Optimal margins and equilibrium prices *

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

This version: January 2015

PRELIMINARY DRAFT

Abstract

We study the interaction between optimal contracts and equilibrium pricing when risk averse agents buy insurance from protection sellers subject to moral hazard. In the optimal contract, information arrival triggers margin calls, depressing equilibrium prices. When risk-aversion is large and moral hazard is severe, margin calls generate downward sloping supply curves which, in turn, can lead to multiple equilibria. Because all players maximize expected utility and contracts are optimal, welfare is well-defined. We therefore analyze the private and social costs and benefits of margins. Privately optimal margins, and the associated fire-sales arising in equilibrium, are larger than their second-best counterparts. This reflects negative externalities: to obtain more insurance, a protection buyer requires larger margins, depressing prices, thus negatively affecting other protection buyers. Capping margins can restore optimality.

JEL classification: G21, G22, D82

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

*We would like to thank audiences at Imperial College, the European Central Bank and the European Winter Finance Conference for their comments and suggestions. Jana Urbankova provided excellent editing assistance. The views expressed do not necessarily reflect those of the European Central Bank or the Eurosystem.

†CNRS CRM, FBF-IDEI Chair on financial markets and investment banking, 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 Introduction

Counterparty risk is a major concern in derivative markets. For example, when Lehman Brothers filed for bankruptcy protection in September 2008, it froze the positions of more than 900,000 derivative contracts (about 5% of all derivative transactions globally) in which Lehman Brothers was a party (Fleming and Sarkar, 2014). The standard tool to mitigate counterparty risk is to use margins, which are a form of collateral in the context of derivative contracts. The immediate response of regulators and law-makers to the financial crisis therefore was to require a significant expansion of the use of margins in derivative activity (Dodd-Frank Act in the US, EMIR in the EU). However, there is a growing awareness that margins can be pro-cyclical (BIS, 2010). Margins calls, which occur when concerns about counterparty risk increase, can lead to assets sales that exert downward pressure on market prices with further adverse consequences for market participants.

This paper evaluates the benefit and cost of margin requirements. The benefit of margins is more risk-sharing in the presence of potential counterparty risk. Potential counterparty risk arises because the seller of insurance via derivatives is subject to moral hazard. The cost of margins is a fire-sale externality. Margin calls lead to asset sales that lower asset prices, which in turn make margin calls more expensive. We evaluate the benefit and cost of margins in a model with optimal contracts and endogenous asset prices. The optimal derivative contract limits moral hazard and reflects the equilibrium market price of assets.

Our model features risk-averse agents who want to insure against a common exposure to risk (protection buyers) and risk-neutral agents who are not directly exposed to the risk of buyers and who offer insurance (protection sellers). A key feature of our model is that new contractible information about the insured risk becomes available during the life of the insurance contract. Such news affect the expected pay-offs of the contracting parties: it makes the contract an asset for one party and a liability for the other.

Protection sellers have a moral hazard problem. They have risky assets and, because of limited liability, can make insurance payments only if these assets are sufficiently valuable. The value of a protection seller's assets depends on her effort to prevent downside risk. Without exerting costly effort, the value of her assets may be so low that she may be unable to make insurance payments. Negative news about the insured risk turns the derivative contract into a liability for a protection seller, and undermines her incentive to exert the risk-prevention effort, exposing the protection buyer to counterparty risk.

The benefit of margins is they ring-fence assets from moral hazard, as in Biais, Heider and Hoerova (2012a). A margin call requires a protection seller to deposit safe, liquid assets in a third-party escrow account (e.g., at a central counterparty). As protection seller's assets are risky, opaque and subject to moral hazard, she has to sell some of her assets in exchange for assets that can be deposited in the margin account. After a margin call, a protection seller has fewer risky assets subject to moral-hazard and the safe, liquid assets on the margin account are outside of her control. This incentive benefit of margins enhances the scope for risk-sharing via derivatives.

The cost of margins is a fire-sale externality, as in Gromb and Vayanos (2002) and Brunnermeier and Pedersen (2009). A margin call triggers the sale of protection sellers' assets and increases the supply of these assets in the market. There is a continuum of arbitrageurs willing to buy these assets but they value them less than the protection sellers (e.g., because they are less skilled at managing them). Asset sales are inefficient as they occur at fire-sale prices. Protection sellers anticipate the cost of selling assets in case of a margin call and demand compensation. This compensation makes the derivative contract more expensive.

Key to our analysis of the benefit and cost of margins is the interaction of optimal contracting and equilibrium pricing. On the one hand, the optimal contract limits moral hazard and rationally anticipates future asset prices. On the other hand, the asset price is determined by market clearing and reflects the supply of the asset that is triggered by margin calls. Thus, there is a rational expectations loop, of which the optimal contract and the equilibrium price are fixed points.

Our model has two important implications that we explore in depth. First, margin requirements may create financial instability via multiple equilibria. Second, we show that the private market outcome is not socially efficient: margins are used too much.

Our model admits multiple, welfare-ranked equilibria. Multiple equilibria arise because the supply of assets in the market, induced by margin calls, can be downward-sloping. The proceeds from margin calls are used to offer more incentive-compatible insurance to the protection buyer after a bad signal. If the price decreases, proceeds decrease taking the size of the margin call as given. If the protection buyer is very risk averse, he wants to offset this by raising the amount of the margin. Multiple equilibria give rise to self-fulfilling asset price drops. If market participants expect the price to be high, they request small margins.

Small asset sales generate enough funds to be deposited on the margin account. Because the asset sale is small, it does not depress the price much, which confirms the initial expectation of high prices. In contrast, if market participants expect the price to be low, they request large margins, which depresses prices via fire-sales, again confirming the initial expectation. If market participants are pessimistic, and focus on the equilibrium with low prices, this amplifies fire-sales. When there are multiple equilibria, protection buyers prefer the high-price equilibrium, but arbitrageurs prefer the low price-equilibrium. Yet, the high-price equilibrium dominates the low-price one in terms of utilitarian welfare. While a low price leads to profits for arbitrageurs, these profits are lower than the utility cost to protection buyers (protection sellers always break-even) because it is inefficient to have arbitrageurs hold the assets of protection sellers.

Because all agents have well-defined expected utility which they maximize using optimal, complete contracts, welfare is well-defined. Thus, we can characterize the second-best and compare it to the market equilibrium. We show that market equilibrium has larger margin calls than the second-best. Over-margining arises because of a negative (pecuniary) externality. When the protection buyer wants to obtain more insurance, he increases the margin in his optimal contract. The seller must liquidate assets to satisfy margin calls. This lowers the equilibrium price, which makes other protection sellers worse off. This negative externality implies that market equilibrium is not second-best, as in Greenwald and Stiglitz (1986). Note that over-margining arises even when the equilibrium is unique. When there are multiple equilibria, they are all different from the second-best and all involve over-margining.

This welfare analysis shows that in equilibrium too much insurance is sold, implying that aggregate margin calls are too large. The second-best can be achieved by imposing a cap on margins. This also eliminates the market instability when there are multiple equilibria. When the regulator/central bank caps margins, the asset market equilibrium is unique.

While our model illustrates how over-margining can arise in equilibrium, we do not argue that in reality markets always choose excessively large margins. In practice other forces are at play which, for simplicity and clarity, we do not include in the present model. For example, if protection buyers 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, 2012b). To avoid this, the regulator can impose a floor on margins. In sum, in the presence of moral hazard, market forces do not always lead to information-constrained efficient amount of margins. Hence, regulatory intervention may be needed, for example via caps and floors on margins.

