The Design of a Central Counterparty*

This version: October, 2021

John Kuong[†] Vincent Maurin[‡]

Abstract

This paper analyzes the optimal allocation of losses via a Central Clearing Counterparty (CCP) in the presence of counterparty risk. A CCP can hedge this risk by providing loss-absorbing capital or by enabling loss mutualization among its members. This protection, however, weakens members' incentives for risk management. When the market is large, a third-party CCP alleviates this tension by acting as a centralized monitor. We endogenize the typical default waterfall of a CCP including the CCP's junior equity tranche. Privately optimal choices of skin-in-the-game capital can be socially inefficient. Our results have implications for the ownership structure of CCPs.

Keywords: Central Clearing, Collateral, Monitoring, Default Waterfall, CCP Regulation

JEL Codes: D86, G23, G28.

* We would like to thank Jean-Edouard Colliard, Chester Spatt, Ernst-Ludwig von Thadden; our discussants Thomas Gherig, David Murphy, Roberto Ricco, Anatoli Segura, Guillaume Vuillemey, Haoxiang Zhu; and seminar and conference participants at CMU Tepper, ICEF HSE, INSEAD, Stockholm School of Economics, UC3M, HKIMR, Vienna, the Asia-Pacific Corporate Finance Online Workshop, the Gerzensee Workshop on Money, Payments, Banking and Finance, the Kelley Junior Finance Conference, the Microstructure Exchange, the NFN Young Scholars Webinar Series, OxFIT, the SEC CFMR Conference, the SGF Conference and Stern Salomon Microstructure Conference for useful comments.

[†]INSEAD and CEPR. E-mail: john.kuong@insead.edu.

[‡]Stockholm School of Economics. E-mail: vincent.maurin@hhs.se.

1 Introduction

The 2007–2008 global financial crisis put the spotlight on counterparty risks in over-the-counter (OTC) markets. To tame these risks, regulators around the world have mandated clearing of many OTC contracts via Central Counterparties (CCPs).[1] For instance, the fraction of centrally cleared interest rate derivatives rose to 60% in 2018 from 15% in 2009 (FSB, 2018).[2] CCPs can manage counterparty risks thanks to collateral requirements, by monitoring clearing members' financial soundness, and via loss mutualization. By standing between every transacting parties (the members), a CCP reduces the impact of any member's default by shouldering the losses or reallocating them to other members.

Regulators view the design of this loss allocation process in CCPs, also known as the *default waterfall*, as critical to financial stability (Yellen, 2013; FSB, 2020). Yet, practitioners disagree about some of its key features. Large institutional investors who are clearing members often request more "meaningful" capital contribution from CCPs to cover losses (ABN-AMRO, 2020). CCPs, meanwhile, resist these calls arguing their business is to pool risks, not to insure them. In this view, members' financial resources, typically pre-funded as collateral, should absorb the bulk of the losses, and capital serves primarily to align CCPs' incentives in risk management (LCH, 2015).

In this paper, we propose a framework to analyze the optimal design of CCPs where central clearing is endogenized. In our model, investors match in pairs to trade. The bilateral exposure to idiosyncratic counterparty risk generates benefits from sharing losses. We

[1] In the U.S., 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. See Spatt (2017) for an in-depth discussion on the regulatory changes in swaps and derivative markets in the U.S.

[2] Another example is the Euro interbank repurchase agreements (repos) market where central clearing has become the norm. Mancini, Ranaldo, and Wrampelmeyer (2015) show that from 2009 to 2013, the share of CCP-based repos increased from 42% to 71%, whereas bilateral repos declined from 50% to 19%.

represent the loss allocation mechanism in central clearing as a multilateral contract signed by investors (or "members") and the CCP. Sharing losses, however, increases the expected liability of investors who can only credibly promise to pay up to a fraction of their future income. Investors can mitigate this limited pledgeability problem via counterparty monitoring and by pledging costly cash collateral. In this context, a CCP can add value by enabling loss mutualization, providing insurance with its capital, or monitoring investors. When the contract has the CCP perform at least one of these roles, we say it requires central clearing.

With these basic ingredients, we achieve four main results. First, central clearing dominates bilateral trading when the cost of collateral is intermediate and market size is large. Second, under similar conditions, it is efficient to delegate and centralize all monitoring tasks to a third-party CCP. Third, such a CCP holds a junior equity tranche in the default waterfall to align its incentives, and contributes capital as "skin-in-the-game" (SITG) at members' request. Fourth, the equilibrium level of CCP capital, when chosen either by the CCP or by the members, can be socially inefficient. Our results have implications for the design of the default waterfall, the ownership structure of a CCP, and CCP regulations.

Our results arise due to two fundamental frictions. The first one is the aforementioned limited pledgeability. As in Biais, Heider, and Hoerova (2016), it stems from a moral hazard problem: Investors would shirk for private benefits and default if their expected liability is too large. The shirking metaphor is meant to capture investors' concerns in practice that their counterparties could under-invest in proper risk management or take actions that expose them to "wrong-way risk".[3] Investors can expand their capacity to share risks by liquidating their asset for cash collateral, which is fully pledgeable but has lower returns.

Asset pledgeability can also be improved by counterparty monitoring, but monitoring

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

requires a costly and unobservable effort. This is the second friction: Monitoring needs to be incentivized. The monitoring effort corresponds to investors' and CCPs' due diligence processes in practice to ascertain the financial soundness of their counterparties and members.[4]

To clearly show how these frictions affect the design of central clearing, we proceed in three steps. We first analyze the frictionless benchmark in which investors' asset is fully pledgeable (and thus monitoring is redundant). We will then add the limited pledgeability friction (with observable monitoring), and finally the friction of unobservable monitoring.

In the frictionless benchmark, investors achieve insurance against counterparty default with collateral, CCP capital, or loss mutualization via the CCP. Loss mutualization is limited by the resources of non-defaulting members. This implies that mutualizing losses can never provide full hedging and is less efficient in small CCPs. In contrast, with enough CCP capital or collateral, investors can fully hedge counterparty default risk. Hence, when the cost of CCP capital or collateral is low, investors hedge with the cheaper of the two. Otherwise, they rely on loss mutualization. The key insight is that without friction a CCP *substitutes for* collateral when collateral is costly, by pledging capital or by enabling loss mutualization.

Collateral becomes instead a necessary input for central clearing when the limited pledgeability friction is introduced. Central clearing requires additional payments from investors, either to compensate the CCP for providing capital, or to cover other members' losses in mutualization. Investors' payment capacity is constrained due to limited pledgeability and can be expanded by pledging collateral. This friction also implies that bilateral monitoring is optimal when monitoring is observable. The CCP has the same technology as investors but compensating its effort in centralized monitoring requires investors' collateral.

[4]For example, ESMA (2020) shows that CCPs monitor members with internal credit rating criteria and examine their books regularly. The rigor and incentive structure behind such processes are first-order issues to regulators and CCPs (see e.g., Coeuré 2015 and LCH 2015).

Our first main result is that central clearing strictly dominates bilateral trading only when the cost of collateral is intermediate. In this case, investors mutualize losses or, if capital is cheap enough, use CCP capital as insurance. The intuition is as follows. If the cost of collateral is low enough, full hedging with collateral is desirable, leaving no losses for CCP capital to absorb or to be mutualized. If instead the cost is high, using collateral to support any insurance either from CCP capital or from loss mutualization is too expensive.

When the second friction of unobservable monitoring is added, loss mutualization gives rise to the classic "insurance vs. incentive" conflict of Holmström (1979). When more losses from counterparty default are shared, an investor benefits less from monitoring her counterparty. Hence, to restore incentives for bilateral monitoring, loss mutualization must be reduced. That is, investors retain more exposure to counterparty risk.

An alternative scheme to overcome the "insurance vs. incentive" tension is to delegate and centralize all the monitoring efforts to the CCP. We interpret the CCP in this case as a third-party, for-profit agent who performs the monitoring tasks for compensation stipulated in the contract. Yet, centralized monitoring is costly for two reasons. First, we recall that compensating the CCP requires collateral. Second, the CCP enjoys an agency rent, receiving compensation over and above the effort cost, because monitoring efforts are unobservable.

Our second main result is that centralized monitoring dominates bilateral monitoring in large markets. As in Diamond (1984), these economies of scale arise endogenously because the agency rent for monitoring decreases with the number of members monitored. This result points to a new force shaping the optimal CCP ownership structure. Under centralized monitoring, the CCP is a third-party agent. Under bilateral monitoring, however, the CCP merely channels transfers among members – an arrangement we interpret as a member-owned CCP. Hence, a large (small) market favors third-party (member-owned) CCPs.

The analysis of the optimal contract under centralized monitoring delivers our third main result, which characterizes the compensation and capital contribution of a third-party CCP. It is optimal to only pay the CCP when no member defaults because such high-powered compensation minimizes the agency rent.[5] The CCP thus holds a junior equity tranche in the default waterfall, absorbing losses right after defaulters' collateral. Furthermore, members recoup the rent by requiring the CCP to contribute capital. The capital is akin to skin-in-the game (SITG) in the sense that the CCP will lose it if any member defaults.

Our last main result follows from the comparison of market participants' choice of CCP capital in third-party CCPs with the socially optimal amount. Members and the CCP haggle over the size of SITG capital because more capital reduces collateral requirements for the former but eats away the profit of the latter. This observation can explain the tension between members (ABN-AMRO, 2020) and CCPs (e.g. LCH 2015) about the desirable size of SITG capital. We show that the choice made either by CCPs or members can be socially inefficient as the planner does not care about this surplus distribution. Our analysis hence provides a rationale for regulations of CCP capital and members' collateral.

Our results rationalize several key features of the default waterfall of CCPs as observed in practice. Defaulters pay first as the CCP seizes their collateral. Collateral in our model represents both Initial Margins and pre-funded Default Fund Contributions. The remaining losses are next absorbed by the CCP's SITG capital and junior tranche (in third-party CCPs) and then by surviving members. We emphasize that both STIG capital and junior tranche mainly play the role of incentives while most losses are to be borne by members. These results echo the observations by regulators and CCPs themselves that CCPs should be primarily "risk poolers, not insurers" (Coeuré, 2015; LCH, 2015).

[5]The result that the optimal incentive contract pays only when no investor defaults and the associated agency rent decreases in market size is standard in contracting (see Tirole 2010).

Finally, our analysis also delivers empirical predictions for collateral demand in central clearing. First, collateral requirements in central clearing can be higher (lower) than those in bilateral trading when collateral is expensive (cheap). Second, an increase in bargaining power of a CCP vis-à-vis its members would decrease CCP SITG capital and increase collateral requirements. Hence, we predict that more competition among CCPs would reduce collateral requirements for central clearing.

Literature Review

The premise of our analysis is the ability of CCPs to manage counterparty risks in OTC markets, as in Koeppl and Monnet (2010) and Biais, Heider, and Hoerova (2016).[6] We analyze the tension between the mutualization of losses and the incentives to identify creditworthy counterparties, a version of the classic insurance vs. incentive trade-off (Stiglitz, 1974; Holmström, 1979).[7] In the context of central clearing, this trade-off is studied in related models by Biais, Heider, and Hoerova (2012) and Antinolfi, Carapella, and Carli (forthcoming). Our analysis of member-owned CCPs thus broadly shares some of their conclusions.[8] Our key innovation is the possibility to delegate monitoring efforts to a third-party CCP. This feature allows us to endogenize the optimal ownership structure of CCPs, the default waterfall of third-party CCPs (including SITG capital) and the CCP's compensation. To the best of our knowledge, endogenizing these various aspects of CCP designs from first principles is new.

Some recent works analyze different elements of the default waterfall of a CCP. Wang,

[6] Vuillemey (2020) provides an empirical analysis of counterparty risk hedging in a 19$^{\text{th}}$ century CCP.

[7] Koeppl (2013) and Palazzo (2016) analyze other incentive problems associated with central clearing.

[8] There are however noteworthy differences. While Biais, Heider, and Hoerova (2012) do not consider collateral, we show it is an important determinant of central clearing benefits. From a methodological point of view, both models assume a continuum of traders for tractability, while we can perform comparative statics with respect to the number of clearing members and derive implications for clearing benefits. Empirically, the number of members varies greatly across CCPs (see Domanski, Gambacorta, and Picillo 2015).

Capponi, and Zhang (forthcoming) also stress the need to align members' risk-management incentives and show that pre-funded contributions to the default fund are superior to initial margins if covering losses ex-post is costly. As we do not make this assumption, such pecking order between types of collateral is absent in our analysis. Instead, we endogenize another key element of the waterfall, CCP SITG capital, as part of a solution to the counterparty monitoring problem. Huang (2019) argues that for a given loss allocation, for-profit CCPs under-supply loss-absorbing capital to shift liabilities to surviving members. We show instead that even when CCPs are optimally given a junior tranche to align their incentives with members', the equilibrium capital contribution can still be socially inefficient. In particular, members can demand too much capital for rent extraction. We thus provide a different rationale for CCP capital and collateral regulations. In Huang and Zhu (2021) loss mutualization is analyzed as an auction for the defaulting members' positions run by the CCP. With our optimal contracting approach, all transfers via and to the CCP are specified ex-ante.

The ownership structure is considered critical in the CCP design discussion (Board, 2010; McPartland and Lewis, 2017). It has been argued that for-profit CCPs may allow too much risk-taking (Huang, 2019) while member-owned utilities in general may deter entry (Hart and Moore, 1996). We instead emphasize the costs and benefits of delegating monitoring to the CCP and predict that third-party CCPs dominate member-owned CCPs in large or opaque markets, thanks to endogenous economies of scale as in Diamond (1984).

Our paper focuses on CCPs' role in mitigating counterparty risks, which is most relevant to the default waterfall design. We thus abstract from other important benefits from central clearing that have been discussed in the literature (see the comprehensive surveys by Pirrong 2011 and Menkveld and Vuillemey 2021). Duffie and Zhu (2011) analyze netting efficiency for central and bilateral clearing. Zawadowski (2013) and Acharya and Bisin (2014) show that central clearing can reduce counterparty risk externalities by increasing transparency.[9] Koeppl, Monnet, and Temzelides (2012) show that a CCP can lower trading costs by deferring settlement and providing credit to clearing members.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 maps our general contracting approach to centrally cleared contracts in practice. In Section 4, we analyze the costs and benefits of central clearing by deriving the optimal contract when monitoring is observable. Section 5 analyzes the full problem when monitoring needs to be incentivized. We compare bilateral monitoring to centralized monitoring and provide a welfare analysis of the optimal contract. We gather practical implications of our model for CCP design in Section 6. Section 7 concludes. All proofs are in the Appendix.

2 A Model of Central Clearing

2.1 The framework

There are two dates $t \in \{0, 1\}$. At date 1, there are two equiprobable aggregate states of the world $S \in \{A, B\}$. We denote $S' = \{A, B\} \backslash S$. The economy is populated by investors and a CCP agent, simply called the CCP. All agents consume one good –"cash".

Investors Investors belong to two groups indexed by $S \in \{A, B\}$, and each group has N homogeneous investors. An S-investor has the following utility function:

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

where c_S is the consumption in state S, $\nu > 1$, and $\hat{c} > 0$. In words, S-investors strictly

[9] See also Leitner (2011) for related arguments about the benefits of having a central intermediary.

prefer consuming in state S until their consumption reaches $\hat{c}$. These preferences imply that investors from different groups gain from trading consumption across states. Per-unit gains from trade are equal to the difference in marginal utility $\nu - 1$. To fix ideas, we say these preferences reflect hedging needs against an aggregate state, with $\hat{c}$ the hedging demand.[10]

Each S-investor is endowed with one unit of a non-tradable asset which pays $2R$ per unit with an exogenous probability $q \in (0, 1)$ in state S' and fails to pay anything otherwise, as shown in Figure 1. The success or failure of the asset is independent across S-investors, conditional on the realization of state S'. Since S-investors have assets that pay in the state in which S'-investors value consumption more, some gains from trade can be realized. Due to assets' idiosyncratic payoff risk, however, an investor who is supposed to pay will fail to do so with probability $1 - q$.

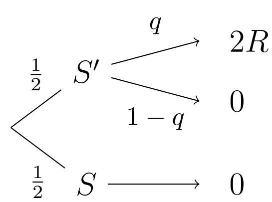

Figure 1. Payoff from an S-investor's asset

Trading is limited by the fact that the asset's cash flow is not fully pledgeable. An investor with asset pledgeability $\tilde{\beta} \in \{0, \beta\}$ can credibly promise to pay no more than $\tilde{\beta}$ in expectation out of the cash flows of the asset. If faced with a larger liability, the investor shirks at date 0 which destroys the asset cash flow, like an asset failure.[11] The limited

[10]An alternative interpretation of these preferences is that investors have different beliefs across groups. Then, investors would trade to bet about the realization of the state of the world they think is more likely. To identify robust principles for clearing, we do not specify a particular hedging/betting instrument. Our model can accommodate one-sided hedging needs as in the Credit Default Swaps (CDS) market.

[11]Limited pledgeability is often motivated with moral hazard, as in Holmström and Tirole (1997). Suppose

pledgeability friction captures investors' concerns for counterparties taking excessive risks or shirking proper risk management effort when their liability becomes large (see footnote 3).

The limited pledgeability problem can be mitigated with monitoring. If monitored, an investor's asset pledgeability is $\beta > 0$. If unmonitored, her asset pledgeability is β with probability $1 - \alpha$ only and 0 otherwise. Monitoring is performed by another investor or the CCP. It costs $\psi > 0$ per investor and the monitoring effort is unobservable to third parties. Monitoring can be seen as a way to ensure an investor's position does not exceed her financial capacity and is considered by CCPs as important defense against counterparty risks (see also footnote 4). It is also relevant in OTC markets where a counterparty's overall risk exposure may be difficult to assess due to the lack of transparency.

CCP The CCP agent is risk-neutral and competitive. It has a large initial endowment E of cash at date 0 and no asset. Its utility function is given by

$$U_C = \nu_C c_0 + c_1 \tag{2}$$

with $\nu_C > 1$. The parameter ν_C is the gross interest rate required by the CCP to substitute date-1 consumption for date-0 consumption. We thus refer to $\nu_C - 1$ as the cost of CCP capital. The CCP can use (cash) capital to satisfy investors' hedging needs. It can also monitor investors but its monitoring effort is as costly as the investors' and it is also unobservable.

