HONG KONG INSTITUTE FOR MONETARY AND FINANCIAL RESEARCH

THE DESIGN OF A CENTRAL COUNTERPARTY

John Kuong and Vincent Maurin

HKIMR Working Paper No.21/2021

September 2021

金融學院

The Design of a Central Counterparty

John Kuong
INSEAD

Vincent Maurin
Stockholm School of Economics

September 2021

Abstract

This paper studies the benefits of central clearing and the design of a central counterparty (CCP) with an optimal contracting approach. Investors sign contracts to hedge an underlying exposure. There is counterparty risk because investors can default on the contract due to idiosyncratic shocks and moral hazard. Mutualization of losses can thus hedge against counterparty risk but demands collateral for preventing moral hazard. The optimal contract involves loss mutualization, which requires central clearing, only when the cost of collateral is intermediate. Furthermore, as loss mutualization dilutes investors' incentives to monitor their counterparties, a third-party CCP can emerge as a centralized monitor and is given a first-loss, equity tranche as incentive compensation. Our results endogenize key features of the default resolution process, known as "default waterfall", in a CCP. Finally, we show that larger user base of a contract favors central clearing (over bilateral trading) and clearing with third-party CCP (over member owned CCP).

Keywords: CCP, Financial Stability, Contracting, Market Design.
JEL classification: D47, D86, G23.

* Email: Kuong: john.kuong@insead.edu and Maurin: vincent.maurin@hhs.se

* The authors thank Jean-Edouard Colliard, Selman Erol, Thomas Gehrig, Deeksha Gupta, David Murphy, Chester Spatt, Ernst-Ludwig von Thadden, Haoxiang Zhu, seminar audiences at INSEAD, Stockholm School of Economics, Higher School of Economics ICEF, Tepper School of Business, Asia-Pacific Corporate Finance Online Workshop, Kelley Junior Faculty Conference, and Workshop on Money, Payments, Banking and Finance by Study Center Gerzensee and SaMMF, SGF 2021, Stern/Salomon Microstructure Meeting 2021, Annual Conference on Financial Market Regulation 2021, UC3M and U of Mannheim for useful comments. Kuong gratefully acknowledges financial support from Hong Kong Institute for Monetary and Financial Research. This paper represents the views of the authors only. Any inaccuracies or omissions are the responsibility of the authors.

1 Introduction

Since the 2007–2008 global financial crisis, there has been a substantial rise in the share of financial contracts cleared by Central Counterparties (CCPs).[1] Post-crisis regulations have been an important driver behind these trends as central clearing became mandatory for many Over-The-Counter instruments.[2] Regulators view CCPs as a financial market infrastructure to mitigate counterparty risks. By standing between every transacting party, or its members, a CCP reduces the impact of any member's default by mutualizing the associated losses between members and absorbing some of the losses itself. Some commentators and academics stressed, however, that central clearing is costly because members are required to post high-quality collateral, in part, as guarantees for the default fund.[3] Furthermore, there is widespread concern that faulty design of CCP resolution and incentive structure could undermine rather than foster financial stability (see Yellen 2013).[4]

In this paper, we propose a general contracting framework to conduct a cost-and-benefit analysis of central clearing and study the optimal design of a CCP. Our primitives are that insurance provision can be limited by counterparty risk, limited pledgeability and insufficient monitoring of counterparties. Collateral can help mitigate these frictions but it is

[1]From 15% in 2009, the fraction of interest rate derivatives cleared by CCPs steadily rose to 60% in 2018 (FSB 2018). In Euro interbank repurchase agreements (repos) market, central clearing has become the norm. Mancini, Ranaldo, and Wrampelmeyer (2015) show that from 2009 to 2013, the market share of CCP-based repos increased from 42% to 71%, whereas the share of bilateral repos declined from 50% to 19%. The share of triparty repos remained relatively constant at around 10%.

[2]In the US, Section 723 of the Dodd-Frank Act mandates central clearing of interest rate swaps and credit default swaps. In the EU, the EMIR regulation introduced similar requirements.

[3]Singh (2010) argues that central clearing will significantly increase the need for costly collateral. Ghamami and Glasserman (2017) provide a detailed cost comparison between centrally cleared trading and purely bilateral trading, showing that collateral costs are a key driver.

[4]CPSS-IOSCO (2012) suggests that a CCP should have enough prefunded resources to sustain the default of two largest members. On incentives, Coeuré (2015) notes that a CCP's own contribution to the loss-mutualization process should be seen as its "skin-in-the-game" to induce proper risk management, rather than to significantly cover its loss exposure.

costly. With these basic ingredients we achieve three main results. First, we show that loss mutualization in central clearing is desirable when the cost of collateral is intermediate and market size is large. Second, we characterize the optimal loss mutualization scheme and endogenize many important features of centrally cleared contracts such as initial margins and default fund contributions. Third, CCPs with different structure arise endogenously to implement the optimal contract. A third-party, for-profit CCP with a junior equity tranche emerges as an efficient solution of the contracting frictions when the market is large enough.[5]

In the model, a finite number of investors are interested in sharing risks across aggregate states. For instance, some traders wish to buy oil futures to hedge against the risk of elevated oil prices while others are willing to sell the futures. Investors are matched bilaterally and sign a contract. Conditional on the realisation of the aggregate states, one investor has to pay her counterparty out of the cash flows of an asset she owns. This asset may fail to pay off which leads the payer to default on the contract. We aim to capture the counterparty risk futures buyers face when sellers are unable to honour the contract after a substantial hike of oil prices. Due to idiosyncratic counterparty risks, investors as a whole can benefit in ex-ante signing a multilateral contract under which, an investor with a defaulted payer ex-post mutualizes the losses with other non-defaulting payers. As we will argue, a multilateral contract amounts to novating and clearing a bilateral contract with a CCP.

Besides counterparty risk, insurance provision and risk-sharing are subject to two fundamental frictions. First, investors cannot promise to pay all the cash flows of their asset to their counterparties or, via loss mutualization, to other investors. This limited pledgeability problem, as in Biais, Heider, and Hoerova (2016), stems from a moral hazard friction: Investors would shirk for private benefits and default if their expected liability is too large.

[5] While third-party CCPs became more prevalent in the past decade, substantial heterogeneity remains in the ownership structure of CCPs around the world (see Section 8 of CPSS 2010).

The shirking and private benefits metaphor are meant to capture investors' concerns in practice that their counterparties could privately take actions that expose them to "wrong-way risk".[6] The futures sellers in our example could further take large short position in oil prices and default in states of high oil prices, leaving the futures buyers unprotected. Second, an investor can monitor her counterparty to mitigate this moral hazard problem but monitoring effort is costly and unobservable. This monitoring effort corresponds to the investors' and CCPs' due diligence processes to ascertain the financial soundness of their counterparties and members. The rigour and incentive structure behind such processes are first-order issues to the regulators and the CCPs (see Coeuré 2015 and ESMA 2020).

Investors can also specify collateral requirement in the contract. Despite having a lower expected return than the asset, cash collateral plays important roles in overcoming the contracting frictions. First, it expands the amount of insurance investors can credibly provide because, unlike the asset, it is fully pledgeable. That is, it has an incentive value. Second, it has a safety value in providing insurance to investors over and beyond what can be achieved with loss mutualization. Third, as collateral supports more insurance, the counterparty creditworthiness is less relevant if investors use a lot of collateral. Hence, it also reduces the need for monitoring.

We start the analysis with only the limited pledgeability friction, that is, monitoring efforts are observable. We call the solution to this partial problem the *optimal* contract. The analysis of the optimal contract shows when central clearing, which allows loss mutualization, dominates bilateral trading. Its implementation also rationalizes the crucial features of a central clearing contract such as the default waterfall and members' contribution to the

[6] In Basel III, wrong-way risk is defined as follows: *a bank is exposed to "specific wrong-way risk" if future exposure to a specific counterparty is highly correlated with the counterparty's probability of default.* See BCBS (2019).

default fund. Then, we bring back the friction of unobservable monitoring efforts. The solution to the full problem is called the *incentive-compatible (IC)* contract. The IC contract shows when a CCP can serve as a centralized monitor and sheds lights on the design of a CCP's organisation and capital structure.

Our first main result from the analysis of the *optimal contract* is that loss mutualization is essential only when the cost of collateral is intermediate. If the cost of collateral is lower than its safety value, the optimal contract is fully collateralized, which leaves no loss to be mutualized. If instead the cost of collateral is higher than the combined value from its safety and incentive functions, using collateral to support any insurance including loss mutualization is too expensive. Finally, when the cost of collateral is intermediate, the optimal contract features complete loss mutualization, under which investors are paid in full unless all contract payers default. The optimal collateral requirement ensures that surviving payers can credibly take on all liability of the defaulting payers. A simple observation from our analysis is that investors use more collateral for insurance when it is cheap, which reduces the need for monitoring counterparties. In the rest of the analysis, we focus on the case in which complete loss mutualization and monitoring are optimal.

There are several implications from the analysis of the optimal contract. First, mandating central clearing is only optimal when the cost of collateral is intermediate. When the collateral cost is too low or too high, the optimal contract can be implemented bilaterally because it features no loss mutualization. Also, central clearing is more desirable in larger markets, that is, for more actively traded contracts. The intuition is that loss mutualization becomes relatively more efficient than full collateralization in mitigating counterparty risk when losses can be shared among more members.

Second, central clearing does not always require more collateral than bilateral trading.

If investors only partially collateralize bilateral contracts, central clearing does require more collateral to support loss mutualization. However, if counterparty risk is high and collateral cost is intermediate, the optimal bilateral contract is fully collateralized while the optimal multilateral contract is only partially collateralized. Intuitively, central clearing can provide cheaper counterparty risk insurance via loss mutualization while investors can only resort to collateral in bilateral contracts.

Third, the optimal contract rationalizes some important features of the loss mutualization scheme of CCPs such as default waterfall. Under an implementation of the optimal contract, the collateral posted by the defaulters is first used to cover the losses before other members contribute. Collateral in our model captures both Initial Margins and Default Fund Contributions in practice, which are indeed the first line of defence in the waterfall. The remaining shortfall is covered by other CCP members. Such default fund contributions by surviving members require collateral to be posted ex-ante. In practice, CCPs indeed require their members to pre-fund such contributions. However, our description of the default waterfall so far misses an important feature, namely, the CCP's own contribution to the default waterfall. We fill this gap in the analysis of the *incentive compatible* contract.

In the second part of the paper, we focus on the case in which central clearing is essential and we now assume that monitoring efforts are unobservable. We first show that the optimal loss mutualization scheme may not be implementable due to the classic "insurance vs incentive" conflict of Holmström (1979). If an investor knows that she will get paid from other payers when her own payer defaults, she finds counterparty monitoring wasteful. To avoid such free-riding on the loss mutualization scheme, investors must be sufficiently exposed to their counterparty risk. Hence, while the optimal contract prescribes complete loss mutualization, incentive compatible loss mutualization can only be incomplete. The required distortion is more severe in larger markets because, as we have argued above, the insurance in the optimal contract improves as the market grows larger.

An alternative scheme to overcome the "insurance vs incentive" conflict is to centralize and delegate all the monitoring efforts to a third-party agent. We interpret this agent as a third-party, for-profit CCP who has no endowment so that he cannot provide insurance and has the same monitoring cost as individual investors. By giving a high-power incentive contract to the CCP to induce monitoring efforts, centralized monitoring can be more efficient than bilateral monitoring by individual investors.[7] Yet, centralized monitoring is costly for the following reasons. First, the CCP enjoys agency rent, receiving compensation over and above the effort cost, because monitoring efforts are unobservable. Second, the CCP's compensation increases the liability of investors, which requires more collateral.

We show that centralized monitoring dominates bilateral monitoring when the market is large enough. A larger market favours centralized monitoring for two reasons. Agency rents in the CCP's compensation decrease with the number of investors to be monitored[8] and, as explained, the loss mutualization distortion under bilateral monitoring is worse in larger markets. We therefore rationalize the CCP's active role of monitoring its members. In practice, CCPs carefully vet members with internal credit rating criteria and examine their books regularly.[9]

The comparison between bilateral and centralized monitoring and the solution for the optimal compensation of CCPs shed light on the optimal design of CCPs. A third-party CCP emerges endogenously as an efficient solution of the incentive frictions when the traded

[7] It can be shown that the optimal incentive contract pays only when no investors default and the associated agency rent decreases in market size. This is a standard result in contracting commonly known as "cross-pledging". For a textbook treatment, see Tirole (2010).

[8] The reduction of agency rent is similar to the (endogenous) economies of scale effect in Diamond (1984)

[9] See, for example, the clearing rules of ICE at https://www.theice.com/publicdocs/clear_credit/ICE_Clear_Credit_Rules.pdf.

contract has a large enough user base, which could be the case for more standardized contracts. In that case, the CCP is given a high-power incentive contract similar to a junior equity tranche. This junior tranche is essentially the CCP's contribution to the default fund which, when some members default, is wiped out before surviving members' default fund contribution. This result is in line with the "skin-in-the-game" interpretation of CCP capital (see e.g Coeuré 2015) and completes our picture of the default waterfall (see Duffie 2015). For less traded contracts, loss mutualization is also beneficial but it is best implemented without a third-party CCP. We interpret this arrangement as a member owned CCP because in this case, all the transfers are ultimately made and received by its members.

