[\alpha^*(\underline{s}) + (1-\alpha^*(\underline{s}))\mathcal{P}\right]\right) - u'\left(E[\tilde{\theta}|\bar{s}] - \frac{\text{prob}[\underline{s}]A\left[\alpha^*(\underline{s})\left(R-C\right) + (1-\alpha^*(\underline{s}))\mathcal{P}\right]}{\text{prob}[\bar{s}]}\right)\right]$$

where we have used the envelope theorem to claim $\frac{\partial\alpha^*(\underline{s})}{\partial p} = 0$. We know that $1-\alpha^*(\underline{s}) > 0$ since $\alpha^*(\underline{s}) < 1$ (Proposition 6). Due to the binding incentive constraint after a bad signal (Proposition 4), the protection buyer's consumption is larger after a good signal than after a bad signal implying that the term in the square brackets above is positive. Since $\mathcal{P} = R - \frac{C}{1-p}$, we have $\frac{\partial\mathcal{P}}{\partial p} < 0$ implying that the expected utility under effort decreases in p when the limited liability constraint in state $(\underline{\theta}, \underline{s})$ is slack.

Now consider the other possibility, i.e., that the limited liability constraint in state $(\underline{\theta}, \underline{s})$ is binding. Then, there is still full risk-sharing conditional on a good signal but there is no longer full risk-sharing conditional on a bad signal. Using Proposition 5, the expected utility of the protection buyer is given by

$$\text{prob}[\bar{s}]u\left(E[\tilde{\theta}|\bar{s}] - \frac{\text{prob}[\underline{s}]A\left[\alpha^*(\underline{s})\left(R-C\right)+(1-\alpha^*(\underline{s}))\mathcal{P}\right]}{\text{prob}[\bar{s}]}\right)+$$
$$\pi\left(1-\lambda\right)u\left(\bar{\theta}+\alpha^*(\underline{s})A-(1-\alpha^*(\underline{s}))A\frac{(1-\underline{\pi})R-\mathcal{P}}{\underline{\pi}}\right)+(1-\pi)\,\lambda u\left(\underline{\theta}+\alpha^*(\underline{s})A+(1-\alpha^*(\underline{s}))AR\right)$$

The derivative of the expected utility with respect to p is given by

$$-\,\text{prob}[\bar{s}]u'\left(E[\tilde{\theta}|\bar{s}] - \frac{\text{prob}[\underline{s}]A\left[\alpha^*(\underline{s})\left(R-C\right)+(1-\alpha^*(\underline{s}))\mathcal{P}\right]}{\text{prob}[\bar{s}]}\right)\frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}A(1-\alpha^*(\underline{s}))\frac{\partial\mathcal{P}}{\partial p}+$$
$$\frac{\pi\left(1-\lambda\right)}{\underline{\pi}}u'\left(\bar{\theta}+\alpha^*(\underline{s})A-(1-\alpha^*(\underline{s}))A\frac{(1-\underline{\pi})R-\mathcal{P}}{\underline{\pi}}\right)A(1-\alpha^*(\underline{s}))\frac{\partial\mathcal{P}}{\partial p}=\text{prob}[\underline{s}]A(1-\alpha^*(\underline{s}))\frac{\partial\mathcal{P}}{\partial p}\times$$
$$\left[u'\left(\bar{\theta}+\alpha^*(\underline{s})A-(1-\alpha^*(\underline{s}))A\frac{(1-\underline{\pi})R-\mathcal{P}}{\underline{\pi}}\right)-u'\left(E[\tilde{\theta}|\bar{s}] - \frac{\text{prob}[\underline{s}]A\left[\alpha^*(\underline{s})\left(R-C\right)+(1-\alpha^*(\underline{s}))\mathcal{P}\right]}{\text{prob}[\bar{s}]}\right)\right]$$

where we used $\frac{\pi(1-\lambda)}{\underline{\pi}}$ =prob$[\underline{s}]$ and we again made use of the envelope theorem to claim $\frac{\partial\alpha^*(\underline{s})}{\partial p}=0$. Since $\alpha^*(\underline{s})<1$ (Proposition 6), $1-\alpha^*(\underline{s})>0$. Using (56), and the fact that the limited liability constraints in states $(\bar{\theta},\bar{s})$ and $(\bar{\theta},\underline{s})$ are always slack (Proposition 2) and the incentive constraint after a bad signal binds (Proposition 4), we have that $u'\big(\bar{\theta}+\tau(\bar{\theta},\underline{s})\big)>u'\big(\bar{\theta}+\tau(\bar{\theta},\bar{s})\big)$ or, equivalently, that the term in the square brackets above is positive. Since $\frac{\partial\mathcal{P}}{\partial p}<0$, the expected utility under effort decreases in p when the limited liability constraint in state $(\underline{\theta},\underline{s})$ is binding.