We also want to issue a warning about the implementation of the margin-cap policy. Suppose some market participants have already entered derivative contracts before the cap is introduced. These contracts entail large insurance payments, made incentive-compatible by high margins. Now suppose the regulator caps margins, as we suggest above. Clearly, the cap should be applied to new contracts, but should it also be applied to the pre-existing ones? If it is applied to both, while leaving contracted transfers unchanged, then the risk-prevention effort may no longer be incentive-compatible, and many protection sellers may end up in default. Therefore, if a margin-cap is introduced, old contracts should either be exempt or the transfers they promise should be revised downwards.

The empirical implications of our theoretical analysis reflect the interaction of optimal contracting and equilibrium asset pricing. For example, our model generates contagion from the market for protection buyer assets to the market for the protection seller assets. The arrival of bad news about protection buyer assets triggers margin calls, and hence fire-sales for protection seller assets. In another example, the greater the risk aversion of protection buyers, the volatility of protection buyer assets, or the moral hazard problem of protection sellers, the larger are the required margins and the associated fire-sales, and the greater is the scope for decreasing supply curves and the emergence of "bad equilibria".

The next section surveys the related literature. Section 3 describes the model and presents the first-best benchmark. Section 4 analyzes optimal margining under moral hazard. The analysis in that section builds on the analysis in Biais, Heider and Hoerova (2012a). Section 5, which derives the market equilibrium and the second-best equilibrium, and compares the two, is the key contribution of the present paper. Section 6 discusses the empirical and policy implications of our analysis. The proofs are in the appendix.

2 Literature

In Biais, Heider and Hoerova (2012a), we show that margins are features of optimal derivative contracts when risk averse agents buy insurance from protection sellers subject to moral hazard. Margins ring-fence assets from moral hazard and are thus privately optimal for two contracting parties. In that analysis, we interpret margins as an institutional arrangement that affects split of a protection seller's balance sheet between transparent assets and opaque investments, which are then subject to moral hazard. In the current paper, we examine the consequences for asset prices when protection sellers sell some of their risky, opaque assets in order to obtain safe, transparent ones for the margin deposit. With endogenous asset prices, margin calls increase the supply of assets and depress asset prices with consequences for the cost of margins for all market participants.

The analysis of the interaction between liquidation induced by financial constraints and equilibrium prices goes back, at least, to Kiyotaki and Moore (1997) and Shleifer and Vishny (1992). In their 2011 survey, Shleifer and Vishny write:

> "a fire-sale is essentially a forced sale of an asset at a dislocated price... The price is dislocated because the highest potential bidders are typically involved in a similar activity as the seller, and ... cannot .... buy the asset ... assets are then bought by nonspecialists who ... are only willing to buy at valuations that are much lower."

The fire-sales and inefficiencies arising in our model are in line with this characterization. Differences between our analysis and that of Shleifer and Vishny (1992) include our focus on i) margins, ii) optimal contracting and iii) information-constrained Pareto optimality.

Gromb and Vayanos (2002) offer the first analysis of how margin/collateral constraints depress prices in financial markets, giving rise to pecuniary externalities, and driving the equilibrium away from information-constrained efficiency.[1] Suppose there is a liquidity shock so that some investors must sell their holdings of an asset. This will generate a drop in the price, unless arbitrageurs step in and buy. Arbitrageurs, however, can't freely do so because they are subject to margin/collateral constraints. More precisely, the amount each can buy is limited by an upper bound, increasing in his own wealth. Now, this wealth is evaluated at current prices. So if the arbitrageur is long in the asset, the lower the price, the tighter the constraint, the less the arbitrageur can buy. This generates pecuniary externalities: If one arbitrageur is constrained and cannot buy, this depresses the price. This depressed price tightens the margin/collateral constraint of the other arbitrageurs. Because of these pecuniary externalities, equilibrium is not efficient. The major difference between this analysis and ours is that, in Gromb and Vayanos (2002), margin/collateral constraints are exogenous, while in our model they are endogenous and emerge as features of the optimal

[1] Gromb and Vayanos (2010) is a very interesting survey of the literature on limits to arbitrage, including an illuminating presentation of a simplified version of Gromb and Vayanos (2002).

contract in the presence of moral hazard. This enables us to study privately and socially optimal margins and the tradeoff between the costs and benefits of margins.

This comment also applies to Brunnermeier Pedersen (2009) where, similarly to Gromb and Vayanos (2002), margin constraints are exogenous. In addition, the economic mechanism linking margins and equilibrium price is different in Brunnermeier and Pedersen (2009) and in our paper. In their analysis, market participants are learning about volatility. When they observe a large price drop, they increase their estimate of the volatility. Because volatility is higher, margins are raised. This triggers fire-sales amplifying the initial price drop. Because our economic mechanism is different, we get different implications: From a normative point of view, modelling private and social costs and benefits of margins yields our implication on over-margining. From a positive point of view, our implication that larger protection buyers' risk aversion or protection sellers' moral hazard increase margins, fire-sales and the scope for equilibrium multiplicity differs from the implications of Brunnermeier and Pedersen (2009).

Lorenzoni (2008) and Hombert (2009) also study pecuniary externalities associated with collateral, but in a different context. In Lorenzoni (2008), entrepreneurs raise funds to invest. Then, if there is a bad aggregate shock, all entrepreneurs need more cash to salvage their projects. Because of limited commitment, entrepreneurs cannot raise new debt at this point, while outside investors cannot credibly promise to insure entrepreneurs against the negative shock. Hence, when the negative shock hits, entrepreneurs must sell assets to raise money to salvage their project. These fire-sales depress the price and are inefficient because they allocate the asset to outside agents who value it less than entrepreneurs. As in Gromb and Vayanos (2002), this gives rise to pecuniary externalities. When one entrepreneur invests a lot initially, this implies he must sell a lot after the bad shock, which depresses the price. This depressed price is costly for the other entrepreneurs, because it forces them to sell more assets to raise the same amount of cash. Because of this negative externality, equilibrium is not efficient, more precisely, equilibrium prices are too low. In contrast, Hombert (2009) show that equilibrium prices can be too high relative to the second-best. He identifies two possible sources of externalities: On the one hand, when firms liquidate their assets, they reduce the pledgeable income of other firms. This collateral effect is similar to that in Lorenzoni (2008). On the other hand, depressed prices offer attractive investment opportunities to entrepreneurs who exerted high effort initially and succeeded. This incentive effect, which differs from that in Lorenzoni, can outweigh the collateral effect, implying that low prices increase welfare. The major difference between the analyses of Lorenzoni (2008) and Hombert (2009) and ours is that they consider real-economy firms borrowing funds against initial collateral. This differs from our analysis of risk-sharing in financial markets with variation margins.

This difference in economic objects also differentiates our analysis from those of Acharya and Viswanathan (2011) and Kuong (2014). Acharya and Viswanathan (2011) study the equilibrium price at which borrowers resell their assets to overcome credit rationing, and analyze the negative externality induced by fire-sales. In Kuong (2014), as in our paper, there is an interaction between equilibrium prices and optimal contracts. Investment in his model (insurance in ours) can lead to collateral liquidation (margin calls in our analysis) and generate fire-sales and the associated negative externality. There are further important differences with both of these papers.