Literature Review

Our paper contributes to the burgeoning literature on central clearing. Menkveld and Vuillemey (forthcoming) provide an excellent survey on various aspects about central clearing we discuss below (and many more).

Our paper's primary focus, namely, the tension between mutualizing losses among traders and preserving their incentives to identify creditworthy counterparties, is also shared by Biais, Heider, and Hoerova (2012) and Antinolfi, Carapella, and Carli (2018).[10] Biais, Heider, and Hoerova (2012) also study the optimal clearing contract in the presence of moral hazard. They emphasize that when there is aggregate risk, loss mutualization has to be incomplete in order to preserve traders' incentive. Antinolfi, Carapella, and Carli (2018) argue that, without borrowers' information acquisition, a CCP cannot know the creditworthiness of lenders and thus cannot efficiently tailor collateral requirements. While the analysis of a member owned CCP in our paper contains a similar trade-off, we also consider the possibility of del-

[10]See Koeppl (2013) and Palazzo (2016) for works analyzing other incentive problems associated with central clearing.

egating monitoring efforts to a third-party CCP. This innovation allows us to deliver several new results. We highlight the cost and benefit of having a third-party CCP, characterize the conditions under which a third-party CCP dominates a member owned CCP, and derive the optimal capital and default waterfall structure of a third-party CCP.

Some recent papers also study CCPs' capital and default waterfall. Huang (2019) argues that because of limited liability and exogenously costly capital, a for-profit CCP tends to have insufficient capital for loss absorption and thus capital requirements may be warranted. We instead emphasize the incentive role of CCP equity, as suggested by Coeuré (2015), and take the optimal contracting approach to endogenize the cost of CCP equity. Wang, Capponi, and Zhang (2019) do not study CCP capital but differentiate initial margin from default fund contribution, by focusing on members' risk-taking incentives. We do not make that distinction but we rationalize CCP equity, a crucial element of the default waterfall.

To the best of our knowledge, our result on the optimal ownership structure of CCPs is new in the literature. A third-party agent can improve upon a member owned CCP thanks to the delegation of risk-mitigation efforts, for reasons similar to the diversification benefits in Diamond (1984). McPartland and Lewis (2017) argue that the ownership structure of CCPs is a central feature when discussing the economic role of CCP capital and the default waterfall. More generally, our analysis is related to discussions about the ownership structure of exchanges (see for instance Hart and Moore 1996).

The literature on CCPs has also discussed various specific aspects of central clearing. Duffie and Zhu (2011) compare the netting efficiency under central and bilateral clearing. Acharya and Bisin (2014) argue that central clearing increases position transparency and reduces counterparty risk externality in OTC markets.[11] Koeppl, Monnet, and Temzelides

[11] See also Zawadowski (2013). Relatedly, Leitner (2011) shows that an intermediary like a CCP can even induce voluntary report of trades by clearing members thanks to a position limit

(2012) show that a CCP can reduce trading costs by deferring settlement and providing credit to their clearing members. We focus on the incentive problem inherent in loss mutualization, and complement the literature by simultaneously characterizing the optimal collateral requirements, the ownership, and the capital structure of a CCP.

The rest of the paper is organised as follows. Section 2 presents the model. Section 3 studies the costs and benefits of central clearing by deriving the optimal contract when monitoring is observable. In Section 4 we derive the incentive compatible contract, comparing bilateral monitoring to centralized monitoring. The practical implications of our model for CCPs are collected in Section 5. Section 6 concludes. All proofs are collected in Appendix A.

2 Model

2.1 The framework

There are two dates $t = \{0, 1\}$ with no time discounting. At date 1, there are two equiprobable aggregate states of the world $S \in \{A, B\}$. All agents consume the same good –"cash". The economy is populated by two groups of investors, also indexed by $S \in \{A, B\}$, and each group has N homogeneous investors. An S-investor has the following utility function:

$$U_S(c_S, c_{S'}) = \frac{1}{2}\mathbb{E}[c_{S'}] + \frac{1}{2}\mathbb{E}\left[c_S + (v-1)\min\{c_S, \hat{c}\}\right] \tag{1}$$

where c_S is the consumption in state S, $v > 1$, and $\hat{c} > 0$. An S-investor's marginal utility of consumption in state S' is one whereas each unit of consumption in state S yields utility $v > 1$ until it reaches $\hat{c}$. These preferences are meant to capture in a simple way insurance needs against an aggregate state, with $\hat{c}$ the demand for insurance.

Each S-investor is endowed with one unit of a divisible asset which pays $2R$ per unit with probability $q \in (0,1)$ in state S' and fails to pay anything otherwise. The success or failure of the asset is independent across S-investors, conditional on the realisation of state S'. Because S-investors (resp. S'-investors) have an asset that pays in state S' (resp. S), they can provide insurance to S'-investors (resp. S-investors) who value consumption more in this state. The per-unit gains from trade are given by the difference in marginal utility, equal to $(v-1)$. The riskiness in asset payoff implies trading is subject to counterparty risk.

Besides counterparty risk, insurance provision is hindered by the fact that the asset's cash flow is not fully pledgeable, due to a moral hazard problem as in Holmström and Tirole (1997). At date 0, an investor can decide to shirk after observing the non-pecuniary private benefit (per-unit) from shirking $\tilde{B} \in \{B, b\}$. Shirking is not observable and causes the asset to fail with probability 1. Following Holmström and Tirole (1997), we define the pledgeable income of the asset as the maximum cash flow an investor can promise to pay without shirking. For an investor with private benefit $\tilde{B}$, the pledgeable income is given by[12]

$$\tilde{\beta} \equiv 2R - 2\frac{\tilde{B}}{q} < 2R.$$

We assume that $B = qR$ and $b \in (q(R - \frac{1}{2}), qR)$, and thus, the pledgeable income of the asset is either 0 or $\beta \equiv 2R - 2\frac{b}{q} \in (0,1)$. An investor with asset pledgeability β (resp. 0) is called a regular (resp. rogue) investor.

Limited pledgeability and counterparty risk give a role for collateral. At date 0, the asset can be liquidated for \$1 cash per unit. Cash is fully pledgeable as collateral, and

[12]To see this, consider an investor's incentive to shirk if she has promised to pay $\hat{\beta}$. She will not shirk if and only if $\frac{1}{2}q(2R - \hat{\beta}) \geq \tilde{B}$, or, $\hat{\beta} \leq 2R - 2\frac{\tilde{B}}{q}$.

because it is more pledgeable than the asset ($1 > \beta$), it increases investors' total pledgeable income. Second, it serves as self-insurance and protects an investor against counterparty default because it is safe. We assume, however, that cash collateral is costly because the expected payoff of the asset qR is higher than 1. The comparison of the (endogenous) values and the costs of cash collateral will determine the optimal risk sharing arrangement and collateral requirement. We make the following assumption to ensure that cash collateral is both necessary and sufficient to satisfy investors' insurance needs.

Assumption 1 (Collateral needs)**.** $2 > \hat{c} > \beta$.

Consider a pair with a S-investor and a S'-investor. The ex-post insurance need to be satisfied is $\hat{c}$. If each investor posts $\frac{\hat{c}}{2} < 1$ units of cash collateral, the insurance need can be met with collateral only. Alternatively, without any cash collateral, even a regular investor can only credibly promise to pay up to β when his asset succeeds. This is less than the insurance need under Assumption 1.

The limited pledgeability problem can be mitigated by monitoring. At date 0, each S-investor is bilaterally matched with a S'-investor. An investor can privately decide to monitor her counterparty for a non-pecuniary cost $\psi > 0$. If she monitors, her counterparty is a regular investor with asset pledgeability β. Without monitoring, her counterparty is a regular investor with probability $\alpha \in [0, 1)$ and a rogue investor otherwise. In what follows, for simplicity, we restrict attention to pure-strategy symmetric monitoring decisions. Hence, a profile of investors' monitoring decisions can be summarized by $m = \{0, 1\}$, where $m = 0$ stands for no monitoring and $m = 1$ for every investor monitors.

The following bound on the monitoring cost ensures that monitoring can be optimal.

Assumption 2 (Monitoring cost)**.** $\psi \leq \psi_{max} \equiv \frac{\beta q(1-q)(1-\alpha)(v-1)}{v(2-\beta\alpha q)(1-\alpha q)} \left(1 - \frac{\hat{c}}{2}\right)$.

The upper bound on the cost will ensure that there is a region of collateral cost for which monitoring is optimal. The characterization of this bound is deferred to Section 4. Intuitively, ψ_{max} is proportional to the increase in the probability that a counterparty is regular if monitored $(1-\alpha)q$, the pledgeable income of a regular investor β and the gains from trade $v-1$.

2.2 Multilateral contract

After each S-investor is matched with a S'-investor, investors collectively sign a multilateral contract that specifies state-contingent transfers.[13] For simplicity, we only consider pooling contracts rather than menus of contracts designed to induce investors to report their private type, rogue or regular.[14] The environment is symmetric and thus, S-investors pay in state S' to S'-investors what they themselves receive in state S from these investors. Henceforth, we drop the reference to the aggregate state and refer to investors by their ex-post role: *payers* or *receivers*. The relevant state variables to index transfer are the number $d \in \{0, 1, ..., N\}$ of defaulting payers, and, for each pair of matched investors, the outcome of the payer's asset $o \in \{s, f\}$, success (s) or failure (f). We define a multilateral contract as follows.

Definition 1. *A contract $\mathcal{C} = \{x, p_o(d), r_o(d)\}$ specifies an amount of collateral $x \in [0,1]$ posted by an investor at date 0, a set of transfers $p_o(d)$ to be made by a payer and a set of transfers $r_o(d)$ to be received by the payer's counterparty at date 1. The variable $d = \{0, 1, ..., N\}$ denotes the total number of defaulting payers and $o \in \{s, f\}$ denotes the outcome for the payer's asset, success (s) or failure (f).*

[13] In practice, investors first join a CCP which specifies rules for trading. Then, when they trade with another CCP member, a contract is signed and cleared by the CCP. With only one round of trading, we can collapse these two stages into one. We discuss the implementation of our contract in Section 3.3 and 5.

[14] This issue is moot when investors are monitored since their type is regular with probability 1. Without monitoring, however, the type is the private information of investors.

In the state with a total of d payers in default, $p_s(d)$ (resp. $p_f(d)$) specifies a payer's transfer to a common pool when its asset succeeds (resp. fails). The transfer to a receiver is $r_s(d)$ (resp. $r_f(d)$) when its matched payer's asset succeeds (resp. fails). Observe that transfers p_s and r_s (p_f and r_f) are not defined when N (0) payers default.

A feasible contract must satisfy the resource constraints of payers, given by

$$p_s(d) \leq x + (1-x)2R, \tag{2}$$

$$p_f(d) \leq x, \tag{3}$$

when the payer succeeds or fails, respectively. To simplify the analysis, we will consider parametrizations of the model in which the resource constraint never binds for a succesful payer, that is (2) is slack. This can be ensured by setting R large enough. A feasible contract must also satisfy the budget constraint state by state,

$$(N-d)r_s(d) + dr_f(d) = Nx + (N-d)p_s(d) + dp_f(d), \quad \forall\, d \in \{0, 1, ..., N\}. \tag{4}$$

Equation (4) says that the total transfer to be received must be equal the total resources available, which consist of the sum of the collateral pledged by the receivers and the contractual transfers made by payers.

2.3 Investors' problem and loss mutualization

After defining the contract, we are now ready to state the investors' problem and the relevant constraints faced by investors. First, we state a useful property of a contract and derive the investor's utility in the following lemma.

Lemma 1. *Under a contract $\mathcal{C} = \{x, p_o(d), r_o(d)\}$,*

$$\mathbb{E}[r_o(d)] = \mathbb{E}[p_o(d)] + x. \tag{5}$$

An investor's utility is given by

$$U = qR + \frac{v-1}{2}\mathbb{E}[\min\{r_o(d), \hat{c}\}] - x(qR - 1), \tag{6}$$

where $\mathbb{E}[\cdot]$ is the expectation operator taken over o and d, for a given monitoring choice.[15]

Equation (5) follows from the budget constraint (4) and says that the contract is actuarially fair in cashflow terms: The amount an investor expects to receive from the contract is the amount she expects to pay plus the cash collateral she posts. Equation (6) highlights the cost and benefit of contracts to an investor. Upon signing a contract, relative to the utility level under autarky qR, an investor benefits from the insurance, measured by the expected gains from trade $\frac{v-1}{2}E[\min\{r_o(d), \hat{c}\}]$, but needs to pledge collateral which costs $x(qR - 1)$.

Next, we turn to the constraints imposed by the frictions of the model. The limited pledgeability problem of the asset implies that investors shirk if the increase in expected payment given the success of the asset exceeds the asset's pledgeable income. A rogue investor always weakly prefers to shirk because his asset is not pledgeable. To avoid shirking by regular investors, however, the payments $p_o(d)$ must satisfy

$$\mathbb{E}_s[p_o(d)] - \mathbb{E}_f[p_o(d)] \leq (1 - x)\beta. \tag{7}$$

[15]The outcome variable o has a Bernoulli distribution, with parameter $q(m)$ equal to the probability of success, where $m \in \{0, 1\}$ is the monitoring choice. As will become clear, under an incentive-compatible contract, $q(0) = \alpha q < q = q(1)$. The variable d has a binomial distribution with parameter $q(m)$ and N.

where $\mathbb{E}_{o'}[.]$ is the expectation conditional on outcome $o' \in \{s, f\}$ for a payer. For a regular investor, the right-hand side of (7) is the pledgeable income from the remaining fraction $1 - x$ of the asset which is not liquidated for cash collateral.

