

Optimal margins and equilibrium prices ^{*}

Bruno Biais[†] Florian Heider[‡] Marie Hoerova[§]

This version: September 2015

Abstract

We study the interaction between contracting and equilibrium pricing when risk-averse hedgers purchase insurance from risk-neutral investors subject to moral hazard. Moral hazard limits risk-sharing. In the individually optimal contract, margins are called (after bad news) to improve risk-sharing. But margin calls depress the price of investors' assets, affecting other investors negatively. Because of this fire-sale externality, there is too much use of margins in the market equilibrium compared to the utilitarian optimum. Moreover, equilibrium multiplicity can arise: In a pessimistic equilibrium, hedgers who fear low prices request high margins to obtain more insurance. Large margin calls trigger large price drops, confirming initial pessimistic expectations. Finally, moral hazard generates endogenous market incompleteness, raises risk premia, and induces contagion between asset classes.

JEL CLASSIFICATION: G21, G22, D82

KEYWORDS: Insurance; Derivatives; Moral hazard; Risk-management; Margin requirements; Contagion; Fire-sales

*Many thanks for helpful comments to seminar and conference participants at Cass Business School, European School of Management and Technology, Imperial College, Oxford University, University of British Columbia, University of Washington, University of Zurich, the European Central Bank, the Bank for International Settlements, the European Winter Finance Conference, the China International Conference in Finance, the European Summer Symposium in Financial Markets (Gerzensee), the Cambridge Corporate Finance Theory Symposium, the European Finance Association (Vienna), the Paul Woolley Centre Conference (London), and especially to our discussants Shiming Fu, Albert Menkveld and Vish Viswanathan. Jana Urbankova provided excellent editing assistance. The views expressed do not necessarily reflect those of the European Central Bank or the Eurosystem.

[†]Toulouse School of Economics (CNRS-CRM, FBF-IDEI Chair on Investment Banking and Financial Markets), email: bruno.biais@univ-tlse1.fr

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

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

1 Introduction

One of the major functions of financial markets is to enable participants to share risk. For example, financial institutions can purchase protection in derivative markets, such as the CDS market, to hedge the risk of their assets. The effectiveness of such risk-sharing trades, however, can be significantly reduced by counterparty risk.¹

Margin deposits mitigate counterparty risk. Accordingly, the immediate response of regulators and law-makers to the financial crisis was to require a significant expansion of the use of margins in derivative activity (Dodd-Frank Act in the US, EMIR in the EU).

But margin calls can also be destabilizing, as first shown by Gromb and Vayanos (2002) in the context of arbitrage activities.² Because of the margin (or collateral) constraint, an arbitrageur's holding of a risky asset cannot exceed a given fraction of her wealth. As this wealth is evaluated at market prices, the arbitrageur must sell some of the asset if there is a price drop and the margin constraint tightens. The asset sale generates a pecuniary externality. When one arbitrageur sells, this further depresses the price and tightens the margin constraints for other arbitrageurs. As shown in Gromb and Vayanos (2002), because of the pecuniary externality, the market equilibrium is not efficient.

The goal of the present paper is to evaluate the costs and the benefits of margins, taking into account the interaction between the determination of asset prices by market clearing and the determination of margins by optimal contracts. The margin requirement in Gromb and Vayanos (2002) is exogenous and therefore one cannot evaluate the benefit of margins or the design of optimal margins in their setting. To examine both the cost and benefit of margin requirements in the context of risk-sharing and derivatives, we extend Biais, Heider, and Hoerova (2015), where contracts are optimal but prices exogenous, to an equilibrium setting, where asset prices are endogenous, too. In that setting we compare the socially optimal allocation to its laissez-faire counterpart, and analyze the asset pricing implications of endogenous margins.

We consider risk-averse hedgers who want to insure against a common exposure to risk, and risk-neutral investors who offer insurance in a derivative market. Investors have limited

¹For example, when Lehman Brothers filed for bankruptcy in September 2008, it froze the positions of more than 900,000 derivative contracts (Fleming and Sarkar, 2014).

²Gromb and Vayanos (2010) is a survey of the literature on the limits to arbitrage, which includes a simplified version of Gromb and Vayanos (2002). Gromb and Vayanos (2015) extend the analysis of financially constrained arbitrageurs to a dynamic context and several assets.

liability. They can make insurance payments only when their assets are sufficiently valuable. Hence, there is counterparty risk for hedgers who buy insurance from investors. To avoid downside risk to their assets and maintain asset value, investors must exert costly effort. The level of such risk-prevention effort is unobservable by other market participants and in combination with limited liability, this generates moral hazard.

After market participants have entered derivative positions, new information about the insured risk is observed, which changes the expected pay-offs of the contracting parties. While investors initially expected to (at least) break even on selling insurance, they expect to make losses on their derivative position after bad news (and similarly, they expect to make gains after good news). The expected liability after bad news creates a debt-overhang problem and reduces investors' incentive to exert risk-prevention effort. As shown in Biais, Heider, and Hoerova (2015), the optimal contract relies on margin calls after bad news to cope with moral hazard and reduce counterparty risk.

The analysis of the equilibrium implications of such contracts is the novel contribution of this paper. To fulfill margin calls and deposit cash on a margin account, investors must sell some of their assets in the market place. The larger the margin call, the greater the fraction of assets that must be sold, and the lower is the market clearing price for these assets. This gives rise to a fire-sale externality, as in Gromb and Vayanos (2002). As all players maximize expected utility and contracts are optimal, welfare is well-defined, and we can conduct a normative analysis of the consequences of this externality. Because of the pecuniary externality, the information-constrained utilitarian optimum differs from the market equilibrium (see also Greenwald and Stiglitz, 1986). With *laissez-faire*, market participants purchase too much insurance, and correspondingly request excessive margins, because they do not internalize the negative externality they generate. The information-constrained utilitarian optimum can be implemented by limits on derivative positions or on margins.

When the risk-aversion of hedgers increases, they become more eager to purchase insurance. But this eagerness can be self-defeating. After bad news about the insured risk, it is the margin call that leads to incentive-compatible insurance. As a consequence, after bad news the consumption of hedgers increases in the proceeds from margin calls, αp , where α is the fraction of assets sold and p is the market price. For a given margin call α , the proceeds are low when the price p is low, which reduces the consumption of hedgers. Lower consumption increases marginal utility, particularly when hedgers' risk-aversion is high. The

increase in the marginal utility of consumption when the price p goes down can lead to an *increase* of the required margin α . As the supply of the asset in the market place comes from asset sales in response to margin calls, the supply can decrease in price when hedgers' risk-aversion is high. This can give rise to multiple equilibria. In a pessimistic equilibrium hedgers anticipate low prices, request large margins, which, if bad news arrive, triggers large asset sales and lowers prices, confirming the initial expectation of low prices. There also exists an optimistic, high-price equilibrium, which leads to higher welfare.

Key to our analysis is the interaction between optimal contracting and equilibrium pricing. On the one hand, market participants rationally anticipate equilibrium prices when they design privately optimal contracts. On the other hand, the market-clearing price reflects the supply of the asset, which results from margin calls, which are part of the optimal contract. Thus, there is a rational expectations loop with the optimal contract and the equilibrium price as the fixed point (and multiple equilibria arise when there are multiple fixed points).

While the same pecuniary externality generates both equilibrium multiplicity *and* equilibrium suboptimality, the two are conceptually different. Even when the equilibrium is unique, it is suboptimal because of the fire-sales externality and over-margining. When there are multiple equilibria, even the equilibrium with the highest price can be suboptimal, again because of fire-sales and over-margining.

Our model generates several implications about asset pricing. First, consider the price of the risky asset held by hedgers. Without moral hazard, hedgers are fully insured against that risk by investors. Hence, they are locally risk-neutral. In contrast, with moral hazard hedgers still bear some risk in equilibrium. Moral hazard generates market incompleteness. Now suppose such partially-insured hedgers meet in a market where they can trade their risky asset. The equilibrium price in this market reflects their valuation for the asset and the corresponding risk-premium in the presence of incentive-compatible insurance (with derivatives). When the moral hazard problem becomes more severe, risk-sharing becomes more limited, hedgers' valuation of the asset decreases and the risk-premium increases.

Second, consider the correlation between the risky asset held by hedgers and the risky asset initially held by investors. Without moral hazard, the prices of these assets are independent. With moral hazard, they are positively correlated. The arrival of bad news about the hedgers' asset (e.g., the realization of a low price) triggers a margin call for investors, which lowers the price of their asset. The larger the variance of the bad news about the

hedgers' asset (e.g., larger price volatility), the larger the margin call, and, correspondingly, the greater is the price impact on the asset held by investors. That is, the larger the (exogenous) variance of the value of hedgers' assets, the larger the (endogenous) variance of the value of investors' assets.

Third, consider the interaction between the risk-aversion of hedgers and the price of the asset initially held by the investors. When hedgers are moderately risk-averse, margin calls are small and generate only limited contagion and price drops for the investors' asset. As the hedgers' risk-aversion increases, they demand larger margins. At some point, there is a switch from a unique equilibrium to multiple equilibria. If hedgers are pessimistic and coordinate on the low-price equilibrium, this triggers a strong drop in the price of the investors' asset. A relatively small increase in risk-aversion of some market participants can therefore generate a large drop in the price of an asset that they do not hold, which can be interpreted as a crash.

Fourth, consider the link between moral hazard and pricing. The greater the opacity, complexity and difficulty of the risk-prevention task faced by investors, the more severe the moral hazard problem, the greater the need for margins, and the larger are the corresponding price drops. So when financial intermediaries' risk-prevention tasks becomes more complex and fraught with moral hazard, asset markets become more unstable.

Our result that margins can be destabilizing echoes Brunnermeier and Pedersen (2009). In their analysis, however, the channel is different. Large price drops raise estimated volatility, trigger margin calls, which generates asset sales and further price drops. As in Gromb and Vayanos (2002), the margin is exogenously given, and price drops and margin calls feed on each other ("spirals"). In our set-up, margin calls are endogenous. They are part of an optimal contract that also takes into account any price impact from margin-induced asset-sales. Moreover, the destabilizing effect of margins in our set-up does not rely on margins triggering further price declines. The initial trigger is bad news about the hedgers' asset. The ensuing price drop occurs in the market for the investors' asset. This price drop does not feed back into further bad news about the hedgers' asset.

There are a number of papers that relate to different aspects of our analysis. But the common difference to us is that they analyse investment problems (by firms or banks) whereas we focus on risk-sharing in financial markets. Correspondingly, these papers generate implications for leverage and credit rationing, while we generate implications for risk-premia, asset

prices and contagion.³ Lorenzoni (2008) and Hombert (2009) study the pecuniary externality associated with selling collateral. Like us, Acharya and Viswanathan (2011) analyze the interaction between contracting and equilibrium prices. Both in their analysis and ours, the incentive-compatibility constraint states that pledgeable income from assets, including margin deposits, cannot be lower than what must be paid to outsiders. Kuong (2014) also endogenizes contracts in the presence of equilibrium prices and obtains equilibrium multiplicity.

Section 2 describes the model and presents the first-best benchmark. Section 3 analyzes optimal margining under moral hazard, building on Biais, Heider and Hoerova (2015). The key contributions of the present paper are in Section 4 and Section 5. Section 4 derives the market equilibrium and the utilitarian optimum, and compares the two. Section 5 discusses the implications of our analysis for asset pricing and for regulation. 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 hedgers and a mass- $(1 + m)$ continuum of investors. At $t = 0$, investors make investment decisions. They can also contract with hedgers. At $t = 1$, margining and trading decisions are made. At $t = 2$, payoffs are received.

Players and assets. Hedgers are identical, with twice differentiable concave utility function u . Each hedger is initially endowed with one unit of an asset with random return $\tilde{\theta}$ at $t = 2$.⁴ 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 hedgers.⁵

Hedgers can seek insurance against the risk $\tilde{\theta}$ from investors with limited liability. Each investor j is initially endowed with one unit cash. She can either keep her initial cash

³A common theme in these papers is that they tend to rule out the financing of investments using long-term, contingent contracts. In contrast, our insurance contracts are fully-contingent and long-term (as in any derivative, the value of the contract depends on intermediate developments in the underlying asset).

⁴The concavity of the objective function of the hedger can reflect institutional, financial or regulatory constraints, such as leverage constraints or risk-weighted capital requirements.