Collateral At date 0, any fraction of an investor's asset can be liquidated one to one for cash. Asset payoff risk and limited pledgeability give a role for cash to be used as collateral as cash is safe and fully pledgeable. First, by holding cash, an investor can use it to consume in her favorite state, thereby reducing her hedging needs. Second, when trading with investors

the investor can shirk for a private benefit $\tilde{B}$ per unit of asset held. Then, an investor with private benefit $\tilde{B} = q\left(R - \frac{\tilde{\beta}}{2}\right)$ can credibly promise to repay no more than $\tilde{\beta}$.

from the other group, cash collateral can protect against counterparty default. Third, as we will show, cash collateral expands investors' aggregate risk-sharing capacity, due to the limited pledgeability friction. Using collateral, however, is costly as we assume the expected payoff of the asset qR is higher than 1. In what follows, we call $k \equiv qR - 1$ the cost of collateral. This cost captures the foregone return on high-return assets compared to assets widely accepted as collateral such as cash or government bonds.[12]

2.2 Contracting

In practice, investors sign a bilateral contract which is then novated to and cleared by a CCP. A cleared contract implicitly specifies contingent transfers among investors and the CCP. In the model, we directly consider a general multilateral contract between investors and the CCP. We discuss the mapping to a cleared contract in practice in Section 3.1.

At date 0, each S-investor matches with an S'-investor, called her counterparty. Although matching is bilateral, contracting is multilateral. A contract specifies transfers, and if necessary, a monitoring scheme: bilateral (counterparty) monitoring or centralized (CCP) monitoring. To streamline the exposition, we only analyze the contract with monitoring in the main text. The optimal contract without monitoring is derived in the proof of Proposition 4 when we characterize conditions for monitoring to be optimal.

With monitoring, all investors have the same asset pledgeability β and thus, a single contract is offered to all investors.[13] The contract specifies state-contingent transfers and an aggregate state in general is defined by $\{S, d\}$, where d is the number of defaulting

[12]In practice, CCPs require members to post a fraction of collateral as cash (Armakolla and Bianchi 2017) and their cash reinvestment policy is limited to safe low-return vehicles (e.g. Article 47 of regulation EMIR).

[13]As we show in Proposition A.1 in the Appendix, even in the case without monitoring, a single (pooling) contract will be offered to investors with heterogeneous asset pledgeability. Separating contracts are not feasible because the single-crossing property fails. In particular, all investors have the same cost of collateral.

investors with $d \in \{0, 1, .., N\}$. As the environment is symmetric, we focus on symmetric contracts. Moreover, S-investors should receive payments and S'-investors should pay only in states $\{S, d\}$ when the former have high marginal utility of consumption. We thus drop the reference to S and label investors by their ex post role: *receiver* or *payer*. We allow an investor's transfer to be contingent on the idiosyncratic outcome $o \in \{s, f\}$ for the payer in the pair, where s stands for success and f for failure (of the asset).[14] The CCP's transfer is indexed by the state d only. The contract is thus defined as follows.

Definition 1. *A contract $\mathcal{C} = \{x, p_o(d), r_o(d), e, \pi(d)\}$ with $o \in \{s, f\}$ and $d \in \{0, ..., N\}$ is a set of non-negative transfers. At date 0, investors post an amount of collateral x and the CCP contributes capital Ne. At date 1, a payer pays $p_o(d)$, a receiver gets $r_o(d)$ and the CCP gets compensation $N\pi(d)$. The contract also specifies a monitoring scheme by the indicator function $\mathbb{1}_{cm}$, which is equal to 1 when the CCP monitors all investors (centralized monitoring) and 0 when each investor monitors her own counterparty (bilateral monitoring).*

Transfers $r_s(N)$, $p_s(N)$ and $r_f(0)$, $p_f(0)$ are set to 0 as they are not well-defined. For instance, there cannot be N defaulting payers if a payer succeeds.

We are now ready to formally define the investors' problem.

[14] It can be argued that in practice, a cleared contract's payments would not be contingent on some (idiosyncratic) outcome of the original counterparty because the counterparty is the CCP after novation. Such restriction on the contract severely undermines investors' incentives to monitor each other, thus strengthening our result that the CCP emerges as a centralized monitor (see Proposition 7).

Investors' Problem.

$$\max_{\{\mathcal{C},\mathbb{1}_{cm}\}} U = qR + \frac{\nu-1}{2}\mathbb{E}[\min\{r_o(d),\hat{c}\}] - xk - (1-\mathbb{1}_{cm})\psi - \frac{1}{2}\left(\mathbb{E}[\pi(d)] - e\right) \tag{3}$$

$$s.\ to \quad \forall\, d, \quad p_s(d) \leq x + (1-x)2R, \tag{4}$$

$$\forall\, d, \quad p_f(d) \leq x \tag{5}$$

$$\forall\, d, \quad (N-d)r_s(d) + dr_f(d) + N\pi(d) = N(x+e) + (N-d)p_s(d) + dp_f(d) \tag{6}$$

$$\mathbb{E}[\pi(d)] \geq \nu_C e + \mathbb{1}_{cm}2\psi \tag{PC$_{CCP}$}$$

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

$$\textit{If } \mathbb{1}_{cm} = 1, \qquad 2\psi \quad \leq \mathbb{E}[\pi(d)|m=1] - \mathbb{E}[\pi(d)|m=0]; \tag{MIC$_{cm}$}$$

$$\textit{If } \mathbb{1}_{cm} = 0, \quad \frac{\psi}{q(1-\alpha)} \leq \frac{1}{2}\Big(\mathbb{E}_s[r_o(d)] - \mathbb{E}_f[r_o(d)]\Big) + \frac{\nu-1}{2}\Big(\mathbb{E}_s[\min\{r_o(d),\hat{c}\}] - \mathbb{E}_f[\min\{r_o(d),\hat{c}\}]\Big) \tag{MIC$_{bm}$}$$

where the expectation operator $\mathbb{E}[\cdot]$ *is taken over the state* d *and* $\mathbb{E}_{o'}[\cdot]$ *is the expectation conditional on an outcome* $o' \in \{s,f\}$ *for the payer (a receiver's counterparty).*

Investors' problem is to maximize their expected utility (3), subject to resource constraints (4)-(6), CCP's participation constraint (PC_{CCP}), investors' limited pledgeability constraint (LP), monitoring incentive constraint under centralized monitoring (MIC_{cm}) or bilateral monitoring (MIC_{bm}). We discuss each elements of the Investors' Problem below.

The objective function, given by equation (3) represents an investor's expected utility. We present the steps to obtain (3) from (1) in the Appendix and provide the intuition here. The first term is the investor's utility under autarky when she uses no collateral, and the remaining terms capture the net benefits of the contract: the expected gain from transferring consumption to investors' preferred state less the collateral cost, the cost of monitoring in a bilateral scheme $\mathbb{1}_{cm} = 0$ and the CCP compensation net of its capital contribution.

A feasible contract satisfies individual resource constraints (4) and (5) for payers and aggregate resource constraint(s) (6). The latter say that in any state, the sum of receivers' transfers and the CCP compensation must equal total resources available: those committed at date 0 by receivers (collateral) and the CCP (capital), and payments by payers at date 1.

The CCP's participation constraint is formalized by equation (PC_{CCP}). The CCP participates in the contract if its expected payoff exceeds its cost per investor pair, which includes the cost of the capital contribution and the monitoring cost when it monitors.

The first key constraint is investors' Limited Pledgeability constraint (LP). The pledgeability problem implies that the additional expected liability upon success relative to that upon failure cannot exceed an investor's pledgeable income from the $1-x$ units of asset.[15]

The second key constraints are the Monitoring Incentive Constraints (MIC_{cm}) or (MIC_{bm}), imposed because monitoring efforts are unobservable. Under the centralized monitoring scheme, the CCP monitors all investors. Equation (MIC_{cm}) ensures that the CCP prefers monitoring everyone to no one. We verify later this is the relevant deviation even if the CCP could also deviate by monitoring a subset of the $2N$ investors. Under the bilateral monitoring scheme, the constraint is given instead by (MIC_{bm}). It says that the utility loss for an investor from the default of her counterparty must be greater than the monitoring cost ψ weighted by its efficacy in reducing the probability of counterparty default $[q(1-\alpha)]^{-1}$.

2.3 Assumptions

In this section, we describe our main assumptions and explain how they affect the analysis.

[15] If a payer's expected liability increases with the number of defaulting payers d, a coordination problem arises as an investor's decision to shirk depends on her expectations about other investors' behavior. We abstract from this coordination problem here to focus on the welfare-maximizing outcome, that is, we only impose that equation (LP) holds under the expectation that other investors behave.

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

Assumption 1 ensures that cash collateral is both necessary and sufficient to satisfy investors' hedging needs. Without any collateral ($x = 0$), by constraint (LP), each payer can at most pay β which is less than each receiver's hedging need $\hat{c}$. If instead each investor posts $\frac{\hat{c}}{2} < 1$ units of cash collateral, a receiver's hedging needs can always be met with collateral from herself and her counterparty.

Assumption 2 (Monitoring cost)**.** $\psi \le \bar{\psi} \equiv \min\left\{\frac{(1-q)(\nu-1)}{\nu(2-\beta\alpha q)(1-\alpha q)}, \frac{1}{2}\right\}\beta q(1-\alpha)\left(1-\frac{\hat{c}}{2}\right)$.

The first part of Assumption 2 ensures that there are parameters such that monitoring is optimal and the CCP plays a role. The expression for this upper bound will be derived in Proposition 7. The second part of Assumption 2 plays a technical role.

Assumption 3 (Resources)**.** $N \le \frac{2R}{\hat{c}}$.

Assumption 3 ensures that the hedging demand $N\hat{c}$ of all receivers can be satisfied even if only one payer's asset succeeds, as the asset pays out $2R$ in this case. This implies the resource constraint (4) is slack for all $d \le N - 1$. Assumption 3 simplifies our analysis in that the only aggregate risk receivers must bear is that of all payers' joint default.[16]

3 Cleared contract and frictionless benchmark

We first provide a result to restrict the set of relevant contracts for our analysis. The following proposition allows us to map our general contract to a centrally cleared contract in practice.

[16] We derive the optimal contract when Assumption 3 fails in Internet Appendix B. In this case, risk sharing is further limited because receivers' hedging needs cannot be satisfied when too few payers survive. We show, however, that the key trade-off identified in the main text continues to hold .

3.1 Sufficient contracts as cleared contracts in practice

Proposition 1. *Contracts with the following properties are optimal*

1. *A receiver with a successful payer gets* $r_s(d) = r_s$. *Otherwise,* $r_f(d) = r_f \leq r_s$ *if at least one (other) payer survives* ($d < N$) *and* $r_f(N) = 2x + e \leq r_f$ *if all payers default.*

2. *A defaulting payer's collateral is seized:* $p_f(d) = x$. *A successful payer's transfer is*

$$p_s(d) = \underbrace{r_s - x - e}_{\textit{Bilateral transfer}} + \underbrace{\frac{d}{N-d}(r_f - 2x - e)}_{\textit{Loss Mutualization transfer}} + \underbrace{\frac{N}{N-d}\pi(d)}_{\textit{CCP compensation}} \tag{7}$$

Proposition 1 says that given a collateral amount x and a CCP contract $\{e, \pi(d)\}$, investors' transfers can be parametrized with two scalars r_f and r_s only. The intuition for this result is as follows. As shown by expression (3), receivers are risk-averse and thus wish to minimize the variability of their transfers. Yet, transfers may be state-contingent for two reasons. First, receivers are exposed to the aggregate risk of a joint payer default. In this state of the world, by budget constraint (6), their transfer $r_f(N)$ cannot exceed precommitted resources $2x + e$ as no payer survives. Second, investors may optimally retain some counterparty risk exposure ($r_s > r_f$) to satisfy the bilateral monitoring constraint (MIC_{bm}). For payers now, it is optimal to set $p_f(d) = x$ because a larger payment in default relaxes investors' pledgeability constraint (LP). This makes larger payments sustainable in case of success. This payment $p_s(d)$ is pinned down residually by budget constraint (6).

Proposition 1 offers an interpretation of the general multilateral contract as a cleared OTC contract. A receiver with a successful payer gets r_s which can be viewed as the face value of the contract. When the payer defaults, the resources available directly to the pair are $2x$ units of collateral and the capital e pledged by the CCP per investor pair. These resources pin down the receiver's consumption $r_f(N)$ when all payer defaults. When some (other) payers survive, they can transfer resources to the receiver whose consumption r_f lies above $r_f(N)$. We call *loss mutualization* the feature that an investor's loss when her counterparty defaults is reduced by transfers across investor pairs. This loss mutualization transfer is captured by the second term of a successful payer's transfer in (7). It corresponds to investors' contributions to a default fund in practice. The first and the third terms of (7) are respectively the purely bilateral transfer and the share of the CCP compensation.

3.2 Frictionless benchmark

To see how frictions affect outcomes, it is useful to characterize the frictionless benchmark. We derive below the solution to the Investors' Problem when the asset is fully pledgeable ($\tilde{\beta} = 2R$). We are interested in the role of the collateral cost k and capital cost $\nu_C - 1$. Note that monitoring is redundant when the asset is fully pledgeable. In particular,

Proposition 2 (No Friction)**.** *The solution to the Investors' Problem with $\tilde{\beta} = 2R$ is*

1. *if $\min\{k, \nu_C - 1\} \leq (\nu - 1)(1 - q)^N$, a full-hedging contract with $r_s = r_f = \hat{c}$ and*
   (a) *$(x, e) = \left(\frac{\hat{c}}{2}, 0\right)$ if $k \leq \nu_C - 1$,*
   (b) *$(x, e) = (0, \hat{c})$ if $k > \nu_C - 1$,*
2. *otherwise, a complete loss mutualization contract with $r_s = r_f = \hat{c}$ and $x = e = 0$.*

Any CCP compensation schedule $\{\pi(.)\}$ such that (PC_{CCP}) binds is optimal.

The intuition for the result is as follows. Absent frictions and under Assumption 3, one successful payer can credibly cover the hedging needs of all receivers. It is thus always optimal to set $r_f = r_s = \hat{c}$ to realize all gains from trade when one or more payers survive.

To further hedge the joint-default state, safe resources, either collateral or CCP capital, need to be pre-committed. Full hedging with the cheaper of the two safe resources is optimal when one resource is cheaper than the benefits of hedging the joint-default state, measured by $(\nu-1)(1-q)^N$. When this condition is not met (Case 2), neither CCP capital nor collateral is used and counterparty risk is only dealt with mutualization. We say loss mutualization is complete because a receiver's transfer is not affected by the default of her counterparty as long as at least one other payer survives.

Proposition 2 shows that the contract assigns two roles to the CCP in the frictionless benchmark. When investors desire full hedging and capital is cheaper than collateral, the CCP uses its capital to hedge investors' joint default risk. Alternatively, the CCP provides hedging by enabling loss mutualization among investors when both capital and collateral are too costly. In both cases, the CCP *substitutes for* collateral. In the next section, we show how limited pledgeability changes the relationship between clearing and collateral.

4 Clearing with observable monitoring

Limited pledgeability and the unobservability of monitoring are the two key frictions in our model. To isolate the effect of the former, in this section we assume away the monitoring friction and solve the Investor's Problem without constraints (MIC_{bm}) or (MIC_{cm}).

The limited pledgeability friction gives collateral a new function beyond hedging against the joint-default state. Collateral now helps satisfy receivers' hedging needs *when payers survive*. To see this, let us consider an investor pair. If each investor pledges x units of collateral ex ante, a non-defaulting payer can credibly pay $x+(1-x)\beta$ in expectation, using Proposition 1 to substitute $p_f(d)=x$ in (LP). Also, the receiver can use her own collateral

x for consumption. Together, a non-defaulting payer's payment capacity in excess of her receiver's needs is

$$EPC(x) = x + (1 - x)\beta - (\hat{c} - x) \tag{8}$$

which increases in x because $\beta < 2$ (Assumption 1).

The limited pledgeability friction makes collateral a *necessary input* for central clearing. This new feature arises because collateral increases the excess payment capacity EPC at the investor-pair level, as shown above. Without collateral, $EPC(0) < 0$ because $\beta < \hat{c}$ (Assumption 1). In words, without collateral an investor's payment capacity already falls short of her counterparty's hedging needs. Therefore, collateral is needed to support any additional payment for loss mutualization or to compensate the CCP for providing capital.

We begin the analysis by showing that the limited pledgeability problem also affects the choice of monitoring scheme.

Lemma 1. *If monitoring is observable, the optimal monitoring scheme is bilateral.*

Lemma 1 states that monitoring by investors dominates monitoring by the CCP. The reason is that when investors' pledgeable income is limited, CCP monitoring entails a collateral cost. Intuitively, to compensate the CCP for the monitoring costs, each payer is expected to pay an additional 2ψ at $t = 1$. These additional payments require costly collateral to be pledged when constraint (LP) binds. Since the CCP has no intrinsic technological advantage as a monitor, bilateral monitoring is superior. As we will show in Section 5, this conclusion can be overturned when monitoring is not observable.

In Section 4.1 we solve for the optimal contract under observable monitoring, called the *OM-contract*. We provide conditions such that monitoring is optimal in Section 4.2.