If counterparty monitoring is to be elicited, the contract must ensure that the investor is better off when monitoring her counterparty because monitoring efforts are unobservable. This bilateral monitoring constraint is given by

$$\frac{\psi}{q(1-\alpha)} \leq \frac{1}{2}\Big(\mathbb{E}_s[r_o(d)] - \mathbb{E}_f[r_o(d)]\Big) + \frac{v-1}{2}\Big(\mathbb{E}_s[\min\{r_o(d), \hat{c}\}] - \mathbb{E}_f[\min\{r_o(d), \hat{c}\}]\Big) \quad (8)$$

The left-hand side of (8) is equal to the cost of monitoring effort ψ divided by the increase in the probability that the counterparty succeeds if monitored, $q(1-\alpha)$. The right-hand side is the utility loss for a receiver from a default of his counterparty. Monitoring is incentive compatible for investors if and only if equation (8) holds.

We can now formally define the investor's problem.

Definition 2 (Investor's problem)**.** *Investors design a multilateral contract $\mathcal{C}$ with monitoring decision $m = \{0, 1\}$ to maximize their utility (6), subject to the resource constraint (3), the budget constraint (4), the limited pledgeability constraint (7), and, if monitoring is to be induced, the monitoring incentive constraint (8).*

2.4 Preliminary Analysis

We establish a preliminary result to simplify the investors' contracting problem and highlight the key difference between multilateral contracts and bilateral contracts.

Lemma 2. *It is weakly, and sometimes strictly, optimal to set $r_s(d) = r_s$, $r_f(d) = r_f \leq r_s$ for any $d \neq N$, and $p_f(d) = x$ for all d. Then, by the budget constraint, $r_f(N) = 2x$ and*

$p_s(d) = r_s + \frac{d}{N-d} r_f + \frac{N+d}{N-d} x.$

Lemma 2 says that it is without loss to consider a simplified contract with three scalars: x as the amount of collateral; r_s and r_f respectively as the transfers received when the counterparty does not fail, and when the counterparty fails but at least one payer does not default. The receiver transfer when all payers default $r_f(N)$ is pinned down by the amount of collateral since no other resource is available in this state of the world. When he fails, a payer transfers the collateral posted, which is his only resource, that is, $p_f(d) = x$. Finally, the payer transfer $p_s(d)$ when his asset succeeds is determined residually from budget constraint (4) to support receivers' transfers.

The intuition behind Lemma 2 is as follows. First, increasing the transfer $p_f(d)$ from a defaulting payer relaxes the limited pledgeability constraint and allows investor to support more insurance, as shown by equation (5). It is thus optimal to saturate the resource constraint (3) and set $p_f(d) = x$. Second, since investors are risk averse, it is desirable to minimize the variation of their transfers received. By assumption, the resource constraint binds only when all payers fail, and, hence, it is always feasible to set constant $r_s(d) = r_s$ and $r_f(d) = r_f$ for any $d \neq N$. Yet, variations in received transfers can arise for two reasons. Even if one single successful payer can satisfy the insurance needs of all receivers, receivers are still exposed to the risk that all payers default. In this state of the world, they can only consume $r_f(N) = 2x$, which may contractually differ from the amount r_f. Finally, while receivers could be hedged against counterparty risk, a quick analysis of constraint (8) suggests that bilateral monitoring may not be sustainable at $r_s = r_f$. Then, setting $r_s > r_f$ may be necessary.

Lemma 2 highlights a key difference between a multilateral contract and a purely bilateral contract: the scope for loss mutualization. In a purely bilateral contract, when her own payer fails, the receiver may only get $2x$, which is the total amount of collateral posted by the pair. In a multilateral contract, this is true only if all other payers fail. Otherwise, other surviving payers can transfer resources to the receiver with a defaulting counterparty. We refer to transfers from other payers when one's own payer defaults as loss mutualization, for which we provide a formal definition below.

Definition 3 (Loss mutualization)**.** *A contract features loss mutualization if*

$$r_f > r_f(N) = 2x. \tag{9}$$

A contract without loss mutualization, such as a bilateral contract, is with $r_f = 2x$. *Loss mutualization is said complete if* $r_s = r_f$.

The definition above is straightforward. When her counterparty defaults and if there are no other payers to mutualize the losses, either because the contract is bilateral or because all other payers fail, a receiver consumes $2x$ – the sum of her own collateral and the collateral posted by her counterparty. Loss mutualization reduces the receiver's consumption loss $(r_s - r_f)$ caused by her counterparty's default. A complete loss mutualization means that as long as there is at least one successful payer, the receiver is not affected.

While loss mutualization provides receivers better insurance against the default of their counterparties, receivers still suffer losses when all payers default. The only contract providing *full insurance*, that is, realising all the gains from trades in all states of the world, is defined below.

Definition 4 (Full insurance)**.** *A full insurance contract is one with* $x = \frac{\hat{c}}{2}$ *and* $r_o = \hat{c}$

The full insurance contract is one with a fully collateralized promised payment of $\frac{\hat{c}}{2}$. Therefore, even when the counterparty defaults, the receiver can fully meet her insurance needs of $\hat{c}$. This contract will prove a useful benchmark for our analysis. A full insurance contract can be implemented bilaterally because there are no losses (to be mutualized) and counterparty monitoring is useless because a rogue investor faces the same cost for posting collateral. Intuitively, however, this contract will be dominated when collateral is costly. Observe that any contract with more collateral than the full insurance contract is suboptimal because all gains from trade are realised. Hence, in what follows, we set $x \leq \frac{\hat{c}}{2}$ without loss.

In sum, the investors' problem is to find the most cost-effective use of collateral (and monitoring) to support insurance. Our focus in the rest of the paper is to find out when loss mutualization is part of the optimal contract and how it can be implemented. Since, in our model, loss mutualization is the *raison d'être* of central clearing, we use these terms intechangeably in what follows.

3 Optimal Clearing

Our model of multilateral contracting has two main frictions: the asset's limited pledgeability and the unobservability of counterparty monitoring effort. In order to isolate the effects of the two frictions, in this section, we focus on the limited pledgeability problem and assume counterparty monitoring is observable. We characterize the optimal contract under observable monitoring, which we call "optimal loss mutualization", in Section 3.1. In Section 3.2, we compare central clearing to bilateral trading. Finally, in Section 3.3, we discuss the implementation of the optimal loss mutualization scheme and relate it to the design of central clearing arrangement in practice.

3.1 Optimal loss mutualization

The investor's contracting problem is as stated in Definition 2 with one difference: Monitoring is observable and thus the bilateral monitoring constraint (8) is ignored. The solution, which we call the optimal loss mutualization, highlights the fundamental trade-off between insurance and the cost of collateral as well as the role played by counterparty monitoring.

It is useful to first build intuition about the trade-off between the insurance benefit and the cost of collateral. The cost of collateral is the foregone return from the superior investment option $qR - 1$. The benefits of cash collateral are twofold. First, because it is safe, collateral provides insurance against the extreme event that all payers default. In this state, receivers get $r_f(N) = 2x$ which is the amount of collateral posted by each investor pair. Collateral also relaxes the limited pledgeability constraint and helps expand insurance provision. To see this, rewrite constraint (7) using the results from Lemma 2. A regular investor cannot promise to pay more than

$$\mathbb{E}_s[p_o(d)] \leq x + (1 - x)\beta \tag{7b}$$

where the right-hand side is the total pledgeable income. Under Assumption 1, posting more collateral relaxes the constraint because $\beta < 1$, which helps support larger transfers r_s and r_f to receivers, as shown by equation (5). This is valuable when investors demand for insurance is not satiated. The formal analysis of this trade-off leads to the following results.

Proposition 1 (Optimal loss mutualization)**.** *Suppose that counterparty monitoring is observable. There exist three thresholds of collateral cost $\underline{k}_N \geq 0$, $k_m \in [\underline{k}_N, \bar{k})$, and $\bar{k}$ such that the optimal monitoring decision is $m^{opt} = \mathbb{1}_{[k \geq k_m]}$ and the optimal contract is*

1. *a full insurance contract if $k \leq \underline{k}_N$. In this case, there is no loss to be mutualized;*

2. *a complete loss mutualization contract with* $r_s^{opt} = r_f^{opt} = \hat{c}$ *and*

$$x = x^{opt}(m^{opt}) \equiv \frac{\left[1-(1-q(m^{opt}))^N\right]\hat{c} - \beta q(m^{opt})}{2\left[1-(1-q(m^{opt}))^N\right] - \beta q(m^{opt})} \in \left(0, \frac{\hat{c}}{2}\right), \tag{10}$$

if $k \in (\underline{k}_N, \bar{k})$*, where* $q(m) = \alpha^{1-m} q$*;*

3. *an uncollateralized contract with* $r_s = \beta$*,* $r_f = x = 0$ *if* $k \geq \bar{k}$*. In this case, there is no loss mutualization.*

Proposition 1 is the first main result of the paper and there are three takeaways. The first one is that the optimal level of insurance decreases with the cost of collateral. When the collateral cost is low enough ($k \leq \underline{k}_N$), it is optimal to completely collateralize the contract and fully insure investors in all states. As the collateral cost increases, insurance against the unlikely event that all payers default becomes too costly. In this intermediate case, it is still optimal to use enough collateral to increase the pledgeable income for receivers to consume $\hat{c}$ in all other states. Loss mutualization is complete. Finally when the collateral cost is high ($k \geq \bar{k}$), any kind of insurance support is too costly: An uncollateralized bilateral contract is optimal. We show in the proof that the higher threshold of collateral cost is given by

$$\bar{k} = \frac{1}{2}(v-1)(2-q\beta). \tag{11}$$

This is the value of collateral when receivers consume less than $\hat{c}$. An extra unit of collateral increases consumption by 1 unit, as self-insurance, and by $1-q\beta$ units as net insurance from payers, since $q\beta$ is the expected pledgeable income of a regular investor's asset.[16]

[16] In the proof of Proposition 1, we show that

$$\underline{k}_N = \min\left\{(v-1)(1-\alpha q)^N, (v-1)(1-q)^N + \frac{2\psi}{\beta q(2-\hat{c})}\left(2\left[1-(1-q)^N\right] - \beta q\right)\right\}.$$

The second takeaway is that loss mutualization is only valuable when the collateral cost is intermediate. When the collateral cost is low full insurance is optimal and there is no loss to be mutualized. When the collateral cost is too high, loss mutualization is too costly and investors choose to be fully exposed to counterparty risk. In both cases, the optimal contract can be implemented bilaterally. For intermediate values of the collateral cost, unless all payers default, loss mutualization bridges the gap between the resources of a pair when the payer defaults, equal to $2x^{opt}(m^{opt}) < \hat{c}$ and the optimal insurance amount $\hat{c}$. Hence, counterparty risk now only affects investors when the aggregate risk that all other payers default materializes.

The third takeaway is that counterparty monitoring allows investors to economize collateral when it is expensive ($k \geq k_m$). We recall that monitoring prevents an investor from becoming rogue, effectively decreasing her default probability from $1-\alpha q$ to $1-q$. Therefore, similar to collateral, it increases an investor's capacity of insurance provision. Monitoring is thus used as a substitute to collateral when collateral is too expensive. Conversely, when the collateral cost drops below k_m, it is more cost-efficient to forgo the monitoring and to use more collateral. The collateral requirement then jumps from $x^{opt}(1)$ to either $x^{opt}(0)$ under the complete loss mutualization contract or $x = \frac{\hat{c}}{2}$ under the full insurance contract.

To sum up, the analysis of the optimal loss mutualization scheme shows that when the collateral cost is intermediate, central clearing is superior to bilateral trading because it improves insurance among investors by mutualizing losses caused by counterparty defaults.

Below $\underline{k}_N$, the optimal contract is full insurance without monitoring. Above $\underline{k}_N$, the optimal contract features complete loss mutualization, either without monitoring or with monitoring. The first case corresponds to the first argument of the *min* above. The threshold is then equal to the insurance value of collateral against the state where all payers default. Note that the probability of default without monitoring is $1 - \alpha q$. The second case corresponds to the second argument of the *min*. Then, for complete loss mutualization to be optimal, it is not enough that the collateral cost equals the insurance value against the joint default state, given by $(v-1)(1-q)^N$, because monitoring is costly.

Our optimal contracting approach ensures that the benefits of central clearing identified here are not due to some ad-hoc restrictions on the contracts or exogenously imposed costs.

3.2 Central Clearing vs. Bilateral Trading

We use our results to compare central clearing with bilateral trading. The optimal multilateral contract can be implemented bilaterally when $k \leq \underline{k}_N$ or $k \geq \bar{k}$. Hence, we first provide comparative statics for the region $[\underline{k}_N, \bar{k}]$ in which central clearing is essential.