We now show that the expected utility of the contract with risk-taking is increasing in p. Under risk-taking, the consumption of the protection buyer is equalized across all states. Therefore, using the optimal transfers from Proposition 10 in (26), the expected utility of the protection buyer under no effort is given by: $u\left(E[\tilde{\theta}]-\text{prob}[\underline{s}](1-p)A\mathcal{P}\right)$. Using $(1-p)A\mathcal{P}=R-C-pR$, we have that the derivative of the expected utility with respect to p is given by

$$\text{prob}[\underline{s}]ARu'\Big(E[\tilde{\theta}]-\text{prob}[\underline{s}](1-p)A\mathcal{P}\Big)>0$$

Lastly, note that as $p\to\frac{R-C}{R}$ (or, equivalently, as $\mathcal{P}\to 0$), the expected utility under risk-taking is strictly higher than the expected utility under effort. This is because the expected utility under risk-taking is approaching $u\left(E[\tilde{\theta}]\right)$, which is the first-best level of utility, while the expected utility under effort is strictly smaller than the first-best level of utility since $A\mathcal{P} < (\pi - \underline{\pi})\Delta\theta$ and hence it is not possible to reach the first-best with effort after bad news (Lemma 1).

In sum, for $p < \max\left\{\frac{R-C-1}{R-1}, \frac{1}{R}\right\}$, the contract with effort is optimal. For $p \to \frac{R-C}{R}$, the contract with risk-taking is optimal. For $\max\left\{\frac{R-C-1}{R-1}, \frac{1}{R}\right\} \leq p < \frac{R-C}{R}$, the expected utility under effort is decreasing in p while the expected utility under risk-taking is increasing in p. Therefore, there exists a threshold value of p, denoted by $\hat{p}$, such that effort after bad news is optimal if and only if $p \leq \hat{p}$.

QED

Proof of Proposition 12

With protection seller effort, there is full risk-sharing conditional on the realization of the signal $\tilde{s}$ and we can write the objective function (6) as

$$U = \text{prob}[\bar{s}]u(E[\tilde{\theta} + \tau^B(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}]) + \text{prob}[\underline{s}]u(E[\tilde{\theta} + \tau^B(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}]). \tag{84}$$

Using the binding incentive and participation constraints, equations (15) and (16) express the expected transfer to protection buyers conditional on the signal, $E[\tau^B(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}]$ and $E[\tau^B(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}]$, as a function of the margin $\alpha(\underline{s})$ (recall that there is no margin call after a good signal). Writing the problem in terms of the expected transfers after a signal simplifies the exposition of the proof.

The first partial derivative of the objective function with respect to the margin is (for notational ease, we drop the reference to the $\underline{s}$ in $\alpha(\underline{s})$):

$$\frac{\partial U}{\partial \alpha} = \text{prob}[\bar{s}]\frac{\partial E[\tau^B(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}]}{\partial \alpha}\bar{u}' + \text{prob}[\underline{s}]\frac{\partial E[\tau^B(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}]}{\partial \alpha}\underline{u}', \tag{85}$$

where $\underline{u}'$ and $\bar{u}$ denote the marginal utility conditional on the bad and the good signal, respectively. The partial derivative of the expected transfer after a bad signal with respect to the margin is

$$\frac{\partial E[\tau^B(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}]}{\partial \alpha} = A\left[1 - \mathcal{P}(\alpha) + (1-\alpha)\mathcal{P}'(\alpha)\right]. \tag{86}$$

When the derivative is positive, margins relax the incentive constraint. Define

$$X \equiv 1 - \mathcal{P}(\alpha) + (1-\alpha)\,\mathcal{P}'(\alpha) \tag{87}$$

The derivative is positive if and only if $X > 0$. This is condition (46) in the text.

The partial derivative of the expected transfer after a good signal with respect to the margin is

$$\frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]}{\partial \alpha} = -\frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} A\left[(R-C-1)+X\right] \tag{88}$$

The derivative is negative when $X > 0$ since $R - C > 1$ (condition (2)). When $X < 0$, then the derivative may either be positive or negative, depending on how X compares to the opportunity cost of margins, $R - C - 1$.