4.1 Optimal contract under observable monitoring (OM-contract)

Proposition 3 (Optimal clearing with observable monitoring)**.** *There exists two thresholds of collateral cost* $\bar{k} = \frac{1}{2}(\nu - 1)(2 - q\beta)$ *and* $\underline{k}_N$ *a continuous function of* $\nu_C - 1$ *such that the contract solving the Investors' Problem without* (MIC$_{bm}$) *and* (MIC$_{cm}$) *is as follows:*

1. *For* $k \leq \underline{k}_N$*, it features full hedging with* $r_s^{OM} = r_f^{OM} = \hat{c}$*, and*

   (a) $(e^{OM}, x^{OM}) = \left(0, \frac{\hat{c}}{2}\right)$ *if* $k < \nu_C - 1$*, with, in this case,* $\underline{k}_N = (\nu - 1)(1 - q)^N$*,*

   (b) $(e^{OM}, x^{OM}) = \left(\frac{q\beta(2-\hat{c})}{2\nu_C - q\beta}, \frac{\hat{c} - e^{OM}}{2}\right)$ *if* $k \geq \nu_C - 1$*.*

   In this case, $\underline{k}_N$ *is strictly decreasing with* $\nu_C - 1$ *and* $\lim_{\nu_C \to 1} \underline{k}_N = \bar{k}$*.*

2. *For* $k \in [\underline{k}_N, \bar{k}]$*, there is complete loss mutualization:* $r_s^{OM} = r_f^{OM} = \hat{c}$*,* $e^{OM} = 0$ *and*

$$x^{OM} \equiv \frac{\left[1 - (1-q)^N\right] \hat{c} - \beta q}{2\left[1 - (1-q)^N\right] - \beta q} \in \left(0, \frac{\hat{c}}{2}\right), \tag{9}$$

3. *For* $k \geq \bar{k}$*, the contract is uncollateralized with* $r_s^{OM} = \beta$*,* $r_f^{OM} = x^{OM} = e^{OM} = 0$*.*

Proposition 3 shows how the limited pledgeability friction changes the economics of a CCP. In the frictionless benchmark (Proposition 2), the CCP's function is to substitute for collateral when collateral is costly enough, either with capital or loss mutualization. Here in contrast, when investors' asset is not fully pledgeable, the CCP can only play a role with the help of collateral. Investors must now pledge collateral to tap into the CCP capital (Case 1b) or to mutualize losses (Case 2). The intuition is that investors must be able to pay to the default fund with loss mutualization or to compensate the CCP for pledging capital with full hedging. Their excess payment capacity can only be expanded by pledging collateral.

When collateral is needed for central clearing, Proposition 3 shows that loss mutualization is no longer optimal if the collateral cost is too high. Above a threshold $\bar{k}$, no collateral is used, receivers do not fully satisfy their hedging needs ($r_s < \hat{c}$), and they are fully exposed to counterparty risk ($r_f = 0$). This threshold $\bar{k}$ captures the total hedging value of collateral starting from the contract in Case 3. A unit of collateral provides 1 unit of self-hedging, $1 - q$ expected units of counterparty risk-insurance and $q(1 - \beta)$ extra units from relaxing the pledgeability constraint (LP), thus increasing the expected incentive-compatible transfer to receivers by $2 - q\beta$. When $k > \bar{k}$, hedging and thus loss mutualization are too costly.[17]

Proposition 3 sheds light on the benefits of having a CCP. We say that a CCP is *essential* if the OM-contract cannot be implemented via a bilateral contract, defined as follows.

Definition 2. *A contract is bilateral if it satisfies* $r_o(d) = p_o(d) + x$ *for all* $d \in \{0, 1, .., N\}$.

Intuitively, with a bilateral contract, an investor pair does not receive transfers from or make payments to other investors or the CCP.[18] Notably, the contracts in Case 1a and Case 3 can be implemented bilaterally. In both cases, CCP capital is too expensive to be used for insurance. In addition, loss mutualization is not used for different reasons. When collateral is cheap (Case 1a), the payer's transfer is fully backed by collateral ($p_o^{OM} = x$) and the receiver is fully hedged ($r_o^{OM} = 2x = \hat{c}$), which leaves no counterparty risk to mutualize. When collateral is expensive (Case 3), loss mutualization, which requires collateral, is too costly. Receivers then optimally remain fully exposed to counterparty risk ($r_f^{OM} = r_f^{OM}(N) = 0$). These observations imply that clearing benefits are hump-shaped in the cost of collateral.

Corollary 1 (Essentiality of CCP)**.** *A CCP is essential, that is, the OM-contract cannot be implemented bilaterally, for* $k \in [\underline{k}^{ess}, \bar{k}]$ *with* $\underline{k}^{ess} = \min\{\underline{k}_N, \nu_C - 1\}$. *The threshold* $\underline{k}^{ess}$ *is weakly increasing in* N.

[17] A similar logic explains why the full-hedging region with CCP capital (Case 1b) shrinks relative to is counterpart in Proposition 2. Full hedging with CCP capital now requires collateral, so the collateral cost becomes a limiting factor: the condition $k \leq \underline{k}_N$ is thus needed in addition to $\nu_C - 1 \leq (\nu - 1)(1 - q)^N$.

[18] By the resource constraint (6) and (PC_{CCP}), a bilateral contract implies that the CCP posts no capital ($e = 0$) and gets no compensation ($\pi(d) = 0$ for all d).

Corollary 1 implies that in the intermediate region of collateral cost, clearing with a CCP strictly dominates bilateral trading as the contract cannot be implemented bilaterally. In addition, this region expands when the market size becomes larger.[19] This is because when there are more investors to share idiosyncratic default risks, the joint-default state becomes less likely and thus full hedging is less desirable relative to loss mutualization.[20]

As central clearing also changes collateral requirements, we compare the demand of collateral in the multilateral contract of Proposition 3 to that in the optimal bilateral contract, which satisfies Definition 2.

Corollary 2 (Bilateral Contract vs. CCP)**.** *When a CCP is essential for some $N \geq 2$, the bilateral contract requires strictly more (less) collateral when k is low (k is high).*

Corollary 2 shows that mandating central clearing of OTC contracts has an ambiguous effect on the demand for collateral. The intuition for the high-k part of Corollary 2 is that a complete loss mutualization contract requires more collateral the more investors can share counterparty risk. When k is close to the upper bound $\bar{k}$, complete loss mutualization is optimal for any N by Proposition 3. The collateral amount increases with N, as shown by (9), to sustain larger loss mutualization transfers. In a bilateral contract, less collateral is needed because it only protects an investor against the default of her counterparty.

When the collateral cost is close to the lower bound $\underline{k}^{ess}$ of the essential CCP region, however, a bilateral contract requires more collateral. This result arises for two different

[19]For this result, it is assumed that investors are monitored. The proof of Corollary 1 also characterizes the lower bound of the essential CCP region when we account for the optimal monitoring decision (see Section 4.2). This lower bound is higher than $\underline{k}^{ess}$, but the comparative statics with respect to N remains valid.

[20]This result does not account for the netting gains – a potentially important benefit of central clearing. When investors have off-setting positions with others for a given contract, multilateral netting via clearing reduces collateral needs. As Duffie and Zhu (2011) argue, however, netting benefits for a given contract in clearing have to be compared with bilateral netting benefits across different contracts. This trade-off is absent in our model as there is a single contract and each investor has one position.

reasons. First, suppose investors require full hedging both with a bilateral and a multilateral contract. Trading via a CCP lowers the need for collateral if CCP capital is cheaper as a hedging tool. The second reason is a consequence of Corollary 1. As N increases, the lower bound $\underline{k}^{ess}$ of the essential CCP region decreases due to larger gains from loss mutualization. Hence, when k is low, investors require full hedging if they can only trade bilaterally while with a CCP, they (only) mutualize losses, which requires less collateral. Intuitively, investors use more collateral when it is the only tool against counterparty risk, as in a bilateral contract.

To summarize, central clearing can reduce the need for collateral to protect against counterparty risk because CCPs provide alternative tools for insurance: their own capital or loss mutualization. The very mutualization of losses, however, requires collateral because CCPs need to make sure investors will deliver when asked 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 2021) with claims that central clearing increases the need for collateral (see e.g. Domanski, Gambacorta, and Picillo 2015).

4.2 Optimal Monitoring

So far, we assumed investors should be monitored. To conclude this section, we provide conditions for monitoring to be optimal. It is clear that monitoring can be suboptimal. Consider for instance the case in which investors fully hedge with collateral (Case 1a of Proposition 3). Monitoring is wasteful because any investor, monitored or not, can post collateral. More generally, monitoring plays the same role as collateral in enhancing the investors' total pledgeable income. In other words, monitoring is a substitute of collateral and it is optimal when the collateral cost is high enough (relative to the cost of monitoring).

Proposition 4. *Monitoring is optimal (when observable) if and only if $k \geq \hat{k}^m$ with $\hat{k}^m$ an increasing function of ψ. The threshold satisfies $\hat{k}^m \in [\underline{k}^{ess}, \bar{k}]$.*

The lower bound of $\hat{k}^m$ confirms the intuition that monitoring is suboptimal when the contract is fully hedged with collateral, that is, when $k < \underline{k}^{ess}$. The upper bound on the monitoring cost in Assumption 2 ensures that $\lim_{\psi \to \bar{\psi}} \hat{k}^m < \bar{k}$, that is, there always exists a region of collateral costs in which monitoring is optimal and the CCP is essential. In the next section, we will restrict our analysis to this parameter region $k \in [\hat{k}^m, \bar{k}]$ to show how the incentive problem in monitoring affects the contract design and the role of the CCP.

We illustrate results from this section in Figure 2. The figure shows the parameter regions which map into the different contracts of Proposition 3. The threshold $\underline{k}_N$ is the frontier between the yellow region on the one hand and the blue and red regions on the other hand. The dashed line represents the monitoring threshold $\hat{k}^m$ of Proposition 4. For a higher value of N, the full-hedging regions in Figure 2 would shrink together with $(\nu - 1)(1 - q)^N$.

5 Clearing with Monitoring Incentives

In this section, we add back the friction of unobservable monitoring and analyze the Investors' Problem in full. The main new insight is that clearing conflicts with investors' incentives to monitor their counterparty and, consequently, the CCP can emerge as the efficient monitor. The analysis also sheds lights on the role of CCP capital as skin-in-the-game and provides new implications about the design of the CCP loss allocation process.

Monitoring incentives matter for the investors' problem only if the *OM-contract* of Proposition 3 is not incentive-compatible with bilateral monitoring. The following lemma describes the parameter region for such a case.

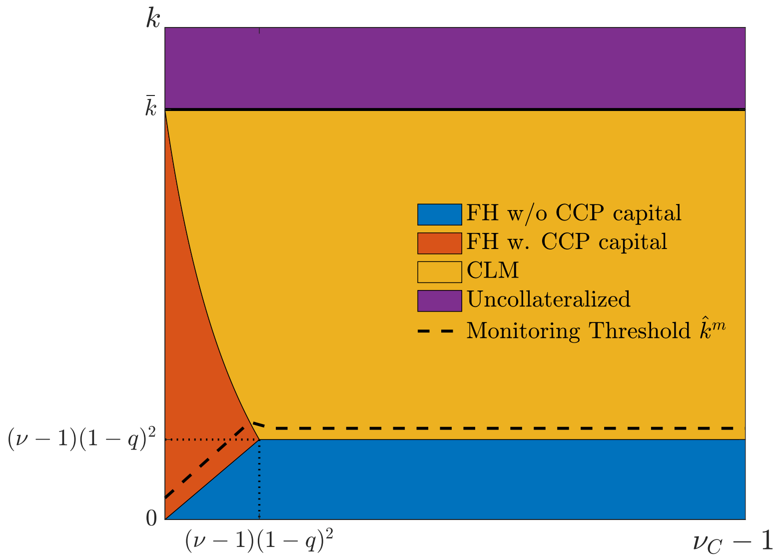

Figure 2. Optimal *OM-contract* with $N = 2$. *The x-axis is the cost of CCP capital k and the y-axis is the cost of collateral $\nu_C - 1$. Parameter values: $\hat{c} = 0.8$, $\nu = 2$, $\beta = 0.6$, $\alpha = 0.4$, $\psi = 3.7 \times 10^{-3}$. FH=Full Hedging. CLM=Complete Loss Mutualization.*

Lemma 2. *The OM-contract violates (MIC$_{bm}$) for $k \leq \underline{k}_N$, and for $k \in (\underline{k}_N, \bar{k})$ when $N > N^*$, where N^* is the largest value of N such that*

$$\frac{\psi}{q(1-\alpha)} \leq \nu(1-q)^{N-1}\left(\frac{\hat{c}}{2} - x^{OM}\right), \qquad \textit{with } x^{OM} \textit{ given by } (9) \tag{10}$$

The intuition for Lemma 2 is as follows. When $k > \bar{k}$, the *OM-contract* is bilateral and uncollateralized. Investors are exposed to sufficient counterparty risk to induce monitoring, provided that monitoring is not too costly (which is guaranteed by Assumption 2). The case $k \leq \underline{k}_N$ is the opposite: The *OM-contract* features full hedging. As investors receive the same transfer irrespective of the payer's outcome, monitoring is privately suboptimal. Finally, the case $k \in (\underline{k}_N, \bar{k})$ with loss mutualization is intermediate as investors retain some exposure to counterparty risk. This exposure and thus investors' incentives to monitor are captured by the right-hand-side of (10), which decreases with N for two reasons. First, the loss given counterparty default is more likely to be mutualized because the state when all payers default becomes less likely. Second, as loss mutualization improves when N increases, the amount of collateral x^{OM} also increases. This reduces the "loss given joint default" $\hat{c} - 2x^{OM}$, which again lowers an investor's expected loss from a counterparty default.

In order to clearly study the consequences of monitoring incentives, we impose some parametric restrictions in Assumption 4. The assumption first ensures that bilateral monitoring in the *OM-contract* is not incentive compatible, but that monitoring can be optimal ($k > \hat{k}^m$). Furthermore, we impose the condition $k > \underline{k}_N$ and relegate the analysis of the full-hedging case to Internet Appendix C. Our motivation for this additional assumption is twofold. First, the full-hedging region $k < \underline{k}_N$ shrinks exponentially as N increases, as shown by Proposition 3. Second, the new rationale for CCP capital and insights about the design of CCP capital structure arise primarily in the case with loss mutualization ($k > \bar{k}_N$).[21]

Assumption 4. $k \in [\max\{\hat{k}^m, \underline{k}_N\}, \bar{k}]$ *and* $N > N^*$. *That is, monitoring is optimal when observable, the OM-contract features complete loss mutualization, and it violates* (MIC$_{bm}$).[22]

The rest of Section 5 proceeds as follows. We derive the optimal contract under bilateral monitoring in Section 5.1 and under centralized monitoring in Section 5.2. We compare the two schemes to show when the CCP emerges as the efficient monitor in Section 5.3. Section 5.4 provides a welfare analysis of the result.

[21] Internet Appendix C shows our key findings are robust. The high-powered compensation contract for the CCP characterized in Proposition 6 remains (weakly) optimal. Second, the result that the CCP can emerge as the efficient monitor (Proposition 7) is strengthened as it always holds with the full-hedging contract.

[22] Assumption 4 only implies that monitoring is optimal when it is observable. When monitoring is unobservable, it becomes more expensive because the monitor(s) must be incentivized. When we characterize the optimal monitoring scheme in Section 5.3, we derive the new (higher) monitoring threshold.

5.1 Bilateral Monitoring

We first consider the bilateral monitoring scheme. The main tension under this scheme is that counterparty risk insurance via loss mutualization reduces an investor's incentive to monitor her counterparty. This is the classic risk exposure and incentive trade-off extended to a multilateral contracting context. To incentivize monitoring, an investor must suffer large enough losses when her counterparty defaults. This can be achieved by distorting the *OM-contract* via either increasing the payoff an investor receives when her counterparty succeeds or decreasing the payoff conditional on counterparty default. We characterize the optimal distortion in the proposition below, in which we use the superscript $*$ for the equilibrium variables of the optimal contract with unobservable monitoring.

Proposition 5 (Optimal contract under bilateral monitoring). *Let $\bar{k}_{bm} = \frac{1-q}{1-q+\nu q}\bar{k}$. Under Assumption 4, the optimal contract with incentive-compatible bilateral monitoring is*

1. *if $k \leq \bar{k}_{bm}$, a contract with a higher payoff upon counterparty success, that is, $r_s^* > r_f^* = \hat{c}$, no CCP capital, $e^* = 0$ and more collateral than in the OM-contract, $x^* > x^{OM}$,*

2. *if $k \in [\bar{k}^{bm}, \bar{k}]$, a contract with lower payoff upon counterparty default, that is, $r_s^* = \hat{c} > r_f^*$, no CCP capital, $e^* = 0$ and less collateral than in the OM-contract, $x^* < x^{OM}$.*

Proposition 5 shows how to efficiently preserve enough counterparty risk exposure to restore incentives for bilateral monitoring. Increasing the transfer received by an investor conditional on counterparty success ($r_s^* > \hat{c}$) is more efficient than decreasing the transfer conditional on counterparty default ($r_f^* < \hat{c}$) when the collateral cost is low enough ($k < \bar{k}^{bm}$). This is intuitive because a larger transfer to receivers requires more collateral to increase investors' excess payment capacity.

The main take-away from the analysis of bilateral monitoring is that counterparty risk cannot be mutualized completely because counterparty risk insurance conflicts with monitoring incentives. This result motivates our analysis of CCP monitoring in the next section

5.2 Centralized Monitoring by the CCP

In this section, we analyze clearing with centralized monitoring. As all monitoring tasks are delegated to the CCP, the incentive problem associated with monitoring no longer interferes with investors' risk-sharing needs. Compensating the CCP for its monitoring service is, however, costly as it increases investors' liability and hence requires additional collateral (Lemma 1). The CCP contract is then designed with the aim to minimize the cost borne by investors while ensuring the CCP has incentives to exert effort. We derive below the optimal amount of capital e and the optimal compensation schedule $\pi(d)$ for the CCP.

Proposition 6 (Centralized monitoring contract)**.** *Under Assumption 4, the optimal contract with centralized monitoring features complete loss mutualization with $r_s^* = r_f^* = \hat{c}$ and $x^* > x^{OM}$. The CCP breaks even; its compensation and capital contribution are given by*

$$\pi^*(0) = \frac{2\psi}{q^N(1-\alpha^N)}, \quad \pi^*(d) = 0 \textit{ for } d > 0, \quad \textit{and} \tag{11}$$

$$e^* = \underline{e} \equiv \frac{1}{\nu_C}\frac{2\psi\alpha^N}{(1-\alpha^N)}. \tag{12}$$

Proposition 6 shows first that investors must post additional collateral $x^* - x^{OM}$ to support the compensation to the CCP. However, investors can still completely mutualize losses as in the OM contract ($r_o^* = r_o^{OM}$). This is made possible by the separation of monitoring and risk-sharing incentives when the CCP monitors. This result contrasts with Proposition 5 in which we showed loss mutualization is distorted to satisfy bilateral monitoring incentives.