Corollary 1. *The range of collateral cost $[\underline{k}_N, \bar{k}]$ for which central clearing is essential is expanding with market size N.*

This result shows that central clearing is more beneficial in large markets. To build some intuition, observe first that the upper bound of the region, $\bar{k}$, is independent from N. As we explained, $\bar{k}$ is the marginal insurance value of collateral starting from the uncollateralized (bilateral) contract. Investors could realise this extra insurance even in a bilateral contract, which is why $\bar{k}$ does not depend on N. The lower bound of the central clearing region, $\underline{k}_N$ is decreasing with N. This threshold $\underline{k}_N$ is the marginal insurance value of collateral starting from the complete loss mutualization contract of Proposition 1. By definition, extra insurance is only valuable in the state where all payers default. Because a joint default is less likely as N increases, investors are only willing to pay for this extra insurance for lower values of collateral cost. Our result thus suggests that clearing of contract is more beneficial in large markets. A critical mass of traders is needed for loss mutualization to be valuable with respect to full insurance.

We now compare the optimal multilateral contract to the optimal bilateral contract. The optimal bilateral contract solves the investors problem under the additional constraint that loss mutualization is not possible, that is, $r_f = 2x$ must hold. Proposition 1 shows that bilateral contracts are only restrictive in the region $[\underline{k}_N, \bar{k}]$ for collateral cost so we focus on this region for our analysis. Motivated by claims that central clearing increases the need for collateral, we compare the amount of collateral in each case.

Corollary 2. *For $N \geq 2$, the optimal bilateral contract requires strictly more (less) collateral than the optimal multilateral contract if $k \in [\underline{k}_N, \underline{k}_1)$ ($k \in (\underline{k}_1, \bar{k}]$).*

The result follows from Corollary 1. We showed that the upper bound $\underline{k}_N$ for the full insurance region is decreasing with N. Hence, when $k \in [\underline{k}_N, \underline{k}_1)$, only the bilateral contract features full insurance and thus requires strictly more collateral. Intuitively, since counterparty risk insurance is not available via loss mutualization in bilateral contracts, investors use collateral instead. In the region $(\underline{k}_1, \bar{k}]$, Proposition 1 and Corollary 1 characterize the optimal loss mutualization for any N, including the bilateral case $N = 1$ as a degenerate case. Since the collateral requirement x^{opt} is increasing with N, as shown by equation (10), the second part of the result follows. Intuitively, with more members, there is a greater potential to mutualize losses but more collateral is needed to support the larger expected transfers induced by loss mutualization because of the limited pledgeability constraint.

To summarize, central clearing reduces the need for collateral to protect against counterparty risk because CCPs act as risk-poolers by mutualizing losses (Coeuré (2015)). However, the very mutualization of losses requires collateral because CCPs need to make sure investors will deliver when called to cover other members' losses. By stressing these two roles of collateral, our result reconciles views that CCPs provide collateral efficiency gains (see Menkveld and Vuillemey (forthcoming)) with claims that central clearing increases the need for collateral (see e.g. Domanski, Gambacorta, and Picillo 2015).

3.3 Central Clearing Implementation

We argue that optimal contract identified in the previous section can be implemented by novating a bilateral contract to a CCP, an arrangement commonly carried out in practice. Consider the following implementation of the complete loss mutualization contract: at $t = 0$, each pair of investors bilaterally signs a contract with a promised payment of $\tau = \hat{c} - x^{opt}$ and novates the contract to a CCP, which requires all its members to post collateral x^{opt}. At $t = 1$, all receivers collect their posted collateral. If her counterparty succeeds, a receiver gets the promised payment and, thus, $r_s = x^{opt} + \tau = \hat{c}$. When her counterparty default, the receiver is given the priority to seize the counterparty's collateral x^{opt}. Then, the loss given default $\tau - x^{opt}$ is mutualized among all successful payers, and, thus she receives $r_f = 2x^{opt} + \tau - x^{opt} = \hat{c}$. In the state in which all payers default, receivers are only left with the collateral $r_f(N) = 2x^{opt}$.

It is useful to also discuss these transfers from the perspective of payers. To do so, consider a state with a total of $d \in (0, N)$ payers defaulting. For the defaulted payers, any collateral posted is immediately seized, that is, $p_f(d) = x^{opt}$. Collateral in our model corresponds both to the Initial Margin (IM) and to the ex-ante contribution to the Default Fund Guarantee (DFG) by members of a CCP, in pratice.[17] In practice, both these pre-committed resources are seized from defaulting payers before any other member contributes, as in our model. Since these resources do not cover the contractual payment to the receivers, the total loss given default $d(\hat{c} - 2x^{opt})$ must be shared among $N - d$ investor pairs with a successful payer. Because surviving members contribute to the default fund, they must post collateral to secure this liability to the CCP. Most CCP guidelines indeed emphasize

[17] Traders also post Variation Margin (VM) reflecting daily or lower frequency variations in the price of the asset underlying the contract. VM could be rationalized in our framework, for instance if there is an additional moral hazard problem at date 1 when uncertainty is realised.

the importance of pre-funded DFG contributions (see e.g. Arnsdorf (2012)).

Our model replicates some important features of the CCP Default Waterfall, the loss allocation process when a member defaults. Most importantly, CCPs use a defaulter-pay model based on collateral and allocate losses to surviving members when the defaulter resources are insufficient. A standard feature of the default waterfall our model cannot yet speak to is the CCP's own contribution. This is not surprising: Under the optimal loss mutualization scheme, a CCP is merely a nexus of contracts. It transfers resources between members but plays not active role. In the next section, we show that a CCP can act as a central monitor when monitoring is unobservable. Endogenizing the CCP incentive pay for monitoring helps fill the gap between our model and default waterfalls observed in practice.

4 Incentive Compatible Clearing

We now derive the optimal allocation when monitoring efforts need to be incentivized as they are not observable. The goal of this analysis is twofold: to show how the incentive problem in monitoring distorts the optimal loss mutualization and to characterize the optimal monitoring arrangement. To ensure such investigations are meaningful, we focus our analysis on the region of collateral cost $k \in [k_m, \bar{k}]$ in which monitoring is desirable and loss mutualization is essential in the optimal loss mutualization scheme. The need to incentivize monitoring effectively makes it more costly and hence the parameter region in which it should be induced will shrink, as we confirm in Section 4.3.

If an investor does not monitor her counterparty, the counterparty is a regular investor with probability α and a rogue investor with probability $1 - \alpha$. A rogue investor's asset is not pledgeable, and thus, she can only honour payments that are collateralized. As a rogue investor behaves the same as a defaulted regular investor, the lack of monitoring effectively increases the counterparty risk from $1 - q$ to $1 - \alpha q$.

We first show that the optimal loss mutualization scheme characterized in Proposition 1 cannot be incentive compatible in large markets. To see this, evaluate the monitoring constraint (8) at the optimal contract derived in Proposition 1. We obtain

$$\frac{\psi}{q(1-\alpha)} \leq \frac{v}{2}(1-q)^{N-1}\left(\hat{c} - 2x^{opt}\right) \tag{12}$$

Constraint (12) becomes tighter when N increases as the left-hand side converges exponentially to 0. There are two related reasons for this result. First, with complete loss mutualization, investors receive the full insurance payment $\hat{c}$ unless all payers default. Hence, a counterparty default only matters if all other payers default, an event with probability $(1-q)^{N-1}$. In addition, the amount of collateral x^{opt}, increases with N, as shown by equation (10). The lower "loss given (joint) default" $\hat{c} - 2x^{opt}$ as N increases also reduces investors' exposure to counterparty risk and thus weakens their incentives to monitor.

In the rest of the section, we focus on the case in which the optimal contract is not incentive compatible. We define N^* as the largest value of N such that (12) holds.

Assumption 3 (Monitoring problem)**.** *The number of investor pairs satisfies $N > N^*$. Hence, optimal loss mutualization is not incentive compatible for collateral cost $k \in [k_m, \bar{k}]$.*

We then derive the optimal incentive-compatible contract, which, to distinguish from the optimal contract, we call the *incentive-compatible (IC)* contract/loss mutualization.

4.1 Incentive-compatible loss mutualization

We start with rewriting the monitoring constraint (8) under a generic, simplified contract

$$\begin{aligned}\frac{\psi}{1-\alpha} \leq\ & \frac{1}{2}\Big[r_s - r_f + (1-q)^{N-1}(r_f - 2x)\Big] \\ & + \frac{v-1}{2}\Big[\min\{r_s, \hat{c}\} - \big(\big[1-(1-q)^{N-1}\big]\min\{r_f, \hat{c}\} + (1-q^{N-1})2x\big)\Big] \qquad \text{(8b)}\end{aligned}$$

We can see that the monitoring constraint can be relaxed by increasing r_s and/or decreasing r_f, which are the receiver transfers when her counterparty succeeds and defaults respectively. Intuitively, an investor has stronger incentives to monitor her counterparty when there is less loss mutualization, that is, when $r_s - r_f$ is bigger. Increasing r_s or reducing r_f to ensure loss mutualization is incentive-compatible induces different types of costs. Increasing r_s is costly because it increases the liability of payers, and, hence, requires more costly collateral to expand the payers' pledgeable income. Since the additional collateral provides insurance in the state in which all payers default, this IC contract features *over-insurance* as there are more gains from trades realised in this solution than in the optimal allocation. Meanwhile, reducing r_f implies that some gains from trade are foregone when the counterparty defaults, and, hence, we call it *under-insurance*. In the following proposition, we characterize the IC loss mutualization scheme. We denote the equilibrium variables with the superscript $*$.

Proposition 2 (IC loss mutualization with monitoring)**.** *For a given $k \in [k_m, \bar{k}]$, the IC loss mutualization with monitoring is incomplete. There exists a threshold $\hat{k} < \bar{k}$ such that the IC contract features*

1. *over-insurance if* $k < \hat{k}$ *with*

$$r_s^{*,oi} > \hat{c}, \quad r_f^{*,oi} = \hat{c}, \quad x^{*,oi} = \frac{\big(1 - (1-q)^{N-1}\left[qv + (1-q)\right]\big)\hat{c} - \beta q + \frac{2\psi}{1-\alpha}}{2 - 2(1-q)^{N-1}\left[qv + (1-q)\right] - \beta q} > x^{opt}; \tag{13}$$

2. *under-insurance if* $k > \hat{k}$ *with*

$$r_s^{*,ui} = \hat{c}, \quad r_f^{*,ui} < \hat{c}, \quad \textit{and} \quad x^{*,ui} = \frac{\hat{c} - q\beta - \frac{2\psi(1-q)}{vq(1-\alpha)}}{2 - q\beta} < x^{opt}. \tag{14}$$

Proposition 2 characterizes the IC contract when the optimal loss mutualization scheme is not incentive compatible. As we argued above, both over- and under-insurance contracts share a common feature: Loss mutualization is incomplete to preserve an investor's incentives to monitor her counterparty. Furthermore, over-insurance (under-insurance) is the preferred distortion when the collateral cost is low (high). To see why, it is instructive to derive the utility loss from these distortions. Over-insurance distorts the optimal contract and makes loss mutualization incomplete by increasing the transfers received by investors when their counterparties do not default, that is, $r_s^{*,oi} > r_s^{opt} = \hat{c}$. This increases the payers' liability and thus requires additional collateral $x^{*,oi} - x^{opt}$. Given that $r_f^{*,oi} = r_f^{opt} = \hat{c}$, the additional collateral supports insurance only when all payers default. The utility loss is

$$U^{opt} - U^{*,oi} = \left[k - (v-1)(1-q)^N\right](x^{*,oi} - x^{opt}), \tag{15}$$

where the term between brackets is the collateral cost net of the expected value of the additional insurance in the state in which all payers default. Hence, over-insurance is less costly when the cost of collateral is low.

In contrast, under-insurance makes loss mutualization incomplete by reducing the transfers received by investors when their counterparties default $r_f^{*,ui} < r_f^{opt} = \hat{c}$. Investors thus forgo valuable insurance while saving $x^{opt} - x^{*,ui}$ units of collateral as the payers' liability decreases. Overall, the utility loss is given by

$$U^{opt} - U^{*,ui} = \left[\bar{k} - k\right] (x^{opt} - x^{*,ui}), \tag{16}$$

with the term between brackets equal to the value of collateral if the investors' insurance needs are not satisfied when their counterparties defaults net of the cost of collateral. Since under-insurance allows investors to economize on the use of collateral, it is preferred when the cost of collateral is high.

To sum up, as loss mutualization improves insurance and in turn weakens investors' incentive to monitor their counterparties, the incentive-compatible contract has to limit the scope of loss mutualization. In the next section, we show that an alternative monitoring arrangement, namely, centralized monitoring, can sustain complete loss mutualization while incurring a different type of distortion.

4.2 Centralized Monitoring

In this section, we consider a different scheme where the monitoring efforts of all investors are centralized and delegated to a single external agent. This third-party agent is risk-neutral, has neither endowment nor asset, and is protected by limited liability. His monitoring effort is as costly as the investors and is also unobservable. By design, he is unable to provide insurance, has no insurance need nor a superior monitoring technology. We call this agent a CCP since his role as a monitor resembles that of an actual CCP, as discussed in Section 5.