⁵At 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 hedgers could insure this risk among themselves, without seeking insurance from investors.

endowment, earning 0 rate of return, or invest it in a risky asset returning \tilde{R}_j per unit at $t = 2$. The distribution of \tilde{R}_j is affected by the investor's risk-management decision at $t = 1$. To model risk-management we assume that investor j can undertake costly effort to make her assets safer, as Holmstrom and Tirole (1997). If she undertakes such risk-prevention effort, the per unit return \tilde{R}_j is R with probability one. If she does not exert the risk-prevention effort, then the return is R with probability $\mu < 1$ and zero with probability $1 - \mu$. Because investors' assets are riskier without costly effort, we also refer to the decision not to exert effort as "risk-taking".⁶

Investors differ in the efficiency of their risk-management skills, and, correspondingly in their cost of effort. Specifically, for a mass-one continuum of investors, risk-management effort costs C per unit of assets under management at $t = 1$. We call such investors "sophisticated". The remaining, mass- m continuum, of investors are "unsophisticated": for them risk-management effort costs $C + \delta > C$ per unit of assets under management. Undertaking effort is efficient,

$$R - C - \delta > \mu R, \tag{1}$$

i.e., the expected net return is larger with effort than without it, for both investor types. We also assume that when a sophisticated investor exerts risk-prevention effort, return on her assets is higher than return on cash while the opposite is the case for an unsophisticated investor,

$$R - C > 1 > R - C - \delta. \tag{2}$$

The risk-management process reflects the unique skills of the investor and is therefore difficult to monitor for outside parties. Combined with limited liability, effort unobservability generates moral hazard. For simplicity, conditional on effort, \tilde{R}_j is independent across investors and independent of hedgers' risk $\tilde{\theta}$. To allow sophisticated investors who exert effort to fully insure hedgers, we assume

$$R > \pi \Delta \theta. \tag{3}$$

We assume sophisticated investors are risk-neutral, which generates potential risk-sharing gains from trade with hedgers. In contrast, we assume unsophisticated investors are infinitely

⁶We assume effort improves returns in the sense of first-order stochastic dominance, as in Holmstrom and Tirole (1997). Alternatively, we could have assumed effort improves returns in the sense of second-order stochastic dominance. In that case, lack of effort would correspond to risk-shifting, as in Acharya and Viswanathan (2011). Our qualitative results would still obtain if we assumed the agent could engage in risk-shifting.

risk-averse. This simplifies our analysis, as it implies unsophisticated investors don't provide any insurance.⁷

Advance information. At the beginning of $t = 1$, before investment and effort decisions are made, a public signal \tilde{s} about hedgers' 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 hedgers' 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,

$$\bar{\pi} = \text{prob}[\bar{\theta}|\bar{s}] = \frac{\lambda\pi}{\lambda\pi + (1-\lambda)(1-\pi)} \text{ and } \underline{\pi} = \text{prob}[\bar{\theta}|\underline{s}] = \frac{(1-\lambda)\pi}{(1-\lambda)\pi + \lambda(1-\pi)}. \quad (4)$$

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.

Contracts and margins. At time 0, the hedger makes a take-it-or-leave-it insurance contract offer to the sophisticated investor. Similar results would hold if, instead, we assumed the investor had some or bargaining power. The contract specifies a transfer τ at time 2 between the sophisticated investor and the hedger. When $\tau > 0$ the investor pays the hedger and vice versa when $\tau < 0$. The transfer τ can be conditional on all observable information: the realization of the risk $\tilde{\theta}$, the return on the sophisticated investor's assets \tilde{R} and the advance signal \tilde{s} . Hence, transfers are denoted by $\tau(\tilde{\theta}, \tilde{s}, \tilde{R})$. The contract also specifies margin requirements. At the beginning of $t = 1$, after the advance signal \tilde{s} was observed, a variation margin can be called.

At time 0, investors could decide to keep a fraction of their initial cash endowment, while investing the remaining fraction in the risky asset generating \tilde{R} . While cash holdings can be used directly to make deposits on the margin account, investment in the risky asset must be liquidated, at price p . Because the focus of this paper is on the interaction between optimal contracting and the equilibrium determination of p , we focus on the case where the sophisticated agents invest all their initial endowment in the risky asset generating \tilde{R} . In

⁷That unsophisticated investors don't participate in the market for insurance can be interpreted as a form of market fragmentation, in line with Gromb and Vayanos (2002).

the supplementary appendix we analyse the case where they can also keep a fraction of their initial cash endowment. We show that there is a strong form of substitutability between the two types of holdings: investors find it optimal to either keep all their initial cash, or invest it entirely in the risky asset generating \tilde{R} . We also provide a sufficient condition, involving only exogenous parameters, under which sophisticated agents find it optimal to invest all their initial endowment in the risky asset. In this context, to satisfy the margin call, a sophisticated investor liquidates a fraction $\alpha(\tilde{s}) \in [0, 1]$ of her risky assets at price p per unit and deposit the proceeds (αp) on her margin account. The equilibrium determination of p is analysed in Section 4 below.⁸

Investors' transfers are constrained by limited liability. When he has invested all his initial cash endowment in the risky asset generating \tilde{R} , an investor 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. Thus,

$$\tau(\tilde{\theta}, \tilde{s}, \tilde{R}) \leq \alpha(s)p + (1 - \alpha(s))R, \quad \forall(\theta, s, R). \quad (5)$$

Margin calls entail an opportunity cost when the asset's liquidation value (p) is lower than their value under the management of the investor ($R - C$). Margins also have advantages, however. Our key assumption is that the cash deposited in the margin account is safe and no longer under the discretion of the investor, i.e., it is ring-fenced from moral hazard. Furthermore, if the investor defaults, the cash on the margin account can be used to pay the hedger. Margin accounts can be implemented as escrow accounts set up by the hedger or via a market infrastructure such as a central counterparty (CCP). Importantly, we assume that margin deposits are observable and contractible, and that contractual provisions calling for margin deposits are enforceable. It is one of the roles of market infrastructures to ensure such contractibility and enforceability.

The sequence of events is summarized in Figure 1.

Insert Figure 1 here

2.2 First-best: observable effort

In the first-best, sophisticated investors exert risk-prevention effort, since doing so is efficient (see (1)). Margins are not used because they are costly (see (2)) and offer no benefit.

⁸This differs from Biais, Heider and Hoerova (2015), where the price was exogenous and normalized to one.

Transfers are chosen to maximize hedgers' utility

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

subject to the limited liability constraints (5), as well as the constraint that sophisticated investors accept the contract. By accepting (and exerting effort) sophisticated investors obtain $R - C - E[\tau(\tilde{\theta}, \tilde{s}, \tilde{R})]$. If they do not sell protection, they obtain $R - C$.⁹ Therefore, a sophisticated investor's participation constraint in the first-best is

$$E[\tau(\tilde{\theta}, \tilde{s}, \tilde{R})] \leq 0. \quad (7)$$

Because hedgers are assumed to have all the bargaining power, (7) binds, so the contract is actuarially fair. In this context, hedgers are fully insured. Correspondingly their marginal rate of substitution for consumption in different states of the world is equal to one, like the risk-neutral investors. We will see below that this is not the case under moral hazard. Finally, unsophisticated investors hold cash, since, by (2), the return on cash is higher than the net return they can generate on the risky asset ($R - C - \delta$).

3 Optimal margins under moral hazard

In this section, we analyze the privately optimal contract between the hedger and the sophisticated investor, under moral hazard. The analysis is similar to that in Biais, Heider and Hoerova (2015).¹⁰ The major difference is that, in Biais, Heider and Hoerova (2015), the price was normalized to one. In contrast, in the present paper, p is determined in equilibrium, as explained in the next section. When effort is unobservable, it is exerted (after observing signal s) only if it is in investors' own interest, that is:

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

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

⁹Without derivative trading, sophisticated investors always exert effort since it is efficient to do so (see condition (1)).

¹⁰Correspondingly, the results in the present section are not formally presented as propositions. The proofs in Biais, Heider and Hoerova (2015) extend directly to the results in the present section.

not incur the cost C . We hereafter focus on contracts for which this incentive compatibility condition always holds. This is optimal if lack of effort generates very low expected output.

Without effort, the investor's assets under management return R with probability μ and zero with probability $1 - \mu$. In order to relax the incentive constraint, the contract requests the largest possible transfer from a sophisticated investor when $\tilde{R} = 0$: $\tau(\tilde{\theta}, \tilde{s}, 0) = \alpha(\tilde{s})p$. This rationalizes the stylized fact that, in case of default of a sophisticated investor, margin deposits are seized and used to pay hedgers.

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

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

Using the notion of “pledgeable return” \mathcal{P} (see Holmström and Tirole, 1997),

$$\mathcal{P} \equiv R - \frac{C}{1 - \mu}, \quad (8)$$

the incentive compatibility constraint simplifies to

$$\alpha(s)p + (1 - \alpha(s))\mathcal{P} \geq E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = s]. \quad (9)$$

Similarly to the incentive constraint in Acharya and Viswanathan (2011), the left-hand side of (9) is the amount the sophisticated investor can promise to pay (or pledge) without undermining her incentives to exert risk-prevention effort. Crucially, the price at which assets are liquidated when margins are called, p , enters the incentive constraint directly. Higher price p relaxes (9). Moreover, as long as the liquidation price is higher than pledgeable income, $p > \mathcal{P}$, higher margins also relax (9). The right-hand side of (9) is what the sophisticated investor expects to pay to hedgers after observing signal s . It is positive when conditional on s , the sophisticated investor expects, on average, to make transfers to the hedger. It is negative if the seller expects, on average, to receive transfers from the hedger.

When

$$\mathcal{P} \geq (\pi - \underline{\pi})\Delta\theta = E[\tilde{\theta}] - E[\tilde{\theta}|s]. \quad (10)$$

pledgeable the income is so high that the incentive-compatibility constraints are not binding at the first-best allocation. We hereafter focus on the case when the first-best cannot be reached as

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

The participation constraint of the sophisticated investor is

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

Because the sophisticated investor exerts effort on the equilibrium path, $\tilde{R} = R$ and again, for brevity, we write $\tau(\tilde{\theta}, \tilde{s}, \tilde{R})$ as $\tau(\tilde{\theta}, \tilde{s})$. Collecting terms, the participation constraint is

$$-E[\tau(\tilde{\theta}, \tilde{s})] \geq E[\alpha(\tilde{s})(R - C - p)], \quad (12)$$

The expected transfers to the sophisticated investor (left-hand-side) must be high enough to compensate her for the expected opportunity cost of margins (right-hand-side). Raising the price p at which assets are liquidated when margins are called relaxes the participation constraint. As will be shown in the next section, in equilibrium $R - C \geq p$. Hence, the right-hand side of (12) is positive. So, if margins are used, the contract is not actuarially fair.

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

$$R > \bar{\pi}\Delta\theta - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} \mathcal{P} \quad (13)$$

$$1 - \frac{\pi\Delta\theta}{R - \mathcal{P}} > \frac{(1 - \pi)R - \mathcal{P}}{\pi + (1 - \pi)R - \mathcal{P}} \quad (14)$$

These assumptions guarantee that limited liability conditions are slack in states $(\underline{\theta}, \bar{s})$ and $(\underline{\theta}, \underline{s})$ (see Biais, Heider and Hoerova, 2015, for details).

As can be seen in (9), margins are not used after a good signal, $\alpha(\bar{s}) = 0$, or if the moral hazard is not severe, i.e., $\mathcal{P} \geq p$. Furthermore, the participation constraint and the incentive constraint after a bad signal are binding, which gives expected transfers conditional on the signal (as a function of $\alpha(\underline{s})$ and p):

$$E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] = \alpha(\underline{s})p + (1 - \alpha(\underline{s}))\mathcal{P} > 0 \quad (15)$$

$$E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}] = -\frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} [\alpha(\underline{s})(R - C) + (1 - \alpha(\underline{s}))\mathcal{P}] < 0. \quad (16)$$

Finally, the optimal contract provides full insurance conditional on the realization of \tilde{s} . For a given signal, the consumption of the hedger at time 2 is independent of the realization of $\tilde{\theta}$. Across signals, however, the consumption of the hedger varies, unlike in the first-best. After a bad signal, risk-sharing is limited by the binding incentive constraint, and the consumption

of the hedger (denoted by \underline{c}) is relatively low. Symmetrically, the hedger's consumption after a good signal is denoted by \bar{c} . We have:

$$\begin{aligned}\underline{c} &= E[\theta|\underline{s}] + E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] = E[\theta|\underline{s}] + \alpha(\underline{s})p + (1 - \alpha(\underline{s}))\mathcal{P} \\ \bar{c} &= E[\theta|\bar{s}] + E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}] = E[\theta|\bar{s}] - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} [\alpha(\underline{s})(R - C) + (1 - \alpha(\underline{s}))\mathcal{P}].\end{aligned}$$

Note that for $p > \mathcal{P}$, higher margin calls $\alpha(\underline{s})$ increase consumption after a bad signal and enable to achieve more incentive-compatible risk-sharing.

Now turn to the determination of the optimal margin call after a bad signal, taking price p as given. To analyze the amount of margin calls, it is useful to consider the ratio of the marginal utility of a hedger after a bad and a good signal. Denoting this marginal rate of substitution by φ , we have

$$\varphi(\alpha(\underline{s}), p) \equiv \frac{u'(\underline{c})}{u'(\bar{c})} = \frac{u'(E[\theta|\underline{s}] + \alpha(\underline{s})p + (1 - \alpha(\underline{s}))\mathcal{P})}{u'\left(E[\theta|\bar{s}] - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} [\alpha(\underline{s})(R - C) + (1 - \alpha(\underline{s}))\mathcal{P}]\right)}. \quad (17)$$

As mentioned above, in the first-best there is full insurance and the marginal rate of substitution of the hedger φ is equal to 1, just like that of the investor. With moral hazard, hedgers are exposed to signal risk. This makes insurance imperfect and drives the hedger's marginal rate of substitution φ above one. Thus, hedgers and investors have different marginal rates of substitution. This reflects the endogenous market incompleteness induced by moral hazard. This is in line with the result in Gromb and Vayanos (2002) that, because of financial constraints, agents' marginal rates of substitution differ.

Since there is full insurance conditional on the signal, we can rewrite the objective of the risk-averse hedger as

$$\text{prob}[\bar{s}]u(\bar{c}) + \text{prob}[\underline{s}]u(\underline{c}). \quad (18)$$

Maximizing (18) with respect to $\alpha(\underline{s})$ and using (15) and (16), while taking the price p as given, the optimal margin after bad news (if it is interior) is given by the following condition:

$$\varphi(\alpha^*(\underline{s}), p) = \frac{R - C - \mathcal{P}}{p - \mathcal{P}} = 1 + \frac{R - C - p}{p - \mathcal{P}}, \quad (19)$$

which is illustrated in Figure 2.