Proposition 6 delivers two new insights regarding the CCP compensation and capital contribution when it plays a monitoring role. Regarding compensation, the CCP should only get paid when no investor defaults. The intuition is as follows. Due to unobservable monitoring and limited liability, the CCP always receives a compensation above its monitoring costs. This agency rent, $\mathbb{E}[\pi(d)] - 2\psi$, is minimized when all compensation is concentrated in the state where no payer defaults ($\pi^*(d) > 0$ only if $d = 0$). This is optimal because the no-default state is most indicative of the fact that CCP has monitored all investors. The optimal compensation is then the minimum value of $\pi(0)$ required to bind (MIC_{cm}). The CCP thus loses all of its promised compensation when one or more payer default. Effectively, it holds a junior tranche and absorbs losses right after the defaulters' pre-committed resources (i.e., collateral) have been exhausted.[23]

The second insight is that there exists a new rationale for CCP capital, beyond its role as counterparty risk insurance. In the *OM-contract*, for the same parameter values, the CCP does not pledge capital. Here, it is *required* to do so by the investors, who have the bargaining power, to capture the agency rent the CCP earns from monitoring. Indeed, (PC_{CCP}) binds at $e^* = \underline{e}$. We also note that its contributed capital is akin to *skin-in-the-game* in the sense that the CCP will lose it when one or more members default. In the proof of Proposition 6, we show that by requiring CCP capital, investors economize on collateral. This result thus implies that when a CCP's outside option or bargaining power improves, it contributes less capital and demand more collateral from investors. As discussed in Section 6.3, this tension reflects the disagreement between CCP members who call for larger capital contributions

[23]In practice, for-profit CCPs also collect non-state-contingent fees from members. In our model, if instead the CCP has bargaining power, it would charge such fees to extract members' benefits from central clearing (formal results are available upon request). In contrast, the high-powered compensation described in Proposition 6 is not affected by the distribution of bargaining power since it is used to efficiently sustain the CCP's monitoring incentives.

(ABN-AMRO, 2020) and CCPs who resist these calls (e.g. LCH, 2015; OCC, 2020).

Our results also reveal endogenous economies of scale in centralized monitoring. As the number of investors N grows, the no-default state becomes more indicative of efforts and hence the rent dissipates.[24] These economies of scale can be seen in the reduction of total CCP capital contribution (Ne^* decreases with N). As we discuss in Section 5.3, this is a crucial force in making the CCP a superior monitor.

Remark 1. *As $\pi^*(0)$ increases exponentially with N, it would violate the resource constraint* (4) *for $d = 0$ if N is large enough. Still, the insight from Proposition 6 that the CCP holds a junior tranche is robust in the following way: after exhausting all the available resources in state $d = 0$ to compensate the CCP, the remaining compensation is paid in the* ***states*** *most indicative of effort, i.e., $d = 1$, then $d = 2$, and so on.*

5.3 Optimal monitoring scheme

Having characterized the optimal contract under both monitoring schemes, we now answer the question: Who should monitor? To illustrate the relevant economic forces, we begin with a numerical example. Figure 3 shows the range of collateral cost and market size in which centralized monitoring is optimal (green region) for two different values of α, a measure of the monitoring incentive friction.

In both panels, we observe that central monitoring tend to be optimal when the cost of collateral is intermediate. The intuition is as follows. If collateral is cheap enough, any form of monitoring is wasteful because counterparty risk can be better dealt with collateral. If collateral is very expensive, bilateral monitoring (blue region) is more efficient than central-

[24]This result is known as "cross-pledging" (see Cerasi and Daltung 2000 and Laux 2001).

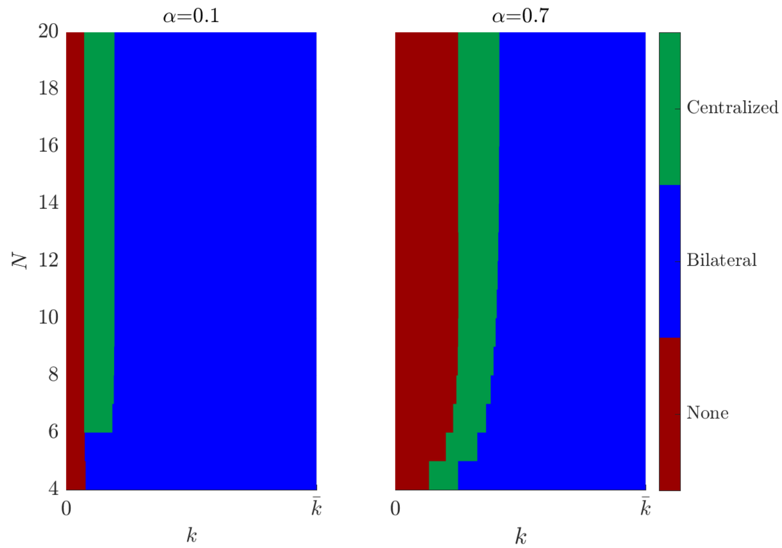

Figure 3. Optimal Monitoring Scheme. Parameter values: $\hat{c} = 0.8$, $\beta = 0.4$, $v = 2$, $q = 0.7$, $\nu_{cm} = 1.9$, $\psi = 5.6 \times 10^{-3}$.

ized monitoring because it requires less collateral, together with reduced loss mutualization (Case 2 of Proposition 5). Therefore, centralized monitoring can only be optimal in the intermediate range of collateral cost.

We further observe that market size N and the severity of the monitoring friction α favor centralized monitoring with respect to bilateral monitoring. A larger N and α require more reduction in loss mutualization to maintain incentives in bilateral monitoring. At the same time, the economies of scale in centralized monitoring becomes more relevant. We note, however, that when N or α increase, loss mutualization also becomes more efficient without monitoring (red region expanded). Hence, the overall effect of these variables on the optimality of centralized monitoring is ambiguous.

To provide analytical support for these observations, we characterize the conditions in which centralized monitoring is optimal when $N \to \infty$. This analysis is subject to the caveat that Assumption 3 cannot hold when N becomes large. We present this result because it is also informative for small values of N: the terms that depend on N in the general condition decrease exponentially (see the proof for details).

Proposition 7. *At the limit $N \to \infty$, when $\alpha > 0$, centralized monitoring is optimal with complete loss mutualization for $k \in [\hat{k}^{cm}, \bar{k}^{cm}]$ where $\hat{k}^{cm} > \hat{k}^{m}$ and $\bar{k}^{cm} < \bar{k}$. This region is non-empty as $\hat{k}^{cm} < \bar{k}^{cm}$ is implied by $\psi < \bar{\psi}$ (Assumption 2).*

Proposition 7 first supports the claim that centralized monitoring is optimal in an intermediate range of collateral. We also confirm the ambiguous effect of monitoring friction by showing that $\hat{k}^{cm}$ and $\bar{k}^{cm}$ both increase with α in the proof.

5.4 Social optimal choice of CCP capital

In this subsection we ask if the equilibrium level of capital and collateral with centralized monitoring is socially optimal. We thus analyze the social planner's choice and then show how it differs in general from privately optimal choices made by either investors or the CCP.

Proposition 8. *Under Assumption 4, there exists $\hat{\nu}_C$ such that the social planner's choice of CCP capital is e^* if $\nu_C \leq \hat{\nu}_C$ and 0 otherwise. In addition, if the CCP had bargaining power, it would not pledge capital.*

Proposition 8 first states that investors may require too much CCP capital compared to the social optimum. CCP capital can replace collateral for hedging, but the substitution is inefficient when capital is costly ($\nu_C \geq \hat{\nu}_C$). For capital level $e \in [0, e^*]$, however, investors do not internalize this cost as the loss from increasing e is borne by the CCP whose monitoring rent decreases. This argument also explains the second statement of Proposition 8: CCPs would not pledge collateral voluntarily if they had bargaining power because doing so eats into their profit.

Putting the results together, the key takeaway of Proposition 8 is that equilibrium choices of CCP capital are generically inefficient. Relative to the socially optimal solution, there is too much (too little) capital if the choice is made by investors (the CCP). The wedge arises because market participants care about the impact of capital contribution on rent distribution but the social planner does not. This result not only echoes the ongoing debate between members and CCP about the suitable amount of capital contribution (see Section 6.3), but also provides a rationale for regulating the loss allocation design in central clearing.

6 Implications for CCP design

In this section, we explain how our results inform the practical design of a CCP. We first discuss the roles of a CCP. We then focus on the CCP's monitoring role and the default waterfall design in this case. Finally, we discuss implications for ownership structure

6.1 Roles of CCP and the determining factors

By studying the efficient management of counterparty risks, our model rationalizes three potential roles of a CCP. First, a CCP can use its capital as insurance against the extreme event of many members defaulting. Second, a CCP can play the role of risk pooler. By ex-ante arranging a loss mutualizing scheme, a CCP pools idiosyncratic member default risks. Third, a CCP can monitor its members to reduce counterparty risks in the first place.

Overall, our results are in line with the view that CCPs are primarily *risk poolers, not insurance providers*—a view widely shared among regulators (Coeuré, 2015; FSB, 2020) and CCPs (LCH, 2015; MRAC, 2021). While we show that CCPs can function as insurance providers, it is only efficient when the CCPs are small and have a lower cost of capital than that of members' collateral. We view the last condition as very restrictive. Instead, we emphasize the CCP's role in facilitating loss mutualization and show that this function is particularly valuable when collateral cost is intermediate and the market size for the cleared contract is large. As we discuss in Section 6.3, however, members will ask the CCP to contribute capital, even when insurance with capital is inefficient.

The monitoring role of CCPs is another emphasis of our paper. Our analysis shows that monitoring is a valuable substitute of costly collateral in mitigating counterparty risk, and that CCPs emerge as efficient monitors of members to facilitate large-scale loss mutualization. Adequate monitoring of members is indeed often cited by many CCPs in practice as their first line of defense against counterparty risks.[25] Monitoring effort in our model represents the costs associated with sound risk management. ESMA (2020) reports that CCPs use internal credit classifications, send mandatory due diligence questionnaires and carry out onsite visits of their members. These tasks require significant investment in data collection and processing capacity as well as in hiring experienced and capable personnel. The provisions of incentives for adequate monitoring is thus paramount and, as we discuss below, have implications for the loss allocation process. Therefore, the two key roles of CCP in our paper are intertwined.

6.2 Default waterfall design

Our analysis of the loss mutualization role of CCPs explains some important features of the loss allocation process, also known as the default waterfall of a CCP. First, due to the limited pledgeability friction, our model endogenizes the commonly observed defaulter-pay principle. Seizing the pledged collateral of defaulting members (akin to initial margin and

[25] For example, in ICE (2019) the CCP writes "The first two levels of protection is the clearing houses' initial and ongoing conservative membership standards... Our clearing houses have developed and implemented a sophisticated review and internal credit rating process that assesses and monitors each clearing member's initial and ongoing credit worthiness."

default fund contribution) efficiently discourages risk taking. Then, the remaining loss will be allocated among surviving members. Their resources pledged in the default fund are thus useful to absorb losses and guarantee further contingent payments at the request of the CCP.

The analysis of monitoring incentives endogenizes another key feature of the default waterfall. As we have explained, a CCP's incentives to monitor its members is best preserved when it holds a first-loss exposure to member's default. This can be achieved by giving the CCP a junior equity tranche in the default waterfall, which absorb losses after the defaulters' resources are exhausted and before surviving members' contributions. This default waterfall structure is very common among CCPs in practice (Duffie, 2015). Regulators and practitioners recognize the importance of this equity tranche to preserve the CCP's incentives.[26]

6.3 The determinants of CCP capital

Three additional, novel implications about the CCP's pre-funded capital, the so-called skin-in-the-game (SITG), follow from our results. First, the CCP's SITG needs not be large and some incentives come in the form of management compensation or loss in value of CCP equity. Our analysis suggests that wiping out CCP's SITG, along with other form of compensation, with only one member default is the most cost-efficient way to provide incentives. Such a thin layer of SITG is common[27] and CCPs in practice make management compensation contingent on the actual usage of SITG to induce risk management effort.[28]

[26]For instance, the Japan Securities Clearing Corporation states that "JSCC should compensate losses before Survivors' Pay, in order to keep incentive for appropriate risk management" (See `https://www.jpx.co.jp/jscc/en/risk/default.html`). See also Coeuré (2015) and FSB (2020) for regulators' views.

[27]We note that in our model when the number of members is very large, the CCP would also receive a positive compensation when more than one member survives. Then, its equity tranche is progressively wiped out as more members default (see Remark 1).

[28]For instance, OCC, a CCP for equity derivatives, says that "OCC will contribute the unvested funds held under its Executive Deferred Compensation Plan (EDCP), on a pro rata basis pari passu with non-defaulting clearing members' default fund contributions." (OCC, 2020) LCH, another CCP, states that besides SITG,

Second, the exact size of SITG would depend on the future profit (agency rent) earned by the CCP for its monitoring service and ultimately the bargaining power of CCP vis-à-vis its members. We showed that a CCP is willing to contribute capital up to a level $\underline{e}$ such that its cost of capital matches its future profit. The higher the future profit is, the larger SITG the CCP is willing to contribute. Meanwhile, the equilibrium capital contribution e^* is a tool for splitting the surplus between the CCP and members. When members have all the bargaining power as in our model, they leave no surplus to the CCP by requesting $e^* = \underline{e}$. The view that SITG is an outcome of bargaining is acknowledged by market participants.[29] Recently, a group of twenty major institutional investors and investment banks has collectively issued a discussion paper (ABN-AMRO, 2020) to request more substantial capital contribution from the CCP. This request illustrates the tug of war between members and the CCPs.

Third, there can be excessive CCP capital demanded by members who do not take its (social) opportunity cost into account, as shown in Section 5.4. In light of the request made in ABN-AMRO (2020) by members, we thus caution against substantially increasing the SITG even for the goal of loss absorption. A similar caution is voiced by regulators and CCPs.[30] Such a change may push the CCP's business model from risk-pooling to insurance provision, as captured by the shift from loss mutualization to full hedging in the contract.

"LCH has further strengthened this incentive structure by linking management compensation directly to usage of the SITG layer." (LCH, 2015)

[29] In a discussion paper written by the International Swap and Derivative Association (ISDA, 2019), for example, ISDA concedes that "The level of SITG is ultimately a judgement call and is still debated between many CCPs and clearing members. We believe that the optimum level of SITG is difficult to agree between CCPs and clearing participants and ask global regulators to develop standards and guidelines as to sizing SITG for CCPs."

[30] FSB (2020) states that "The oversight and/or supervisory authorities acknowledge that this portion is not calibrated with a view to constituting a significant amount of loss absorbing resources. Rather, it is calibrated to provide confidence in the risk management incentives of the CCP." In LCH (2015), LCH argues "A CCP exists to guarantee the default losses of its clearing members and collects sufficient financial resources from its members to provide this default insurance. The CCP does not provide the financial resources necessary to absorb the potential default losses themselves, and so these losses are not a component of CCP capital."

6.4 CCP ownership structure

Private CCPs in practice are either owned by their members or by a third-party, for-profit corporation. CCP ownership structure is relevant to the design of the default waterfall. For instance, in a member-owned CCP, the seniority of funding sources in the default fund is less pertinent, as its SITG and equity are ultimately funded by members. To the extent that our model reveals some key economic principles in default waterfall design, it also has implications for CCP ownership structure.

Consider the case in which the CCP functions as a loss mutualization mechanism.[31] Under bilateral monitoring, the CCP takes transfers from some members and redistributes to others. It neither pledges capital nor receives compensation from members. We interpret this arrangement as a member-owned CCP.[32] Under centralized monitoring, the CCP also acts as an agent who monitors members. It contributes capital and receives an equity-like compensation paid by members. In this sense, it is a third-party, for-profit service provider.

Our results suggest that a third-party CCP is preferable to a member-owned CCP when the number of clearing members is large. In larger CCPs, discipline is better maintained by a third-party agent who is liable for any default. In smaller CCPs, members prefer a "mutualization-light" regime with discipline maintained via bilateral counterparty exposure.

[31] As we showed, the CCP can also act as an insurance provider in the restrictive case where CCP capital is cheaper than members' collateral. In this case, however, the discussion of ownership structure becomes trivial in our model because a member-owned CCP cannot have a lower cost of capital than its members.

[32] Several commentators including McPartland and Lewis (2017) use the terminology "mutualized CCP" for the arrangement without a third-party owner. Since there is also loss mutualization between members in our third-party CCP, we refer to the former CCPs as member-owned.

7 Conclusion

In this paper, we characterized the optimal allocation of losses in a CCP when contracts are subject to counterparty risk. The mutualization of losses hedges investors against their counterparty's default, but this protection lowers market discipline because investors' incentives to trade with creditworthy counterparties become weaker. When the market is large, we show that a third-party CCP can mitigate these inefficiencies by acting as a centralized monitor. Our model endogenizes the typical default waterfall of a CCP with defaulter's collateral, a CCP junior equity tranche and surviving members' default fund contributions. Members and the CCP disagree about the size of the skin-in-the-game capital, and either choice can be socially inefficient.

One of our paper's conclusions is that regulating the capital structure of CCP may be necessary as private choices of market participants can be suboptimal. We believe a dynamic extension of this model could shed light on richer aspects of the regulatory problem, in particular the optimal resolution of CCPs. To understand the basic determinants of the default waterfall, we assumed one CCP clears all trades. In practice, several third-party CCPs may compete for the market. Introducing several CCPs would allows us to analyze the relationship between competition and CCP stability. Relatedly, we also believe that competing CCPs may cater to different clienteles in a model with heterogeneous investors (see e.g., Santos and Scheinkman (2001)). We leave these venues for future research.