To induce the CCP to monitor, the contract must also specify an explicit compensation scheme which is contingent on the outcomes of the payers. As the payers are homogeneous, it is without loss to consider a per-payer compensation scheme that depends on the number of defaulted counterparties $\pi(d)$. The CCP prefers monitoring all investors to none if

$$\mathbb{E}[(N-d)\pi(d)|m=1] - 2N\psi \geq \mathbb{E}[(N-d)\pi(d)|m=0] \tag{17}$$

The left-hand side of (17) is the expected compensation of the CCP given monitoring and the right-hand side is the one without. In the state with d defaulting payers, each of the remaining $(N-d)$ payers transfers $\pi(d)$ to the CCP as compensation. Monitoring all investors costs $2N\psi$ to the CCP. We call the above inequality the centralized monitoring constraint.[18]

Under centralized monitoring, there are two changes to the investor's problem: the addition of the design of the CCP's compensation $\pi(d)$ per payer and the replacement of the bilateral monitoring constraint by the centralized monitoring constraint. Importantly, contracting under centralized monitoring disentangles two key aspects of the investors' problem: the insurance needs and the monitoring incentives. As long as the CCP is properly incentivized, the investors can completely mutualize losses. In other words, we are left to solve the optimal design of the CCP's compensation.

The CCP's compensation contracting problem is the standard principal-agent problem with multiple efforts. Because of limited liability and the effort unobservability, the CCP earns an agency rent, that is, receiving a compensation strictly more than the effort costs. To minimize the agency rent, if there are enough resources, the CCP should only be rewarded in the state that is most indicative of all efforts being exerted. This is the state in which

[18]The CCP can also shirk on any number of efforts but the CCP achieves higher utility by shirking on all efforts than by shirking on some. Thus, the relevant incentive constraint is given by equation (18).

no payers default because monitoring efforts reduce counterparty risks. The centralized monitoring constraint then becomes

$$q^N N\pi(0) - 2N\psi \geq (\alpha q)^N N\pi(0) \tag{18}$$

As the payers are responsible for the compensation to the CCP at date 1, the resource constraint (2) when no payer defaults ($d = 0$) now reads as

$$\pi(0) + p_s(0) \leq x + (1-x)2R, \tag{19}$$

It is optimal for the investors to minimize the CCP's compensation and thus to bind the centralized monitoring constraint.

Lemma 3. *The optimal compensation scheme for the CCP is*

$$\pi^*_{cm}(d) = \begin{cases} \frac{2\psi}{q^N(1-\alpha^N)} & \textit{if} \quad d = 0 \\ 0 & \textit{if} \quad d > 0 \end{cases} \tag{20}$$

if the resources constraint (19) *holds for* $\pi^*_{cm}(0)$.

Constraint (19) could be violated by the compensation contract of Lemma 3 when N is too large. In this case, the optimal compensation contract would also specify a positive payment in the state $d = 1$, which is the second most indicative of monitoring efforts. Because the results are qualitatively similar, we will focus on the case in which equation (19) holds for $\pi^*_{cm}(d)$ for simplicity. In Section 5, we come back to this assumption when we discuss the relationship between the CCP's compensation in the model and the first-loss equity tranche of central counterparties in practice.

As we mentioned, the advantage of centralized monitoring is that it does not require distortion in loss mutualization. However, the need to compensate the CCP increases the payers' liability and thus demands more collateral to support the scheme. Indeed, the investor's pledgeability constraint (7b) now reads

$$\mathbb{E}_s[p_o(d)] + q^{N-1}\pi^*_{cm}(0) \leq x + (1-x)\beta, \tag{21}$$

where the second term on the left-hand side is the expected contribution to the compensation of the CCP given that the investor is a successful payer. We now present the IC contract with centralized monitoring.

Proposition 3 (IC loss mutualization with centralized monitoring)**.** *Given that Assumption 1-3 hold, the IC contract under centralized monitoring involves the CCP's compensation contract $\pi^*_{cm}(d)$ given by (20) and the multilateral contract with $r^*_{s,cm} = r^*_{f,cm} = \hat{c}$ and*

$$x^*_{cm} = x^{opt} + \frac{2\psi}{(1-\alpha^N)(2\big[1-(1-q)^N\big] - \beta q)} \tag{22}$$

Proposition 3 shows that with centralized monitoring, loss mutualization remains complete. The distortions required to make loss mutualization incentive compatible come in the form of additional collateral, which is used to support the CCP's compensation paid by the payers. This extra collateral requirement is proportional to the per-payer expected compensation of the CCP

$$\mathbb{E}[\pi^*_{cm}(d)] = \frac{2\psi}{1-\alpha^N}. \tag{23}$$

The cost of centralized monitoring can be represented by the investor's utility loss relative to the optimum. In the proof of Proposition 3, we show

$$U^{opt} - U^*_{cm} = \left[k - (v-1)(1-q)^N\right](x^*_{cm} - x^{opt}) + \frac{\alpha^N}{1-\alpha^N}\psi \tag{24}$$

The first component is equal to net cost of collateral multiplied by extra amount of collateral $x^*_{cm} - x^{opt}$ needed to implement the allocation. Similar to the case of over-insurance in bilateral monitoring, the additional collateral is costly on net because it only supports more gains from trade in the state when all payers default. The second term is the agency rent $\frac{1}{2}\mathbb{E}[\pi^*_{cm}(d)] - \psi$ per investor paid to the CCP due to the unobservability of efforts.

We note that when $\alpha > 0$, the agency rent component of centralized monitoring decreases in the market size N. These endogenous economies of scale in incentivizing efforts are similar to the diversification benefits in Diamond (1984).[19] In particular, when N is large, the agency rent tends to zero, as if the CCP's centralized monitoring efforts are observable.

4.3 Optimal monitoring arrangement

Having characterized the incentive-compatible allocation under bilateral and centralized monitoring, we can now answer the questions: Which monitoring arrangement is optimal? And when is monitoring optimal?

We show below that when the moral hazard problem is not degenerate ($\alpha > 0$), larger market and lower cost of collateral favor centralized monitoring over bilateral monitoring. We recall that bilateral monitoring could feature over- or under-insurance. The result stated above follows from comparing centralized monitoring first to bilateral monitoring with over-insurance and then to bilateral monitoring with under-insurance.

[19] It is also known as the *cross-pledging* benefits of contracting with multiple projects. See Tirole (2010) for a textbook treatment on the topic.

Relative to the optimal loss mutualization, the distortions in centralized monitoring and in bilateral monitoring with over-insurance are similar. Both feature the use of additional collateral, which is used to support the incentive payment to the CCP in centralized monitoring and the additional counterparty risk exposure, a form of incentive payment, to the investors in bilateral monitoring. Centralized monitoring is thus preferred when the required incentive cost to the CCP is smaller than the implicit incentive cost to the investors in bilateral monitoring. These costs, which can be seen in the respective monitoring constraints (18) and (8), are proportional to $\frac{\psi}{1-\alpha^N}$ under centralized monitoring and $\frac{\psi}{1-\alpha}$ under bilateral monitoring. Hence, under centralized monitoring, the incentive cost is lower than that under bilateral monitoring and decreases in the market size. This is the consequence of the endogenous economies of scale from centralized monitoring mentioned above. Therefore, when the market is larger, the agency rent reduction benefits are more significant, making centralized monitoring more desirable (relative to bilateral monitoring).

Next, vis-à-vis bilateral monitoring with under-insurance, centralized monitoring provides more insurance at the cost of larger collateral requirement. As a result, for centralized monitoring to dominate, the collateral cost cannot be too high. Putting the arguments together, we reach the conclusion that larger market and lower cost of collateral favor centralized monitoring.

We can only claim that centralized monitoring is optimal, however, if monitoring itself is optimal. The intuition for the optimal monitoring choice here is similar to the case of observable monitoring. As monitoring is costly, it is desirable only when the collateral is expensive enough. When monitoring is unobservable, it becomes effectively costlier because it has to be incentivized by distorting the allocation. Hence, as shown below, the new condition on the collateral cost for optimal monitoring is tighter than the condition $k \geq k_m$

derived in Proposition 1.

In the next proposition, we characterize the precise conditions for centralized monitoring to be optimal when the market becomes infinitely large. In the Appendix, we provide analytical conditions for finite N. The limit case $N \to \infty$ is useful for two reasons. First, we obtain simple analytical expressions highlighting the role of the model parameters. More importantly, as we show in the Appendix, the terms that depend on N in the general condition decay exponentially with N. Hence, the limit analysis is in fact instructive for small values of N, as also shown by our numerical example below.

Proposition 4 (Optimal monitoring arrangement)**.** *At the limit $N \to \infty$, when $\alpha > 0$, centralized monitoring is optimal for $k \in [k'_m, k_{cm}]$, with*

$$k'_m \equiv \frac{2-\beta q}{\frac{\beta q(1-\alpha)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right)-\psi}\frac{\psi}{2} > k_m, \qquad k_{cm} \equiv \frac{(1-q)\bar{k}}{1-q+vq(1-\alpha)} < \bar{k}, \tag{25}$$

and $k'_m < k_{cm}$ is implied by Assumption 2.

We illustrate these results with a numerical exercise in Figure 1. Both panels show the range of collateral cost and market size in which centralized monitoring is optimal. The right panel corresponds to a higher value of α. We first observe that the characterization of the centralized monitoring region as an intermediate range of collateral costs, also applies for finite values of N. In fact, this range does not change substantially as N increases. As we argued, the conditions for the limit case $N \to \infty$ in Proposition 4 are a good approximation even for small values of N. Next, when comparing the two panels, we see that increasing α has an ambiguous effect on the region in which centralized monitoring is optimal: Centralized monitoring dominates bilateral monitoring for larger values of k but simultaneously monitoring itself is dominated for a larger range of values of k. This numerical result is confirmed analytically in the limit case of Proposition 4: k'_m and k_{cm} both increase with α. Increasing α favors centralized monitoring relative to bilateral monitoring due to the agency-cost-reduction benefits of centralized monitoring. Meanwhile, as α increases, the benefits from monitoring decrease since the default probability of an unmonitored investor $1 - \alpha q$ is lower.

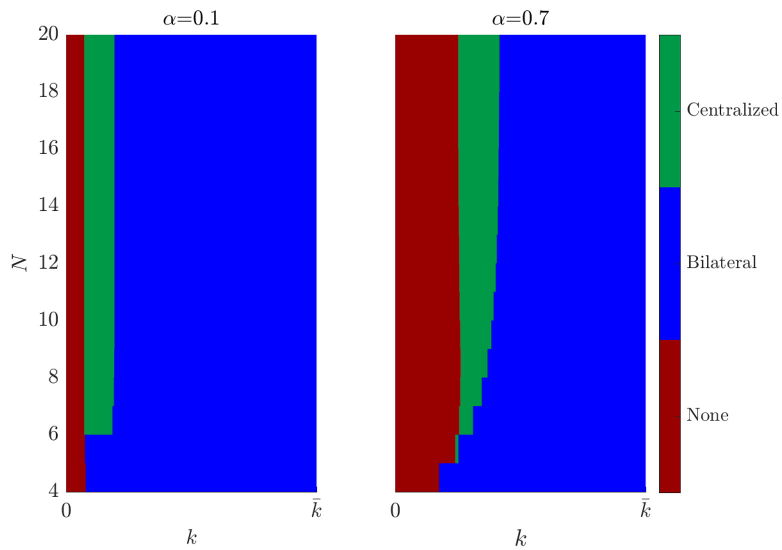

Figure 1. Incentive-compatible monitoring. *Parameter values:* $\hat{c} = 0.8$, $\beta = 0.4$, $v = 2$, $q = 0.7$, $\psi = 5.6 \times 10^{-3}$.

5 Implications for CCP design

In this section, we explain how our results relate to the design of a central counterparty in practice. Section 3.2 and 3.3 already discussed the implementation of our multilateral contract as a novated contract and some features of the CCP default waterfall, respectively. Hence, we focus on the novel implications derived in Section 4. We first derive implications for the CCP ownership structure. Then, we show that the optimal compensation schedule of the agent under centralized monitoring relates to the loss absorption capacity of a CCP.

CCP Ownership Structure In our analysis of the incentive-compatible contracts, we compared two different schemes. Under centralized monitoring, monitoring efforts are expended by a single third-party agent rather than by individual clearing members. The monitoring effort can be interpreted as the costly process of vetting members and ensuring that risk management practice is adequate. This third-party agent, who is only responsible for members' creditworthiness, but neither receives nor provides insurance, can be described as a third-party CCP. When bilateral monitoring is optimal, there is no need to involve a third-party. In this case, we interpret the optimal arrangement as a member owned CCP.

Our results suggest that a third-party CCP is preferable to a member owned CCP when the number of clearing members is large. We showed that free-riding benefits from lack of monitoring are higher in large CCPs where losses are more efficiently shared among members. Hence, discipline is better maintained via centralization of monitoring efforts. Monitoring is indeed one of the key roles of CCPs and can take many forms. Among other due diligence exercises, ESMA (2020) reports that CCPs must use an internal credit classification, send mandatory due diligence questionaires and perform onsite visits of their members. Finally, by highlighting the role of factors such as market size or the cost of collateral for the ownership structure, our model also informs the policy debate about the optimal ownership structure (see e.g. Board (2010)).

CCP Equity Tranche Default by a seller in our model is a sign that the third-party CCP did not exert due diligence when vetting clearing members. Under the optimal compensation contract, the third-party CCP only receives a payment when no seller defaults.