In our analysis, optimal contracts feature variation margins, called after the arrival of new information (unavailable at the time at which the contract was signed), but generating asset liquidation before effort is exerted. Thus, in our context, margins are beneficial because they directly relax the incentive-compatibility constraint of the agent, although they tighten the participation constraint of the principal. In contrast, in Kuong (2014), as in Acharya and Viswanathan (2011), collateral is supplied when the contract is signed, but liquidated after effort is exerted. Thus, collateral has no direct effect on incentives. Its direct effect is to relax the participation constraint of creditors. This reduces the repayment they request which, in turn, relaxes the incentive constraint of the borrower. Another difference between our analysis and those of Kuong (2014) and Acharya and Viswanathan (2011) is that they don't study the second-best, while the characterization of the second-best and the finding that it differs from the equilibrium are key contributions of the present paper. Relatedly, another difference between our analysis and that of Kuong (2014) pertains to policy implications. In Kuong (2014), high price equilibria can coexist with less efficient low price equilibria, implying that a commitment by the central bank to support prices will improve efficiency by eliminating the bad (low price) equilibria. While our analysis may also feature multiple equilibria, we additionally show that market equilibria are inefficient as they involve excessive margining, implying that a cap on margins will improve efficiency, eliminate multiplicity and, in fact, achieve the second-best.

3 Model and First-Best Benchmark

3.1 The model

There are three dates, $t = 0, 1, 2$, a mass-one continuum of protection buyers, a mass-one continuum of protection sellers, a mass-one of arbitrageurs. At $t = 0$, each protection buyer is matched with a protection seller and they contract. At $t = 1$, margining and trading decisions are made. At $t = 2$, payoffs are received.

Players and assets. Protection buyers are identical, with twice differentiable concave utility function u, and are endowed with one unit of an asset with random return $\tilde{\theta}$ at $t = 2$.[2] For simplicity, we assume $\tilde{\theta}$ can only take on two values: $\bar{\theta}$ with probability π and $\underline{\theta}$ with probability $1 - \pi$, and we denote $\Delta\theta = \bar{\theta} - \underline{\theta}$. The risk $\tilde{\theta}$ is the same for all protection buyers.[3]

Protection buyers seek insurance against the risk $\tilde{\theta}$ from protection sellers who are risk-neutral and have limited liability. Each protection seller j has an initial endowment of one unit of a risky asset returning $\tilde{R}_j$ at $t = 2$. This payoff is affected by a protection seller's risk-management decision at $t = 1$. To model risk-management in the simplest possible way, we assume that each seller j can undertake a costly effort to make her assets safer. If she undertakes 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 $1 - \mu$. The risk-management process reflects the unique skills of the protection seller and is therefore difficult to observe and monitor by outside parties. Combined with limited liability, effort unobservability generates moral hazard.

Exerting the risk-prevention effort costs C per unit of assets under management at $t = 1$.[4] Because protection seller assets are riskier without costly effort, we also refer to the decision

[2] The concavity of the objective function of the protection buyer can reflect institutional, financial or regulatory constraints, such as leverage constraints or risk-weighted capital requirements. For an explicit modeling of hedging motives see Froot, Scharfstein and Stein (1993).

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

[4] We show in Biais, Heider and Hoerova (2012a) that the qualitative results are unchanged when C is convex in the amount of assets under management.

not to exert effort as "risk-taking".[5] Undertaking effort is efficient,

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

i.e., the expected net return is larger with effort than without it. We also assume that when a protection seller exerts risk-prevention effort, return on her assets is higher than one (return on cash),

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

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

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

Each arbitrageur k values one unit of the risky asset at $v_k < R - C$. We assume v_k is distributed over $[x, 1]$.[6] Arbitrageurs do not insure protection buyers (e.g., because they are infinitely risk averse or because they do not have the information and trading technology to do so). For simplicity, we assume all protection buyers value the asset at x, so that it is not optimal that they buy it or obtain it from the protection sellers and then hold it.

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

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

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

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

We assume that $\lambda \geq \frac{1}{2}$. If $\lambda = \frac{1}{2}$, then $\bar{\pi} = \pi = \underline{\pi}$ and the signal is completely uninformative. If $\lambda > \frac{1}{2}$, then $\bar{\pi} > \pi > \underline{\pi}$, i.e., observing a good signal $\bar{s}$ increases the

[5]Here effort improves returns in the sense of first-order stochastic dominance. We have checked that our results are robust when effort improves returns in the sense of second-order stochastic dominance, so that lack of effort corresponds to risk-shifting.

[6]Alternatively, we could assume the upper bound of the support of v_k is $R - C$. The reason we assume the upper bound is 1 is that this setting nests Biais, Heider, Hoerova (2012).

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 protection buyer makes a take–it–or–leave–it contract offer to the protection seller. Similar results would hold if, instead, we assumed the protection seller had (some or all the) bargaining power. The contract specifies a transfer τ at time 2 between the protection seller and the protection buyer. When $\tau > 0$ the protection seller pays the protection buyer 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 seller'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. To satisfy the margin call, a protection seller can liquidate a fraction $\alpha(\tilde{s}) \in [0, 1]$ of her assets by selling them at price p per unit and deposit the resulting cash on a margin account. The cost of such deposits is that their liquidation value is lower than what it could have been had the assets remained under the management of the protection seller $R - C > 1 > p$. In Biais, Heider and Hoerova (2012a), the price was exogenous and normalized to one. The analysis in the present paper considers endogenous prices, set by market clearing conditions.

Yet margins also have advantages. Our key assumption is that the cash deposited in the margin account is safe and no longer under the discretion of the protection seller, i.e., it is ring-fenced from moral hazard. Furthermore, if the protection seller defaults, the cash on the margin account can be used to pay the protection buyer.

Margin accounts can be implemented as escrow accounts set up by the protection buyer 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.

Transfers from protection sellers are constrained by limited liability,

$$\tau(\tilde{\theta}, \tilde{s}, \tilde{R}) \leq \alpha(s)p + (1 - \alpha(s))R, \quad \forall(\theta, s, R). \tag{4}$$

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

Asset market. The supply of the asset, denoted by S, is given by the asset liquidations of protection sellers due to margin calls:

$$S = \alpha\left(s\right). \tag{5}$$

The demand for the asset, denoted by $D\left(p\right)$, comes from a mass of arbitrageurs with valuation $v_k \geq p$:

$$D\left(p\right) = 1 - F\left(p\right) \tag{6}$$

Market clearing at $t = 1$ requires that the supply of the asset is equal to the demand for the asset at price p:

$$p = F^{-1}\left(1 - \alpha\left(s\right)\right). \tag{7}$$

Note that the price $p \leq 1$ since $v_k \leq 1$. In the special case of the uniform distribution of v_k over $[x, 1]$, we have $D\left(p\right) = \frac{1-p}{1-x}$ so that

$$p = 1 - \left(1 - x\right)\alpha\left(s\right).$$

The case where $p = 1$, analyzed in Biais, Heider and Hoerova (2012a) arises in the limit when x goes to 1. If there were only protection sellers and arbitrageurs, protection sellers would always keep the asset, since their value for the asset would exceed that of the arbitrageurs, as $R - C > 1$. The sequence of events is summarized in Figure 1.