Combining (86), (87) and (88), we can write (85) as

$$\frac{\partial U}{\partial \alpha} = \text{prob}[\underline{s}] A\bar{u}'\left[\frac{\underline{u}'}{\bar{u}'} - \left(\frac{R-C-1}{X}+1\right)\right] X$$

When $X > 0$ then $\frac{\partial U}{\partial \alpha} = 0$ yields the condition for an optimal interior margin in the proposition (when $X < 0$ then $\frac{\partial U}{\partial \alpha} < 0$ for sure since $\frac{\underline{u}'}{\bar{u}'} \geq 1$). (Note that as in the linear cost case, it may be optimal not to use margins).

We now show that when $\gamma < 0$, then the optimization problem may not be well-behaved. The second partial derivative of the objective function (84) with respect to margins is

$$\begin{aligned}\frac{\partial^2 U}{\partial \alpha^2} &= \text{prob}[\bar{s}]\left[\bar{u}''\left(\frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]}{\partial \alpha}\right)^2 + \bar{u}'\frac{\partial^2 E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]}{\partial \alpha^2}\right] \\ &+ \text{prob}[\underline{s}]\left[\underline{u}''\left(\frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}]}{\partial \alpha}\right)^2 + \underline{u}'\frac{\partial^2 E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}]}{\partial \alpha^2}\right]\end{aligned}$$

The first term in each squared bracket is negative (because of concave utility). A sufficient condition for a local maximum is therefore

$$\text{prob}[\bar{s}]\bar{u}'\frac{\partial^2 E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]}{\partial \alpha^2} + \text{prob}[\underline{s}]\underline{u}'\frac{\partial^2 E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}]}{\partial \alpha^2} \leq 0$$

Using (86), (87) and (88) the condition becomes

$$\text{prob}[\underline{s}] A\frac{\partial X}{\partial \alpha}\left(\underline{u}' - \bar{u}'\right) \leq 0.$$

Since $\underline{u}' - \bar{u}' \geq 0$ (protections buyers may bear signal risk), the sufficient condition holds when $\frac{\partial X}{\partial \alpha} \leq 0$ or, equivalently, when $\gamma \geq 0$. When $\gamma < 0$ we cannot be sure that the first-order condition identifies a local maximum.

Finally, note that when $\gamma \geq 0$ then $1 > R - \frac{c}{1-p}$ is sufficient for $X > 0$ for all α.

QED

Proof of Proposition 13

The first-order condition stipulates $\frac{\partial U(\alpha^*,\gamma)}{\partial \alpha} = 0$ (for simplicity we consider only interior solutions, $\alpha^* \in (0,1)$). After total differentiation of this implicit function we obtain

$$\frac{d\alpha^*}{d\gamma} = -\frac{\frac{\partial^2 U}{\partial\alpha\partial\gamma}}{\frac{\partial^2 U}{\partial\alpha^2}}$$

When α^* is a local maximum, then a more convex cost of effort leads to larger optimal margins, $\frac{d\alpha^*}{d\gamma} > 0$, if and only if $\frac{\partial^2 U}{\partial\alpha\partial\gamma} > 0$. This cross-partial derivative is

$$\begin{aligned}\frac{\partial^2 U}{\partial\alpha\partial\gamma} &= \text{prob}[\bar{s}]\left[\bar{u}''\frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]}{\partial\gamma}\frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]}{\partial\alpha} + \bar{u}'\frac{\partial^2 E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]}{\partial\alpha\partial\gamma}\right] \\ &+ \text{prob}[\underline{s}]\left[\underline{u}''\frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}]}{\partial\gamma}\frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}]}{\partial\alpha} + \underline{u}'\frac{\partial^2 E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}]}{\partial\alpha\partial\gamma}\right]\end{aligned}$$

Using (86), (87) and (88), the cross-partial derivative becomes

$$\begin{aligned}&\frac{\partial^2 U}{\partial\alpha\partial\gamma} = \text{prob}[\underline{s}]A\times \\ &\quad\left[-\bar{u}''\frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]}{\partial\gamma}[(R-C-1)+X] + \underline{u}''\frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}]}{\partial\gamma}X + \frac{\partial X}{\partial\gamma}\left(\underline{u}' - \bar{u}'\right)\right]\end{aligned}$$

Moreover,

$$\begin{aligned}\frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]}{\partial\gamma} &= \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}\frac{(1-\alpha)^2 A^2}{1-p} > 0 \\ \frac{\partial E[\tau^B(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}]}{\partial\gamma} &= -\frac{(1-\alpha)^2 A^2}{1-p} < 0 \\ \frac{\partial X}{\partial\gamma} &= 2\frac{(1-\alpha)A}{1-p} > 0\end{aligned}$$

When $\gamma \geq 0$ then α^* is a local maximum and $R - \frac{c}{1-p} < 1$ is sufficient for $X > 0$. And when $X > 0$, the cross-partial derivative is positive.