By concentrating the CCP's payoff in this state of the world, his incentives are preserved at the minimal cost. With this first-loss exposure, the CCP contract resembles a junior equity tranche, which is a typical feature of default waterfalls in practice (see e.g. Duffie (2015)). As in our model, the description of the default waterfall of the Japan Securities Clearing Corporation (JSCC) explicitly refers to the CCP incentives.[20]

> *JSCC should compensate losses before Survivors' Pay, in order to keep incentive for appropriate risk management*

We observed in Section 3.3 that a very high-powered incentive contract for the CCP may not be feasible because it requires a very large payment to the CCP when all payers survive. Optimal contracting predicts that the CCP would also be compensated in the next most informative states, that is, when all but one member survives. Following our line of interpretation, the CCP equity tranche would not be wiped out if only one member defaults.

Our optimal incentive compatible contract also implies that surviving members should be exposed to the default risk of other members, even if they are not directly responsible. Hence, in the words of Coeuré (2015), *CCPs are risk poolers, not insurance providers.* In fact, we show that the optimal CCP may be owned by its members, in which case, surviving members are directly exposed after the resources of a defaulting member are exhausted (McPartland and Lewis (2017) make this point informally).

6 Conclusion

This paper characterizes the optimal loss mutualization scheme in a Central Clearing Counterparty. Loss mutualization hedges insurance buyers against the counterparty risk of in-

[20] See https://www.jpx.co.jp/jscc/en/risk/default.html

surance providers but it also lowers market discipline since buyers have less incentives to search for creditworthy counterparties. We show that a third-party CCP can mitigate these inefficiencies by acting as centralized monitor. We predict that third-party CCPs are more likely to arise for contracts with a large user base because of (endogenous) economies of scale. Less traded contracts are more efficiently cleared with a member owned CCP. Our paper is one of the first to consider the ownership structure of a CCP and to endogenize the junior equity tranche of third-party CCPs.

Our aim is to understand the basic determinants of the default waterfall of CCPs but our framework could be extended to discuss more complex aspects of the capital structure of CCPs. Additionally, while in our paper a CCP is efficiently run by design, there has been a growing concern that some of these market players became "too-big-to-fail" and that they impose externalities on financial markets. Related to this issue, the trade-off discussed in banking between competition and stability seems to apply to CCPs as well. We leave these interesting venues for future research.

We end with a discussion of some limitations of our analysis. In particular, our model does not consider netting efficiency, an important benefit of having a CCP. There are two types of netting that can be performed via a CCP. First, when a member has two opposite positions on the same contract with two different members, a CCP can net out the gross positions and impose collateral requirement on the net exposure. Second, if members trade different risks or contracts with different members, a CCP who clears all these contracts can compute the total risk exposure and charge the margin at a *portfolio level*. Diversification of risks would thus net out some of the risk exposure and margin requirement. These netting benefits are not captured in our model because each investor only takes one position and there is only a single source of risk. We choose not to consider netting benefits in order to focus on the loss mutualization benefits and the associated design aspects of the CCP. As a result, the benefit of central clearing identified in this paper should be considered as a lower bound. For a comprehensive survey on various benefits and costs of central clearing, see Menkveld and Vuillemey (forthcoming).

Appendix

A Proofs

A.1 Proof of Lemma 1

With a slight abuse of notation, denote $q(m)$ the probability the asset of an investor succeeds in the relevant aggregate state, with $m \in \{0,1\}$ the monitoring decision. By definition, $q(0) = \alpha q$ and $q(1) = q$. For a given monitoring choice m, the number of defaulting payers among k is a random variable with a binomial distribution $\mathcal{B}(k, q(m))$. Taking expectation over budget constraints (4), we thus obtain

$$\begin{aligned}
\mathbb{E}_s[p_o(d)] &= \sum_{d=0}^{N-1} (1-q(m))^d q(m)^{N-1-d} \binom{N-1}{d} \left[r_s(d) + \frac{d}{N-d}(r_f(d) - p_f(d)) - \frac{N}{N-d}x \right] \\
&= \mathbb{E}_s[r_o(d)] + \sum_{d=1}^{N-1} (1-q(m))^d q(m)^{N-1-d} \binom{N-1}{d-1} (r_f(d) - p_f(d)) - x \sum_{d=0}^{N-1} (1-q(m))^d q(m)^{N-1-d} \binom{N}{d} \\
&= \mathbb{E}_s[r_o(d)] + \frac{1-q(m)}{q(m)} \sum_{l=0}^{N-1} (1-q(m))^l q(m)^{N-1-l} \binom{N-1}{l} (r_f(l+1) - p_f(l+1)) - \frac{(1-q(m))^N}{q(m)} x \\
&\quad - x \sum_{d=0}^{N-1} (1-q(m))^d q(m)^{N-1-d} \binom{N}{d} \\
&= \mathbb{E}_s[r_o(d)] + \frac{1-q(m)}{q(m)} (\mathbb{E}_f[r_o(d)] - \mathbb{E}_f[p_o(d)]) - \frac{x}{q(m)}
\end{aligned}$$

The last line is equivalent to equation (5).

An investor utility with contract $\mathcal{C}$ and monitoring choice $m \in \{0,1\}$ is given by

$$U = \frac{1}{2}\Big(q(1-x)2R + x - \mathbb{E}[p_o(d)]\Big) + \frac{1}{2}\Big(\mathbb{E}[r_o(d)] + (v-1)\mathbb{E}\big[\min\{r_o(d), \hat{c}\}\big]\Big)$$

Substituting $\mathbb{E}[p_o(d)]$ thanks to equation (5), we obtain

$$U = qR + \frac{1}{2}x - qRx + \frac{1}{2}x + \frac{v-1}{2}\mathbb{E}\big[\min\{r_o(d), \hat{c}\}\big]$$

which is equivalent to equation (6).

A.2 Proof of Lemma 2

In the proof, we use the remark following Definition 4 that $x \leq \frac{\hat{c}}{2}$ without loss.

Proof that resource constraint (3) binds

From equation (4), increasing $p_f(d)$ for $d < N$ allows investors to increase $r_s(d)$ in this state. Such a change may only relax limited pledgeability constraint (7). Since investors' utility (6) is weakly increasing with $r_s(d)$, it is optimal to saturate resource constraint (3). Hence, $p_f(d) = x$ for all $d < N$.

For state $d = N$, suppose (3) is slack and consider increasing $p_f(N)$ by $\Delta p_f(N) \in (0, x - p_f(N)]$. Denote $\Delta\mathbb{E}_f[p_o(d)]$ the corresponding increase in $\mathbb{E}_f[p_o(d)]$. Let us also increase $\mathbb{E}_s[p_o(d)]$ by $\Delta\mathbb{E}_s[p_o(d)] = \Delta\mathbb{E}_f[p_o(d)]$ in order to ensure limited plegeability constraint (7) still holds. Consider then a joint increase in $r_f(N)$ and $\Delta\mathbb{E}_s[r_o(d)]$ such that

$$\Delta r_f(N) \leq \Delta p_f(N), \qquad \Delta\mathbb{E}_s[r_o(d)] \geq v\mathbb{E}_f[r_o(d)], \qquad \Delta\mathbb{E}_s[r_0(d)] \leq \Delta\mathbb{E}_s[p_o(d)]$$

The first constraint ensures that resource constraint (3) is still satisfied following the perturbation. The second constraint ensures that bilateral monitoring constraint (8) is satisfied after the perturbation if needed. The last constraint ensures that budget constraint (4) is still satisfied. Since $\Delta p_f(N) > 0$ and $\mathbb{E}_s[r_0(d)] > 0$, by construction, such a perturbation exists. Since

$$c_f(N) \leq p_f(N) + x < 2x < \hat{c},$$

where the inequalities follow from our assumptions, this perturbation strictly increases investors' utility (6).

Proof that $r_s(d) = r_s$ for all $d < N$

Let two states (d, d') such that $r_s(d) > r_s(d')$. We argue that the following perturbation weakly increases investors' utility: decrease $r_s(d)$ and $p_s(d)$ and increase $r_s(d')$ and $p_s(d')$ such that $\mathbb{E}_s[r_o(d)]$ and $\mathbb{E}_s[p_o(d)]$ are unchanged. This perturbation is feasible because it does not affect constraint (7) and it weakly relaxes bilateral monitoring constraint (8). It is (weakly) profitable because the objective function (6) is concave in $r_s(d)$ and $r_s(d')$ and it is strictly profitable if $r_s(d) > \hat{c} > r_{s'}(d')$.

Proof that $r_f(d) = r_f$ for all $d < N$

Let two states (d, d') such that $r_f(d) > r_f(d')$. The argument used above also applies here if $r_f(d) > r_f(d') \geq \hat{c}$ or if $r_f(d') < r_f(d) \leq \hat{c}$. Hence, we are left to analyze the case in which $r_f(d') < \hat{c} < r_f(d)$. For $\epsilon > 0$ small enough, consider the following perturbation

$$(\Delta r_f(d'), \Delta r_f(d)) = \left(\epsilon, -\frac{f(d')}{f(d)}v\epsilon\right)$$

with $f(d)$ the probability that d payers default among $N-1$. The perturbation is designed such that the right-hand side of incentive constraint (8) is unchanged. To satisfy budget constraint (4) in state d and d', set $\Delta p_s(d) = \frac{1-q}{q}\Delta r_f(d)$ and $\Delta p_s(d') = \frac{1-q}{q}\Delta r_f(d')$. The limited pledgeability constraint (7) still holds after the perturbation because the expected payment $\mathbb{E}_s[p_o(d)]$ increases by

$$\Delta\mathbb{E}_s[p_o(d)] = -\frac{1-q}{q}(v-1)f(d')\epsilon$$

The perturbation strictly increases the objective function (6) which is concave in r_f.

A.3 Proof of Proposition 1

In the proof, we first derive the optimal contract for a given monitoring choice $m \in \{0,1\}$ (Step 1) and then derive the optimal monitoring decision (Step 2). We use the notation introduced in the proof of Lemma 2 and let $q(m)$ be the probability a payer succeeds given monitoring decision m.

Step 1. Optimal contract

We first derive a simplified version of the investor problem in Definition 2 thanks to the results from Lemma 2. We have

$$\begin{aligned}\mathbb{E}_f[p_o(d)] &= x\\ q(m)\Big(\mathbb{E}_s[p_o(d)] - \mathbb{E}_f[p_o(d)]\Big) &= \mathbb{E}_s[r_o(d)] + \frac{1-q(m)}{q(m)}(\mathbb{E}_f[r_o(d)] - \mathbb{E}_f[p_o(d)]) - \frac{x}{q(m)}\\ &= q(m)r_s + (1-q(m))\big[1-(1-q(m))^{N-1}\big]r_f - \Big[1-(1-q(m))^N\Big]2x\end{aligned}$$

Given $m \in \{0,1\}$, a contract $\mathcal{C}$ is optimal if x, r_s, r_f solve the following problem

$$\max_{x,r_s,r_f} \frac{v-1}{2}\left[q(m)\min\{r_s,\hat{c}\} + (1-q(m))\Big(\big[1-(1-q(m))^{N-1}\big]\min\{r_f,\hat{c}\} + (1-q(m))^{N-1}2x\Big)\right] - x(qR-1)$$

$$\text{subject to}\quad q(m)r_s + (1-q(m))\big[1-(1-q(m))^{N-1}\big]r_f \leq q(m)\beta + \Big(2 - q(m)\beta - 2(1-q(m))^N\Big)x \tag{A.1}$$

where (A.1) is the limited pledgeability constraint (7) expressed as a function of r_s, r_f and x only, using the equations above.

The objective function is strictly increasing with r_s and r_f for all $r_s \leq \hat{c}$ and $r_f \leq \hat{c}$ and constant otherwise. It is thus weakly optimal to set $r_s \leq \hat{c}$ and $r_f \leq \hat{c}$. Two cases are then possible. Either $r_s = r_f = \hat{c}$ or constraint (A.1) binds. In the first case, the derivative of the objective function with respect to the collateral x is given by

$$\frac{\partial U}{\partial x} = (v-1)(1-q(m))^N - (qR-1) = \hat{k}_1(m) - (qR-1)$$

If $qR - 1 \leq \hat{k}_1(m)$, then x should be increased until $r_f(N) = 2x$ reaches $\hat{c}$. This implies

$x = \frac{\hat{c}}{2}$. Otherwise, x should be decreased until constraint (A.1) binds. We are thus left to consider the case in which (A.1) binds. Then, plugging (A.1) into the objective function, the maximization problem is equivalent to:

$$\max_x \left[\frac{v-1}{2}(2-q(m)\beta)-(qR-1)\right]x = \left[\hat{k}_2(m)-(qR-1)\right]x$$

$$\text{subject to} \quad 0 \le x \le x(m) \equiv \frac{\left[1-(1-q(m))^N\right]\hat{c}-\beta q(m)}{2\left[1-(1-q(m))^N\right]-\beta q(m)}$$

where the upper bound on x is obtained by setting $r_s = r_f = \hat{c}$ in (A.1). If $qR-1 \le \hat{k}_2(m)$, then $x = (m)$. Otherwise, if, $qR-1 \ge \hat{k}_2(m)$, then $x = 0$ is optimal. In this case, any contract such that

$$r_s + \frac{1-q(m)}{q(m)}\left[1-(1-q(m))^{N-1}\right]r_f = \beta$$

is optimal. Observe also that $\hat{k}_2(m) > \hat{k}_1(m)$.