Insert Figure 1 here

3.2 First-best: observable effort

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

In the first-best, protection sellers are requested to exert risk-prevention effort when offering protection since doing so increases the resources available for risk-sharing (see (1)). Margins are not used because they are costly (see (2)) and offer no benefit when risk-prevention effort is observable. The transfers are chosen to maximize buyers' utility

$$E[u(\tilde{\theta} + \tau(\tilde{\theta}, \tilde{s}, \tilde{R})] \tag{8}$$

subject to the limited liability constraints (4), as well as the constraint that protection sellers accept the contract. By accepting (and exerting effort) sellers obtain $R - C - E[\tau(\tilde{\theta}, \tilde{s}, \tilde{R})]$. If they do not sell protection, they obtain $R-C$.[7] Therefore, a protection seller's participation constraint in the first-best is

$$E[\tau(\tilde{\theta}, \tilde{s}, \tilde{R})] \leq 0. \tag{9}$$

In the first-best, protection sellers exert risk-prevention effort. In this context, the return $\tilde{R}$ is always equal to R and we drop the reference to the return in the transfers τ for ease of notation. As shown in Biais, Heider and Hoerova (2012a), the optimal contract provides full insurance, is actuarially fair and does not react to the signal. Margins are not used and the transfers are given by

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

The first-best insurance contract is actuarially fair since the expected transfer from protection sellers is zero, $E[\tau(\tilde{\theta}, \tilde{s})] = 0$.

4 Optimal margins under moral hazard

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

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

The left-hand side is a protection seller'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 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, her 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 protection seller when $\tilde{R}=0$: $\tau(\tilde{\theta}, \tilde{s}, 0) = \alpha(\tilde{s})p$. This rationalizes

[7] Without derivative trading, protection sellers always exert effort since it is efficient to do so (see condition (1)).

the stylized fact that, in case of default of a protection seller, margin deposits are ceized and used to pay protection buyers.

With effort, protection seller 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

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

or, using the notion of "pledgeable return" $\mathcal{P}$ (see Holmström and Tirole, 1997),

$$\mathcal{P} \equiv R - \frac{C}{1-\mu}, \tag{10}$$

the incentive compatibility constraint rewrites as

$$\alpha(s)p + (1-\alpha(s))\mathcal{P} \geq E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = s]. \tag{11}$$

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

For sufficiently high levels of $\mathcal{P}$, the incentive-compatibility constraints are not binding at the first-best allocation. As shown in Biais, Heider and Hoerova (2012a), even if effort is not observable, the first-best can be achieved if and only if the pledgeable income is high enough, in the sense that

$$\mathcal{P} \geq (\pi - \underline{\pi})\Delta\theta = E[\tilde{\theta}] - E[\tilde{\theta}|\underline{s}]. \tag{12}$$

In what follows, we will focus on the case when the first-best cannot be reached, i.e., when

$$\mathcal{P} < (\pi - \underline{\pi})\Delta\theta = E[\tilde{\theta}] - E[\tilde{\theta}|\tilde{s} = \underline{s}]. \tag{13}$$

The participation constraint of the protection seller 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 protection sellers exert effort on the equilibrium path, we have $\tilde{R} = R$ 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

$$-E[\tau(\tilde{\theta}, \tilde{s})] \geq E\left[\alpha(\tilde{s})(R - C - p)\right], \tag{14}$$

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

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

$$R > \bar{\pi}\Delta\theta - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}\mathcal{P} \tag{15}$$

$$1 - \frac{\underline{\pi}\Delta\theta}{R - \mathcal{P}} > \frac{(1-\underline{\pi})R - \mathcal{P}}{\underline{\pi} + (1-\underline{\pi})R - \mathcal{P}} \tag{16}$$

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, 2012a, for details).

As shown in Biais, Heider and Hoerova (2012a), 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):

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

$$E[\tau(\tilde{\theta}, \tilde{s})|\tilde{s} = \bar{s}] = -\frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}\left[\alpha(\underline{s})\,(R - C) + (1 - \alpha(\underline{s}))\mathcal{P}\right]. \tag{18}$$

Finally, as also shown in Biais, Heider and Hoerova (2012a), the optimal contract provides full insurance conditional on the signal: For a given realization of the signal, the consumption of the protection buyer at time 2 is independent of the realization of θ. More precisely, the transfers are given by

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

The key difference to the first-best contract is that the transfers now depend on the signal. To preserve the seller's incentives to exert effort, the buyer must reduce the amount of insurance after a bad signal, $\tau(\underline{\theta},\underline{s}) < \tau(\underline{\theta},\bar{s})$, and thus accept incomplete risk-sharing. Hence, the protection buyer bears signal risk. Conditional on the signal, the optimal contract provides full insurance against the underlying risk $\tilde{\theta}$:

$$\tau(\underline{\theta},\bar{s}) - \tau(\bar{\theta},\bar{s}) = \tau(\underline{\theta},\underline{s}) - \tau(\bar{\theta},\underline{s}) = \Delta\theta > 0 \tag{19}$$

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

$$\text{prob}[\bar{s}]u(E[\theta|\bar{s}] + E[\tau(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]) + \text{prob}[\underline{s}]u(E[\theta|\underline{s}] + E[\tau(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}]). \tag{20}$$

To simplify the notation, let

$$\begin{aligned}\bar{c} &\equiv E[\theta|\bar{s}] + E[\tau(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]\\ \underline{c} &\equiv E[\theta|\underline{s}] + E[\tau(\tilde{\theta},\tilde{s})|\tilde{s}=\underline{s}]\end{aligned}$$

We now turn to the determination of the optimal margin call after a bad signal. To analyze the amount of margin calls, it is useful to consider the ratio of the marginal utility of a protection buyer after a bad and a good signal. Denoting this ratio by φ, and using (17) and (18), we have

$$\varphi(\alpha(\underline{s}),p) = \frac{u'(\underline{c})}{u'(\bar{c})} = \frac{u'\left(E[\theta|\underline{s}] + \alpha(\underline{s})p + (1-\alpha(\underline{s}))\mathcal{P}\right)}{u'\left(E[\theta|\bar{s}] - \frac{\text{prob}[\underline{s}]}{\text{prob}[\bar{s}]}\left[\alpha(\underline{s})\left(R-C\right) + (1-\alpha(\underline{s}))\mathcal{P}\right]\right)}. \tag{21}$$

In the first-best, there is full insurance, margins are not used and φ is equal to 1. With moral hazard, protection buyers are exposed to signal risk. This makes insurance imperfect and drives φ above one. By (17), $E[\tau(\tilde{\theta},\tilde{s})|\tilde{s}=\bar{s}]$ is decreasing in $\alpha(\underline{s})$. Hence the denominator of φ is increasing in $\alpha(\underline{s})$. On the other hand, the numerator of φ is decreasing in $\alpha(\underline{s})$. Hence, φ is decreasing in $\alpha(\underline{s})$. Higher margins reduce φ, as they reduce the wedge between consumption after a good and a bad signal, i.e., they improve insurance against signal risk. Optimal margins trade-off this benefit with their cost: assets in the margin account are less profitable than under the management of a protection seller exerting effort. Note that it is never optimal to liquidate all assets of a protection seller, i.e., the optimal margin is always smaller than one (see Biais, Heider and Hoerova, 2012).