Insert Figure 2 here

As illustrated in the figure, for a given price p , $p > \mathcal{P}$, the marginal rate of substitution φ is decreasing in $\alpha(\underline{s})$. This is because higher margins reduce the wedge between consumption

after a good and a bad signal, i.e., they improve insurance. (19) states that the optimal margin after a bad signal, $\alpha^*(\underline{s})$, is such that the marginal rate of substitution $\varphi(\alpha^*(\underline{s}), p)$ is equal to one (its value in the first best) plus a positive markup, reflecting the deviation from perfect risk-sharing. Higher p reduces the opportunity cost of margins, $R - C - p$, and increases the extent to which margins relax the incentive constraint, $p - \mathcal{P}$. Thus, higher p lowers the right-hand side of (19), corresponding to more attractive margins. On the left-hand side, higher price decreases the marginal rate of substitution $\varphi(\alpha^*(\underline{s}), p)$. Ceteris paribus, this implies that margins are needed less. Because of these countervailing effects of p on the left-hand-side and right-hand-side of (19), optimal margin can either increase or decrease with price. In the next section we further investigate this mechanism and its equilibrium consequences.

4 Equilibrium and optimality

4.1 Market equilibrium

In the first best there is no need to trade the asset held by investors, but, with moral hazard, when margins are called, the sophisticated investors must liquidate a fraction $\alpha(s)$ of their holdings of the risky asset with final payoff \tilde{R} . Thus, the supply of the asset at time 1 is $\alpha(s)$.

The only market participants able to provide liquidity in the market at that point in time are the unsophisticated investors, who value the asset at $R - C - \delta$, and overall have an amount of cash m available to buy the asset. For prices above $R - C - \delta$, their demand is zero. Below that threshold, total demand from unsophisticated investors is

$$D(p) = \frac{m}{p}, \quad \forall p \leq R - C - \delta. \quad (20)$$

Market clearing at $t = 1$ requires that the supply of the asset is equal to the demand. After a good signal, there are no margin calls, and therefore no reason to trade. At any price between $R - C$ and $R - C - \delta$ the market clears, with 0 volume. After a bad signal, the supply is $\alpha^*(\underline{s})$. While at $t = 1$ this is a fixed number, at $t = 0$ margin calls are optimally set by contracting parties, rationally anticipating the equilibrium price. Thus, for each possible anticipated price p , there is an optimal amount of margin calls after a bad signal, $\alpha^*(\underline{s})$. This is the supply function, $S(p) = \alpha^*(\underline{s})$.

After a bad signal, when margins are strictly positive ($\alpha^*(\underline{s}) > 0$), the market clearing price is:

$$p = \frac{m}{\alpha^*(\underline{s})}, \quad (21)$$

whenever $\frac{m}{\alpha^*(\underline{s})} \leq R - C - \delta$. (21) underscores that there is cash-in-the-market pricing (as in Allen and Gale, 1994, and Acharya and Viswanathan, 2011), since the demand of the unsophisticated investors is constrained by the amount of cash they hold. They would like to buy more of the asset, since the price is lower than their valuation ($R - C - \delta$), but they can buy only up to their cash holdings (m). They cannot borrow to buy more of the asset, because, at time 1, no other market participant has cash to lend them: sophisticated investors have invested all their initial endowment in the risky asset with final payoff \tilde{R} at time 2, and the return of the hedgers's asset ($\tilde{\theta}$) will also occur at time 2 only.

4.1.1 Existence

When parties anticipate a price lower than the pledgeable income, so that margins don't relax the incentive compatibility condition, they choose not to use margins. Thus, for any $p < \mathcal{P}$, $S(p) = 0$. Denote by \hat{p} the price such that $\varphi(0) = \frac{R-C-\mathcal{P}}{p-\mathcal{P}}$, i.e., such that, when margins are zero, the marginal rate of substitution of the hedger is exactly equal to the ratio on the right-hand-side of (19). That is

$$\hat{p} = \mathcal{P} + \frac{R - C - \mathcal{P}}{\varphi(0)}. \quad (22)$$

When the price is lower than \hat{p} , the cost of margins is so high that, starting from 0 margins, the hedger does not want to raise $\alpha(\underline{s})$. In contrast, when $p > \hat{p}$, starting from 0 margins, the hedger would be better off raising $\alpha(\underline{s})$ at least a bit above 0. Hence, for any $p < \hat{p}$, there are no margins, and therefore zero supply, while, for $p \geq \hat{p}$, hedgers request $\alpha^*(\underline{s}) > 0$. Because of these margin calls, supply is positive after a bad signal, for prices above \hat{p} . On the other hand, demand is decreasing and goes to 0 for $p > R - C - \delta$. Thus, if demand is positive at \hat{p} , it crosses supply at some price above \hat{p} , which is an equilibrium. This happens when $R - C - \delta > \hat{p}$, or equivalently:

$$\varphi(0) > \frac{R - C - \mathcal{P}}{R - C - \mathcal{P} - \delta}. \quad (23)$$

(23) holds when the need for risk-sharing, measured by $\varphi(0)$, is high, which implies the benefits of margins are high, while δ is relatively low, which limits the cost of margins. Thus we obtain the following proposition:

Proposition 1 (Existence) *Equilibrium exists. If (23) does not hold, then there are no margins, and the market clears (with 0 volume) at any price between $R - C - \delta$ and \hat{p} . Otherwise, if (23) holds, the optimal margin is*

$$\alpha^*(\underline{s}) = \varphi^{-1} \left(\frac{R - C - \mathcal{P}}{p - \mathcal{P}} \right) \in (0, 1), \quad (24)$$

while the market clearing price is

$$p^* = \frac{m}{\alpha^*(\underline{s})} > \hat{p}. \quad (25)$$

4.1.2 Uniqueness

While the demand curve is decreasing, the supply curve can be non-monotonic: Using equation (24), we show in the appendix that $S(p)$ is increasing if and only if:

$$\alpha^*(\underline{s}) = S(p) < \frac{1}{\rho(\underline{c})(p - \mathcal{P})}, \quad (26)$$

where $\rho(\underline{c})$ denotes the coefficient of the absolute risk aversion. Thus, if (26) holds, higher price p leads to an increase in the supply of the asset. Conversely, if

$$\alpha^*(\underline{s}) = S(p) \geq \frac{1}{\rho(\underline{c})(p - \mathcal{P})} \quad (27)$$

holds, higher price p leads to a decrease in the supply of the asset $\alpha(p)$. These two cases are illustrated in Figure 3, panels A and B. In Panel A, (26) holds for all $p \in [\hat{p}, R - C - \delta]$. In Panel B, (26) initially holds for relatively low values of p , but then, for larger values of p , (27) holds and supply decreases.

Insert Figure 3 here

To offer an example where supply can be increasing or non-monotonic, consider the case of the exponential utility with absolute risk-aversion parameter ρ . In that case $\varphi(\alpha^*(\underline{s}), p)$ is equal to

$$\exp \left[\rho \left\{ E[\theta|\bar{s}] - E[\theta|\underline{s}] - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} [\alpha(\underline{s})(R - C) + (1 - \alpha(\underline{s}))\mathcal{P}] - [\alpha(\underline{s})p + (1 - \alpha(\underline{s}))\mathcal{P}] \right\} \right].$$

Relying on this formula, the following propositions spell out the consequences of ρ for supply and equilibrium uniqueness.

Proposition 2 (Sufficient condition for uniqueness) *Suppose utility is exponential. If the coefficient of the absolute risk aversion of hedgers is sufficiently small, $\rho < \frac{1}{1-\mathcal{P}}$, then the supply is non-decreasing and the market equilibrium is unique.*

Proposition 3 (Necessary condition for multiplicity) *Suppose utility is exponential. For each price p , there exists a threshold value of the coefficient of the absolute risk aversion of hedgers ρ , ρ^* , such that if $\rho > \rho^*$, then $\alpha^*(\underline{s}) \geq \frac{1}{\rho(p-\mathcal{P})}$ and the supply of the asset is decreasing in p .*

The intuition for supply non-monotonicity is as follows. When hedgers are very risk-averse, they care a lot about their consumption after bad news, \underline{c} . They are happy to increase margins if this raises \underline{c} . They realize margins carry an opportunity cost, but this cost is paid with consumption after good news, \bar{c} . When hedgers are very risk averse, they are willing to tradeoff consumption after good news for consumption after bad news. To see more precisely how this mechanism operates, note that, as shown in the previous section, \underline{c} is equal to

$$E[\theta|\underline{s}] + \alpha(\underline{s})p + (1 - \alpha(\underline{s}))\mathcal{P}.$$

When the price p goes down, $\alpha(\underline{s})p$ decreases, and so does \underline{c} . If the hedger is very risk-averse, she finds it optimal to increase $\alpha^*(\underline{s})$ to counter the impact of the decrease in the price p . This gives rise to non-monotonic supply. In addition, with exponential utility, we can pin down the impact on supply of the risk-aversion parameter ρ :

Proposition 4 *Suppose utility is exponential. If ρ increases, supply increases.*

Thus, there can be two regimes in the market, depending on risk aversion. When risk aversion is low, supply is relatively low and upward-sloping, and equilibrium is unique, with a relatively high price and low margins. When risk aversion gets higher, however, supply increases, which lowers the price. In addition, supply can become non-monotonic. Correspondingly, there may be multiple equilibria. With multiple equilibria, if market participants expect the price to be reasonably high, they don't need to request large margins to generate enough pledgeable income after bad news. Because margins are small, prices are not severely depressed after bad news, confirming the initial expectation. In contrast, if market participants expect very low prices, they request large margins, which depress prices via fire-sales, confirming the initial expectation. These two regimes, and the possibility of multiple

equilibria, are illustrated in Figure 4, panels A and B, which depict the exponential utility case.

Insert Figure 4 here

4.2 Utilitarian optimum

We now compare the laissez-faire regime to the allocation chosen by a benevolent central planner, putting all the weight on the hedgers (subject to the participation constraints of the other agents), to which we hereafter refer as the utilitarian optimum. Hedgers benefit from high prices, which enable them to raise their consumption after bad news. Sophisticated investors are indifferent, because their participation constraint binds, and they always obtain $R - C$, irrespective of p . This leads to the next proposition.

Proposition 5 *If there are multiple equilibria, utilitarian welfare is larger for equilibria with higher prices.*

In the market equilibrium, hedgers individually choose margins to maximize their objective (18). Margins, in turn, determine supply, and therefore the equilibrium price at time 1. Because they are competitive, individual hedgers don't take into account the aggregate effect of their individual margins on the market clearing price. Yet, when one hedger increases the margin she requests, she exerts a negative pecuniary externality on the others, by pushing the price down (as in Gromb and Vayanos, 2002). Under symmetric information, this pecuniary externality would not reduce welfare, but under information asymmetry it can (Greenwald and Stiglitz, 1986). Thus, as shown below, the market equilibrium differs from the utilitarian optimum, which internalizes pecuniary externalities.

The utilitarian optimum is obtained by maximizing (18) with respect to $\alpha(\underline{s})$, substituting the optimal transfers (15), (16), and setting a price for which the market clears with volume $\alpha(\underline{s})$. For prices strictly lower than $R - C - \delta$, corresponding to margins strictly above $m/(R - C - \delta)$, the market clearing condition is $\alpha(\underline{s})p = m$. Substituting in the hedger's expected utility, the latter writes as

$$\text{prob}[\bar{s}]u\left(E[\theta|\bar{s}] - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}P - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} \alpha(\underline{s})(R - C - P)\right) + \text{prob}[\underline{s}]u(E[\theta|\underline{s}] + P + m - \alpha(\underline{s})P),$$

which is decreasing in α . This is because raising margins, in addition to generating productive inefficiencies, fails to increase the hedger's consumption after bad news. Indeed, any increase

in αp due to an increase in α is exactly offset by the corresponding decrease in p , so that αp remains equal to m . Consequently, from the point of view of the social planner maximizing the welfare of hedgers, any pair (p, α) such that $p < R - C - \delta$ and $\alpha(\underline{s}) > m/(R - C - \delta)$ is dominated by setting $p = R - C - \delta$ and $\alpha(\underline{s}) = m/(R - C - \delta)$. In contrast with the social planner, individual traders take p as given in αp . From their competitive point of view, αp is increasing in α . Thus a hedger could want to raise α above $\frac{m}{R-C-\delta}$, hoping this will enable her to increase her consumption after good signal. When they all follow this logic, this results in an equilibrium price lower than $R - C - \delta$, annihilating the effort of each hedger to raise his consumption after bad news. This yields the next proposition.