QED

Figure 1: Centralized clearing

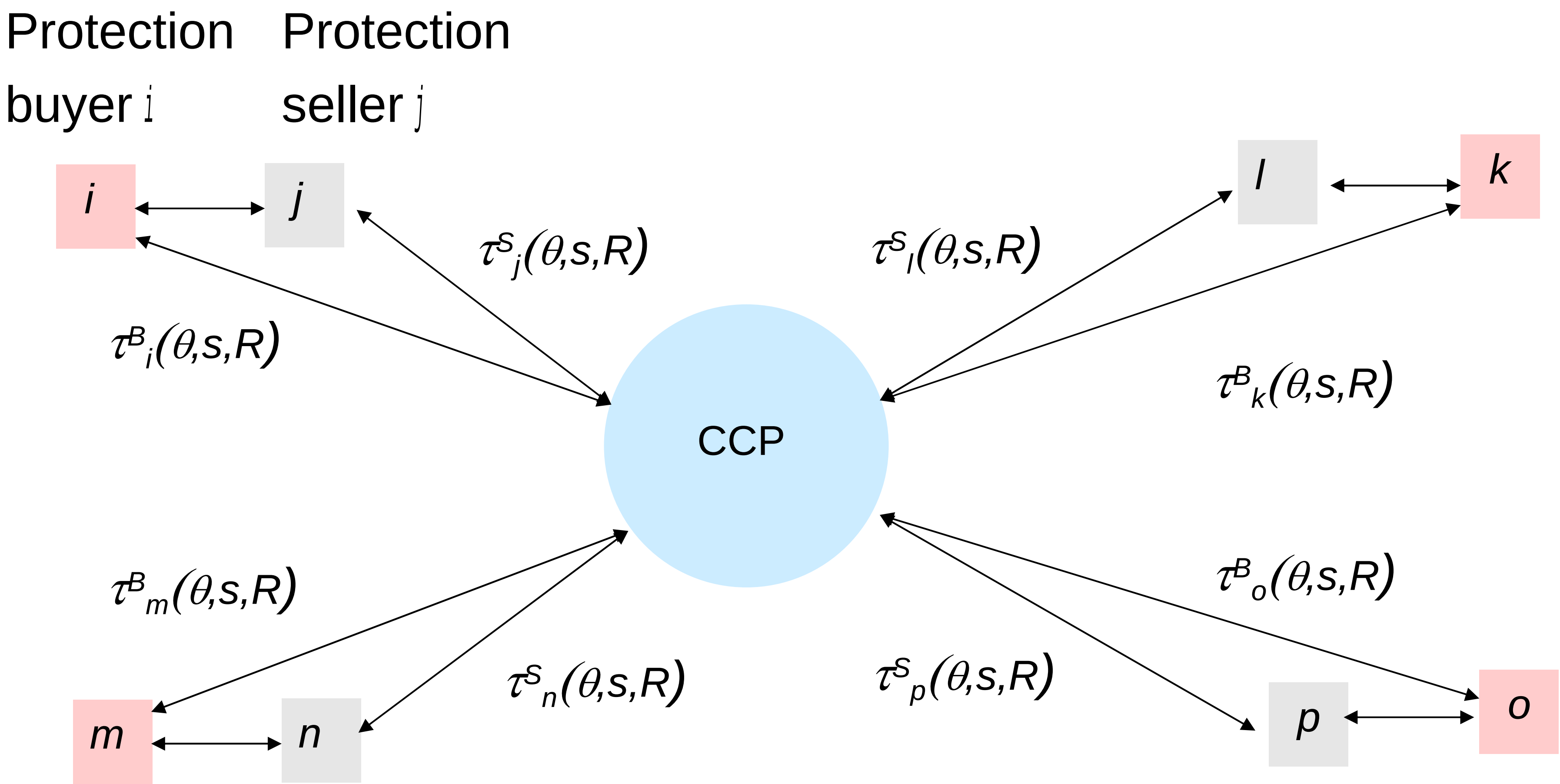

Figure 2: Timing

$t = 0$	$t = 1$	$t = 2$
Optimal CCP design	Signal s about θ observed by all	θ & R_j realize
Participation decisions	Margin $\alpha(s)$	Transfers τ^S τ^B (contingent on all observables: s, all R_j, θ)
	Effort or not	