Maximizing (20) with respect to $\alpha(\underline{s})$ and using (17) and (18), while taking the price p as given, the optimal margin after bad news (if it is interior) is implicitly given by the following condition:

$$\varphi(\alpha^*(\underline{s}), p) - 1 - \frac{R - C - p}{p - \mathcal{P}} = 0. \tag{22}$$

The last term on the left-hand side of (22) reflects the trade-off between the cost and benefit of margins. The numerator, $R - C - p$, is the opportunity cost of depositing a margin. The denominator goes up as $\mathcal{P}$ decreases, i.e., as the incentive problem gets more severe. Equation (22) is illustrated in Figure 2. Also, it will be useful hereafter to note that Equation (22) re-writes as

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

Insert Figure 2 here

5 Equilibrium and optimality

We first study the market equilibrium with optimal margins. We then derive the second-best with margins, and compare it with the market equilibrium.

5.1 Market equilibrium

5.1.1 Existence

The supply of the asset at time 1 is zero after a good signal, and the equilibrium price is $p^* = 1$. After a bad signal, the supply is the amount of margin calls $\alpha^*(\underline{s})$. While, at $t = 1$, the supply is a fixed number, at $t = 0$, margin calls are optimally set by contracting parties rationally anticipating the equilibrium price. 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, $S(p) = \alpha^*(\underline{s})$.

When parties anticipate a price lower than the pledgeable income, they choose not to use margins. Thus, for any $p < \mathcal{P}$, $S(p) = 0$. Denote by $\hat{p}$ the price such that $\varphi(0) = \frac{R-C-\mathcal{P}}{p-\mathcal{P}}$,

$$\hat{p} \equiv \mathcal{P} + \frac{R - C - \mathcal{P}}{\frac{u'(\underline{c})}{u'(\overline{c})}}. \tag{24}$$

Now, $\alpha^*(\underline{s}) = 0$ whenever

$$\varphi(0) \leq \frac{R - C - \mathcal{P}}{p - \mathcal{P}} \tag{25}$$

because and the right-hand side of (25) is decreasing in p. Hence, for any $p < \hat{p}$, we still have $S(p) = 0$. For $p \geq \hat{p}$, $\alpha^*(\underline{s}) > 0$, with $\alpha^*(\underline{s})$ given by (23). As shown in the appendix, building on this analysis, one obtains the following characterization of the equilibrium.

Proposition 1 (Existence) *Equilibrium exists. If $D(\hat{p}) > 0$, the optimal margin is interior, and given by*

$$\alpha^*(\underline{s}) = \varphi^{-1}\left(\frac{R - C - \mathcal{P}}{p - \mathcal{P}}\right) \tag{26}$$

while the market clearing price is

$$p^* = F^{-1}\left(1 - \alpha^*(\underline{s})\right) > \hat{p}. \tag{27}$$

5.1.2 Uniqueness

We now investigate if equilibrium is unique. The demand curve is decreasing, but, as shown below, the supply curve can be non–monotonic. This can generate multiplicity, and we discuss its economic interpretation.

Using equation (26) and using the implicit function theorem we show in appendix (see (32)) that the supply function $S(p)$ is increasing if and only if:

$$\frac{R - C - \mathcal{P}}{(p - \mathcal{P})^2} > \alpha^*(\underline{s})\rho(\underline{c})\,\varphi(\alpha^*(\underline{s}), p). \tag{28}$$

where $\rho(\underline{c})$ denotes the coefficient of the absolute risk aversion. Using (23), (28) is equivalent to

$$\alpha^*(\underline{s}) = S(p) < \frac{1}{\rho(\underline{c})\,(p - \mathcal{P})}. \tag{29}$$

Thus, if (29) holds, then higher price p leads to an increase in the supply of the asset $\alpha(p)$. Conversely, if

$$\alpha^*(\underline{s}) = S(p) \geq \frac{1}{\rho(\underline{c})\,(p - \mathcal{P})} \tag{30}$$

holds, then higher price p leads to a decrease in the supply of the asset $\alpha(p)$. These two cases are illustrated in Figure 3. In Panel A, (29) holds for all $p \in [\hat{p}, 1]$. In Panel B, (29) initially holds for relatively low values of p, but then, for larger values of p, (30) 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 given by

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

and we state, in the following propositions, that if is low supply is increasing and equilibrium unique, while if is large supply can be non–monotonic, giving rise to multiple equilibria.

Proposition 2 (Sufficient condition for uniqueness) *Suppose utility is exponential. If the coefficient of the absolute risk aversion of protection buyers 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 protection buyers* 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*,* $\frac{\partial\alpha}{\partial p} < 0$*.*

The intuition is as follows. When protection buyers are very risk-averse, they care a lot about their consumption after bad news, $\underline{c}$, which is determined by margins. Although margins carry an opportunity cost, this cost is paid with consumption after good news, $\bar{c}$, which is less important for risk-averse protection buyers. When the price p goes down, $\alpha(\underline{s})p$ decreases, and so does consumption after bad news (17). Therefore, if the protection buyer 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 of risk aversion on supply.

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, again confirming the initial expectation. These two regimes, and the possibility of multiple equilibria, are illustrated in Figure 4.

Insert Figure 4 here

The next proposition states that, when there are multiple equilibria, they are Pareto-ranked.

Proposition 5 *If there are multiple equilibria, they are Pareto-ranked from the point of view of protection buyer-protection seller pair, with the high price-low margin equilibrium being the preferred one.*

A lower price is preferred from the point of view of arbitrageurs as a higher price decreases their payoff. The welfare of arbitrageurs is given by $(1 - F(p))(E[v \mid v > p] - p)$, with the underlying utility of arbitrageurs given by $\max[0, v - p]$. Yet, the equilibrium with the highest price dominates the other equilibria in terms of utilitarian welfare. This is because while arbitrageurs make profits thanks to low prices, these profits are lower than the utility cost to the other market participants because it is inefficient to have arbitrageurs buy the asset.

5.2 Second-best

In equilibrium, protection buyers maximize their objective, (20) at time 0, to determine margins. Margins, in turn, determine supply, and therefore the equilibrium price, at time 1. Because they are competitive, individual protection buyers don't take into account the aggregate effect of their individual margins on the market clearing price. Yet, when one protection buyer increases the margin she requests, she exerts a negative externality on the others, by pushing the price down. Under symmetric information, this pecuniary externality would not reduce welfare, but under information asymmetry it does, as in Greenwald and Stiglitz (1986). Thus, as we'll show below, the equilibrium is not information–constrained Pareto optimal. In contrast to the market equilibrium, the second–best entails socially optimal margins, internalizing pecuniary externalities.

More precisely, the second best is obtained by maximizing (20) with respect to $\alpha(\underline{s})$, substituting the optimal transfers (17), (18), and the market clearing price (7). When the optimal margin is interior, it is pinned down by the following optimality condition:

$$\varphi(\alpha^{SB}(\underline{s})) = \frac{R - C - \mathcal{P}}{p^{SB} - \mathcal{P} + \alpha^{SB}(\underline{s})\frac{\partial p^{SB}}{\partial \alpha}} \tag{31}$$

where $p^{SB} = F^{-1}\left(1 - \alpha^{SB}\right)$. (31) is very similar to (23). The difference is that, in (31), there is an additional term: $\frac{\partial p^{SB}}{\partial \alpha}$, capturing the external effect of margins on prices.

We can now state our key result that compares the second-best equilibrium outcome with the market equilibrium outcome.