Appendix

A Proofs

A.1 Derivation of Equation (3)

We first derive a relationship between expected transfers, given by

$$\mathbb{E}[r_o(d)] + \mathbb{E}[\pi(d)] = \mathbb{E}[p_o(d)] + x + e \tag{A.1}$$

As a payer succeeds with probability q, and default is idiosyncratic the number of defaulting payers among k payers is a random variable with a binomial distribution $\mathcal{B}(k, 1-q)$. Taking expectations over (6), we thus obtain

$$\begin{aligned}
\mathbb{E}_s[p_o(d)] &= \sum_{d=0}^{N-1} (1-q)^d q^{N-1-d} \binom{N-1}{d} \left[r_s(d) + \frac{d}{N-d}(r_f(d) - p_f(d)) - \frac{N}{N-d}(x + e - \pi(d)) \right] \\
&= \mathbb{E}_s[r_o(d)] + \sum_{d=1}^{N-1} (1-q)^d q^{N-1-d} \binom{N-1}{d-1} (r_f(d) - p_f(d)) - (x+e) \sum_{d=0}^{N-1} (1-q)^d q^{N-1-d} \binom{N}{d} \\
&\quad + \sum_{d=0}^{N-1} (1-q)^d q^{N-1-d} \binom{N}{d} \pi(d) \\
&= \mathbb{E}_s[r_o(d)] + \frac{1-q}{q} \sum_{l=0}^{N-2} (1-q)^l q^{N-1-l} \binom{N-1}{l} (r_f(l+1) - p_f(l+1)) \\
&\quad - \frac{(x+e)}{q} \left[1 - (1-q)^N \right] + \frac{1}{q} \left[\mathbb{E}[\pi(d)] - (1-q)^N \pi(N) \right] \\
&= \mathbb{E}_s[r_o(d)] + \frac{1-q}{q} (\mathbb{E}_f[r_o(d)] - \mathbb{E}_f[p_o(d)]) - \frac{x+e}{q} + \frac{\mathbb{E}[\pi(d)]}{q}
\end{aligned}$$

where to obtain the last line, we used (6) for $d = N$. The last line is equivalent to (A.1).

Using equation (1), we can now derive equation (3). We have

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

Substituting $\mathbb{E}[p_o(d)]$ thanks to equation (A.1), we obtain

$$U = qR + \frac{1}{2}x - qRx + \frac{1}{2}(x+e) - \frac{1}{2}\mathbb{E}[\pi(d)] + \frac{\nu - 1}{2}\mathbb{E}\big[\min\{r_o(d), \hat{c}\}\big] - (1 - \mathbb{1}_{cm})\psi$$

which is equivalent to equation (3).

A.2 Proof of Proposition 1

We prove the results in several steps. Step 1 proves that resource constraint (5) binds. Step 2 proves that for all $d < N$, $r_s(d)$ is constant. Step 3 proves that for all $d < N$, $r_f(d)$ is a constant lower than $\hat{c}$ and r_s. In Step 4, we prove that we can focus on contract with $2x + e \leq \hat{c}$ without loss of generality. Finally, in Step 5, we prove $r_f > r_f(N)$. For some arguments in this proof, we will refer to certain contracts introduced later in the main text.

Step 1. Resource constraint (5) binds: $p_f(d) = x$

From equation (6), increasing $p_f(d)$ for $d < N$ allows investors to increase $r_s(d)$ in this state. Such a change may only relax constraints (LP) and (MIC_{bm}). Because investors' utility (3) is weakly increasing with $r_s(d)$, it is thus optimal to set $p_f(d) = x$ for all $d < N$.

For state $d = N$, suppose (5) 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 pledgeability constraint (LP) still holds. Consider then a joint increase in $r_f(N)$ and $\mathbb{E}_s[r_o(d)]$ such that

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

The first constraint ensures that resource constraint (5) is still satisfied following the perturbation. The second constraint ensures that bilateral monitoring constraint (MIC_{bm}) is satisfied after the perturbation if needed. The last constraint ensures that budget constraint (6) is still satisfied. Since $\Delta p_f(N) > 0$ and $\Delta\mathbb{E}_s[r_o(d)] > 0$, by construction, such a perturbation exists and it is weakly optimal because investors' utility weakly increases with $r_o(d)$. Hence, $p_f(N) = x$ is optimal.

Step 2. $r_s(d) = r_s$ **for all** $d < N$

Suppose instead there are 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 (LP) and it weakly relaxes bilateral monitoring constraint (MIC_{bm}) (strictly if $r_s(d) > \hat{c} > r_{s'}(d')$). It is (weakly) profitable because objective function (3) is concave in $r_s(d)$ and $r_s(d')$.

Step 3. $r_f(d) = r_f \leq \min\{r_s, \hat{c}\}$ **for all** $d < N$

We first show that setting $r_f(d) = r_f$ for all $d < N$ is optimal. Suppose instead there are two states (d, d') such that $r_f(d) > r_f(d')$. The argument used in Step 2 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

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

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 (MIC_{bm}) is unchanged. To satisfy budget constraint

(6) 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 (LP) still holds after the perturbation as the expected payment $\mathbb{E}_s[p_o(d)]$ increases by

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

The perturbation strictly increases the objective function (3) which is concave in r_f.

We then show that $r_f \leq \min\{r_s, \hat{c}\}$ is optimal. The result $r_f \leq \hat{c}$ follows from two observations. First, the objective function (3) is independent of r_f when $r_f > \hat{c}$ and increasing r_f does not relax any constraint but it tightens constraint (MIC_{bm}).

For the second part of the result, suppose $r_f > r_s$ and consider the following perturbation:

$$\Delta r_f < 0, \qquad \Delta r_s = -\frac{1-q-(1-q)^N}{q}\Delta r_f, \qquad \text{such that} \quad r_f + \Delta r_f = r_s + \Delta r_s$$

Let $\Delta p_s(d)$ be the perturbation to $p_s(d)$ needed in state $d < N$ to satisfy the budget constraint (6) while keeping other variables constant. The perturbation is designed such that $\mathbb{E}[p_s(d)]$ does not change, as can be seen from (A.1). This implies constraint (LP) still holds. Hence, the perturbation is feasible under constraint (LP) and (MIC_{bm}) because the right-hand side of the latter constraint is increasing with r_s and decreasing with r_f. With this perturbation, $\mathbb{E}[r_o(d)]$ is unchanged, which means investors' utility is unchanged. Hence, it is weakly optimal to set $r_s \geq r_f$ and it can be strictly optimal if it relaxes (MIC_{bm}).

Step 4. Proof that $r_f(N) = 2x + e \leq \hat{c}$

To prove this statement, we first rely on properties of the CCP's compensation contract shown later in the text. Proposition 6 shows that it is optimal not to compensate the CCP in state $d = N$. Hence, we set $\pi(N) = 0$. Using the result in Step 1, we can rewrite budget constraint (6) in state $d = N$ as $r_f(N) \leq 2x + e$. Setting $r_f(N) \leq \hat{c}$ is weakly optimal by the same argument used in Step 3 for r_f. Hence, we are left to show that we can focus on contracts such that $2x + e \leq \hat{c}$. We proceed by contradiction considering a "candidate" contract such that $2x + e > \hat{c}$.

Case 1. $k \leq \nu_C - 1$

In this case, the candidate contract is dominated by the full-hedging contract described in Proposition 2. Because this contract does not require monitoring, it is enough to show that the candidate contract is more costly since hedging benefits are lower. The combined cost of collateral and CCP capital with the candidate contract is given by

$$xk + \frac{1}{2}e(\nu_C - 1) > \frac{\hat{c}}{2}k + \frac{1}{2}e(\nu_C - 1 - k) > \frac{\hat{c}}{2}k$$

The last expression is the cost for the the full-hedging contract. Hence, the candidate contract cannot be optimal.

Case 2. $k > \nu_C - 1$

We first rewrite the limited pledgeability constraint (LP) using the result from Step 1, 2 and 3, the participation constraint of the CCP (PC_{CCP}) and the budget constraint (6)

$$qr_s+(1-q)\left[1-(1-q)^{N-1}\right]r_f \leq q\beta+\Big(2-q\beta-2(1-q)^N\Big)x-\left[\nu_C - \left\{1-(1-q)^N\right\}\right]e-2\psi\mathbf{1}_{cm} \quad \text{(A.2)}$$

By assumption about our candidate contract, $r_f(N)$ satisfies $r_f(N) \leq \hat{c} < 2x+e$. In what follows, we treat separately the two monitoring schemes.

Consider first the bilateral monitoring scheme. Consider a perturbation to the candidate contract with $\Delta x < 0$ and $\Delta e = \frac{2-q\beta-2(1-q)^N}{\nu_C-\{1-(1-q)^N\}}\Delta x$. By construction, this perturbation leaves the right-hand-side of (A.2) unchanged. Hence, the same levels for r_s and r_f as in the candidate contract can be financed. Besides, because $\hat{c} < 2x+e$, it is also possible to finance any receiver transfer $r_f(N) \leq \hat{c}$ for a small enough changes in x and e. Because CCP capital and collateral are costly, it is feasible and optimal to decrease e and r_f until either $2x+e = \hat{c}$, $e = 0$ or $x = 0$. In the first case, the desired result follow directly. In the second case, the candidate contract is dominated by the full-hedging contract described above. In the third case, a contract with $x = 0$ and $e > 0$ is dominated by a contract with $x = 0$ and $e = 0$ because CCP capital is costly and reduces pledgeability, as shown formally in Lemma 1. This proves the result with bilateral monitoring.

We now consider the centralized monitoring scheme. In Proposition 6, we show that the CCP capital should be no lower than $\underline{e} = \frac{2\psi\alpha^N}{1-\alpha^N}$. This implies that the candidate contract satisfies $e \geq \underline{e}$ by definition of $\underline{e}$, as it could be improved otherwise. We now prove that the candidate contract is dominated by the (optimal) contract characterized in Proposition 6 which has the following features: $e = \underline{e}$ such that $r_s = r_f = \hat{c}$ and $2x+e \leq \hat{c}$. Hence, this optimal contract fully satisfies investors' hedging needs except in the joint payer default state. The candidate contract can only dominate that contract if it provides full hedging or if it is cheaper.

In the first case, however, starting from the optimal contract, it is better to satisfy $2x+e = \hat{c}$ in order to provide full hedging, because collateral and CCP capital are costly. In the second case, because $e > \underline{e}$ for the candidate contract, it can only be cheaper than the optimal contract if it uses less collateral. But then start again from the optimal contract for which the limited pledgeability constraint (A.2) binds. Reducing x and increasing e thus implies lowering the expected transfer to receivers. If this change increases investors' utility, x should be lowered to 0 by linearity. But then a contract with $x = 0$ and $e > \underline{e}$ is dominated by a contract with $x = 0$ and $e = \bar{e}$ because CCP capital is costly and tightens constraint (A.2). This proves the result with centralized monitoring.

Step 5. Proof that $r_f \geq r_f(N)$

We consider again the centralized monitoring scheme and the bilateral monitoring scheme in turn. Consider first the centralized monitoring scheme. Either $r_s = r_f = \hat{c}$ or limited pledgeability constraint (LP) binds. In the first situation, $r_f(N) = 2x+e \leq \hat{c} = r_f$ by Step 4. In the second situation, two cases are again possible. If $\frac{\nu-1}{2}(2-q\beta) \geq k$, then increasing x to increase r_s and r_f until they are equal to $\hat{c}$ is optimal. The result follows again. If instead $\frac{\nu-1}{2}(2-q\beta) > k$, it is optimal to decrease x until it reaches 0 so that

$$\mathbb{E}[r] = q\beta - (\nu_C - 1)e - 2\psi$$

But then, it should be optimal to switch to bilateral monitoring with $e = 0$ because it increases the right-hand side and thus the transfers of the left-hand-side of the equality above. Bilateral monitoring is incentive-compatible with contract $r_s = \beta$, $r_f = 0$ and $x = 0$ under Assumption 2 as we will show in Lemma 2. Again, the desired result holds.

Consider now the bilateral monitoring scheme. With a similar argument, we can focus on the case in which the limited pledgeability constraint binds. The argument when $\frac{\nu-1}{2}(2 - q\beta) > k$ is similar to that above. Suppose then $\frac{\nu-1}{2}(2 - q\beta) \leq k$. This implies that x should be increased until $r_s = \hat{c}$. Increasing r_f, however, entails an additional cost because the monitoring constraint (MIC_{bm}) needs to be satisfied. Hence, to increase r_f, one must also increase r_s. Two cases are possible. First, if the cost of collateral is low, r_f should be increased until it reaches $\hat{c}$ and the proof follows by Step 4. Otherwise, r_f should be set such $r_s = \hat{c}$ and (LP) and (MIC_{bm}) hold as equality. This contract is the contract considered in Case 2 of Proposition 5 and, as we show there, it satisfies $r_f \geq 2x + e$ under Assumption 2. This concludes the proof.

A.3 Proof of Proposition 2

Using Lemma 1, we derive a simplified version of the investor's problem in the absence of friction. Recall that monitoring is redundant if the asset is fully pledgeable. The investors solve

$$\max_{x,e,r_s,r_f} \frac{\nu-1}{2}\left[q\min\{r_s,\hat{c}\} + (1-q)\Big(\left[1-(1-q)^{N-1}\right]\min\{r_f,\hat{c}\} + (1-q)^{N-1}(2x+e)\Big)\right] \\ - x(qR-1) - \frac{1}{2}e(\nu_C - 1) \tag{A.3}$$

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 it is constant otherwise. Hence, it is optimal to set $r_s = r_f = \hat{c}$. To determine the optimal values of x and e, compute the derivative of the objective function with respect to these variables:

$$U'(e) = \frac{1}{2}(\nu-1)(1-q)^N - \frac{1}{2}(\nu_C - 1) \tag{A.4}$$

$$U'(x) = (\nu-1)(1-q)^N - k \tag{A.5}$$

It is thus optimal to set $r_f(N) = 2x + e$ equal to $\hat{c}$ if and only if $\min\{k, \nu_C - 1\} \leq (\nu-1)(1-q)^N$. In this condition holds, investors choose $(x,e) = \left(\frac{c}{2}, 0\right)$ if $k \leq \nu_C - 1$ and $(x,e) = (0,\hat{c})$ otherwise. This concludes the proof.

A.4 Proof of Proposition 3

Step 1. Limited Pledgeability Constraint

We first rewrite the limited pledgeability constraint (LP). We showed in Proposition 1 that $\mathbb{E}_f[p_o(d)] = x$, and in Lemma 1 that $\mathbb{1}_{cm} = 0$. Using these results together with the binding participation constraint of the CCP (PC_{CCP}) and equation (A.1), we obtain

$$q\Big(\mathbb{E}_s[p_o(d)] - \mathbb{E}_f[p_o(d)]\Big) = \mathbb{E}_s[r_o(d)] - 2x + (\nu_C - 1)e$$
$$= qr_s + (1-q)\big[1-(1-q)^{N-1}\big]r_f - \Big[1-(1-q)^N\Big](2x+e) + \nu_C e$$

We can thus rewrite (LP) as a function of (r_s, r_f, e, x).

$$qr_s + (1-q)\big[1-(1-q)^{N-1}\big]r_f \le q\beta + \Big(2 - q\beta - 2(1-q)^N\Big)x - \big[\nu_C - \big\{1-(1-q)^N\big\}\big]\,e \quad \text{(A.6)}$$

Investors thus solve the problem described in (A.3) under constraint (A.6).

Step 2. Analysis

We first show two results about CCP capital e. First, CCP capital may be used only if $k < \nu_C - 1$. If this condition does not hold, we showed in Proposition 2 that collateral is preferred to CCP capital in the frictionless benchmark. This conclusion still applies under limited pledgeability because x (resp. e) relaxes (resp. tightens) constraint (A.6). Second, if CCP capital is used, it must be that (A.6) binds. Otherwise, it is optimal to increase e and decrease x while keeping $r_f(N) = 2x + e$ constant. With a small enough change, constraint (A.6) still holds and the objective function increases because $k < \nu_C - 1$ must hold if CCP capital is used, as we just showed.

We now argue we can consider two different cases for the analysis: Either $r_s = r_f = \hat{c}$ or constraint (A.6) binds. This observation follows from Proposition 2 where we showed $r_s = r_f = \hat{c}$ is optimal in the absence of constraint (A.6). Besides, the relative weight on these two variables is the same in the objective function (A.3) and in constraint (A.6).

Suppose first that $r_s = r_f = \hat{c}$. We now derive conditions such that $r_f(N) = 2x + e = \hat{c}$.

Optimality of $r_f(N) = \hat{c}$

Case 1a. $k \le (\nu - 1)(1-q)^N$

Increasing x until $r_f(N) = 2x + e = \hat{c}$ is then optimal by (A.5) and because increasing x relaxes constraint (A.6). If in addition $k < \nu_C - 1$, CCP capital should not be used as shown above. In this case, the contract is given by $r_s^{OM} = r_f^{OM} = \hat{c}$, $x^{OM} = \frac{\hat{c}}{2}$ and $e^{OM} = 0$. This corresponds to Case 1a of Proposition 3.

If instead $k > \nu_C - 1$, CCP capital should be used and, as shown above, constraint (A.6) should bind. Hence, the contract is given by $r_s^{OM} = r_f^{OM} = \hat{c}$ and x^{OM} and e^{OM} such that $r_f^{OM}(N) = 2x^{OM} + e^{OM} = \hat{c}$ and (A.6) binds. This corresponds to Case 1b of Proposition 3.