Figure 3: Optimal margins

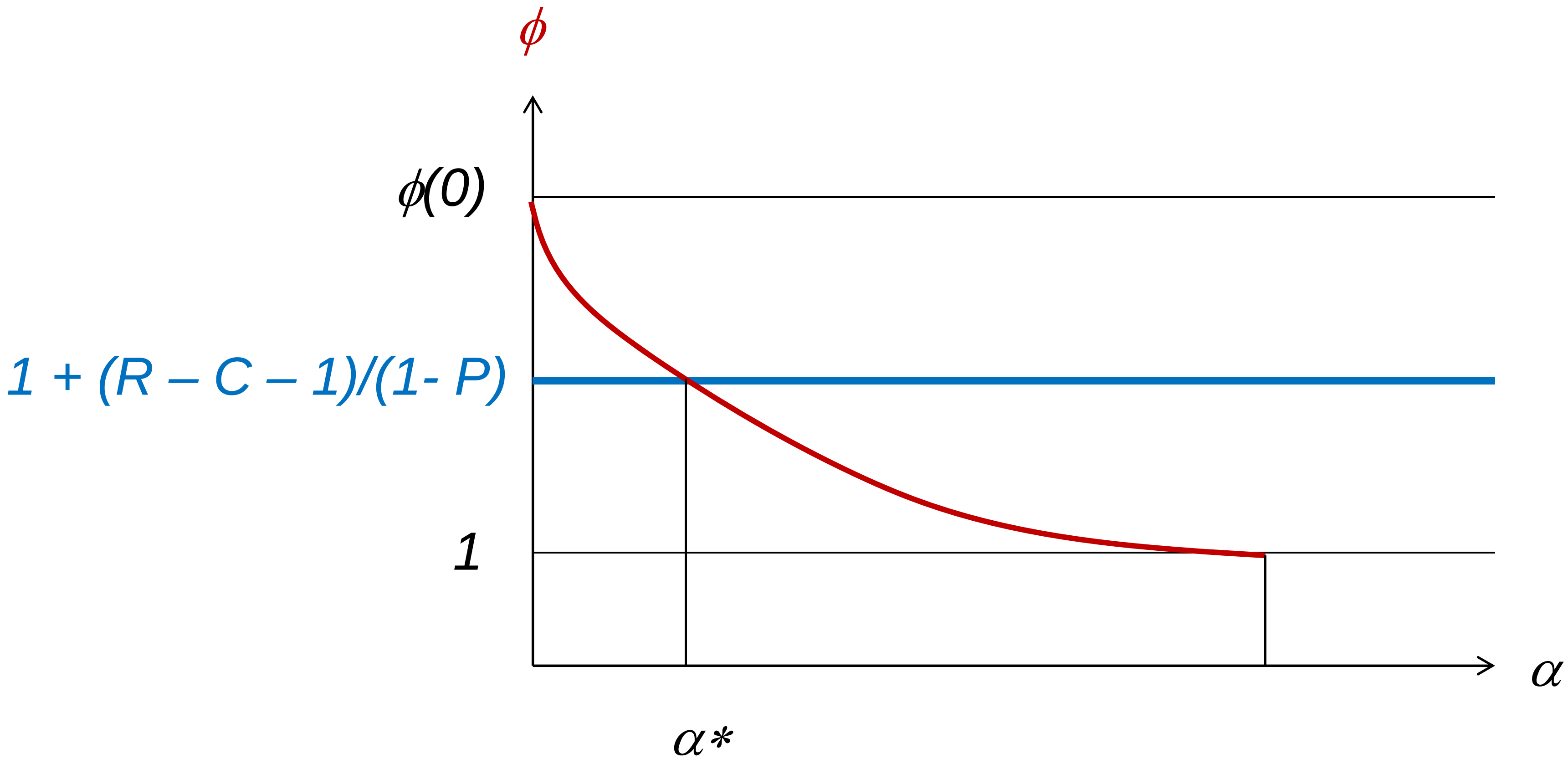

Figure 4: Optimal effort level