Proposition 6 (Over-margining) *In the market equilibrium with $\alpha^*(\underline{s}) > 0$, margining is excessive compared to the second-best equilibrium, $\alpha^*(\underline{s}) > \alpha^{SB}(\underline{s})$.*

What leads to excessive margining is the contractual externality: When one hedger wants to obtain more insurance, he raises the margin in his optimal contract. By doing so, he increases supply. This lowers the equilibrium price, which makes other hedgers worse off. This negative externality implies that the market equilibrium is not constrained-efficient. Importantly, over-margining arises even when the market equilibrium is unique. When there are multiple equilibria, they are all different from the second-best, and all involve over-margining.

6 Empirical and policy implications

Empirical implications. The empirical implications of our theoretical analysis reflect the interaction between optimal contracting and equilibrium pricing.

First, our model generates contagion from the market for the protection buyer's asset towards the market for the protection seller's asset. The arrival of bad news about protection buyer's assets worsens seller's incentives, margins are called to restore incentives, and this affects the (market) value of protection seller's assets.[8]

Second, higher risk aversion of protection buyers or more severe moral hazard problem of protection sellers create a need for larger margins, which increases the associated fire-sales.

[8] This is a different form of contagion than in Biais, Heider and Hoerova (2012). In that paper, contagion arises in case protection sellers don't do effort after bad news. Here, contagion arises even when protection sellers always do effort.

In other words, contagion is more pronounced. Similarly, higher volatility of the protection buyer's risk, $\Delta\theta$, necessitates larger equilibrium margins and depresses the price of protection seller's assets.

Moreover, as equilibrium margins increase, there can be a switch from increasing supply and equilibrium uniqueness, to non-monotonic supply curve and multiple equilibria. Therefore, the scope for the emergence of "bad equilibria" is higher when the risk aversion of protection buyers or the volatility of their assets is higher, or when the agency problems of protection sellers are more severe.

Policy implications. Our welfare analysis shows that in equilibrium too much insurance is sold by protection sellers, implying that aggregate margin calls are too large. The second-best can be achieved by imposing a cap on margins.

A cap on margins also solves the market instability problem caused by multiplicity: when the regulator/central bank caps margins, equilibrium in the protection seller's asset market is unique.

Margins caps are a form of macro-prudential policy. Since the scope for multiplicity and market instability is higher when risk aversion increases, regulators must impose margin caps in this case. This implies that margins should be countercyclical.

Our analysis also highlights that phasing-in of the margin-cap policy must be carefully designed; otherwise margin caps can lead to suboptimal outcomes. For example, suppose some market participants have already contracted, before the cap is introduced. In these contracts, protection sellers have promised large insurance payments, made incentive-compatible by large margins. Now suppose the regulator caps margins, as we suggested above. Clearly, the cap should be applied to new contracts, but should it also be applied to the old contracts? If it is, while keeping transfers promised in the old contracts, then effort may no longer be incentive-compatible, and many protection sellers end up defaulting. So, if a margin cap is introduced, either old contracts should keep high margins and be exempt from the cap, or their transfers should be revised downwards.

Our model implies that, in the presence of moral hazard, margining can be excessive. However, we are not arguing that in reality markets always choose excessively large margins. In practice other forces are at play which, for simplicity and clarity, we don't include in the present model. For example, if protection buyers 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, 2012b). To avoid this, the regulator can impose a floor on margins. In sum, when agency problems lead to margining practices which are not in line with information-constrained efficiency, a regulatory intervention may be needed, for example via caps and floors on margins.

Selected references

Acharya, V. and S. Viswanathan, 2011, "Leverage, Moral Hazard, and Liquidity," *Journal of Finance*, 99-38.

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, 2012a, "Risk–sharing or Risk–taking? Counterparty Risk, Incentives and Margins," ECB Working Paper No. 1413.

Biais, B., F. Heider, and M. Hoerova, 2012b, "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.

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.

Shleifer, A. and R. Vishny, 1992, "Liquidation Values and Debt Capacity: A Market Equilibrium Approach," *Journal of Finance* 47, 1343-1366.

Shleifer, A. and R. Vishny, 2011, "Firesales in Finance and Macroeconomics," *Journal of Economic Perspectives* 25, 29-48.

Yellen, J., 2013, "Interconnectedness and Systemic Risk: Lessons from the Financial Crisis and Policy Implications," speech at the AEA/AFA Joint Luncheon, San Diego.

Appendix

Proof of Proposition 1 The first step is to show that the supply function $S(p) = \alpha^*(\underline{s})$ is continuous in p and increasing in p on a non-empty interval $(\hat{p}, \tilde{p})$, $\tilde{p} \leq 1$. We first investigate how the optimal interior margin $\alpha^*(\underline{s}) > 0$ changes as the price p changes. Denoting the left-hand side of (22) 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

$$\frac{R-C-\mathcal{P}}{(p-\mathcal{P})^2} > \alpha^*(\underline{s})\rho(\underline{c})\,\varphi(\alpha^*(\underline{s}), p). \tag{32}$$

Using (23) in (32), we get that if

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

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

$$\alpha^*(\underline{s}) \geq \frac{1}{\rho(\underline{c})\,(p-\mathcal{P})} \tag{34}$$

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

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 since 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 $p = \hat{p}$. Using (32), we also have that $\frac{\partial \alpha}{\partial p} > 0$ at $\alpha^*(\underline{s}) = \varphi^{-1}\left(\frac{R-C-\mathcal{P}}{\hat{p}-\mathcal{P}}\right)$. Any interior $\alpha^*(\underline{s})$ is determined by (23) where function φ is continuous in p.

The second step is to show that the demand for the asset lies above the supply at $p = \mathcal{P}$, while it lies below the supply at $p = 1$. At $p = \mathcal{P}$, $S(p) = \alpha^*(\underline{s}) = 0$ while $D(p) = 1 - F(\mathcal{P}) > 0$. At $p = 1$, $D(p) = 0$. As for the supply, there are two possibilities. Either the supply function is increasing for any $p > \hat{p}$, implying that $S(p = 1) > 0$. 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 (34)). Therefore, at $p = 1$, $D(p = 1) = 0 < S(p = 1) > 0$.

In sum, both the demand for and the supply of the asset are continuous in p. The demand is decreasing in p, and lies above the supply at $p = \mathcal{P}$, while it lies below the supply at $p = 1$. It follows that the equilibrium exists.

Proof of Proposition 2

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

Taking logs and using (23), we get

$$\begin{aligned}
\alpha^*(\underline{s}) &= \frac{\text{prob}[\bar{s}]}{\mathcal{P} - \text{prob}[\bar{s}]p - \text{prob}[\underline{s}]\left(R-C\right)}\left[\frac{1}{\rho}\ln\left(\frac{R-C-\mathcal{P}}{p-\mathcal{P}}\right) + \frac{\mathcal{P}}{\text{prob}[\bar{s}]} - \left(E[\theta|\bar{s}] - E[\theta|\underline{s}]\right)\right] \\
&= \frac{\text{prob}[\bar{s}]}{-\text{prob}[\bar{s}]\left(p-\mathcal{P}\right) - \text{prob}[\underline{s}]\left(R-C-\mathcal{P}\right)}\left[\frac{1}{\rho}\ln\left(\frac{R-C-\mathcal{P}}{p-\mathcal{P}}\right) + \frac{\mathcal{P} - \left(\pi - \underline{\pi}\right)\Delta\theta}{\text{prob}[\bar{s}]}\right] \qquad (35)
\end{aligned}$$

where the last term follows from:

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

Note that $\mathcal{P} < \left(\pi - \underline{\pi}\right)\Delta\theta$ holds since we are not in the first-best. Moreover, the denominator of the first fraction in (35) 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 $p \leq 1$.