Proposition 6 (Over-margining) *In any market equilibrium with $p < R - C - \delta$, margining is excessive compared to the utilitarian optimum, $\alpha^*(\underline{s}) > \alpha^U(\underline{s})$.*

Importantly, the over-margining result of Proposition 6 holds irrespective of whether market equilibrium is unique or not. The analysis above implies the social planner must choose between setting $p = R - C - \delta$ and having positive margins or setting $p > R - C - \delta$ and having no margins. In the latter case the expected utility of the hedger is

$$\text{prob}[\bar{s}]u\left(E[\theta|\bar{s}] - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}P\right) + \text{prob}[\underline{s}]u(E[\theta|\underline{s}] + P + m).$$

In the former case, the planner must choose $\alpha \in [0, \frac{m}{R-C-\delta}]$ (unsophisticated investors are indifferent since the price is just equal to their valuation). Therefore, the hedger's expected utility can be written as

$$\begin{aligned} &\text{prob}[\bar{s}]u\left(E[\theta|\bar{s}] - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}P - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} \epsilon \frac{m}{R-C-\delta}(R-C-P)\right) \\ &+ \text{prob}[\underline{s}]u\left(E[\theta|\underline{s}] + P + \epsilon \frac{m}{R-C-\delta}(R-C-\delta-P)\right) \end{aligned} \quad (28)$$

where $\epsilon \in [0, 1]$, with $\epsilon = 0$ corresponding to $\alpha(\underline{s}) = 0$ and $\epsilon = 1$ corresponding to $\alpha(\underline{s}) = m/(R - C - \delta)$. The socially optimal level of margins is determined by choosing $\epsilon \in [0, 1]$ to maximize (28). Taking the first-order condition, we obtain the next proposition.

Proposition 7 (Utilitarian optimal margin) *The social planner maximizing hedger's welfare sets margin to*

$$\alpha(\underline{s}) = \epsilon \frac{m}{R - C - \delta}$$

such that

$$\frac{u' \left(E[\theta|\underline{s}] + \mathcal{P} + \epsilon \frac{m(R-C-\delta-\mathcal{P})}{R-C-\delta} \right)}{u' \left(E[\theta|\bar{s}] - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} \mathcal{P} - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} \epsilon \frac{m(R-C-\mathcal{P})}{R-C-\delta} \right)} = \frac{R - C - \mathcal{P}}{R - C - \delta - \mathcal{P}}. \quad (29)$$

The left-hand-side of (29) is the marginal rate of substitution $\varphi(\alpha(\underline{s}) = \epsilon \frac{m}{R-C-\delta})$, which reflects the benefits from margins. The right-hand-side reflects the cost of margins, which is increasing in δ . Building on Proposition 7, one can check under what condition margins are strictly positive in the utilitarian optimum. This is the case when the left-hand side of (29), evaluated at $\epsilon = 0$, is strictly greater than the right hand-side, that is

$$\varphi(\alpha(\underline{s}) = 0) > \frac{R - C - \mathcal{P}}{R - C - \mathcal{P} - \delta}. \quad (30)$$

Condition (30) is exactly the same as (23): the condition under which market equilibrium entails positive margins is the same as the condition under which the utilitarian optimum entails positive margins. This is because, on the vertical segment of the demand curve (except at the top), margins can be increased without changing the price, hence there is no externality, so that, locally, individual optimality and utilitarian optimality coincide. Correspondingly, market equilibrium implements the utilitarian optimum only in the special case when the market equilibrium is such that $p = R - C - \delta$ (at which point the demand curve is vertical).

While the analysis above is made starker by cash-in-the-market-pricing, we have checked that our excessive margining result also obtains with a more general demand function stemming, for example, from a population of unsophisticated investors with heterogeneous cost of effort.

5 Implications

The implications of our theoretical analysis reflect the interaction between optimal contracting, driven by the demand for risk-sharing, and equilibrium pricing.

5.1 Asset pricing

As shown above, because of endogenous market incompleteness, hedgers cannot fully hedge the risk of their initial endowment. Such imperfect hedging affects their valuation for the asset with final payoff $\tilde{\theta}$, and the corresponding risk-premium. To evaluate that risk-premium,

consider a market for $\tilde{\theta}$, between hedgers, at time 0, after they have contracted with investors. Similarly to the Lucas (1978) model, there is no trade between hedgers, because they are all identical, and the equilibrium price is such that the market clears with 0 volume.¹¹ Yet, this price, which we denote by q , reflects the marginal valuation of the hedgers for the asset and the corresponding risk-premium.

In this context, a hedger considering buying z units of the asset solves:

$$\max_z Eu(c(\bar{s}) + z(\tilde{\theta} - q)).$$

The first element, $c(\bar{s})$, is the consumption resulting from the contract between the hedger and the investor. The second one, $z(\tilde{\theta} - q)$, results from the purchase, at price q , of z units of the asset with final value $\tilde{\theta}$. Considering explicitly the four possible realizations of $(\tilde{s}, \tilde{\theta})$, the maximization problem writes as

$$\begin{aligned} & \max_z \pi \lambda u(c(\bar{s}) + z(\bar{\theta} - q)) + \pi(1 - \lambda)u(c(\underline{s}) + z(\bar{\theta} - q)) \\ & + (1 - \pi)(1 - \lambda)u(c(\bar{s}) + z(\underline{\theta} - q)) + (1 - \pi)\lambda u(c(\underline{s}) + z(\underline{\theta} - q)). \end{aligned}$$

The first-order condition yields:

$$\begin{aligned} 0 &= \pi(\bar{\theta} - q)[\lambda u'(c(\bar{s}) + z(\bar{\theta} - q)) + (1 - \lambda)u'(c(\underline{s}) + z(\bar{\theta} - q))] \\ &+ (1 - \pi)(\underline{\theta} - q)[(1 - \lambda)u'(c(\bar{s}) + z(\underline{\theta} - q)) + \lambda u'(c(\underline{s}) + z(\underline{\theta} - q))]. \end{aligned}$$

As mentioned above, since all hedgers are identical, there are no gains from trade among them, and in equilibrium $z = 0$. Thus, in equilibrium, the first-order condition simplifies to

$$0 = \pi(\bar{\theta} - q)[\lambda u'(c(\bar{s})) + (1 - \lambda)u'(c(\underline{s}))] + (1 - \pi)(\underline{\theta} - q)[(1 - \lambda)u'(c(\bar{s})) + \lambda u'(c(\underline{s}))].$$

We can thus solve for the price

$$q = \frac{\pi[\lambda u'(c(\bar{s})) + (1 - \lambda)u'(c(\underline{s}))]\bar{\theta} + (1 - \pi)[(1 - \lambda)u'(c(\bar{s})) + \lambda u'(c(\underline{s}))]\underline{\theta}}{\pi[\lambda u'(c(\bar{s})) + (1 - \lambda)u'(c(\underline{s}))] + (1 - \pi)[(1 - \lambda)u'(c(\bar{s})) + \lambda u'(c(\underline{s}))]},$$

and obtain the next proposition.

Proposition 8 *When hedgers can trade the asset with final payoff $\tilde{\theta}$ at time 0 after contracting with sophisticated investors, the market clearing price is*

$$q = \hat{\pi}\bar{\theta} + (1 - \hat{\pi})\underline{\theta},$$

¹¹Note that zero trading volume in the hedgers' asset and contracting in derivatives on the hedgers' asset is consistent with the empirical observation of high liquidity in CDS markets and low liquidity in the underlying bond market.

where

$$\hat{\pi} = \frac{\pi[\lambda u'(c(\bar{s})) + (1 - \lambda)u'(c(\underline{s}))]}{\pi[\lambda u'(c(\bar{s})) + (1 - \lambda)u'(c(\underline{s}))] + (1 - \pi)[(1 - \lambda)u'(c(\bar{s})) + \lambda u'(c(\underline{s}))]}.$$

The probability $\hat{\pi}$ can be interpreted as the price of the Arrow-Debreu security paying one unit of consumption in state $\bar{\theta}$. Dividing the numerator and the denominator by $u'(c(\bar{s}))$, $\hat{\pi}$ rewrites as a function of the marginal rate of substitution φ ,

$$\hat{\pi} = \frac{\pi[\lambda + (1 - \lambda)\varphi]}{\pi[\lambda + (1 - \lambda)\varphi] + (1 - \pi)[(1 - \lambda) + \lambda\varphi]}.$$

When there is no moral hazard, then $\varphi = 1$ and $\hat{\pi} = \pi$. That is, although hedgers are risk-averse, without moral hazard, they are perfectly insured by the investors, and therefore do not demand any risk premium for holding the asset. With moral hazard, however, $\varphi > 1$ and $\hat{\pi} < \pi$.¹² So, there is a risk premium. In fact, $\hat{\pi}$ decreases with φ . As moral hazard gets more severe, risk-sharing gets more imperfect, the market becomes more incomplete, and the risk premium increases.

5.2 Risk aversion, cash-in-the-market and prices

The risk appetite of hedgers in our model can reflect regulatory constraints that reduce the amount of risk they can bear. It can also proxy for increases in macro-economic risk not captured by other variables in our model. When hedgers' risk appetite goes down, they demand more insurance from sophisticated investors. To make this incentive compatible, they request larger margins. The corresponding increase in supply $S(p)$ (see Proposition 4) worsens the drop in the price of the sophisticated investors' asset after bad news on $\tilde{\theta}$. Furthermore, as hedgers' risk appetite decreases, there is a switch from an increasing supply curve and equilibrium uniqueness, to a non-monotonic supply curve and multiple equilibria (see Proposition 3). This is illustrated in Figure 5.

Insert Figure 5 here

In this context, a small change in risk aversion, which triggers a switch to multiple equilibria, can lead to a large jump in the market price if the low-price equilibrium is chosen. This is illustrated in Figure 6.

Insert Figure 6 here

¹² $\hat{\pi} < \pi$ is equivalent to $\frac{(1-\lambda)+\lambda\varphi}{\lambda+(1-\lambda)\varphi} > 1$. That is $(\varphi - 1)(2\lambda - 1) > 0$, which holds since $\lambda > \frac{1}{2}$.

Similarly, when the cash endowment of unsophisticated investors (m) decreases, this decreases the demand for sophisticated investors' assets, $D(p)$, and increases the fire-sale discount.

5.3 Contagion

When bad news hit the market, following the same logic as for Proposition 8, the price of the asset with final payoff $\tilde{\theta}$ becomes

$$q(\underline{s}) = \underline{\pi}\bar{\theta} + (1 - \underline{\pi})\underline{\theta}.$$

Risk-neutral pricing prevails in the market at that time because \underline{c} does not depend on the final realization of $\tilde{\theta}$ in the optimal contract. Using (4) and Proposition 8, it is easy to check that $\underline{\pi} < \hat{\pi}$. Consequently, $q(\underline{s}) < q$, i.e., the price for the hedgers' asset $\tilde{\theta}$ drops after bad news on $\tilde{\theta}$, consistent with intuition. The arrival of bad news about $\tilde{\theta}$ also triggers margin calls, and leads to a drop in the market valuation of the investors's asset, from $R - C$ to $p \leq R - C - \delta$. The prices of the asset generating $\tilde{\theta}$ and that of the asset generating \tilde{R} are correlated: after a bad signal, both prices drop. This is in contrast with the frictionless case, where these two prices are independent. Thus, moral hazard generates endogenous correlation, which can be interpreted as a form of contagion from the valuation of the hedgers' asset towards that of the investors' asset.

There is also contagion from the moral hazard problem associated with the investors' asset to the valuation of the hedgers' asset. The greater the moral hazard problem, the lower the amount of risk-sharing, the larger the marginal rate of substitution φ , and, by Proposition 8, the lower the market valuation of the hedgers' asset.

Such contagion effects are more pronounced when the hedgers' asset risk ($Var(\tilde{\theta})$) is larger. An increase in $Var(\tilde{\theta})$ increases hedgers' demand for insurance, which leads to higher margins and larger drops in the price of the investors' asset after bad news on $\tilde{\theta}$. An increase in the risk aversion of hedgers or in the severity of the moral hazard problem of investors also deepens the contagion effects.

5.4 Policy

As shown above (see Proposition 6), an equilibrium typically entails excessive margining relative to the utilitarian optimum. This is because hedgers are too eager to buy protection against the $\tilde{\theta}$ risk, and investors correspondingly sell too much insurance which requires

large margins, without internalizing the negative effect that such margins have on the price p . To correct this market failure, the regulator could set position limits for investors, or cap margins.¹³ If better capitalized investors have smaller moral hazard problems, such regulatory constraints should be more stringent for investors with low capital ratios. While position limits (or margin caps) can help narrow the wedge between equilibrium and utilitarian optimum, they can also reduce the market instability problem caused by multiplicity: By imposing the right constraints, the regulator/central bank can rule out inefficient low-price equilibria. Such regulatory intervention is a form of macro-prudential policy. Now, the scope for multiplicity and market instability is higher when risk aversion increases. Hence, the macro-prudential policy should be tighter when risk-aversion is larger. This goes against the received wisdom that prudential regulation should be countercyclical.

6 Conclusion

Risk-sharing is one of the main functions of financial markets, and the structure of risk-exposures is one of the main drivers of asset pricing. This paper offers a framework to analyse how moral-hazard limits risk-sharing, i.e., induces endogenous market incompleteness, which in turns affects asset pricing. Central to our analysis is the interaction between optimal contracting and equilibrium pricing. Our theoretical model delivers both positive and normative implications. On the positive front, our analysis shows how moral hazard increases risk-premia and generates endogenous contagion, especially when risk aversion is high. On the normative front, we show that margin calls, although individually optimal, generate negative externalities. Because of these externalities, there is a gap between market equilibrium and utilitarian optimum. To bridge that gap, the regulator can limit positions and cap margins, especially for poorly capitalized institutions.

¹³While our model implies that in the presence of moral hazard margining can be excessive, we are not arguing that in reality markets *always* choose excessively large margins. In practice other forces are at play. For example, if hedgers are insured against counterparty default by a central clearing counterparty, then they prefer not to use margins, which undermines incentives (see Biais, Heider and Hoerova, 2012). Our implication for position limits seems more robust, as it would not be ruled out by the incentive problems analyzed in Biais, Heider and Hoerova, 2012.