We thus fully characterized the optimal contract for a given monitoring decision m. If $qR-1 \le \hat{k}_1(m)$, the full insurance contract is optimal. If $qR-1 \in [\hat{k}_1(m), \hat{k}_2(m)]$, the complete loss mutualization contract with $r_s = r_f = \hat{c}$ is optimal. In this case, we showed the collateral amount is given by (10) for a given monitoring choice m. Finally, if $qR-1 \ge \hat{k}_2(m)$, the optimal amount of collateral is given by $x(m) = 0$ and the optimal contract can be implemented bilaterally with $r_s = \beta$ and $r_f = 0$ is optimal.

Step 2. Optimal monitoring decision

An allocation is characterized by a type of contract and a monitoring decision. For any $m \in \{0,1\}$, denote $\mathcal{C}^{*,m}$ the optimal contract with $\mathcal{C}^{*,m} \in \{FI^m, CM^m, NC^m\}$. FI stands for Full Insurance, CM for Complete (Loss) Mutualization and NC for No Collateral. Observe that since $q(0) < q(1)$, we have $\hat{k}_i(1) < \hat{k}_i(0)$ for $i = 1,2$ by definition of these thresholds.

We first show that the benefit of monitoring is increasing with the collateral cost $qR-1$. S If $qR-1 \le \hat{k}_1(1)$, the full insurance contract is optimal for any $m \in \{0,1\}$. Hence, the net benefit from monitoring is strictly negative and independent of the collateral cost. Suppose now $qR-1 \in [\hat{k}_1(1), \hat{k}_2(1)]$ so that $\mathcal{C}^{*,1} = CM^1$ and $\mathcal{C}^{*,0} \in \{FI^0, CM^1\}$. If $\mathcal{C}^{*,0} = FI^0$, the net benefit of monitoring is given by

$$U(\mathcal{C}^{*,1}) - U(\mathcal{C}^{*,0}) = \frac{1}{2}\left[qR-1-(v-1)(1-q)^N\right](\hat{c}-2x(1)) - \psi$$

with $x(1)$ the collateral in the CM^1 contract. The benefit is strictly increasing with $qR-1$ since $2x(1) < \hat{c}$. If instead $\mathcal{C}^{*,0} = LM$, we have

$$U(\mathcal{C}^{*,1}) - U(\mathcal{C}^{*,0}) = \frac{1}{2}\left[qR-1-(v-1)(1-q)^N\right](\hat{c}-2x(1)) - \frac{1}{2}\left[qR-1-(v-1)(1-\alpha q)^N\right](\hat{c}-2x(0)) - \psi$$

The collateral $x(m)$ in the CM^m contract is independent of $qR-1$. Hence, the expression above is increasing in $qR-1$ because $x(0) > x(1)$.

We are left to consider the case $qR-1 \geq \hat{k}_2(1)$. Then, $\mathcal{C}^{*,1} = NC^1$ and $\mathcal{C}^{*,0} \in \{FI, CM^0, NC^0\}$. If $\mathcal{C}^{*,0} = NC^0$, no collateral is used for any $m \in \{0,1\}$ which implies that the benefit from monitoring is constant as a function of the collateral cost. Suppose then that $\mathcal{C}^{*,0} = CM^0$. Then, the net benefit of monitoring is equal to

$$U(\mathcal{C}^{*,1}) - U(\mathcal{C}^{*,0}) = U(NC^1) - U(CM^1) + \left[U(CM^1) - U(CM^0)\right]$$

We already showed that the second term between brackets is increasing with $qR-1$. The first term is also increasing in $qR-1$ because the NC^1 contract requires no collateral while the LM^1 contract collateral requirement is positive independent of $qR-1$. If $\mathcal{C}^{*,0} = FI$, we can conclude with a similar argument. We thus proved that the value of monitoring is increasing with $qR-1$.

Denote k_m the threshold, if any, above which monitoring is optimal. Since the benefit of monitoring is increasing in $qR-1$, this threshold is unique if it exists. Our analysis above already shows that $k_m > \hat{k}_1(1)$. We now want to show that $k_m \leq \hat{k}_2(1)$. Monitoring is optimal for $qR-1 = \hat{k}_2(1)$ if and only if

$$\begin{aligned}
&0 \leq U(\mathcal{C}^{*,1})_{|qR-1=\hat{k}_2(1)} - U(\mathcal{C}^{*,0})_{|qR-1=\hat{k}_2(1)} \\
&0 \leq \frac{v-1}{2}q\beta - \psi - \left[v-1-\bar{k}\right]\frac{\hat{c}}{2} - \frac{1}{2}\max\left\{0, \alpha\frac{\hat{k}_2(1) - (v-1)(1-\alpha q)^N}{2\left[1-(1-\alpha q)^N\right] - \beta\alpha q}\right\}\beta q(2-\hat{c}) \\
&\psi \leq \frac{v-1}{2}q\beta\left(1-\frac{\hat{c}}{2}\right) - \alpha\beta q\frac{v-1}{2}\left(1-\frac{\hat{c}}{2}\right)\max\left\{0, \frac{2-q\beta-2(1-\alpha q)^N}{2\left[1-(1-\alpha q)^N\right]-\beta\alpha q}\right\}
\end{aligned}$$

To derive the second line, we used $\mathcal{C}^{*,1} = NC^1$ and $\mathcal{C}^{*,0} \in \{FC^0, LM^0\}$ for $qR-1 = \hat{k}_2(1)$. In the last equation, observe that the argument of the max is increasing in N. Hence, the inequality above holds if it holds in the limit $N \to \infty$. We have

$$\lim_{N\to\infty} \frac{2-q\beta-2(1-\alpha q)^N}{2\left[1-(1-\alpha q)^N\right]-\beta\alpha q} = \frac{2-\beta q}{2-\beta\alpha q}$$

Hence, the last inequality holds for all N if

$$\psi \leq \frac{\beta q(1-\alpha)(v-1)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right)$$

It is straightforward to verify that the right-hand side is lower than ψ_{max} and thus that this condition holds under Assumption 2. This implies that $k_m \leq \hat{k}_2(1)$.

To conclude we need to define the thresholds $\underline{k}_N$ and $\bar{k}$ and show that the monitoring threshold k_m lies in $[\underline{k}_N, \bar{k}]$. Define

$$\bar{k} \equiv \hat{k}_2(1) = \frac{v-1}{2}(2 - q\beta)$$

This implies $k_m \leq \bar{k}$ since we showed that monitoring is optimal for $k = \bar{k}$ and the benefit from monitoring is increasing with the collateral cost. To define $\underline{k}_N$, let us first derive the threshold $\hat{k}_m$ such that the CM^1 contract delivers the same utility as the FI^0 contract. Our analysis above shows this threshold exists and it is defined implicitly by

$$\begin{aligned} 0 = U(CM^1)_{\hat{k}_m} - U(FI^0)_{\hat{k}_m} &= \frac{1}{2}\big[\hat{k}_m - (v-1)(1-q)^N\big](\hat{c} - 2x^{opt}(1)) - \psi \\ &= \frac{\hat{k}_m - (v-1)(1-q)^N}{2\big[1-(1-q)^N\big] - \beta q}\beta q\left(1 - \frac{\hat{c}}{2}\right) - \psi \end{aligned} \tag{A.2}$$

where we used equation (10) to substitute for $x^{opt}(1)$. Hence, we obtain

$$\hat{k}_m = (v-1)(1-q)^N + \frac{2\psi}{\beta q(2-\hat{c})}\Big(2\big[1-(1-q)^N\big] - \beta q\Big)$$

We ca now define $\underline{k}_N$ as

$$\underline{k}_N = \min\Big\{(v-1)(1-\alpha q)^N, \hat{k}_m\Big\}$$

If $\underline{k}_N = \hat{k}_m$, then by definition $k_m = \underline{k}_N$. If $\underline{k}_N = (v-1)(1-\alpha q)^N$, this implies that contract CM^1 is dominated by contract FI for $k = \underline{k}_N$ and hence that the monitoring threshold satisfies $k_m \geq \underline{k}_N$. This concludes the proof.

A.4 Proof of Corollary 1

Equation (11) shows that $\bar{k}$ does not depend on N. To prove the result for $\underline{k}_N$, consider equation (A.2) in the proof of Proposition 1. Let $g : y \mapsto \frac{\bar{k}+(v-1)y}{2+2y-\beta q}$. We have

$$g'(y) = \frac{(v-1)(2-\beta q) - 2\underline{k}_N}{[2+2y-\beta q]^2} = \frac{2(\bar{k} - \underline{k}_N)}{[2+2y-\beta q]^2} \geq 0$$

where the last inequality follows from Proposition 1. Sincee $y = -(1-q)^N$ is increasing with N, the term on the right-hand side of (A.2) is increasing with N. Since this term is also increasing with $\underline{k}_N$, by the Implicit Function Theorem, $\underline{k}_N$ is decreasing with N.

A.5 Proof of Corollary 2

By Corollary 1, $\underline{k}_1 > \bar{k}_N$. Hence, by Proposition 1, the optimal bilateral contract uses strictly more collateral than the optimal multilateral contract for $k \in [\underline{k}_N, \underline{k}_1]$. For $k \in [\underline{k}_1, \bar{k}]$, the collateral requirement in the optimal bilateral contract is given by $x^{opt}_{|N=1} = \frac{1-\beta}{2-\beta}$ independently of the optimal monitoring decision m^{opt} for $N = 1$. Hence, since $x^{opt}(m)$ is strictly increasing with N for $m \in \{0, 1\}$, it follows that the collateral requirement is strictly lower in the optimal bilateral contract.

A.6 Proof of Proposition 2

The bilateral monitoring constraint (8) must bind. Otherwise the second-best allocation can be implemented which is a contradiction with Assumption 3. The pledgeability constraint (7) must also bind. If not, it is optimal to decrease x until (7) binds. To see this, compute the marginal effect of a decrease in x

$$\frac{\Delta U}{\Delta x} = -(v-1)(1-q)^N + qR - 1 = -\underline{k} + qR - 1 \geq 0$$

where the first term captures the effect of decreasing x on the payment $r_f(N) = 2x$. The inequality above follows from Assumption 3.

Hence, the IC contract maximizes the investors' utility (6) under the binding pledgeability constraint (7b) and the binding monitoring constraint (8b). Since $r_f \leq \hat{c}$ is optimal without constraint (8b) by Proposition 1 and increasing r_f tightens constraint (8b), it follows that $r_f \leq \hat{c}$. The opposite argument implies that $r_s \geq \hat{c}$. Using these results, we can solve for r_s and r_f as a function of the collateral cost x only. From (7b) and (8b), we have

$$\begin{aligned} qr_s + (1-q)\big[r_f - (1-q)^{N-1}(r_f - 2x)\big] &= (2 - q\beta)x + q\beta \\ r_s - v\big[r_f - (1-q)^{N-1}(r_f - 2x)\big] &= \frac{2\psi}{1-\alpha} - (v-1)\hat{c} \end{aligned}$$

Hence, we obtain

$$(1-q)\Big[r_f - (1-q)^{N-1}(r_f - 2x)\Big] = \frac{(1-q)\big[(2-q\beta)x + q\beta\big] - q(1-q)\left[\frac{2\psi}{1-\alpha} - (v-1)\hat{c}\right]}{qv + (1-q)}$$

Since the investors utility (6) is independent of r_s when $r_s \geq \hat{c}$, the effect of marginal increase in x is given by

$$\frac{\partial U}{\partial x} = \frac{v-1}{2}\frac{(1-q)(2-q\beta)}{qv + 1 - q} - k$$

Let $\hat{k}$ denote the threshold for the collateral cost equal to the first term on the right-hand side. Two cases are possible. First if $k > \hat{k}$, the IC contract is such that $r_s = \hat{c}$. Using the binding constraints (7b) and (8b), the values of x and r_f in equation (14) obtain.

If $k < \hat{k}$, collateral should be increased until $r_f = \hat{c}$. Using once again the binding constraints (7b) and (8b), we obtain the remaining contract variables r_s and x as in (13).

Finally, since $(1-q) < qv + 1 - q$, $\hat{k} < \bar{k} = \frac{v-1}{2}(2 - q\beta)$ as stated in Proposition 2.

A.7 Proof of Proposition 3

We first prove Proposition 3 and then provide the derivations leading to equation (24).

By construction, the compensation contract given in (20) is the cheapest incentive-compatible contract for the agent. Since monitoring is not done bilaterally by investors, the results in Proposition 1 for $m^{opt} = 1$ apply here. Investors optimally consume $\hat{c}$ unless all payers default, that is $r^*_{s,cm} = r^*_{f,cm} = \hat{c}$. Given the characterization of contracts in Lemma 2, we are thus left to derive the collateral requirement x^*_{cm}

Using the binding budget constraint (5) and the limited pledgeability constraint (7) as well as the compensation schedule given by (20), we obtain

$$\hat{c} - (1-q)^N(\hat{c} - 2x^*_{cm}) = q\left(x^*_{cm} + (1 - x^*_{cm})\beta - \frac{2\psi}{q(1-\alpha^N)}\right) + (1-q)x^*_{cm} + x^*_{cm}$$

We thus find

$$x^*_{cm} = \frac{\left[1-(1-q)^N\right]\hat{c} - \beta q + \frac{2\psi}{1-\alpha^N}}{2\left[1-(1-q)^N\right] - \beta q}$$

which can be rewritten as (22).