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

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

so that (33) holds. But then, the supply is increasing in p, a contradiction.

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

Proof of Proposition 3 The optimal $\alpha^*(\underline{s}) \in [0,1]$. For $\alpha^*(\underline{s}) = 1$, the claim in the proposition is straightforward. An interior $\alpha^*(\underline{s})$ is given by equation (35). Therefore, we need to show that

$$\frac{\text{prob}[\bar{s}]}{-\text{prob}[\bar{s}]\,(p-\mathcal{P}) - \text{prob}[\underline{s}]\,(R-C-\mathcal{P})}\left[\frac{1}{\rho}\ln\left(\frac{R-C-\mathcal{P}}{p-\mathcal{P}}\right) + \frac{\mathcal{P}-(\pi-\underline{\pi})\,\Delta\theta}{\text{prob}[\bar{s}]}\right] \geq \frac{1}{\rho\,(p-\mathcal{P})}. \tag{36}$$

Consider $\rho \to \infty$. We have $\frac{1}{\rho}\ln\left(\frac{R-C-\mathcal{P}}{p-\mathcal{P}}\right) \to 0$ in (36) and, therefore, $\alpha^*(\underline{s}) > 0$. The right-hand side of (36), $\frac{1}{\rho(p-\mathcal{P})} \to 0$. Now consider $\rho \to 0$. We have that the left-hand side of (36) $\to -\infty$, while the right-hand side $\to \infty$.

Hence, at $\rho \to 0$, the left-hand side of (36) is below the right-hand side of (36), while at $\rho \to \infty$, it is the other way around. Since the left-hand side of (36) is increasing in ρ, while the right-hand side of (36) is decreasing in ρ, the claim in the proposition follows.

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 protection buyers (protection sellers are held at their participation constraints). Let $\bar{p}$ denote a higher and $\underline{\mathrm{p}}$ a lower price, respectively, $\bar{p} > \underline{\mathrm{p}}$. Let $EU\,(p,\alpha\,(p))$ denote the value of the expected utility of a protection buyer at an equilibrium with the price p and the corresponding margin $\alpha\,(p)$. Then, we have that

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

where the first inequality follows from the fact that $\alpha^*\left(\underline{\mathrm{p}}\right)$ was not chosen for $\bar{p}$ and the second inequality follows from the fact that, given the same α, a protection buyer always prefers to get a higher price for the asset.

Proof of Proposition 4 By (35), $\frac{\partial \alpha^*(\underline{s})}{\partial \rho} > 0$ for all p. Higher risk aversion ρ leads to a higher supply.

Proof of Proposition 6 In the market equilibrium,

$$\varphi(\alpha^*(\underline{s}), p^*)\left[p^* - \mathcal{P}\right] = R - C - \mathcal{P}$$

while in the second-best equilibrium

$$\varphi(\alpha^{SB}(\underline{s}), p^{SB})\left[p^{SB} - \mathcal{P} + \alpha^{SB}(\underline{s})\frac{\partial p^{SB}}{\partial \alpha}\right] = R - C - \mathcal{P}.$$

Therefore, we have

$$\frac{\varphi(\alpha^*(\underline{s}), p^*)}{\varphi(\alpha^{SB}(\underline{s}), p^{SB})} = \frac{p^{SB} - \mathcal{P} + \alpha^{SB}(\underline{s})\frac{\partial p^{SB}}{\partial \alpha}}{p^* - \mathcal{P}}. \tag{37}$$

First, we show that $\alpha^*(\underline{s}) \neq \alpha^{SB}(\underline{s})$ whenever $\alpha^*(\underline{s}) > 0$. We prove the claim by contradiction. Suppose that $\alpha^*(\underline{s}) = \alpha^{SB}(\underline{s}) > 0$. Since $\alpha^*(\underline{s}) = \alpha^{SB}(\underline{s})$, we also have that $p^* = p^{SB}$ so that $\varphi(\alpha^*(\underline{s}), p^*) = \varphi(\alpha^{SB}(\underline{s}), p^{SB})$. Hence, the left-hand side of (37) is equal to 1 implying that

$$\alpha^{SB}(\underline{s})\frac{\partial p^{SB}}{\partial \alpha} = 0$$

must hold. However, $\alpha^{SB}(\underline{s}) > 0$ while $\frac{\partial p^{SB}}{\partial \alpha} < 0$. A contradiction.

Second, we show that $\alpha^*(\underline{s}) > \alpha^{SB}(\underline{s})$. We prove the claim by contradiction. Suppose that $\alpha^{SB}(\underline{s}) > \alpha^*(\underline{s})$. Then, $p^{SB} < p^*$ (since the equilibrium price is decreasing in the supply of the asset). Also, expected utility in the second-best is necessarily higher than in the market equilibrium (the market allocation is feasible for the planner but it is not chosen), i.e.:

$$\begin{aligned}
&\mathrm{pr}[\bar{s}]u\left(E[\theta|\bar{s}] - \frac{\mathrm{prob}[\underline{s}]}{\mathrm{prob}[\bar{s}]}\left[\alpha^{SB}(\underline{s})\left(R - C\right) + (1 - \alpha^{SB}(\underline{s}))\mathcal{P}\right]\right) + \\
&\qquad + \mathrm{pr}[\underline{s}]u\left(E[\theta|\underline{s}] + \alpha^{SB}(\underline{s})p^{SB} + (1 - \alpha^{SB}(\underline{s}))\mathcal{P}\right) \\
&> \mathrm{pr}[\bar{s}]u\left(E[\theta|\bar{s}] - \frac{\mathrm{prob}[\underline{s}]}{\mathrm{prob}[\bar{s}]}\left[\alpha^*(\underline{s})\left(R - C\right) + (1 - \alpha^*(\underline{s}))\mathcal{P}\right]\right) \\
&\qquad + \mathrm{pr}[\underline{s}]u\left(E[\theta|\underline{s}] + \alpha^*(\underline{s})p^* + (1 - \alpha^*(\underline{s}))\mathcal{P}\right)
\end{aligned}$$

implying that

$$\mathrm{pr}[\bar{s}]\left[u\left(\bar{c}\left(\alpha^{SB}(\underline{s}),p^{SB}\right)\right)-u\left(\bar{c}\left(\alpha^{*}(\underline{s}),p^{*}\right)\right)\right]>\mathrm{pr}[\underline{s}]\left[u\left(\underline{c}\left(\alpha^{*}(\underline{s}),p^{*}\right)\right)-u\left(\underline{c}\left(\alpha^{SB}(\underline{s}),p^{SB}\right)\right)\right] \tag{38}$$

where we used our short-hand notation for consumption after good and bad news, $\bar{c}$ and $\underline{c}$. Note that

$$\bar{c}\left(\alpha^{SB}(\underline{s}),p^{SB}\right)<\bar{c}\left(\alpha^{*}(\underline{s}),p^{*}\right) \tag{39}$$

since $\alpha^{SB}(\underline{s})>\alpha^{*}(\underline{s})$. Since u is increasing, the left-hand side of (38) is negative, implying that the right-hand side is negative and

$$\underline{c}\left(\alpha^{*}(\underline{s}),p^{*}\right)<\underline{c}\left(\alpha^{SB}(\underline{s}),p^{SB}\right). \tag{40}$$