References

- Acharya, V. and S. Viswanathan, 2011, "Leverage, Moral Hazard, and Liquidity," *Journal of Finance*, 99-38.
- Allen, F. and D. Gale, 1994, "Limited Market Participation and Volatility of Asset Prices," *American Economic Review*, 84, 933–955.
- Bank for International Settlements, 2010, "The Role of Margin Requirements and Haircuts in Procyclicality," Committee on the Global Financial System Paper No. 36.
- Biais, B., F. Heider, and M. Hoerova, 2015, "Risk-sharing or Risk-taking? Counterparty Risk, Incentives and Margins," forthcoming *Journal of Finance*.
- Biais, B., F. Heider, and M. Hoerova, 2012, "Clearing, Counterparty Risk and Aggregate Risk," *IMF Economic Review* 60, 193-222.
- Brunnermeier, M. and L. Pedersen, 2009, "Market Liquidity and Funding Liquidity," *Review of Financial Studies* 22, 2201-2238.
- Fleming, M. and A. Sarkar, 2014, "The Failure Resolution of Lehman Brothers," *Economic Policy Review*, 20 (2).
- Greenwald, B., and J. Stiglitz, 1986, "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics* 101, 229-264.
- Gromb, D. and D. Vayanos, 2010, "Limits of Arbitrage: The State of the Theory", NBER Working paper No. 15821.
- Gromb, D. and D. Vayanos, 2002, "Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs," *Journal of Financial Economics* 66, 361-407.
- Gromb, D. and D. Vayanos, 2015, "The Dynamics of Financially Constrained Arbitrage", Working paper, LSE.
- Holmstrom, B. and J. Tirole, 1997, "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics* 112, 663–692.
- Hombert J., 2009, "Optimal Financial Structure and Asset Prices," Working Paper, HEC.
- Kuong, J., 2014, "Self-fulfilling Fire Sales: Fragility of Collateralized Short-term Debt Markets," Working Paper, INSEAD.
- Lorenzoni, G., 2008, "Inefficient Credit Booms," *Review of Economic Studies* 75, 809-833.
- Lucas, R., 1978, "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429–1446.

Appendix

Proof of Proposition 1 and of condition (26) The first step is to show that the demand is continuous in p . For $p \in (0, R - C - \delta)$, the demand is given by $D(p) = \frac{m}{p} > 0$. At $p = R - C - \delta$, it is given by $D(p = R - C - \delta) \in [0, \frac{m}{R-C-\delta}]$. For any $p > R - C - \delta$, $D(p) = 0$. At $p = R - C - \delta$, the limit from the left is given by $\frac{m}{R-C-\delta}$, while from the right, the limit is given by 0. Thus, demand is continuous.

The second step is to show that the supply function $S(p) = \alpha^*(\underline{s})$ is continuous in p . For any $p \leq \hat{p}$, the supply function is continuous (and equal to zero). By construction, at $p = \hat{p}$, the supply is given by $\alpha^*(\underline{s}) = \varphi^{-1}\left(\frac{R-C-\mathcal{P}}{\hat{p}-\mathcal{P}}\right) = 0$. The supply is continuous at $p = \hat{p}$. This is because limit from the left is zero: we have shown that $\alpha^*(\underline{s}) = 0$ for any $p \leq \hat{p}$. Limit from the right is also zero as $\varphi^{-1}\left(\frac{R-C-\mathcal{P}}{p-\mathcal{P}}\right)$ is continuous and equal to zero at $p = \hat{p}$. Any interior $\alpha^*(\underline{s})$ is determined by (19) where function φ is continuous in p .

The third step is to offer a necessary and sufficient condition for $\alpha^*(\underline{s})$ to be increasing in p and to show that this condition holds for prices just above \hat{p} . Consider the case where $\alpha^*(\underline{s})$ is interior. Denoting the left-hand side of (19) by F , we have by the implicit function theorem that $\frac{\partial \alpha}{\partial p} = -\frac{\partial F}{\partial p} / \frac{\partial F}{\partial \alpha}$. Then,

$$\begin{aligned} \frac{\partial F}{\partial p} &= \frac{u''(\underline{c}) \alpha(\underline{s})}{u'(\bar{c})} + \frac{R - C - \mathcal{P}}{(p - \mathcal{P})^2} = \left(-\frac{\alpha(\underline{s}) u'(\underline{c})}{u'(\bar{c})} \right) \left(-\frac{u''(\underline{c})}{u'(\underline{c})} \right) + \frac{R - C - \mathcal{P}}{(p - \mathcal{P})^2} \\ &= -\alpha(\underline{s}) \rho(\underline{c}) \varphi(\alpha^*(\underline{s}), p) + \frac{R - C - \mathcal{P}}{(p - \mathcal{P})^2} \end{aligned}$$

where $\rho(\underline{c})$ denotes the coefficient of the absolute risk aversion. Also,

$$\begin{aligned} \frac{\partial F}{\partial \alpha} &= \frac{u''(\underline{c}) (p - \mathcal{P}) u'(\underline{c}) + u'(\underline{c}) u''(\bar{c}) \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} (R - C - \mathcal{P})}{[u'(\bar{c})]^2} \\ &= \frac{u''(\underline{c})}{u'(\bar{c})} (p - \mathcal{P}) + \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} \frac{u'(\underline{c})}{u'(\bar{c})} \frac{u''(\bar{c})}{u'(\bar{c})} (R - C - \mathcal{P}) \\ &= -\left[-\frac{u''(\underline{c})}{u'(\underline{c})} \right] \frac{u'(\underline{c})}{u'(\bar{c})} (p - \mathcal{P}) - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} \frac{u'(\underline{c})}{u'(\bar{c})} \left[-\frac{u''(\bar{c})}{u'(\bar{c})} \right] (R - C - \mathcal{P}) \\ &= -\varphi(\alpha^*(\underline{s}), p) \left[\rho(\underline{c}) (p - \mathcal{P}) + \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} \rho(\bar{c}) (R - C - \mathcal{P}) \right] \end{aligned}$$

Hence,

$$\frac{\partial \alpha}{\partial p} = -\frac{\partial F}{\partial p} / \frac{\partial F}{\partial \alpha} = \frac{-\alpha(\underline{s}) \rho(\underline{c}) \varphi(\alpha^*(\underline{s}), p) + \frac{R-C-\mathcal{P}}{(p-\mathcal{P})^2}}{\varphi(\alpha^*(\underline{s}), p) \left[\rho(\underline{c}) (p - \mathcal{P}) + \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]} \rho(\bar{c}) (R - C - \mathcal{P}) \right]}$$

It follows that $\frac{\partial \alpha}{\partial p} > 0$ if and only if

$$\frac{R - C - \mathcal{P}}{(p - \mathcal{P})^2} > \alpha^*(\underline{s}) \rho(\underline{c}) \varphi(\alpha^*(\underline{s}), p), \quad (31)$$

which holds at $\alpha^*(\underline{s}) = 0$. So we know that, at \hat{p} , since margin is zero, supply is increasing. Using (19) in (31), we get that if

$$\alpha^*(\underline{s}) < \frac{1}{\rho(\underline{c})(p - \mathcal{P})} \quad (32)$$

holds, then higher price p leads to an increase in the supply of the asset $S(p)$. Conversely, if

$$\alpha^*(\underline{s}) \geq \frac{1}{\rho(\underline{c})(p - \mathcal{P})} \quad (33)$$

holds, then higher price p leads to a decrease in the supply of the asset $S(p)$.

The fourth step is to show that the supply and demand intersect. Consider first the case when $\hat{p} \geq R - C - \delta$. Then, the supply and demand intersect at $\alpha^*(\underline{s}) = 0$ and $p^* = R - C - \delta$ as $S(p) = 0$ for any $p \leq \hat{p}$ and $D(p) = 0$ for any $p \geq R - C - \delta$. Now, consider the case when $\hat{p} < R - C - \delta$. Then, supply and demand intersect at $\alpha^*(\underline{s}) > 0$. To see this note that the supply of the asset lies below the demand for the asset at $p = \hat{p}$, while it lies above the demand at $p = R - C - \delta$. At $p = \hat{p}$, $S(\hat{p}) = \alpha^*(\underline{s}) = 0$ while $D(\hat{p}) = \frac{m}{\hat{p}} > 0$ since $\hat{p} < R - C - \delta$. At $p = R - C - \delta$, there are two possibilities for the supply. Either the supply function is increasing for any $p > \hat{p}$, implying that $S(p = R - C - \delta) > 0$. Or the supply is decreasing over some range of $p > \hat{p}$ but then we have that $S(p) = \alpha^*(\underline{s}) \geq \frac{1}{\rho(\underline{c})(p - \mathcal{P})} > \frac{1}{\rho(\underline{c})(1 - \mathcal{P})} > 0$ (by (33)). So in both cases we have $S(p = R - C - \delta) > 0$. As for the demand at $p = R - C - \delta$, it is given by $D(p = R - C - \delta) \in [0, \frac{m}{R - C - \delta}]$. That is, $S(p = R - C - \delta) > 0 = \min \{D(p = R - C - \delta)\}$.

It follows that the equilibrium exists.

Proof of Proposition 2 Consider the case of the exponential utility with absolute risk-aversion parameter ρ . In this case $\varphi(\alpha^*(\underline{s}), p)$ is given by:

$$\exp \left[\rho \left\{ E[\theta|\bar{s}] - E[\theta|\underline{s}] - \frac{\mathcal{P}}{\text{prob}[\bar{s}]} + \frac{\alpha(\underline{s})}{\text{prob}[\bar{s}]} [\mathcal{P} - \text{prob}[\bar{s}]p - \text{prob}[\underline{s}](R - C)] \right\} \right].$$

Taking logs and using (19), we get

$$\begin{aligned} \alpha^*(\underline{s}) &= \frac{\text{prob}[\bar{s}] \left[\frac{1}{\rho} \ln \left(\frac{R - C - \mathcal{P}}{p - \mathcal{P}} \right) + \frac{\mathcal{P}}{\text{prob}[\bar{s}]} - (E[\theta|\bar{s}] - E[\theta|\underline{s}]) \right]}{\mathcal{P} - \text{prob}[\bar{s}]p - \text{prob}[\underline{s}](R - C)} \\ &= \frac{\text{prob}[\bar{s}] \left[\frac{1}{\rho} \ln \left(\frac{R - C - \mathcal{P}}{p - \mathcal{P}} \right) + \frac{\mathcal{P} - (\pi - \pi)\Delta\theta}{\text{prob}[\bar{s}]} \right]}{-\text{prob}[\bar{s}](p - \mathcal{P}) - \text{prob}[\underline{s}](R - C - \mathcal{P})} \end{aligned} \quad (34)$$

where the last term follows from:

$$\begin{aligned} \text{prob}[\bar{s}] (\bar{\pi} - \underline{\pi}) \Delta\theta &= [\text{prob}[\bar{s}]\bar{\pi} - (1 - \text{prob}[\underline{s}])\underline{\pi}] \Delta\theta \\ &= [\text{prob}[\bar{s}]\bar{\pi} - (1 - \text{prob}[\underline{s}])\underline{\pi}] \Delta\theta = [\pi\lambda + \pi(1 - \lambda) - \underline{\pi}] \Delta\theta = (\pi - \underline{\pi}) \Delta\theta. \end{aligned}$$

Note that $\mathcal{P} < (\pi - \underline{\pi}) \Delta\theta$ holds since we are not in the first-best. Moreover, the denominator of (34) is negative.

Suppose, contrary to the claim in the proposition, that $\rho < \frac{1}{1-\mathcal{P}}$ and the supply is decreasing in p . Since $\rho < \frac{1}{1-\mathcal{P}}$, we have

$$1 < \frac{1}{\rho(1-\mathcal{P})} < \frac{1}{\rho(p-\mathcal{P})}$$

where the last inequality follows from $\frac{1}{\rho(p-\mathcal{P})}$ being decreasing in p and $p \leq R - C - \delta < 1$.

Since $\alpha^*(\underline{s}) < 1$, it follows that

$$\alpha^*(\underline{s}) < \frac{1}{\rho(p-\mathcal{P})}$$

so that (32) holds. But then, the supply is increasing in p , a contradiction.

By Proposition 1, equilibrium exists so that the supply and demand intersect. Since the supply is non-decreasing while the demand is decreasing, they intersect exactly once.

Proof of Proposition 3 The optimal $\alpha^*(\underline{s}) \in [0, 1]$. When $\alpha^*(\underline{s}) = 1$, the claim in the proposition is straightforward. When the optimal margin is interior, $\alpha^*(\underline{s})$ is given by (34). Therefore, we need to show that

$$\frac{\text{prob}[\bar{s}] \left[\frac{1}{\rho} \ln \left(\frac{R-C-\mathcal{P}}{p-\mathcal{P}} \right) + \frac{\mathcal{P} - (\pi - \underline{\pi}) \Delta\theta}{\text{prob}[\bar{s}]} \right]}{-\text{prob}[\bar{s}] (p - \mathcal{P}) - \text{prob}[\underline{s}] (R - C - \mathcal{P})} \geq \frac{1}{\rho(p - \mathcal{P})}. \quad (35)$$

Consider $\rho \rightarrow \infty$. We have $\frac{1}{\rho} \ln \left(\frac{R-C-\mathcal{P}}{p-\mathcal{P}} \right) \rightarrow 0$ in (35) and, therefore, $\alpha^*(\underline{s}) > 0$. The right-hand side of (35), $\frac{1}{\rho(p-\mathcal{P})} \rightarrow 0$.

Now consider $\rho \rightarrow 0$. We have that the left-hand side of (35) $\rightarrow -\infty$, while the right-hand side $\rightarrow \infty$.