Case 1b. $k > (\nu - 1)(1-q)^N$

Then, it is optimal to decrease x until constraint (A.6) binds because $U'(x) < 0$. Equation (A.4) shows that increasing e until $r_f(N) = 2x + e = \hat{c}$ can still be optimal if $\nu_C - 1 \le (\nu - 1)(1-q)^N$. To determine the sufficient condition, we need to account for the effect of e on constraint (A.6) when computing the total derivative of the objective function with respect to e. Maintaining r_s and r_f

constant in equation (A.6), we have

$$\frac{\partial x}{\partial e}_{|r_f=r_s=\hat{c},\ (A.6)\text{ binds}} = \frac{\nu_C - 1 + (1-q)^N}{2 - q\beta - 2(1-q)^N} \tag{A.7}$$

We thus obtain

$$\begin{aligned} U'(e)_{|r_f=r_s=\hat{c},\ (A.6)\text{ binds}} &= \frac{\partial U}{\partial e} + \frac{\partial U}{\partial x}\frac{\partial x}{\partial e}_{|r_f=r_s=\hat{c},\ (A.6)\text{ binds}} \\ &= \frac{1}{2}\left[(\nu-1)(1-q)^N - (\nu_C - 1)\right] + \left[(\nu-1)(1-q)^N - k\right]\frac{\nu_C - 1 + (1-q)^N}{2 - q\beta - 2(1-q)^N} \end{aligned}$$

This term is positive if

$$k \leq \underline{k}_N \equiv (\nu-1)(1-q)^N + \frac{1}{2}\frac{2 - q\beta - 2(1-q)^N}{\nu_C - 1 + (1-q)^N}\max\left\{(\nu-1)(1-q)^N - (\nu_C - 1), 0\right\} \tag{A.8}$$

If this inequality holds, $r_f(N) = \hat{c}$ is optimal, and thus the OM-contract is given by $r_s^{OM} = r_f^{OM} = \hat{c}$ and x^{OM} and e^{OM} such that $r_f^{OM}(N) = 2x^{OM} + e^{OM} = \hat{c}$ and (A.6) binds. Hence, we characterized all cases in which $r_f(N) = \hat{c}$ is optimal.

Optimality of $r_s = r_s = \hat{c}$ and $r_f(N) < \hat{c}$

Suppose now that condition (A.8) does not hold while still assuming $r_s = r_f = \hat{c}$. Then the analysis above shows that setting $e = 0$ is optimal. Since $k > \underline{k}_N$ and thus $k > (\nu-1)(1-q)^N$, the collateral amount x is pinned down by saturating constraint (A.6) with $e = 0$. In addition, a contract with the conjectured properties is optimal if decreasing x when (A.6) binds decreases the objective function. We have in this case

$$U'(x) = \frac{\nu-1}{2}\frac{\partial \mathbb{E}[r_o(d)]}{\partial x}_{|e=0,(A.6)\text{ binds}} - k = \frac{\nu-1}{2}(2 - q\beta) - k \equiv \bar{k} - k \tag{A.9}$$

The conjecture is thus optimal if $k \in [\underline{k}_N, \bar{k}]$. This corresponds to Case 2 of Proposition 3.

Optimality of $r_s, r_f < \hat{c}$

Suppose finally that $k \geq \bar{k}$. Then, again setting $e = 0$ is optimal because $\bar{k} \geq \underline{k}_N$ with equality only for $\nu_C = 1$. But in this case, it is also optimal to set x to 0 since the marginal benefit of collateral is given by (A.9). The optimal contract is then characterized by $e^{OM} = 0$, $x^{OM} = 0$. The values of r_s and r_f are pinned down by the binding pledgeability constraint (A.6), that is,

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

In particular the contract such that $r_s = \beta$ and $r_f = 0$ is optimal, which corresponds to Case 3 of Proposition 3. This concludes the proof.

A.5 Proof of Corollary 1

We prove the result here in the case where monitoring is imposed. The proof for the case where investors can choose whether to monitor is in Internet Appendix D.1. We verify that the OM-contracts of Proposition 3 satisfy Definition 2 only in Cases 1a and 3.

For Case 1a, we have $r_o(d) = 2x = p_o(d) + x$ for all d. For Case 3, we have $r_s(d) = p_s(d) = \beta$ and $r_f(d) = 0 = p_f(d)$. Hence, both contracts satisfy Definition 2.

The contract in Case 1b requires CCP capital and thus cannot bee implemented bilaterally. Indeed, the bilateral restriction in Definition 2 and the CCP's participation constraint (PC_{CCP}) imply $e = 0$. For Case 2, we have $r_f^{OM}(d) = \hat{c} > p_f^{OM}(d) + x^{OM}$ for all $d < N$, and thus this contract also violates Definition 2.

It follows that the upper bound for the essential CCP region is given by $\bar{k}$. The lower bound $\underline{k}^{ess}$ corresponds to the upper bound of the region for Case 1a and it is thus given by $\underline{k}^{ess} = \min\{(\nu - 1)(1 - q)^N, \nu_C - 1\}$. This concludes the proof.

A.6 Proof of Corollary 2

The optimal bilateral contract is obtained from Proposition 3 with monitoring and A.1 without monitoring respectively, setting $N = 1$ and imposing $e = 0$.

We first show that when k is close to the upper bound $\bar{k}$ of the essential CCP region, the bilateral contract requires strictly less collateral. By Proposition 3, for k lower but close to $\bar{k}$, the optimal contract is given by Case 2 of Proposition 3 for all $N \geq 1$. Equation (9) shows that the collateral requirement x^{OM} is increasing in N because $\hat{c} \leq 2$ under Assumption 1. This proves that a bilateral contract requires less collateral for k close to $\bar{k}$.

We now show that for k above but close to $\underline{k}_{ess}$ the bilateral contract requires strictly more collateral. Remember that $\underline{k}_{ess} = \min\{\nu_C - 1, (\nu - 1)(1 - q)^N\}$. If $\underline{k}_{ess} = \nu_C - 1$, the optimal multilateral contract is given by Case 1b of Proposition 3 while the optimal bilateral contract is given by Case 1a. The latter requires strictly more collateral, which proves the result in this case. Suppose instead that $\underline{k}_{ess} = (\nu - 1)(1 - q)^N$. Then, there exists a range of collateral cost $[(\nu - 1)(1 - q)^N, (\nu - 1)(1 - q)]$ such that the optimal multilateral contract is given by Case 2 of Proposition 3 while the optimal bilateral contract is given by Case 1a. Again, the latter contract requires strictly more collateral, which concludes the proof.

A.7 Proof of Proposition 4

We first derive the optimal contract without monitoring in Section A.7.1 and then derive the optimal monitoring decision in Section A.7.2.

A.7.1 Optimal Contract without Monitoring

We first establish that a single (pooling) contract is offered although investors may have different ex-post types. Without monitoring, each investor has pledgeability β with probability α or 0 with probability $1 - \alpha$. With unobservable types, a menu of contracts could be used to screen investors.

In our environment, however, screening is not possible due to a failure of the Mirrless-Spence sorting condition. The investor type changes the asset pledgeability but investors' utility (3) does not depend on the type. This implies that investors always agree on the best contract in a menu and separation is not possible.

The result above greatly simplifies the analysis of the optimal contract without monitoring. As only one contract is offered, we can consider investors ex-ante, that is before their pledgeability type is realized. It follows that lack of monitoring simply increases the probability of default of an investor from $1-q$ to $1-\alpha q$. The collateral cost k, however, is the same because the asset succeeds with probability q, independently of the investor type.

If follows from these observations that we can derive the optimal contract without monitoring by adapting Proposition 3 substituting q with αq (while keeping $k = qR - 1$). We use the superscript $\not{m}$ to indicate that investors are not monitored.

Proposition A.1. *Suppose investors are not monitored. There are two thresholds of collateral cost*

$$\underline{k}_N^{\not{m}} = (\nu - 1)(1-\alpha q)^N + \frac{1}{2}\frac{2-\alpha q\beta - 2(1-\alpha q)^N}{\nu_C - 1 + (1-\alpha q)^N} \max\left\{(\nu-1)(1-\alpha q)^N - (\nu_C - 1), 0\right\},$$

$$\bar{k}^{\not{m}} = \frac{1}{2}(\nu - 1)(2 - \alpha q \beta)$$

such that

1. *if $k \leq \underline{k}_N^{\not{m}}$, a fully collateralized contract is optimal with*
   (a) *no CCP capital and collateral $x^{OM,\not{m}} = \frac{\hat{c}}{2}$ if $\nu_C - 1 < k$*
   (b) *CCP capital $e_C^{OM,\not{m}} = \frac{\alpha q\beta(2-\hat{c})}{2\nu_C - \alpha q \beta}$ and collateral $x^{OM,\not{m}} = \frac{\nu_C \hat{c} - \alpha q \beta}{2\nu_C - \alpha q\beta}$ if $\nu_C - 1 \geq k$*
2. *if $k \in [\underline{k}_N^{\not{m}}, \bar{k}^{\not{m}}]$, a complete LM contract is optimal with $r_s^{OM} = r_f^{OM} = \hat{c}$ and*

$$x^{OM,\not{m}} \equiv \frac{\left[1-(1-\alpha q)^N\right]\hat{c} - \beta q}{2\left[1-(1-\alpha q)^N\right] - \beta \alpha q} \in \left(0, \frac{\hat{c}}{2}\right), \tag{A.10}$$

3. *if $k \geq \bar{k}^{\not{m}}$, the contract in Case 3 of Proposition 3 is optimal.*

A.7.2 Optimal monitoring decision

We first prove that monitoring is optimal if the collateral cost is above a threshold $\hat{k}^m$, if it exists. We then characterize $\hat{k}^m$ to prove the properties listed in Proposition 4.

Step 1. Threshold condition

The argument relies on three claims.

The first claim is that for a given monitoring choice, the difference in investor's utility across consecutive contracts is strictly increasing with k. A contract is consecutive to a reference contract if it is the next optimal contract when increasing k. For example, with monitoring the contract consecutive to the Case 1a contract is the Case 1b contract if $\nu_C - 1 \leq (\nu - 1)(1-q)^N$, and the Case 2 contract if $\nu_C - 1 > (\nu - 1)(1-q)^N$ (see Figure 2). We only present the argument when investors monitor because the argument without monitoring is similar. For each Case of Proposition 3, the contract terms do not depend on k. Hence to prove the claim, it is enough to show that a consecutive contract uses strictly less collateral. This result is straightforward for all cases except for Case 1b contract, with consecutive contract the Case 2 contract. Using Proposition 3 to compare collateral requirements, the desired result also holds for this case because

$$\frac{\nu_C \hat{c} - q\beta}{2\nu_C - q\beta} \geq \frac{\hat{c} - q\beta}{2 - q\beta} > \frac{\hat{c}\big[1-(1-q)^N\big] - q\beta}{2\big[1-(1-q)^N\big] - q\beta}$$

where the leftmost (rightmost) term is x^{OM} for Case 1b (Case 2). Both inequalities follow from the observation that the mapping $x \mapsto \frac{x\hat{c} - q\beta}{2x - q\beta}$ is strictly increasing because $\hat{c} < 2$.

The second claim is that for a given contract type, the collateral requirement is lower when investors monitor. A direct comparison between Proposition 3 and A.1 shows the desired inequality holds strictly in all cases except Case 1a when both contracts are the same and thus require the same amount of collateral.

The third claims is that the thresholds between consecutive contracts are strictly higher under no monitoring. The comparison between $\bar{k}$ and $\bar{k}^{\not m}$ shows immediately that $\bar{k} < \bar{k}^{\not m}$. We now compare $\underline{k}_N$ to $\underline{k}_N^{\not m}$. First, if $\nu_C - 1 \geq (\nu - 1)(1 - \alpha q)^N$, we have

$$\underline{k}_N = (\nu - 1)(1-q)^N < (\nu - 1)(1 - \alpha q)^N = \underline{k}_N^{\not m}$$

Next, if $\nu_C - 1 \in [(\nu - 1)(1-q)^N, (\nu - 1)(1 - \alpha q)^N]$, we have $\underline{k}_N = (\nu - 1)(1-q)^N$ while $\underline{k}_N^{\not m} \geq (\nu - 1)(1 - \alpha q)^N$. Finally, when $\nu_C - 1 \leq (\nu - 1)(1-q)^N$,, observe that both thresholds are linearly decreasing functions of $\nu_C - 1$. Besides,

$$\lim_{\nu_C \to 1} \underline{k}_N = \bar{k} < \bar{k}^{\not m} = \lim_{\nu_C \to 1} \underline{k}_N^{\not m}$$

This proves the result in all possible cases.

These three claims together imply that the benefit from monitoring is strictly increasing with k except when $k \leq \underline{k}^{ess}$ where it is constant and equal to $-\psi$. Indeed, in this latter case, the contract is the same with or without monitoring.

Step 2. Characterization of threshold $\hat{k}^m$

The results in Step 1 show that, if it exists, the collateral cost threshold $\hat{k}^m$ above which monitoring is optimal satisfies $\hat{k}^m > \underline{k}^{ess}$ for $\psi > 0$. For the degenerate case $\psi = 0$, any value in $[0, \underline{k}^{ess}]$ is admissible.

Since $\underline{k}^{ess} < \bar{k}$ by Corollary 1, to show that the threshold exists, it is enough to show that monitoring is optimal for $k = \bar{k}$. When $k = \bar{k}$, by Proposition A.1, the optimal contract without monitoring is given by Case 1b or 2. In the first case, that is when $\underline{k}_N^{\not m} \geq \bar{k}$, investors' utility is given by

$$
\begin{aligned}
U^{\not m}_{|k=\bar{k}} &= qR + \Big[(\nu-1)-\bar{k}\Big]\frac{\hat{c}}{2} + \frac{1}{2}\Big[\bar{k}-(\nu_C-1)\Big]e^{OM,\not m} \\
&= qR + q\beta\frac{\nu-1}{2}\frac{\hat{c}}{2} + \frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta}{2\nu_C - q\alpha\beta}\max\left\{0,(2-q\beta)-\frac{2(\nu_C-1)}{\nu-1}\right\}
\end{aligned}
$$

An upper bound for $U^{\not m}_{|k=\bar{k}}$ is obtained by letting $\nu_C \to 1$. We get

$$
U^{\not m}_{|k=\bar{k}} \le qR + q\beta\frac{\nu-1}{2}\frac{\hat{c}}{2} + \frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta}
$$

When $\underline{k}^{\not m}_N \le \bar{k}$, the optimal contract without monitoring is the Case 2 contract, and

$$
\begin{aligned}
U^{\not m}_{|k=\bar{k}} &= qR + \Big[(\nu-1)-\bar{k}\Big]\frac{\hat{c}}{2} + (k-\underline{k}^{\not m}_N)\left(\frac{\hat{c}}{2}-x^{OM,\not m}\right) \\
&= qR + q\beta\frac{\nu-1}{2}\frac{\hat{c}}{2} + \frac{\nu-1}{2}\frac{2-q\beta-2(1-\alpha q)^N}{2\big[1-(1-\alpha q)^N\big]-\beta\alpha q}\beta\alpha q\left(1-\frac{\hat{c}}{2}\right)
\end{aligned}
$$

The second term of the last expression is increasing in N. Hence, an upper bound for $U^{\not m}_{|k=\bar{k}}$ is obtained by letting $N \to \infty$, that is,

$$
U^{\not m}_{|k=\bar{k}} \le qR + q\beta\frac{\nu-1}{2}\frac{\hat{c}}{2} + \frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta}
$$

which is the same upper bound we obtained in the first case.

Hence, the utility without monitoring is lower for $k=\bar{k}$ if

$$
\begin{aligned}
0 &\le U_{k=\bar{k}} - \left\{qR + q\beta\frac{\nu-1}{2}\frac{\hat{c}}{2} + \frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta}\right\} \\
&\le qR + \frac{\nu-1}{2}q\beta - \psi - \left\{qR + q\beta\frac{\nu-1}{2}\frac{\hat{c}}{2} + \frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta}\right\} \\
&\le \frac{\nu-1}{2}q\beta\left(1-\frac{\hat{c}}{2}\right) - \frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta} - \psi \\
&\le \frac{\beta q(1-\alpha)(\nu-1)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right) - \psi
\end{aligned}
$$

It is easy to verify that the first term on the right-hand-side of the last inequality is strictly above the upper bound $\bar{\psi}$ for the monitoring cost. Hence, under Assumption 2, monitoring is optimal for $k=\bar{k}$, and thus the monitoring threshold $\hat{k}^m$ exists and it lies strictly below $\bar{k}$. This concludes the proof.

A.8 Proof of Lemma 2

Suppose first $k \in [\hat{k}^m, \underline{k}_N]$. In this case, by Proposition 3, the *OM-contract* is given by Case 1, with $r_s^{OM} = r_f^{OM} = r_f^{OM}(N)$. This implies the bilateral monitoring constraint (MIC_{bm}) is violated. Suppose now that $k \geq \underline{k}_N$. Under Assumption 4, the *OM-contract* is given by Case 2 of Proposition 3, with $r_s^{OM} = r_f^{OM} = \hat{c}$, $e^{OM} = 0$ and x^{OM} given by equation (9). Plugging these variables into the bilateral monitoring constraint (MIC_{bm}), we obtain condition (10).

A.9 Proof of Proposition 5

We first rewrite the bilateral monitoring constraint (MIC_{bm}) using the results from Proposition 1.

$$\begin{aligned}\frac{\psi}{1-\alpha} \leq \; & \frac{1}{2}\Big[r_s - r_f + (1-q)^{N-1}\left(r_f - (2x+e)\right)\Big] \\ & + \frac{\nu-1}{2}\Big[\min\{r_s, \hat{c}\} - \Big(\big[1-(1-q)^{N-1}\big]\min\{r_f, \hat{c}\} + (1-q^{N-1})((2x+e))\Big)\Big] \quad \text{(A.11)}\end{aligned}$$

The optimal contract under bilateral monitoring solves problem (A.3) under limited pledgeability constraint (A.6) and constraint (A.11) which correspond respectively to constraints (LP) and (MIC_{bm}) in the Investor's Problem. In Step 1, we show that constraints (LP) and (MIC_{bm}) bind. In Step 2, we derive the threshold $\bar{k}_{bm}$. Finally in Step 3, we characterize the optimal distortion to the *OM-contract* of Proposition 3.

Step 1. (LP) and (MIC_{bm}) bind

Under Assumption 4, constraint (A.11) binds because the *OM-contract* in Proposition 3 violates (A.11). The limited pledgeability constraint (LP) must also bind. If it does not, decrease x while keeping r_s and r_f constant. This change relaxes constraint (A.11). Hence, the marginal effect on investors' utility from this perturbation is given by $-U'(x)$ in equation (A.5), which is positive because $k > \bar{k}_N$ by Assumption 4.