By (39), $u'(\bar{c}\left(\alpha^{SB}(\underline{s}),p^{SB}\right))>u'\left(\bar{c}\left(\alpha^{*}(\underline{s}),p^{*}\right)\right)$. By (40), $u'(\underline{c}\left(\alpha^{*}(\underline{s}),p^{*}\right))>u'\left(\underline{c}\left(\alpha^{SB}(\underline{s}),p^{SB}\right)\right)$. Therefore,

$$\varphi(\alpha^{*}(\underline{s}),p^{*})=\frac{u'\left(\underline{c}\left(\alpha^{*}(\underline{s}),p^{*}\right)\right)}{u'\left(\bar{c}\left(\alpha^{*}(\underline{s}),p^{*}\right)\right)}>\frac{u'\left(\underline{c}\left(\alpha^{SB}(\underline{s}),p^{SB}\right)\right)}{u'\left(\bar{c}\left(\alpha^{SB}(\underline{s}),p^{SB}\right)\right)}=\varphi(\alpha^{SB}(\underline{s}),p^{SB}) \tag{41}$$

or, equivalently,

$$\frac{\varphi(\alpha^{*}(\underline{s}),p^{*})}{\varphi(\alpha^{SB}(\underline{s}),p^{SB})}>1.$$

Using (37), it follows that

$$\alpha^{SB}(\underline{s})\frac{\partial p^{SB}}{\partial\alpha}>p^{*}-p^{SB} \tag{42}$$

must hold. Since $\alpha^{SB}(\underline{s})>\alpha^{*}(\underline{s})$, $p^{SB}<p^{*}$ so that the right-hand side of (42) is positive. However, $\alpha^{SB}(\underline{s})>0$ while $\frac{\partial p^{SB}}{\partial\alpha}<0$, a contradiction.

Figure 1: Timing

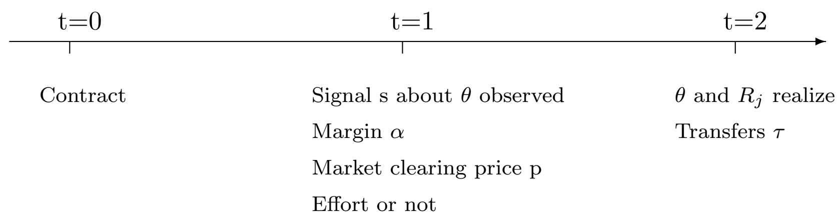

Figure 2: Optimal margin

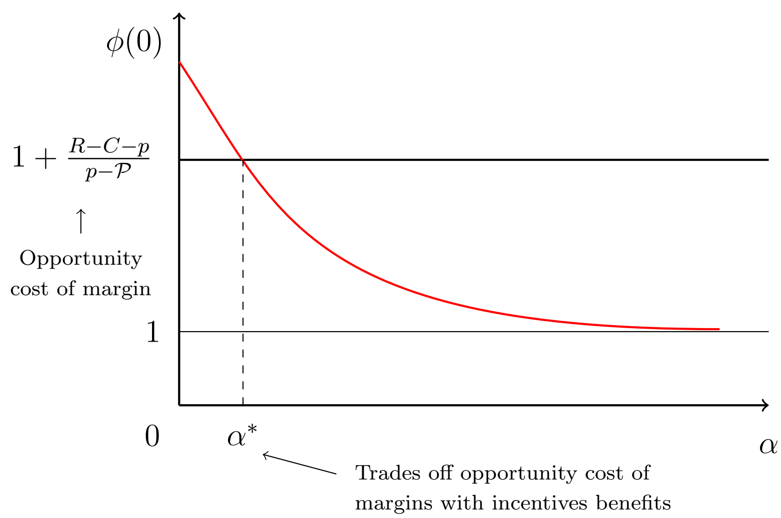

Figure 3, Panel A: Increasing supply curve

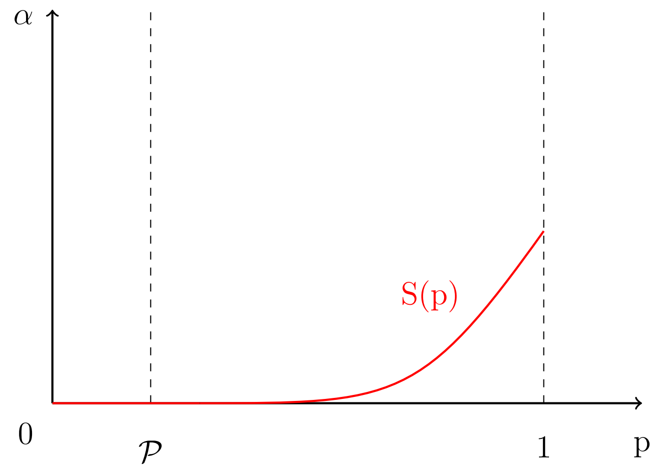

Figure 3, Panel B: Non-monotonic supply curve

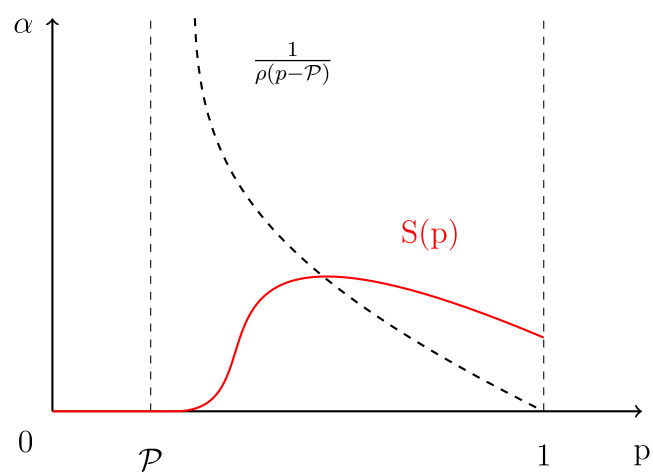

Figure 4, Panel A: Unique equilibrium

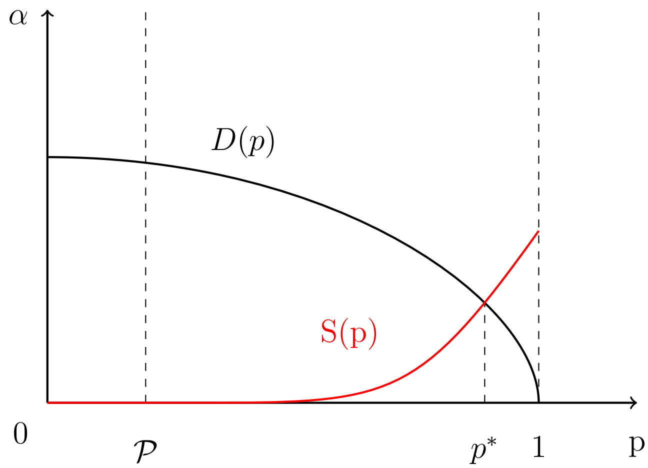

Figure 4, Panel B: Multiple equilibria