Hence, at $\rho \rightarrow 0$, the left-hand side of (35) is below the right-hand side of (35), while at $\rho \rightarrow \infty$, it is the other way around. Since the left-hand side of (35) is increasing in ρ , while the right-hand side of (35) is decreasing in ρ , the claim in the proposition follows.

Proof of Proposition 4 By (34), $\frac{\partial \alpha^*(\underline{s})}{\partial \rho} > 0$ for all p . Higher risk aversion ρ leads to a higher supply.

Proof of Proposition 5 We claim that an equilibrium with a higher price is preferred to an equilibrium with a lower price from the point of view of hedgers (sophisticated investors are held at their participation constraints). Let \bar{p} denote a higher and \underline{p} a lower price, respectively, $\bar{p} > \underline{p}$. Let $EU(p, \alpha(p))$ denote the value of the expected utility of a hedger at an equilibrium with the price p and the corresponding margin $\alpha(p)$. Then, we have that

$$EU(\bar{p}, \alpha^*(\bar{p})) > EU(\bar{p}, \alpha^*(\underline{p})) > EU(\underline{p}, \alpha^*(\underline{p}))$$

where the first inequality follows from the fact that $\alpha^*(\underline{p})$ was not chosen for \bar{p} and the second inequality follows from the fact that, given the same α , a hedger always prefers a higher asset price p (a higher price increases his consumption after a bad signal, \underline{c}).

Proof of Proposition 6 Let $\bar{\alpha}$ denote the “kink” in the demand curve, i.e., $\bar{\alpha} \equiv \frac{m}{R-C-\delta}$. Consider a market equilibrium that lies on the downward-sloping part of the demand curve, $D(p) = \frac{m}{p}$, so that $p < R - C - \delta$ and $\bar{\alpha} < \alpha^*$. Therefore, we have that $p = \frac{m}{\alpha}$ and, from the point of view of the social planner, the consumption after a bad signal is given by

$$\underline{c} = E[\theta|\underline{s}] + \alpha(\underline{s})\frac{m}{\alpha(\underline{s})} + (1 - \alpha(\underline{s}))\mathcal{P}.$$

Then, the first-order condition with respect to $\alpha^U(\underline{s})$ is given by

$$-u'(\bar{c})(R - C - \mathcal{P}) - u'(\underline{c})\mathcal{P} + \frac{\mu_0 - \mu_1}{\text{prob}[\underline{s}]} = 0$$

where μ_0 and μ_1 denote the Lagrange multipliers on the constraints $\alpha = \bar{\alpha}$ and $\alpha = 1$, respectively. It follows immediately that $\mu_0 > 0$ must hold and that the planner will never choose a margin higher than $\bar{\alpha}$.

Intuitively, margins are there to increase $\underline{c} = E[\theta|\underline{s}] + \alpha(\underline{s})p + (1 - \alpha(\underline{s}))\mathcal{P}$. When the demand is given by $p = \frac{m}{\alpha}$, any increase in α is exactly offset by the decrease in price p so that $\alpha(\underline{s})p = m$ is constant. And any increase in α lowers \underline{c} by lowering $(1 - \alpha(\underline{s}))\mathcal{P}$. Therefore, the planner who takes into account the impact on the equilibrium price never chooses to be on the decreasing part of the demand curve.

We conclude that in the market equilibrium with $p < R - C - \delta$ and $\alpha^* > \bar{\alpha}$, there is over-margining compared to what is socially optimal, $\alpha^* > \alpha^U$.

Proof of Proposition 7 The socially optimal level of margins is determined by choosing $\epsilon \in [0, 1]$ to maximize (28). Taking the first-order condition, we get

$$\begin{aligned} & \frac{u' \left(E[\theta|\bar{s}] + \mathcal{P} + \epsilon \frac{m(R-C-\delta-\mathcal{P})}{R-C-\delta} \right)}{u' \left(E[\theta|\bar{s}] - \frac{\text{prob}[\bar{s}]}{\text{prob}[\underline{s}]} \mathcal{P} - \frac{\text{prob}[\bar{s}]}{\text{prob}[\underline{s}]} \epsilon \frac{m(R-C-\mathcal{P})}{R-C-\delta} \right)} + \\ & + \frac{(\mu_0 - \mu_1)(R-C-\delta)}{u' \left(E[\theta|\bar{s}] - \frac{\text{prob}[\bar{s}]}{\text{prob}[\underline{s}]} \mathcal{P} - \frac{\text{prob}[\bar{s}]}{\text{prob}[\underline{s}]} \epsilon \frac{m(R-C-\mathcal{P})}{R-C-\delta} \right) m(R-C-\delta-\mathcal{P})} = \frac{R-C-\mathcal{P}}{R-C-\delta-\mathcal{P}} \end{aligned}$$

where μ_0 and μ_1 are the Lagrange multipliers on the constraints $\epsilon \geq 0$ and $\epsilon \leq 1$, respectively. When ϵ is interior, $\mu_0 = \mu_1 = 0$ and equation (29) follows. For $\mu_0 > 0$, $\epsilon = 0$ and so

$$\frac{u' \left(E[\theta|\bar{s}] + \mathcal{P} + \epsilon \frac{m(R-C-\delta-\mathcal{P})}{R-C-\delta} \right)}{u' \left(E[\theta|\bar{s}] - \frac{\text{prob}[\bar{s}]}{\text{prob}[\underline{s}]} \mathcal{P} - \frac{\text{prob}[\bar{s}]}{\text{prob}[\underline{s}]} \epsilon \frac{m(R-C-\mathcal{P})}{R-C-\delta} \right)} > \frac{R-C-\mathcal{P}}{R-C-\delta-\mathcal{P}}$$

must hold.

Figure 1: Timing

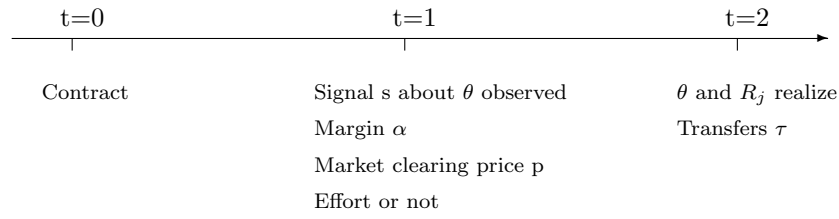


Figure 2: Optimal margin

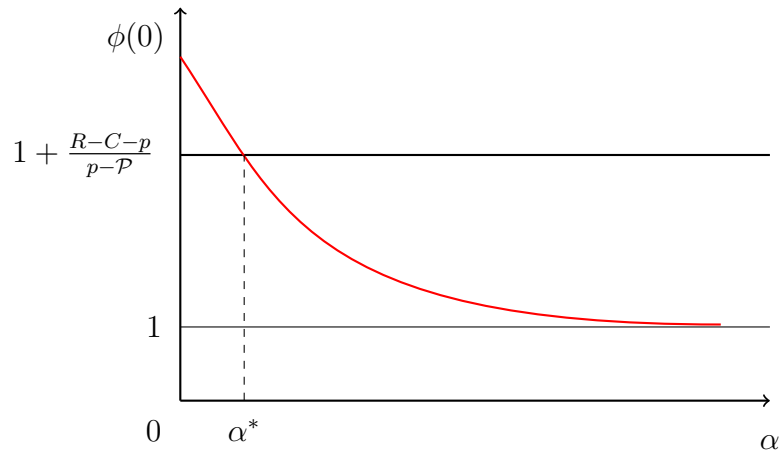


Figure 3, Panel A: Increasing supply curve

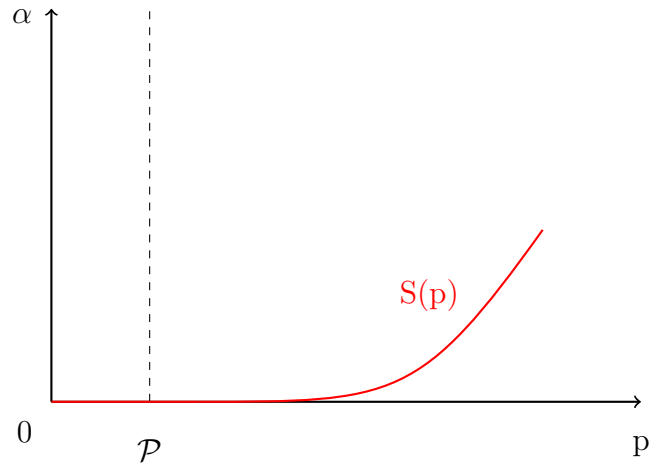


Figure 3, Panel B: Non-monotonic supply curve

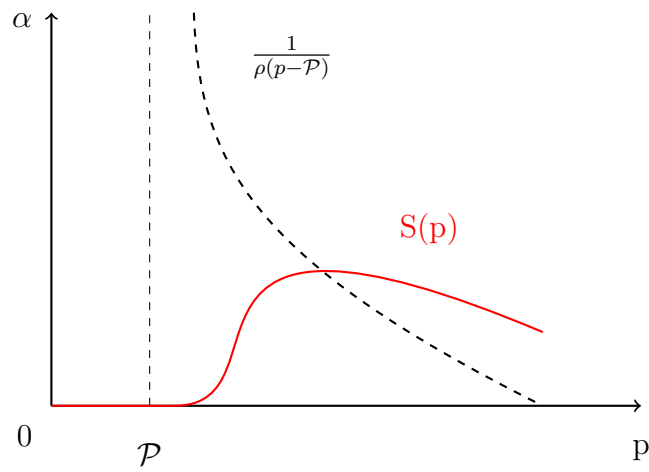


Figure 4, Panel A: Unique equilibrium

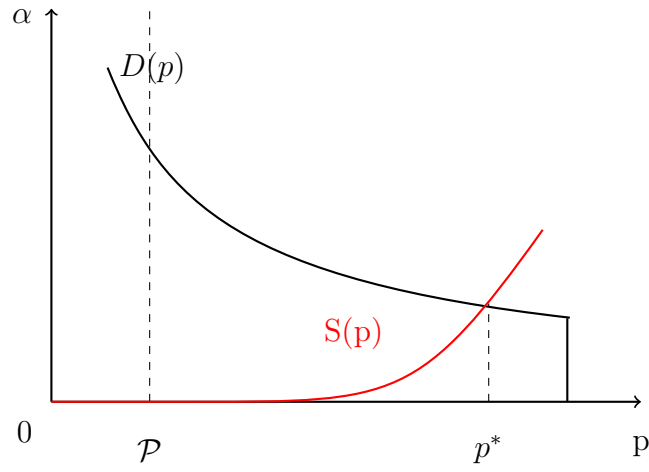


Figure 4, Panel B: Multiple equilibria

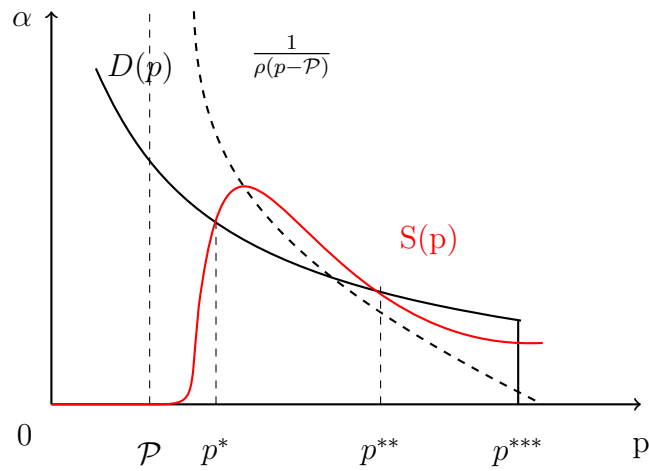


Figure 5: Demand and supply (with different risk aversion ρ)

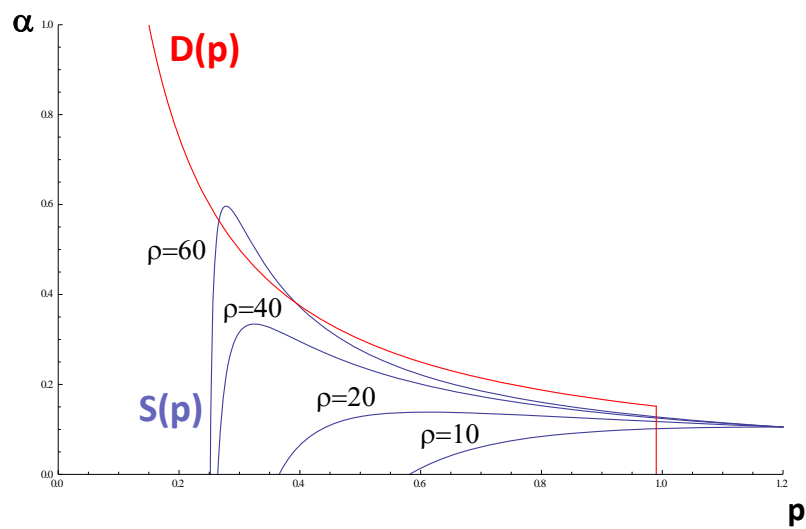
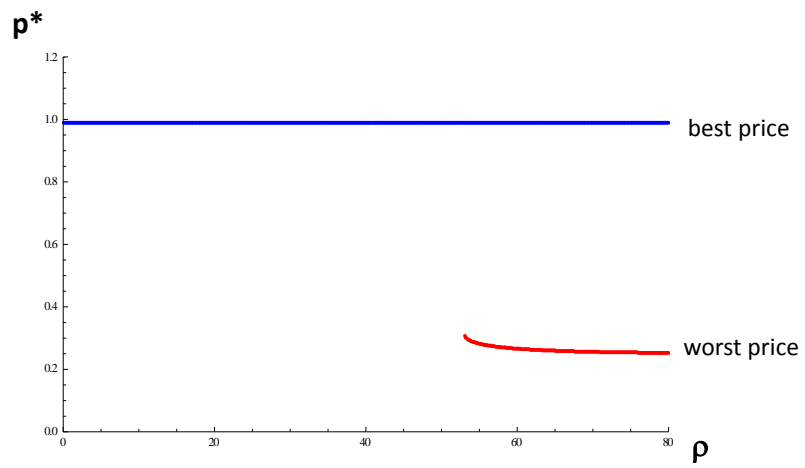


Figure 6: Equilibrium price as a function of risk aversion ρ



Supplementary appendix: Optimal margins and equilibrium prices

Bruno Biais* Florian Heider[†] Marie Hoerova[‡]

September 2015

Abstract

In this supplementary appendix we study the case in which investors, who may face margin calls in the future, can keep a fraction of their initial endowment as cash. They invest the rest of their endowment in the risky asset. Holding cash avoids the sale of assets at fire-sale prices to satisfy margin calls. We derive a condition under which investors find it optimal to invest their entire initial endowment in the risky asset and do not invest in cash.