Step 2. Threshold $\bar{k}_{bm}$ and optimal contract

We now derive the optimal distortion to the Case 2 contract of Proposition 3. By Proposition 3, it is optimal to set $r_s \geq \hat{c}$ under Assumption 4 when constraint (A.11) is not imposed. Hence, it is still optimal under additional constraint (A.11) because increasing r_s relaxes this constraint. It is also optimal to increase r_f until (A.11) binds. Under Assumption 4, this value denoted $\underline{r}_f$ must lie strictly below $\hat{c}$.

The optimal value of r_f, and thus the optimal contract itself, depend on the marginal value of increasing r_f when $r_f \in [\underline{r}_f, \hat{c}]$. From (A.6) and (A.11), we have (for given x and e).

$$qr_s + (1-q)\big[r_f - (1-q)^{N-1}(r_f - 2x - e)\big] = (2-q\beta)x + q\beta - (\nu_C - 1)e \quad \text{(A.12)}$$

$$r_s - v\big[r_f - (1-q)^{N-1}(r_f - 2x - e)\big] = \frac{2\psi}{q(1-\alpha)} - (\nu-1)\hat{c} \quad \text{(A.13)}$$

Hence, we obtain

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

We can plug this relationship into the expression for investors' utility in (A.3). Because $r_s \geq \hat{c}$, the utility U is then a function of x and e only. It follows that increasing x to increase r_f above $\underline{r}_f$ is profitable if and only if

$$k \leq \frac{\nu-1}{2}\frac{1-q}{1-q+\nu q}(2-q\beta)=\frac{1-q}{1-q+\nu q}\bar{k}\equiv \bar{k}_{bm}<\bar{k}$$

Step 3. Optimal distortion

CCP capital e tightens monitoring constraint (A.11). This observation implies that setting $e=0$ remains optimal when $k>\underline{k}_N$, as in the observable monitoring case. The analysis in Step 2 then shows that only two contracts are possible depending on the ranking between k and $\bar{k}_{bm}$.

Case i) $k \leq \bar{k}_{bm}$

In this case, $r_f^*=\hat{c}$. Setting $e^*=0$ and solving for x using (A.12) and (A.13), we obtain

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

Hence,

$$x^*=\frac{\hat{c}\Big(1-(1-q)^{N-1}\big[\nu q+1-q\big]\Big)-q\beta+\frac{2\psi}{1-\alpha}}{2-2(1-q)^{N-1}\big[\nu q+1-q\big]-\beta q}>x^{OM} \tag{A.14}$$

It can easily be verified that the conjecture $2x^*\leq\hat{c}$ holds under Assumption 2.

Case ii) $k \geq \bar{k}_{bm}$

In this case, $r_s^*=\hat{c}$. We then use equations (A.12) and (A.13) to solve for r_f^* and x^* setting again $e^*=0$. We obtain

$$x^*=\frac{\hat{c}-q\beta-\frac{2\psi(1-q)}{q\nu(1-\alpha)}}{2-q\beta}<x^{OM} \tag{A.15}$$

$$r_f^*=\frac{\hat{c}-2(1-q)^{N-1}x^*-\frac{2\psi}{\nu q(1-\alpha)}}{1-(1-q)^{N-1}} \tag{A.16}$$

This concludes the proof.

A.10 Proof of Proposition 6

We first show the results related to the CCP compensation (Step 1). We then derive the optimal contract (Step 2).

Step 1. CCP compensation schedule

We first show that the CCP should only be compensated in state $d = 0$. Define the incentive power of a state $d \in \{0, 1, ...N\}$ as

$$IC(d) = 1 - \frac{\mathbb{P}[d|\text{ shirk}]}{\mathbb{P}[d|\text{ effort}]}$$

with $\mathbb{P}[d|\text{ a}]$ the probability of state d under action a. We have $\mathbb{P}[d|\text{ effort}] = \binom{N}{d}(1-q)^d q^{N-d)}$ while the term $\mathbb{P}[d|\text{ shirk}]$ depends on the number of investor pairs the CCP does not monitor. If it deviates by monitoring only $n_m \in [|0, N-1|]$ investors,

$$\mathbb{P}[d|\text{ shirk}] = \sum_{d_m=0}^{d} \binom{n_m}{d_m}\binom{N-n_m}{d-d_m}(1-q)^{d_m} q^{n_m-d_m}(1-\alpha q)^{d-d_m}(\alpha q)^{N-n_m-d+d_m}$$

After some manipulation, we obtain

$$\frac{\mathbb{P}[d|\text{ shirk}]}{\mathbb{P}[d|\text{ effort}]} = \frac{\sum_{d_m=0}^{d}\binom{n_m}{d_m}\binom{N-n_m}{d-d_m}\left[\frac{1-\alpha q}{\alpha(1-q)}\right]^{d-d_m}}{\binom{N}{d}} = \sum_{d_m=0}^{d} w_{n_m}(d_m)\left[\frac{1-\alpha q}{\alpha(1-q)}\right]^{d-d_m}$$

where $\sum_{d_m=0}^{d} w_{n_m}(d_m) = 1$ by Vandermonde's identity. Because $\frac{1-\alpha q}{\alpha(1-q)} > 1$, the ratio above is minimized by setting $d = 0$ and the minimum is strict. Hence, $IC(d)$ is maximized for $d = 0$.

We will now define $\underline{\pi}(0)$ as the incentive payment such that (MIC_{cm}) holds as an equality. It is defined by

$$Nq^N \underline{\pi}(0) - 2N\psi = \max_{n_m \in [|0,..N-1|]} \left\{Nq^N \alpha^{N-n_m}\underline{\pi}(0) - 2n_m\psi\right\} \tag{A.17}$$

where on the right-hand-side, n_m is the number of investor pairs the CCP monitors when it deviates. The relevant deviation, however, is to monitor no investor. To prove this statement, we need to show that the mapping $g : y \to y(1 - e^{y log(\alpha)})^{-1}$ is increasing with y for $y \geq 1$. We have

$$g'(y) \propto 1 - \alpha^y + y\alpha^y log(\alpha) \geq 1 - \alpha(1 - log(\alpha))$$

The inequality obtains because $y \geq 1$ and $\alpha \leq 1$. We thus have $g'(y) \geq 0$ because $\alpha \mapsto \alpha(1-log(\alpha))$ is increasing and $\lim_{\alpha\to 1}\alpha(1 - log(\alpha)) = 1$. Setting $n_m = 0$ on the right-hand side of (A.17), we find that $\underline{\pi}(0)$ is given by (11). With $\underline{\pi}(0)$, $\underline{e}$ given by (12) is the amount of capital such that (PC_{CCP}) binds.

Step 2. Optimal Contract

Observe first that the expected compensation to the CCP is a fixed cost. Hence, under Assumption 4, the complete loss mutualization contract of Proposition 3 is still optimal under unobservable monitoring. We thus have $r_s^* = r_f^* = \hat{c}$, and we are left to determine x^* and e^*.

Step 2.i) $e^* = \underline{e}$

Building on the proof of Proposition 3, we need to determine the marginal value of e on the investors' utility function when $r_s^* = r_f^* = \hat{c}$ and constraint (LP) binds. The key observation is that the CCP's participation constraint (PC_{CCP}) is slack for any $e \in [0, \underline{e}]$ when using the minimum compensation contract given by (11). When e is increased over $\underline{e}$, however, (PC_{CCP}) is tight, and any increase in CCP capital requires an increase in expected compensation by a factor ν_C. Using formulation (A.6) of constraint (LP), we obtain the following result

$$U'(e)_{|r_s^*=r_f^*=\hat{c},\,(\text{LP})binds} = \frac{\partial U}{\partial e} + \frac{\partial U}{\partial x}\frac{\partial x}{\partial e}$$
$$= \begin{cases} \frac{\nu-1}{2}(1-q)^N - [(\nu-1)(1-q)^N - k]\frac{1-(1-q)^N}{2-2(1-q)^N-\beta q} & \text{if} \quad e \leq \underline{e} \\ [\underline{k}_N - k]\frac{\nu_C - 1 + (1-q)^N}{2-q\beta-2(1-q)^N} & \text{if} \quad e > \underline{e} \end{cases}$$

Since $k > \underline{k}_N$, $U'(e) \geq 0$ if and only if $e \leq \underline{e}$. It follows that the optimal choice of CCP capital is $e^* = \underline{e}$. Note that $\frac{\partial x}{\partial e} < 0$, that is, the amount of collateral decreases with e for $e < \underline{e}$, as claimed in the main text.

We are thus left to determine the optimal collateral amount. To solve for x^*, we saturate the limited pledgeability constraint (LP) to obtain

$$\hat{c}\left[1-(1-q)^N\right] + (1-q)^N(2x^* + e^*) + \mathbb{E}[\pi^*] = q\beta + (2-q\beta)x^* + e^* \tag{A.18}$$

We obtain

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

Finally, we need to verify our conjecture that $2x^* + e^* \leq \hat{c}$. Using the first expression for x^* above, this inequality is equivalent to

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

The right-hand side is increasing with N. Hence, the condition above holds for all N if it holds for $N = 1$. This latter condition is implied by Assumption 2.

A.11 Proof of Proposition 7

We first compare centralized monitoring to no monitoring. To avoid confusion, we add a superscript cm to variables for the optimal centralized monitoring contract. For large N, Proposition A.1 shows that the *OM-contract* without monitoring is given by Case 2. This is because, the condition $k \leq \bar{k}$ in Assumption 4 implies $k \leq \bar{k}^{\cancel{m}}$, and the lower bound of the region $\underline{k}_N^{\cancel{m}}$ converges to 0 as N grows large. Using Proposition 6 and A.1, we derive the following expressions for investors' utility:

$$U^{*,cm} = qR + \Big[\nu - 1 - k\Big]\frac{\hat{c}}{2} + \Big[k - (\nu-1)(1-q)^N\Big]\left(\frac{\hat{c}}{2} - x^{*,cm}\right)$$

$$- \frac{1}{2}\Big[(\nu_{cm} - 1) - (\nu - 1)(1-q)^{N-1}\Big]e^* - \psi \tag{A.19}$$

$$U^{OM,nm} = qR + \Big[\nu - 1 - k\Big]\frac{\hat{c}}{2} + \Big[k - (\nu-1)(1-\alpha q)^N\Big]\left(\frac{\hat{c}}{2} - x^{OM,nm}\right) \tag{A.20}$$

From Proposition 6 and A.1 again, we have

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

$$\frac{\hat{c}}{2} - x^{OM,nm} = \frac{\beta\alpha q}{2\big[1-(1-\alpha q)^N\big] - \beta\alpha q}\left(1 - \frac{\hat{c}}{2}\right)$$

When $N \to \infty$, e^* converges to 0 at an exponential rate by Proposition 6. The second term of $\frac{\hat{c}}{2} - x^{cm,*}$ above also converges at an exponential rate as $N \to \infty$. In the limit, centralized monitoring dominates no monitoring, that is, $U^{*,cm} \geq U^{OM,nm}$ if and only if

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

Under Assumption 2,

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

Hence, the condition can be expressed as a lower bound $\hat{k}^{cm}$ on k with

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

We now turn to the comparison between centralized monitoring and bilateral monitoring. We first consider Case 1 of Proposition 5. In this case, investors' utility can be written as

$$U^* = qR + \Big[\nu - 1 - k\Big]\frac{\hat{c}}{2} + \Big[k - (\nu-1)(1-q)^N\Big]\left(\frac{\hat{c}}{2} - x^*\right) - \psi \tag{A.21}$$

Using equations (A.19) and (A.21), centralized monitoring dominates Case 1 of bilateral monitoring if and only if

$$\Big(k-(\nu-1)(1-q)^N\Big)\left(x^{cm,*}-x^{OM}\right)+\frac{1}{2}\Big((\nu_{cm}-1)-(\nu-1)(1-q)^N\Big)e^* \leq \Big(k-(\nu-1)(1-q)^N\Big)\left(x^*-x^{OM}\right)$$

Using the expression for the collateral requirement in (A.14), we obtain

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

We thus obtain the following condition

$$\frac{1}{2}\Big((\nu_{cm}-1) - (\nu-1)(1-q)^N\Big)e^* \le \left[k - (\nu-1)(1-q)^N\right](x^* - x^{*,cm}) \tag{A.22}$$

$$\frac{1}{2}\Big((\nu_{cm}-1) - (\nu-1)(1-q)^N\Big)e^* \le \frac{k-(\nu-1)(1-q)^N}{2(1-(1-q)^N) - \beta q}\left[\frac{2\psi}{1-\alpha} - \frac{2\psi}{1-\alpha^N} - \frac{vq(1-q)^{N-1}\Big(\beta q(2-\hat{c}) - \frac{4\psi}{1-\alpha}\Big)}{2\left[1-(1-q)^{N-1}(vq+1-q)\right] - \beta q}\right] \tag{A.23}$$

Observe that the terms which depend on N are exponential in N. 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, we turn to the comparison between centralized monitoring and Case 2 of Proposition 5 for bilateral monitoring. Centralized monitoring dominates if and only if

$$\Big(k-(\nu-1)(1-q)^N\Big)(x^{cm,*} - x^{OM}) + \frac{1}{2}\Big((\nu_{cm}-1)-(\nu-1)(1-q)^N\Big)e^* \le \left[\frac{\nu-1}{2}(2-q\beta) - k\right](x^{OM} - x^*)$$

Using equation (9) for x^{OM} and equation (A.15) for x^*, we obtain

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

We observe again that the terms which depend on N are exponential in N. Taking the limit $N \to \infty$, the condition for centralized monitoring to dominate Case 2 of bilateral monitoring becomes

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

This condition holds if and only if $k \le \bar{k}^{cm}$ with

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

Finally, we are left to derive the maximum value of the monitoring cost ψ such that the interval $[\hat{k}^{cm}, \bar{k}^{cm}]$ is non-empty. Observe that $\bar{k}^{cm}$ is independent of ψ while $\hat{k}^{cm}$ is strictly increasing with

ψ. Solving for the value of ψ such that $\hat{k}^{cm} = \bar{k}^{cm}$, we get

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

This is the first argument of the min in the expression for the upper bound on ψ given by Assumption 2. Hence for any $\psi < \bar{\psi}$, the interval $[\hat{k}^{cm}, \bar{k}]$ is non-empty.

A.12 Proof of Proposition 8

We first prove that a CCP would never pledge capital if it had the bargaining power. We then shown that a planner maximizing total surplus may choose a lower level of capital than investors.

The first result follows from our analysis of the *OM-contract* in Proposition 3 and the contracts with unobservable monitoring in Proposition 5 and 6. We showed that under Assumption 4 the net value of CCP capital to investors is negative when its cost is ν_C. Suppose then the CCP has the bargaining power and consider an allocation without CCP capital. For every unit it pledges, the CCP must earn an extra profit at least equal to ν_C which is above the investors' willingness to pay for capital. Hence, the CCP prefers not to pledge capital.

To prove the second result, consider the allocation in the proof of Proposition 6, indexed by the amount of capital $e \in [0, e^*]$ with e^* the investors' choice. By linearity, it is enough to compare the allocations with $e = 0$ and $e = e^*$. Let $U(e)$ denote the investor's utility as a function of $e \in [0, e^*]$,

$$
U(e) = qR + \frac{\nu-1-k}{2}\hat{c} - \frac{1}{2}\mathbb{E}[\pi^*] + \Big[k - (\nu-1)(1-q)^N\Big]\left(\frac{\hat{c}}{2} - x - \frac{e}{2}\right) + \frac{1}{2}\big[1+k\big]e
$$

where x is a function of e given implicitly by equation (A.18) replacing e^* with $e \in [0, e^*]$. With $e = e^*$ the CCP breaks even, while with $e = 0$, the CCP's profit is equal to $N\nu_C e^*$. Hence, for a planner maximizing total surplus, the allocation with $e = e^*$ dominates if and only if

$$
\begin{aligned}
& 0 \leq 2NU(e^*) - (2NU(e=0) + N\nu_C e^*_C) \\
\Leftrightarrow \quad & 0 \leq \Big[k - (\nu-1)(1-q)^N\Big]\left(x(e=0) - x^* - \frac{e^*}{2}\right) - \Big[\nu_C - 1 - k\Big]\frac{e^*_C}{2} \\
\Leftrightarrow \quad & 0 \leq \left\{\frac{k-(\nu-1)(1-q)^N}{\bar{k}-(\nu-1)(1-q)^N}\beta q(\nu-1) - \Big[\nu_C - 1 - k\Big]\right\}\frac{e^*_C}{2}
\end{aligned}
$$

When ν_C is high enough, this condition does not hold, which implies the planner's choice is $e = 0$. This is lower than the investors' choice who always prefer $e = e^*$.

Internet Appendix

B Contract with binding resource constraint

We relax Assumption 3 to analyze the situation in which the resource constraint (4) may bind. To clearly highlight the effect of this assumption, we focus on the parametric case in which CCP capital is not used, that is, we impose $\nu_C - 1 \geq k$. As shown in Proposition 3, this implies that $e = 0$ and $\pi(\cdot) = 0$. We further assume monitoring is costless ($\psi = 0$), which means it is optimal and (bilaterally) incentive-compatible. This implies we can set $r_s(d) = r_f(d)$ for all $d \in \{1, ..., N-1\}$ without loss of generality (see the discussion following Lemma 1).

We first define $\bar{r}_N(d, x)$ as the maximum receiver transfer given a collateral amount x and a state $d \in \{0, 1, ..N-1\}$. Using budget constraint (6) and the resource constraints (4)-(5), we have

$$\bar{r}_N(d, x) \equiv 2x + \frac{N-d}{N}(1-x)2R$$

Assumption 3 is equivalent to $\bar{r}_N(N-1, 0) \geq \hat{c}$. We also note that $\bar{r}_N(N-1, 0) \geq \hat{c}$ implies $\bar{r}_N(d, 0) \geq \hat{c}$ for all $d \in \{0, 1, ..N-1\}$ because $\bar{r}_N(d, x)$ is decreasing with d. When Assumption 3 does not hold, that is, when $\bar{r}_N(N-1, 0) < \hat{c}$, define $\hat{x}_N(N-1) \in \left(0, \frac{\hat{c}}{2}\right)$ such that $\bar{r}(N-1, \hat{x}_N(N-1)) = \hat{c}$. This threshold exists because $\bar{r}_N(d, x)$ is increasing with x and $\bar{r}_N(d, 1) = 2 > \hat{c}$ by Assumption 1.