We now provide the derivations for equation (24). Note that the in the expression for the investor utility the monitoring cost ψ is replaced by his expected contribution to the agent compensation, equal to $\frac{1}{2}\bar{\pi}^*_{cm}$. Hence, we obtain

$$U^*_{cm} = qR + \frac{v-1}{2}\Big(\hat{c} - (1-q)^N(\hat{c} - 2x^*_{cm})\Big) - \frac{1}{2}\bar{\pi} - x^*_{cm}(qR-1) \tag{A.3}$$

Subtituting $\bar{\pi}^*_{cm} = \frac{2c}{1-\alpha^N}$ and subtracting (A.3) to the utility with the optimal contract, we get (24).

A.8 Proof of Proposition 4

We first compare centralized monitoring to no monitoring. For large N, the optimal contract without monitoring features complete loss mutualization, as shown in Proposition 1. To express the condition that centralized monitoring dominates no monitoring, we derive for each allocation the utility gain with respect to the full collateral allocation. We have

$$U_{dm} = qR + \Big[v - 1 - k\Big]\frac{\hat{c}}{2} + \Big[k - (v-1)(1-q)^N\Big]\left(\frac{\hat{c}}{2} - x^*_{cm}\right) - \frac{\psi}{1-\alpha^N}$$
$$U^{opt}_{\emptyset m} = qR + \Big[v - 1 - k\Big]\frac{\hat{c}}{2} + \Big[k - (v-1)(1-\alpha q)^N\Big]\left(\frac{\hat{c}}{2} - x^{opt}_{\emptyset m}\right)$$

where the subscript $\emptyset m$ is used to denote no monitoring. From Proposition 1 and 3, we have

$$\frac{\hat{c}}{2} - x^*_{dm} = \frac{\beta q\left(1 - \frac{\hat{c}}{2}\right) - \frac{2\psi}{1-\alpha^N}}{2\big[1 - (1-q)^N\big] - \beta q}$$
$$\frac{\hat{c}}{2} - x^{opt}_{\emptyset m} = \frac{\beta\alpha q}{2\big[1 - (1-\alpha q)^N\big] - \beta\alpha q}\left(1 - \frac{\hat{c}}{2}\right)$$

Hence, delegated monitoring dominates no monitoring, that is, $U_{dm} \geq U^{opt}_{\emptyset m}$ if and only if

$$\frac{qR - 1 - (v-1)(1-q)^N}{2\big[1-(1-q)^N\big] - \beta q}\left[\beta q\left(1 - \frac{\hat{c}}{2}\right) - \frac{2\psi}{1-\alpha^N}\right] - \frac{\psi}{1-\alpha^N} \geq \frac{qR - 1 - (v-1)(1-\alpha q)^N}{2\big[1-(1-\alpha q)^N\big] - \beta\alpha q}\beta\alpha q\left(1 - \frac{\hat{c}}{2}\right)$$

Taking the limit when $N \to \infty$, we obtain

$$\frac{k}{2 - \beta q}\left[\beta q\left(1 - \frac{\hat{c}}{2}\right) - 2\psi\right] - \psi \geq \frac{k}{2 - \beta\alpha q}\beta\alpha q\left(1 - \frac{\hat{c}}{2}\right)$$

Since, under Assumption 2,

$$\psi \leq \frac{\beta q(1-\alpha)}{2 - \beta\alpha q}\left(1 - \frac{\hat{c}}{2}\right)$$

the condition can be expressed as a lower bound k'_m on k with

$$k'_m = \frac{2 - \beta q}{\frac{\beta q(1-\alpha)}{2-\beta\alpha q}\left(1 - \frac{\hat{c}}{2}\right) - \psi}\frac{\psi}{2}$$

We now turn to the comparison between centralized monitoring and bilateral monitoring. Using equations (15) and (24), delegated monitoring dominates bilateral monitoring with over-insurance if and only if

$$\Big(k - (v-1)(1-q)^N\Big)\left(x^*_{cm} - x^{opt}\right) + \frac{\alpha^N}{1-\alpha^N}c \leq \Big(k - (v-1)(1-q)^N\Big)\left(x^{*,oi} - x^{opt}\right)$$

Rewriting equation (13), we obtain

$$x^{*,oi} - x^{opt} = \frac{2\psi}{[1-\alpha]\left[2(1-(1-q)^N) - \beta q\right]} - \frac{vq(1-q)^{N-1}}{2(1-(1-q)^N) - \beta q}(\hat{c} - 2x^{*,oi})$$
$$= \frac{2\psi}{[1-\alpha]\left[2(1-(1-q)^N) - \beta q\right]} - \frac{vq(1-q)^{N-1}}{2(1-(1-q)^N) - \beta q}\frac{\beta q(2-\hat{c}) - \frac{4\psi}{1-\alpha}}{2\left[1-(1-q)^{N-1}(vq+1-q)\right] - \beta q}$$

We thus obtain the following condition

$$\frac{\alpha^N}{1-\alpha^N}\psi \leq \left[k - (v-1)(1-q)^N\right](x^{*,oi} - x^*_{dm}) \tag{A.4}$$
$$\frac{\alpha^N}{1-\alpha^N}\psi \leq \frac{k-(v-1)(1-q)^N}{2(1-(1-q)^N) - \beta q}\left[\frac{2\psi}{1-\alpha} - \frac{2\psi}{1-\alpha^N} - vq(1-q)^{N-1}\frac{\beta q(2-\hat{c}) - \frac{4\psi}{1-\alpha}}{2\left[1-(1-q)^{N-1}(vq+1-q)\right] - \beta q}\right]$$

which we refer to as $F^{OI} \geq 0$ for simplicity. Taking the limit $N \to \infty$, the left-hand side converges to 0, while the right hand side converges to a strictly positive number if and only if $\alpha > 0$. If $\alpha = 0$, the right-hand side converges to 0.

Finally, centralized monitoring dominates bilateral monitoring with under-insurance if and only if

$$\left(k - (v-1)(1-q)^N\right)\left(x^*_{cm} - x^{opt}\right) + \frac{\alpha^N}{1-\alpha^N}\psi \leq \left[\frac{v-1}{2}(2-q\beta) - k\right]\left(x^{opt} - x^{*,ui}\right)$$

Rewriting equation (14), we obtain

$$x^{opt} - x^{*,ui} = \frac{2\psi(1-q)}{vq(1-\alpha)(2-q\beta)} - \frac{\beta q(2-\hat{c})(1-q)^N}{[2-q\beta]\left[2(1-(1-q)^N) - \beta q\right]}$$

Hence, we can rewrite the condition as follows

$$\frac{\frac{v-1}{2}(2-q\beta) - k}{2-q\beta}\left[\frac{2\psi(1-q)}{vq(1-\alpha)} - \frac{\beta q(2-\hat{c})(1-q)^N}{2(1-(1-q)^N) - \beta q}\right] \geq \frac{k-(v-1)(1-q)^N}{2(1-(1-q)^N) - \beta q}\frac{2\psi}{1-\alpha^N} + \frac{\alpha^N}{1-\alpha^N}\psi$$

which we refer to as $F^{UI} \geq 0$. Taking the limit $N \to \infty$, we obtain

$$\frac{\frac{v-1}{2}(2-q\beta) - k}{2-q\beta}\frac{2\psi(1-q)}{vq(1-\alpha)} \geq \frac{k}{2-\beta q}2\psi$$

This condition holds if and only if $k \leq k_{om}$ with

$$k_{om} \equiv \frac{1-q}{1-q+vq(1-\alpha)}\bar{k} < \bar{k}$$

Finally, we are left to derive the maximum value of the monitoring cost ψ such that the interval $[k'_m, k_{om}]$ is non-empty. Observe that k_{om} is independent of ψ while k'_m is strictly increasing with ψ. Solving for $k'_m(\psi) = k_{om}$, we get

$$\begin{aligned}
0 &= \frac{1-q}{1-q+vq(1-\alpha)}\frac{v-1}{2}(2-q\beta) - \frac{2-\beta q}{\frac{\beta q(1-\alpha)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right)-\psi}\frac{\psi}{2} \\
0 &= (1-q)(v-1)\frac{\beta q(1-\alpha)}{2-\beta\alpha q} - (1-q)(v-1)\psi - \psi\big[1-q+vq(1-\alpha)\big] \\
\psi &= \frac{\beta q(1-q)(1-\alpha)(v-1)}{v(2-\beta\alpha q)(1-\alpha q)}\left(1-\frac{\hat{c}}{2}\right)
\end{aligned}$$

This is the expression for the upper bound on ψ in Assumption 2.

References

Acharya, V., and A. Bisin, 2014, "Counterparty risk externality: Centralized versus over-the-counter markets," Journal of Economic Theory, 149, 153–182.

Antinolfi, G., F. Carapella, and F. Carli, 2018, "Transparency and collateral: central versus bilateral clearing," FEDS Working Paper 2018-017.

Arnsdorf, M., 2012, "Central Counterparty Risk," Journal of Risk Management in Financial Institutions, 5, 273–287.

BCBS, 2019, "Internal models method for counterparty credit risk," .

Biais, B., F. Heider, and M. Hoerova, 2012, "Clearing, counterparty risk, and aggregate risk," IMF Economic Review, 60, 193–222.

Biais, B., F. Heider, and M. Hoerova, 2016, "Risk-sharing or risk-taking? Counterparty risk, incentives, and margins," The Journal of Finance, 71, 1669–1698.

Board, T. F. S. E., 2010, "Financial Stability Report," Discussion paper, Bank of England.

Coeuré, B., 2015, "Ensuring an adequate loss-absorbing capacity of central counterparties," Special invited lecture at Federal Reserve Bank of Chicago 2015 Symposium on Central Clearing, Chicago.

CPSS, 2010, "Market structure developments in the clearing industry: implications for financial stability," Discussion paper, BIS.

CPSS-IOSCO, 2012, "Principles for financial market infrastructure," Discussion paper, BIS.

Diamond, D. W., 1984, "Financial Intermediation and Delegated Monitoring," The Review of Economic Studies, 51, 393–414.

Domanski, D., L. Gambacorta, and C. Picillo, 2015, "Central clearing: trends and current issues," Discussion paper, BIS.

Duffie, D., 2015, "Resolution of Failing Central Counterparties," in Kenneth E. Scott, Thomas H. Jackson, and John B. Taylor (ed.), Making Failure Feasible . chap. 4, Hoover Institution, Stanford University.

Duffie, D., and H. Zhu, 2011, "Does a Central Clearing Counterparty Reduce Counterparty Risk?," The Review of Asset Pricing Studies, 1, 74–95.

ESMA, 2020, "CCPs' Membership Criteria and Due Diligence," .

FSB, 2018, "Incentives to centrally clear over-the-counter (OTC) derivatives," Discussion paper, BIS.

Ghamami, S., and P. Glasserman, 2017, "Does OTC derivatives reform incentivize central clearing?," Journal of Financial Intermediation, 32, 76–87.

Hart, O., and J. Moore, 1996, "The governance of exchanges: members' cooperatives versus outside ownership," Oxford Review of Economic Policy, 12, 53–69.

Holmström, B., 1979, "Moral hazard and observability," The Bell journal of economics, pp. 74–91.

Holmström, B., and J. Tirole, 1997, "Financial intermediation, loanable funds, and the real sector," Quarterly Journal of economics, 112, 663–691.

Huang, W., 2019, "Central counterparty capitalization and misaligned incentives," BIS Working Paper 767.

Koeppl, T., 2013, "The Limits Of Central Counterparty Clearing: Collusive Moral Hazard And Market Liquidity," Discussion paper, Economics Department, Queen's University.

Koeppl, T., C. Monnet, and T. Temzelides, 2012, "Optimal clearing arrangements for financial trades," Journal of Financial Economics, 103, 189 – 203.

Leitner, Y., 2011, "Inducing Agents to Report Hidden Trades: A Theory of an Intermediary," Review of Finance, 16, 1013–1042.

Mancini, L., A. Ranaldo, and J. Wrampelmeyer, 2015, "The Euro Interbank Repo Market," The Review of Financial Studies, 29, 1747–1779.

McPartland, J., and R. Lewis, 2017, "The Goldilocks problem: how to get incentives and default waterfalls "just right"," Federal Reserve Bank of Chicago, Economic Perspectives, 41.

Menkveld, A. J., and G. Vuillemey, forthcoming, "The Economics of Central Clearing," Annual Review of Financial Economics.

Palazzo, F., 2016, "Peer monitoring via loss mutualization," Bank of Italy Temi di Discussione (Working Paper) No, 1088.

Singh, M., 2010, "Collateral, Netting and Systemic Risk in the OTC Derivatives Market," IMF Working Papers 10/99, International Monetary Fund.

Tirole, J., 2010, The theory of corporate finance, Princeton University Press.

Wang, J. J., A. Capponi, and H. Zhang, 2019, “Central Counterparty and Collateral Requirements,” working paper.

Yellen, J., 2013, “Interconnectedness and Systemic Risk: Lessons from the Financial Crisis and Policy Implications,” At the American Economic Association/American Finance Association Joint Luncheon, San Diego, California.

Zawadowski, A., 2013, “Entangled financial systems,” The Review of Financial Studies, 26, 1291–1323.