*CNRS-CRM, Paul Woolley Research Initiative-IDEI, Toulouse School of Economics, email: bruno.biais@univ-tlse1.fr

[†]Financial Research Division, European Central Bank, email: florian.heider@ecb.int

[‡]Financial Research Division, European Central Bank, email: marie.hoerova@ecb.int

1 Adding cash to the model

Compared to our baseline model, we now allow investors to keep a fraction of their initial endowment in cash. We let β denote the fraction of the endowment kept in cash at $t = 0$. The remaining fraction $1 - \beta$ is invested in the risky asset. Cash holdings are observable and do not require costly risk-management effort. Hence, there is no moral hazard associated with investors' cash holdings. Feasibility requires

$$0 \leq \beta \leq 1. \quad (1)$$

Investors' transfers are constrained by limited liability. An investor cannot make transfers larger than what is returned by cash holdings β , the fraction $(1 - \beta)(1 - \alpha(s))$ of assets under her management and by the fraction $(1 - \beta)\alpha(s)$ of assets she deposited on the margin account. Thus,

$$\tau(\theta, s, R) \leq \beta + (1 - \beta)[\alpha(s)p + (1 - \alpha(s))R], \quad \forall(\theta, s, R). \quad (2)$$

With cash, a sophisticated investor's incentive constraint for risk-prevention effort is

$$\beta + (1 - \beta)(\alpha(s)p(s) + (1 - \alpha(s))\mathcal{P}) \geq E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = s]. \quad (3)$$

The pledgeable return of cash is equal to its physical return of one. Cash holdings do not depend on the realization of the signal \tilde{s} at $t = 1$ because the decision to hold cash occurs at $t = 0$, before the realization of the signal \tilde{s} .¹

With cash, the participation constraint of a sophisticated investor is

$$-E[\tau(\tilde{\theta}, \tilde{s})] \geq E[\beta(R - C - 1) + (1 - \beta)\alpha(\tilde{s})(R - C - p(\tilde{s}))], \quad (4)$$

as the opportunity cost of using cash is $R - C - 1 > 0$.²

2 Optimal contract under moral hazard

In this section, we derive the privately optimal contract between the hedger and the sophisticated investor, taking the price p as given. The next proposition characterizes basic properties of the optimal contract under moral hazard.

¹As in the case without cash, the derivation of the incentive constraint takes into account the maximal relaxation of the constraint in case of investor default when all remaining resources are transferred to hedgers, $\tau(\tilde{\theta}, \tilde{s}, 0) = \beta + \alpha(s)p(s)$.

²It follows that cash is not used in the first-best as cash holdings offer no benefit but entail an opportunity cost.

Proposition 1 *The incentive constraint after a bad signal and the participation constraint are binding. The incentive constraint after a good signal as well as the limited liability constraints in states $(\bar{\theta}, \bar{s})$ and $(\bar{\theta}, \underline{s})$ are slack. Margins are not used after a good signal, $\alpha(\bar{s}) = 0$.*

The results in Proposition 1 are similar to those obtained in the case where the sophisticated investor is assumed not to hold cash. When all limited liability constraints are slack, the expected transfers conditional on the signal (as a function of β , $\alpha(\underline{s})$, and p) are given by

$$E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = \underline{s}] = \beta + (1 - \beta) (\alpha(\underline{s})p + (1 - \alpha(\underline{s}))\mathcal{P}) > 0 \quad (5)$$

and

$$\begin{aligned} E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}] &= - \frac{\beta (R - C - 1 + \text{prob}[\underline{s}])}{\text{prob}[\bar{s}]} \\ &\quad - \frac{\text{prob}[\underline{s}] (1 - \beta)}{\text{prob}[\bar{s}]} [\alpha(\underline{s}) (R - C) + (1 - \alpha(\underline{s}))\mathcal{P}] < 0, \end{aligned} \quad (6)$$

where we dropped the reference to the signal s in the price p because margins are not used after a good signal.

Thus, conditional on the realization of the signal \tilde{s} , the optimal contract provides full insurance and we can write a hedger's consumption after a bad and a good signal as

$$\underline{c} = E[\theta|\underline{s}] + \beta + (1 - \beta) (\alpha(\underline{s})p + (1 - \alpha(\underline{s}))\mathcal{P}) \quad (7)$$

and

$$\bar{c} = E[\theta|\bar{s}] - \frac{\beta (R - C - 1 + \text{prob}[\underline{s}])}{\text{prob}[\bar{s}]} - \frac{\text{prob}[\underline{s}] (1 - \beta) [\alpha(\underline{s}) (R - C) + (1 - \alpha(\underline{s}))\mathcal{P}]}{\text{prob}[\bar{s}]} \quad (8)$$

The expected utility of a hedger is

$$\text{prob}[\bar{s}]u(\bar{c}) + \text{prob}[\underline{s}]u(\underline{c}),$$

where \underline{c} and \bar{c} are given above in (7) and (8). The derivative of the hedger's expected utility with respect to α is (weakly) positive if

$$\varphi(\alpha(\underline{s}), p) \equiv \frac{u'(\underline{c})}{u'(\bar{c})} \geq - \frac{\text{prob}[\bar{s}]}{\text{prob}[\underline{s}]} \frac{\frac{d\bar{c}}{d\alpha}}{\frac{d\underline{c}}{d\alpha}} = 1 + \frac{R - C - p}{p - \mathcal{P}}. \quad (9)$$

The numerator of the fraction in the last term on the right-hand side is the opportunity cost of margins, the denominator is the incentive benefit of margins. The derivative of the hedger's expected utility with respect to β is (weakly) positive if

$$\varphi(\alpha(\underline{s}), p) \equiv \frac{u'(\underline{c})}{u'(\bar{c})} \geq -\frac{\text{prob}[\underline{s}] \frac{d\bar{c}}{d\beta}}{\text{prob}[\underline{s}] \frac{d\underline{c}}{d\beta}} = 1 + \frac{(R - C - 1) - \text{prob}[\underline{s}] \alpha(\underline{s})(R - C - p)}{\text{prob}[\underline{s}](1 - \mathcal{P} - \alpha(\underline{s})(p - \mathcal{P}))}. \quad (10)$$

The numerator of the fraction in the last term on the right-hand side is the opportunity cost of cash holdings relative to the opportunity cost of margins. The denominator is the incentive benefit of cash relative to that of margins.

An interior solution, $(\alpha, \beta) \in (0, 1)^2$, would require that (9) and (10) hold as equalities. This is generically not the case, however, because the right-hand sides of (9) and (10) are generically not equal. To see why, note the following: Because the consumptions in (7) and (8) are linear in α and β , the right-hand side of (9) is independent of α and that of (10) is independent of β . Furthermore, because in (7) and (8) the terms involving α are proportional to $(1 - \beta)$, the right-hand side of (9) is also independent of β . Finally, because the ratio of the slopes in the hyperbola in $\frac{d\bar{c}}{d\beta}$ is equal to the fraction in $\frac{d\bar{c}}{d\alpha}$, α cancels when one equates the right-hand side of (9) to that of (10).

This mathematical result reflects the economic fact that α and β are substitutes: Both instruments serve the same purpose (relaxing the incentive constraint) and have the same type of drawback (the opportunity cost of not investing optimally). Either α or β has the more attractive ratio of costs to benefits, and the instrument with the worse ratio of the two is never used in the optimal contract. This is stated in the next proposition.

Proposition 2 *For $\mathcal{P} \geq 1 (> p)$, neither cash nor margins are used, $\alpha^* = \beta^* = 0$. For $\mathcal{P} < 1$, the following holds. When*

$$\text{prob}[\underline{s}] \frac{R - C - p}{p - \mathcal{P}} \leq \frac{R - C - 1}{1 - \mathcal{P}} \quad (11)$$

holds, cash is not used, $\beta^ = 0$. When condition (11) is reversed margins are not used, $\alpha^* = 0$. When (11) holds as an equality, the optimal contract is indifferent between cash and margins.*

Inequality (11) states that the ratio of opportunity costs to incentive benefits is better for margins than for cash. Thus, when (11) holds, cash holdings are never used, while in the opposite case margins are never used. Inequality (11) can be rewritten as

$$p \geq \frac{\mathcal{P}(R - C - 1) + \text{prob}[\underline{s}](R - C)(1 - \mathcal{P})}{(R - C - 1) + \text{prob}[\underline{s}](1 - \mathcal{P})} \in (\mathcal{P}, 1), \quad (12)$$

which shows that if p (the price at which assets are liquidated following the margin call) is high enough, margins offer a more attractive cost-benefit ratio than cash holdings.

3 Zero cash holdings in the market equilibrium

We can now characterize the set of parameters for which cash is not used in the market equilibrium.

Proposition 3 *If*

$$\delta \leq \frac{(R - C - 1)(R - C - \mathcal{P})}{R - C - 1 + \text{prob}[\underline{s}](1 - \mathcal{P})} \quad (13)$$

holds, optimal cash holdings in the market equilibrium are zero, $\beta^ = 0$.*

Inequality (13) states that δ is not too large, i.e., the inefficiency generating by real-locating the assets to the unsophisticated investors is not too large. Correspondingly, the unsophisticated investors buy the assets at a relatively high price. Hence, the cost-benefit ratio of margins is attractive, and margins are used instead of cash holdings.

Proofs

Proof of Proposition 1 Form the Lagrangian using the objective

$$E[u(\tilde{\theta} + \tau(\tilde{\theta}, \tilde{s}, \tilde{R}))], \quad (14)$$

the limited liability constraints (2) with multipliers $\mu_{LL}(\theta, s)$, the feasibility constraints on margins ($0 \leq \alpha \leq 1$) with $\mu_0(s)$ for $\alpha(s) \geq 0$ and $\mu_1(s)$ for $\alpha(s) \leq 1$, the feasibility constraints on cash holdings (1) with μ_0^β for $\beta \geq 0$ and μ_1^β for $\beta \leq 1$, the participation constraint (4) with multiplier μ and the incentive compatibility constraints (3) with multipliers $\mu_{IC}(s)$.

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

$$\text{prob}[\theta, s]u'(\theta + \tau(\theta, s)) - \mu \text{prob}[\theta, s] + \mu_{LL}(\theta, s) + \text{prob}[\theta|s]\mu_{IC}(s) = 0 \quad \forall(\theta, s). \quad (15)$$

Rearranging, we obtain

$$u'(\theta + \tau(\theta, s)) = \mu + \frac{\mu_{LL}(\theta, s)}{\text{prob}[\theta, s]} + \frac{\mu_{IC}(s)}{\text{prob}[s]} \quad \forall(\theta, s) \quad (16)$$

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

We conjecture, and verify later, that the limited liability constraints in $(\bar{\theta}, s)$ states are always slack. That is, $\mu_{LL}(\bar{\theta}, s) = 0$. As for the limited liability constraints in $(\underline{\theta}, s)$ states, in what follows we focus on the set of parameters under these constraints are slack.

We now show by contradiction that the participation constraint (4) binds. Suppose not. Plugging $\mu = 0$ and $\mu_{LL}(\bar{\theta}, s) = 0$ into (16) implies that $\mu_{IC}(s) > 0$ for all s . Hence, both incentive constraints bind, $E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = s] = \beta + (1 - \beta)(\alpha(s)p(s) + (1 - \alpha(s))\mathcal{P})$ for $s = \bar{s}, \underline{s}$. Therefore,

$$E[\tau(\tilde{\theta}, \tilde{s})] = E[E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s}]] = E[\beta + (1 - \beta)(\alpha(\tilde{s})p(\tilde{s}) + (1 - \alpha(\tilde{s}))\mathcal{P})] \quad (17)$$

From the participation constraint, we have

$$\begin{aligned} 0 &\leq -E[\tau(\tilde{\theta}, \tilde{s})] - \beta(R - C - 1) - (1 - \beta)E[\alpha(\tilde{s})(R - C - p(\tilde{s}))] \\ &= -E[\beta + (1 - \beta)(\alpha(\tilde{s})p(\tilde{s}) + (1 - \alpha(\tilde{s}))\mathcal{P})] - \beta(R - C - 1) - (1 - \beta)E[\alpha(\tilde{s})(R - C - p(\tilde{s}))] \\ &= -E[\beta(R - C) + (1 - \beta)\alpha(\tilde{s})(R - C) + (1 - \beta)(1 - \alpha(\tilde{s}))\mathcal{P}], \end{aligned}$$

where we used (17) to go from the first to the second line. The expression on the right-hand side of the last line is strictly negative since $R - C > \mathcal{P} > 0$, $0 \leq \beta \leq 1$, and $0 \leq \alpha(\tilde{s}) \leq 1$. A contradiction. Hence, $\mu > 0$ and the participation constraint binds.

We now show that the incentive constraint after a good signal is slack while the incentive constraint after a bad signal is binding. First note that it cannot be that both incentive constraints are slack since we assume that the first-best is not attainable, $\mathcal{P} < (\pi - \underline{\pi})\Delta\theta$. It also cannot be that both incentive constraints bind (see the argument showing that the participation constraint binds above).

We now show by contradiction that the incentive constraint following a bad signal binds. Suppose not and hence $\mu_{IC}(\underline{s}) = 0$. Since we are considering a set of parameters under which the limited liability constraints are slack, we have by (16) that

$$\begin{aligned} u'(\theta + \tau(\theta, \bar{s})) &= \mu + \frac{\mu_{IC}(\bar{s})}{\text{prob}[\bar{s}]} \\ u'(\theta + \tau(\theta, \underline{s})) &= \mu \end{aligned}$$

so that there is full risk-sharing conditional on the signal and hence

$$\tau(\underline{\theta}, s) - \tau(\bar{\theta}, s) = \Delta\theta > 0 \quad \forall s. \quad (18)$$

Moreover, since $\mu_{IC}(\bar{s}) \geq 0$, it follows that $u'(\theta + \tau(\theta, \underline{s})) \leq u'(\theta + \tau(\theta, \bar{s}))$, and thus

$$\tau(\theta, \underline{s}) \geq \tau(\theta, \bar{s}) \quad \forall \theta. \quad (19)$$

From the binding participation constraint

$$- \left[\text{prob}[\bar{s}] E[\tau(\tilde{\theta}, \tilde{s}) | \tilde{s} = \bar{s}] + \text{prob}[\underline{s}] E[\tau(\tilde{\theta}, \tilde{s}) | \tilde{s} = \underline{s}] \right] = E[\beta(R - C - 1) + (1 - \beta)\alpha(\tilde{s})(R - C - p(\tilde{s}))].$$

Since the right-hand side is non-negative, we know that

$$- \left[\text{prob}[\bar{s}] E[\tau(\tilde{\theta}, \tilde{s}) | \tilde{s} = \bar{s}] + \text{prob}[\underline{s}] E[\tau(\tilde{\theta}, \tilde{s}) | \tilde{s} = \underline{s}] \right] \geq 0$$

while $E[\tau(\tilde{\theta}, \tilde{s}) | \tilde{s} = \bar{s}] > 0$ (binding incentive constraint after a good signal). This implies that

$$E[\tau(\tilde{\theta}, \tilde{s}) | \tilde{s} = \underline{s}] < 0 \quad (20)$$

Using (18), (19) and $\bar{\pi} > \underline{\pi}$, we can write

$$\begin{aligned} 0 &< E[\tau(\tilde{\theta}, \tilde{s}) | \tilde{s} = \bar{s}] \equiv \bar{\pi}\tau(\bar{\theta}, \bar{s}) + (1 - \bar{\pi})\tau(\underline{\theta}, \bar{s}) \\ &< \underline{\pi}\tau(\bar{\theta}, \bar{s}) + (1 - \underline{\pi})\tau(\underline{\theta}, \bar{s}) \\ &\leq \underline{\pi}\tau(\bar{\theta}, \underline{s}) + (1 - \underline{\pi})\tau(\underline{\theta}, \underline{s}) \equiv E[\tau(\tilde{\theta}, \tilde{s}) | \tilde{s} = \underline{s}] \end{aligned}$$

But this contradicts (20). Thus, the incentive constraint after a good signal is slack while the incentive constraint after a bad signal binds.

We now show that there is no margin call after a good signal, $\alpha(\bar{s}) = 0$. Since we are considering a set of parameters under which the limited liability constraints are slack, the first-order conditions of the Lagrangian with respect to $\alpha(s)$ write as

$$\mu_0(s) - \mu_1(s) + (1 - \beta)\mu_{IC}(s)(p(s) - \mathcal{P}) = \mu(1 - \beta)\text{prob}[s](R - C - p(s)) \quad \forall s. \quad (21)$$

Because the incentive constraint after a good signal is slack, we have $\mu_{IC}(\bar{s}) = 0$. If $\beta < 1$, then $\mu_0(\bar{s}) - \mu_1(\bar{s}) > 0$, which implies $\alpha(\bar{s}) = 0$. If $\beta = 1$, then we can set $\alpha(\bar{s}) = 0$ without loss of generality.

When limited liability constraints are slack, there is full-risk-sharing conditional on the realization of the signal \tilde{s} :

$$u'(\theta + \tau(\theta, \tilde{s})) = \mu + \frac{\mu_{IC}(\tilde{s})}{\text{prob}[\tilde{s}]} \quad (22)$$

so that (18) holds.

Combining the binding participation constraint and the incentive constraint after a bad signal with (18) gives the following characterization of the optimal transfers:

$$\begin{aligned} \tau(\bar{\theta}, \bar{s}) = & -(1 - \bar{\pi})\Delta\theta - \frac{\beta(R - C - 1 + \text{prob}(\underline{s}))}{\text{prob}[\bar{s}]} \\ & - \frac{\text{prob}[\underline{s}](1 - \beta)[\alpha(\underline{s})(R - C) + (1 - \alpha(\underline{s}))\mathcal{P}]}{\text{prob}[\bar{s}]} \end{aligned} \quad (23)$$

$$\begin{aligned} \tau(\underline{\theta}, \bar{s}) = & \bar{\pi}\Delta\theta - \frac{\beta(R - C - 1 + \text{prob}(\underline{s}))}{\text{prob}[\bar{s}]} \\ & - \frac{\text{prob}[\underline{s}](1 - \beta)[\alpha(\underline{s})(R - C) + (1 - \alpha(\underline{s}))\mathcal{P}]}{\text{prob}[\bar{s}]} \end{aligned} \quad (24)$$

$$\tau(\bar{\theta}, \underline{s}) = -(1 - \underline{\pi})\Delta\theta + \beta + (1 - \beta)(\alpha(\underline{s})p + (1 - \alpha(\underline{s}))\mathcal{P}) \quad (25)$$

$$\tau(\underline{\theta}, \underline{s}) = \underline{\pi}\Delta\theta + \beta + (1 - \beta)(\alpha(\underline{s})p + (1 - \alpha(\underline{s}))\mathcal{P}) \quad (26)$$

It is immediate that the limited liability constraint in state $(\bar{\theta}, \bar{s})$ is slack because $\tau(\bar{\theta}, \bar{s}) < 0$ (as $R - C > 1$). To show that the limited liability constraint in state $(\bar{\theta}, \underline{s})$ is slack, we substitute $\tau(\bar{\theta}, \underline{s})$ into the limited liability constraint, which then simplifies to

$$-(1 - \underline{\pi})\Delta\theta \leq (1 - \beta)(1 - \alpha)(R - \mathcal{P}),$$

which always holds. QED

Proof of Proposition 2 Considering again the set of parameters such that limited liability constraint are slack, the first-order conditions of the Lagrangian with respect to β and $\alpha(\underline{s})$ are

$$\mu_0^\beta - \mu_1^\beta + \mu_{IC}[1 - (\alpha p + (1 - \alpha)\mathcal{P})] = \mu[(R - C - 1) - \text{prob}[\underline{s}]\alpha(R - C - p)] \quad (27)$$

$$\mu_0 - \mu_1 + (1 - \beta)\mu_{IC}(p - \mathcal{P}) = \mu(1 - \beta)\text{prob}[\underline{s}](R - C - p) \quad (28)$$

where we dropped the reference to \underline{s} in p , α and the Lagrange multipliers.

If $\beta = 1$, then we can set $\alpha(\underline{s}) = 0$ without loss of generality (as we want to show that either cash or margins are used). So we proceed with $\beta < 1$ (and hence $\mu_1^\beta = 0$).

Next, as we only consider parameters for which all limited liability constraints are slack, it must be that $\alpha < 1$. Suppose not, $\alpha = 1$. Then, using (26), $\tau(\underline{\theta}, \underline{s}) = \underline{\pi}\Delta\theta + \beta + (1 - \beta)p > \beta + (1 - \beta)p$. But this violates the limited liability constraint.

We now show that when $\mathcal{P} \geq p$, then margins are not used, $\alpha(\underline{s}) = 0$. For $\beta = 1$, we already know that margins are not used. Consider $\beta < 1$. The right-hand side of (28) is

strictly positive since $R - C > 1 > p$, $\mu > 0$ and $\beta < 1$. If $\mathcal{P} \geq p$, then $\mu_0 > 0$ must hold and $\alpha(\underline{s}) = 0$.

Next, we show that when pledgeable income is high enough, $\mathcal{P} > 1$, cash is not used. First note that if $\mathcal{P} > 1$, then $\mathcal{P} \geq p$ and $\alpha(\underline{s}) = 0$. Then, (27) simplifies to

$$\mu_0^\beta - \mu_1^\beta + \mu_{IC}(1 - \mathcal{P}) = \mu(R - C - 1). \quad (29)$$

The right-hand side of (29) is positive since $R - C > 1$ and $\mu > 0$. On the left-hand side, as $1 - \mathcal{P} < 0$ we then have $\mu_0^\beta > 0$ and hence, $\beta = 0$.

It remains to characterize what happens when $\mathcal{P} < p < 1$. Since $\mu > 0$, $\alpha < 1$, $\beta < 1$ and $\mathcal{P} < p < 1$, we can write (27) and (28) as

$$\frac{\mu_{IC}}{\text{prob}[\underline{s}]\mu} = \frac{(R - C - 1) - \text{prob}[\underline{s}]\alpha(R - C - p)}{\text{prob}[\underline{s}][1 - (\alpha p + (1 - \alpha)\mathcal{P})]} - \frac{\mu_0^\beta}{\text{prob}[\underline{s}]\mu[1 - (\alpha p + (1 - \alpha)\mathcal{P})]} \quad (30)$$

$$\frac{\mu_{IC}}{\text{prob}[\underline{s}]\mu} = \frac{R - C - p}{p - \mathcal{P}} - \frac{\mu_0}{\text{prob}[\underline{s}]\mu(1 - \beta)(p - \mathcal{P})} \quad (31)$$

and hence

$$\frac{(R - C - 1) - \text{prob}[\underline{s}]\alpha(R - C - p)}{\text{prob}[\underline{s}][1 - (\alpha p + (1 - \alpha)\mathcal{P})]} - \hat{\mu}_0^\beta = \frac{R - C - p}{p - \mathcal{P}} - \hat{\mu}_0 \quad (32)$$

where $\hat{\mu}_0 = \frac{\mu_0}{\text{prob}[\underline{s}]\mu(1 - \beta)(p - \mathcal{P})}$ and $\hat{\mu}_0^\beta = \frac{\mu_0^\beta}{\text{prob}[\underline{s}]\mu[1 - (\alpha p + (1 - \alpha)\mathcal{P})]}$. After some manipulation, (32) can be written as

$$\frac{(R - C - 1)(p - \mathcal{P}) - \text{prob}[\underline{s}](R - C - p)(1 - \mathcal{P})}{\text{prob}[\underline{s}](p - \mathcal{P})[1 - (\alpha p + (1 - \alpha)\mathcal{P})]} = \hat{\mu}_0^\beta - \hat{\mu}_0. \quad (33)$$

First, note that the denominator of (33) is strictly positive. So if

$$\frac{R - C - 1}{1 - \mathcal{P}} > \text{prob}[\underline{s}]\frac{R - C - p}{p - \mathcal{P}} \quad (34)$$

the left-hand side of (33) is strictly positive. Then $\hat{\mu}_0^\beta > 0$ (as $\hat{\mu}_0 \geq 0$) and hence, $\beta = 0$.

Similarly, when the inequality in (33) is reverse, we have $\alpha = 0$. Finally, when (33) holds as an equality, then there is no difference between cash and margins. QED

Proof of Proposition 3 Suppose, contrary to the claim that (13) holds and yet $\beta^* > 0$. Condition (13) is equivalent to

$$\frac{\mathcal{P}(R - C - 1) + \text{prob}[\underline{s}](R - C)(1 - \mathcal{P})}{(R - C - 1) + \text{prob}[\underline{s}](1 - \mathcal{P})} \leq R - C - \delta. \quad (35)$$

When $\beta^* > 0$, then $\alpha^* = 0$ and the market clearing price p must be such that (Proposition 2)

$$p < \frac{\mathcal{P}(R - C - 1) + \text{prob}[\underline{s}](R - C)(1 - \mathcal{P})}{(R - C - 1) + \text{prob}[\underline{s}](1 - \mathcal{P})}. \quad (36)$$

Since the supply of assets is 0, the market must clear with 0 trade. This implies the price must be such that

$$R - C - \delta \leq p. \quad (37)$$

Together, (35), (36) and (37) imply

$$p < \frac{\mathcal{P}(R - C - 1) + \text{prob}[\underline{s}](R - C)(1 - \mathcal{P})}{(R - C - 1) + \text{prob}[\underline{s}](1 - \mathcal{P})} \leq R - C - \delta < p,$$

a contradiction. QED