Observe next that Assumption 3 is only sufficient for resource constraint (4) to be slack at the optimal contract of Proposition 3. In fact, in Cases 1 and 3, the resource constraint (4) holds even without Assumption 3. In Case 2, constraint (4) still holds for $d = N-1$ even when Assumption 3 fails if $\hat{x}_N(N-1) < x^{OM}$ with x^{OM} the optimal collateral requirement in (9). Hence, our analysis will only differ from that in the main text if both Assumption 3 and this latter condition are relaxed.

In what follows, we consider the case $N = 3$, which is the smallest value of N such that the resource constraint may bind at the optimal contract. We thus impose $\bar{r}_3(2, 0) < \hat{c}$ and $\hat{x}_3(2) \geq x^{OM}_{|N=3}$ which can be written in a compact form as

$$R < \frac{3}{2} \min \left\{ \hat{c}, \frac{\beta q}{1-(1-q)^3} \right\} \tag{A3n}$$

We now derive the optimal contract for $N = 3$ under (A3n). The possibility that resource constraint (4) binds has two effects. First, as the maximum receiver transfer $\bar{r}_3(2, x)$ increases with x, collateral has an additional hedging value in the state of the world with two payers defaults. By the pledgeability constraint, however, if transfers from payers are reduced due to a lack of resources, less collateral is needed for incentives. The result below shows how these two effects interact.

Proposition B.1. *Let $N = 3$, $\psi = 0$ and $\nu_C - 1 > k$. Under Assumption (A3n), there exists a threshold*

$$\underline{k}_3(2) = \underline{k}_3 + (\nu - 1)q(1-q)^2(3-R) \in (\underline{k}_3, \bar{k})$$

such that the optimal contract is

1. *the contract of Proposition 3 if* $k < \underline{k}_3$ *or* $k > \bar{k}$*,*

2. *if* $k \in [\underline{k}_3, \underline{k}_3(2)]$*, the optimal amount of collateral is given by* $\tilde{x}^{OM} = \hat{x}_3(2) > x^{OM}$*, and if* $k \in [\underline{k}_3(2), \bar{k}]$*, it is given by*

$$\tilde{x}^{OM} = \frac{\left[q^3 + 3q^2(1-q)\right]\hat{c} - q\beta + 2q(1-q)^2 R}{2\left[1-(1-q)^3\right] - 2q(1-q)^2(3-R) - q\beta} < x^{OM} \tag{B.1}$$

Proof. As explained above, the resource constraint in state $d = 2$ may only bind in Case 2 of Proposition 3. Hence, the optimal contract is the same as in Proposition 3 for $k \notin [\underline{k}_3, \bar{k}]$.

For the case $k \in (\underline{k}_3, \bar{k})$, we need to determine the collateral amount x_{IC} such that constraint (LP) binds. By construction, under condition (A3n), this level satisfies $x_{IC} < \hat{x}_3(2)$. Building on the argument in Proposition 3, it is optimal to set the receiver transfer to its maximum value when the pledgeability constraint (LP) is slack. Hence, we can determine x_{IC} by saturating (LP) and setting $r(0) = r(1) = \hat{c}$, $r(2) = \bar{r}_3(2,x)$ and $r(3) = 2x$. Using budget constraint (6), we obtain

$$\mathbb{E}[r_o(d)] = \left[q^3 + 3(1-q)q^2\right]\hat{c} + 3q(1-q)^2\bar{r}_3(2,x) + (1-q)^3 2x = x(2-q\beta) + q\beta \tag{B.2}$$

Solving for x in (B.2), we find x_{IC} as given by equation (B.1). The inequality $x_{IC} < x^{OM}$ obtains because the proof of Proposition 3 shows that x_{IC} solves the same equation as x^{OM} substituting $\bar{r}_3(2,x)$ for $\hat{c} > \bar{r}_3(2,x)$.

The optimal amount of collateral $\tilde{x}^{OM}$ when $k \in [\underline{k}_3, \bar{k}]$ is given either by x_{IC} or $\hat{x}_3(2)$ because the marginal value of collateral is piecewise constant, and it jumps only at these points. Totally differentiating (3) with respect to x, we obtain

$$U'(x) = \begin{cases} (\nu - 1)\Big[(1-q)^3 + q(1-q)^2(3-R)\Big] - k & \text{if} \quad x \in [x_{IC}, \hat{x}_2(3)] \\ \underline{k}_2 - k & \text{if} \quad x \in [\hat{x}_2(3), \frac{\hat{c}}{2}] \end{cases}$$

To obtain the derivative $\frac{\partial \mathbb{E}[r_o(d)]}{\partial x}$ for the first expression, we use the middle term of equation (B.2). By definition of $\underline{k}_3(2)$, this first term is equal to $\underline{k}_3(2) - k$. Hence, as stated in the result, $\tilde{x}^{OM} = \hat{x}_2(3)$ is optimal when $k \in [\underline{k}_3, \underline{k}_3(2)]$ while $\tilde{x}^{OM} = x_{IC}$ is optimal when $k \in [\underline{k}_3(2), \bar{k}]$

□

Case 2 of Proposition B.1 shows the effect of the resource constraint on the optimal contract. When Assumption 3 does not hold, a single payer cannot cover the hedging needs of three receivers if no collateral is pledged. Hence, collateral has a hedging value in the states where all 3 payers default and 2 out of 3 payers default. By contrast, when Assumption 3 holds, this insurance value is only enjoyed in the worst default state. This explains why investors optimally post more collateral than in the optimal contract of Proposition 3 when collateral is relatively cheap.

When the collateral cost is higher, however, that is when $k \in [\underline{k}_2(3, \bar{k}]$ investors post less collateral than in the benchmark. If collateral is expensive, investors forgo this hedging value(s) of collateral. The collateral requirement is then determined by the pledgeability constraint. Since payers' transfers are lower when the resource constraint binds, less collateral is needed.

C Optimal Monitor with Full Hedging

In this section, we analyze the optimal monitoring scheme when monitoring is unobservable and the *OM-contract* features full hedging with CCP capital, that is when $k \in [\hat{k}^m, \bar{k}_N]$. In this case, as shown by Lemma 2, the *OM-contract* is not incentive-compatible under bilateral monitoring.

The analysis proceeds in three steps, as in the main text. First, we derive the optimal contract under bilateral monitoring in Section C.1. Section C.2 then characterizes the optimal contract under centralized monitoring. Finally, we compare these two schemes in Section C.3 to show that centralized monitoring dominates.

C.1 Bilateral Monitoring

Proposition C.1 (Optimal contract under bilateral monitoring)**.** *When $k \in [\hat{k}^m, \bar{k}_N]$, there exists a threshold $\underline{k}_N^{bm}$ such that the optimal contract with incentive-compatible bilateral monitoring is*

1. *if $k \in [\hat{k}^m, \underline{k}_N^{bm}]$, $r_s^* = \hat{c} + \frac{2\psi}{q(1-\alpha)}$, $r_f^* = \hat{c}$, $e^* = e^{OM} - \frac{\frac{4\psi}{1-\alpha}}{2\nu_C - q\beta}$ and $x^* = x^{OM} + \frac{\frac{2\psi}{1-\alpha}}{2\nu_C - q\beta}$;*
2. *if $k \in [\underline{k}_N^{bm}, \underline{k}_N^{bm}]$, the loss mutualization contract of Proposition 5.*

Proof. The proof is in two Steps. First, we show that when monitoring is unobservable, the frontier between the full-hedging contract and the loss mutualization contract shifts from $\underline{k}_N$ to $\underline{k}_N^{bm}$. This explains the second case of Proposition C.1. Second, we characterize the full-hedging contract of case 1 under unobservable monitoring.

Step 1. Threshold $\underline{k}_N^{bm}$

By arguments similar to that of Proposition 3 and 5, we can establish that both constraints (LP) and (MIC_{bm}) must bind. Like $\underline{k}_N$ in Proposition 3, the threshold $\underline{k}_N^{bm}$ is the value of k such that the total marginal effect of CCP capital on investors' utility is equal to 0 when $r_f = \hat{c}$. Consider then a perturbation Δe. Equations (A.12) and (A.13) imply that the following relationships must hold for (LP) and (MIC_{bm}) to hold:

$$q\Delta r_s + (1-q)^N (2\Delta x + \Delta e) = (2 - q\beta)\Delta x - (\nu_C - 1)\Delta e,$$
$$\Delta r_s + \nu(1-q)^{N-1} (2\Delta x + \Delta e) = 0$$

We thus obtain the following relationship between Δx and Δe

$$\Delta e = \frac{2 - q\beta - 2(1-q)^{N-1}\left[\nu q + 1 - q\right]}{\nu_C - 1 + (1-q)^{N-1}\left[\nu q + 1 - q\right]}\Delta x \tag{C.1}$$

Hence, the total derivative of U with respect to e is given by

$$U'(e) = \frac{1}{2}\Big[(\nu - 1)(1-q)^N - (\nu_C - 1)\Big] + \Big[(\nu - 1)(1-q)^N - k\Big]\frac{\Delta x}{\Delta e}$$

which is positive if and only if

$$k \leq \underline{k}_N^{bm} \equiv (\nu-1)(1-q)^N + \frac{1}{2}\frac{2-q\beta-2(1-q)^{N-1}\big[\nu q+1-q\big]}{\nu_C-1+(1-q)^{N-1}\big[\nu q+1-q\big]} \max\left\{\Big[(\nu-1)(1-q)^N-(\nu_C-1)\Big], 0\right\}$$

The claim $\underline{k}_N^{bm} \leq \underline{k}_N$ follows from the observation that the ration $\frac{\Delta x}{\Delta e}$ in equation (C.1) is higher that that in equation (A.7)under observable monitoring. As one unit of CCP capital requires more collateral when monitoring is unobservable, the maximum collateral cost for which CCP capital is profitable must be lower.

Step 2. Optimal contract

When $k > \underline{k}_N^{bm}$, it is optimal to set $e^* = 0$ and the optimal contract is thus given by Proposition 5. When $k \leq \underline{k}_N^{bm}$, the optimal contract features full hedging. It is obtained by setting $r_s = r_f = \hat{c}$ and $2x+e = \hat{c}$ in constraints (A.12) and (A.13). From the second equation, we obtain the expression for r_s^*. Solving for e^* and x^*, we obtain the expression in case 1 of Proposition C.1. Finally, we need to verify $e^* \geq 0$. This requires

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

which follows from Assumption 2. This concludes the proof.

□

C.2 Centralized Monitoring

Proposition C.2 (Centralized monitoring contract)**.** *When $k \in [\hat{k}^m, \underline{k}_N]$, an optimal contract with centralized monitoring is a full-hedging contract with $x^* > x^{OM}$, $e^* = \hat{c} - 2x^* \in (\underline{e}, e^{OM})$, and any $\{\pi^*(d)\}$ such that (PC_{CCP}) binds and (MIC_{cm}) holds.*

Proof. By the same argument as that in the proof of Proposition 6, the full hedging remains optimal with unobservable monitoring in this parameter region. This is again because the compensation cost to the CCP is a fixed cost which affects each contract symmetrically.

We show in Proposition 6 that investors achieve less than full hedging with the minimum incentive compatible CCP capital $\underline{e}$. Hence, given full-hedging is desirable when $k \in [\hat{k}^m, \underline{k}_N]$, the optimal level of CCP capital satisfies $e^* > \underline{e}$. The amount of CCP capital and collateral are pinned down by the condition $r_f^*(N) = 2x^* + e^* = \hat{c}$ and the binding pledgeability constraint (LP)

$$\hat{c} + 2\psi = q\beta + (2-q\beta)x^* - (\nu_C - 1)e^*$$

We obtain

$$x^* = \frac{\nu_C \hat{c} - q\beta + 2\psi}{2\nu_C - q\beta} > x^{OM} \tag{C.2}$$

$$e^* = \frac{q\beta(2 - \hat{c}) - 4\psi}{2\nu_C - q\beta} < e^{OM} \tag{C.3}$$

We are left to characterize the compensation of the CCP. Let $\pi^*(d) = 0$ for $d > 0$ and set $\pi^*(0)$ such that (PC_{CCP}) binds with $e = e^*$, that is

$$\pi^*(0) = \frac{2\psi + \nu_C e^*}{q^N} > \frac{2\psi + \nu_C e}{q^N} = \pi(0) \tag{C.4}$$

where e and $\pi(0)$ are here the expressions derived in Proposition 6. The inequality follows from the result above that more capital is used in this case than in the contract of Proposition 6. Equation (C.4) shows that (MIC_{cm}) is slack with the contract such that $\pi^*(d) = 0$ for $d > 0$ and $\pi^*(0)$ given by (C.4). Hence, since it saturates (PC_{CCP}), this contract is weakly optimal. The slack in (MIC_{cm}) implies that there exist other optimal contracts with $\pi(d) > 0$ for $d > 0$. □

C.3 Optimal Monitor

We are now equipped to characterize the optimal monitor when $k \in [\hat{k}^m, \underline{k}_N]$, that is, when the *OM-contract* of Proposition 3 under observable monitoring features full hedging with CCP capital.

Proposition C.3. *When the optimal contract features full hedging under both monitoring schemes, that is, when $k \in [\hat{k}^m, \underline{k}_N^{bm}]$, centralized monitoring is optimal.*

Proof. Investors realize the same hedging benefits under these monitoring schemes. Hence, the optimal scheme is that which minimizes the combined cost of collateral and CCP capital, given by $xk + e(\nu_C - 1)$. Furthermore, because they feature full hedging, each contract satisfies $2x + e = \hat{c}$. Hence, because $\nu_C - 1 \leq k$ when $k \in [\hat{k}^m, \underline{k}_N^{bm}]$, the best contract is that which uses more CCP capital. The result the follows from the comparison between equation (C.3) and its counterpart in Proposition C.1. □

Proposition C.3 shows the CCP is always the efficient monitor when investors desire full hedging. This result complements and strengthens our finding about optimality of centralized monitoring in the main text. The intuition is as follows. With full hedging, investors have no exposure to counterparty risk. Hence, in order to monitor, they must receive an incentive payment equal to the full agency rent from monitoring $\frac{2\psi}{1-\alpha}$. This incentive payment materializes as an extra collateral cost as the payers' liabilities increase. Under centralized monitoring, the agency rent is given to the CCP. Because the CCP pledges capital, however, investors can recoup part of this rent and lower their collateral requirement under centralized monitoring. This explains the result.

D Additional Proofs

D.1 Proof of Corollary 1 with optimal monitoring

We prove Corollary 1 accounting for the optimal monitoring decision analyzed in Section A.7 below. This proves our claim in the main text that the comparative statics with respect to N in Corollary 1 remains valid in this case.

The upper bound of the essential CCP region is again given by $\bar{k}$. For $k > \bar{k}$, monitoring is optimal as shown in Section A.7.2, and the optimal contract without monitoring can be implemented bilaterally. For k lower than but close to $\bar{k}$, monitoring and loss mutualization are optimal, which means the upper bound is $\bar{k}$. This observation also implies there exists a lower bound $\underline{k}^{ess,m} < \bar{k}$ of the essential CCP region.

By Proposition 3 and A.1, we have $\underline{k}^{ess,m} \geq \underline{k}^{ess}$ because the region with full hedging and without CCP capital in which a CCP is not essential is larger without monitoring. We now characterize the threshold $\underline{k}^{ess,m}$ by considering three different cases. Statements about optimality of contracts below are always conditional on $k \in [\underline{k}^{ess}, \bar{k}]$.

Case 1. $(\nu - 1)(1-q)^N > \nu_C - 1$

Proposition 3 shows that the optimal contract with monitoring features either CCP capital or complete loss mutualization Because $1 - \alpha q > 1 - q$, the same result holds for the contract without monitoring by Proposition A.1. Hence, in this case, a CCP is always essential and $\underline{k}^{ess,m} = \nu_C - 1$.

Case 2. $\nu_C - 1 \in [(\nu - 1)(1-q)^N, (\nu - 1)(1-\alpha q)^N]$

In this region, complete loss mutualization is optimal with monitoring by Proposition 3. Without monitoring, CCP capital or complete loss mutualization is optimal for $k \geq \nu_C - 1$ and full hedging with only collateral is optimal otherwise. This leaves two possibilities: either $\underline{k}^{ess,m} = \nu_C - 1$ or if $\hat{k}^m < \nu_C - 1$ then $\underline{k}^{ess,m} = \hat{k}^m$. In the latter case, $\hat{k}^m$ is the value of the collateral cost such that investors are indifferent between the complete LM contract with monitoring and the full-hedging contract without monitoring. Hence, $\hat{k}^m$ solves

$$\begin{aligned}
0 &= U_{k=\hat{k}^m} - U^{\not{m}}_{|k=\hat{k}^m} \\
&= qR + \left[\nu - 1 - \hat{k}^m\right]\frac{\hat{c}}{2} - (\hat{k}^m - \underline{k}_N)\left(\frac{\hat{c}}{2} - x^{OM}\right) - \psi - \left\{qR + \left[\nu - 1 - \hat{k}^m\right]\frac{\hat{c}}{2}\right\} \\
&= \beta q \left(1 - \frac{\hat{c}}{2}\right) \frac{\hat{k}^m - \underline{k}_N}{2\left[1 - (1-q)^N\right] - \beta q} - \psi
\end{aligned} \tag{D.1}$$

Case 3. $\nu_C - 1 \geq (\nu - 1)(1 - \alpha q)^N$

Proposition 3 shows that complete loss mutualization is optimal with monitoring. Without monitoring the optimal contract features complete loss mutualization for $k \geq (\nu - 1)(1 - \alpha q)^N$, and full hedging with collateral only otherwise. Hence, by an argument similar to that in Case 2, $\underline{k}^{ess,m} = \min\{\hat{k}^m, (\nu - 1)(1-\alpha q)^N\}$.

Monotonicity of $\underline{k}^{ess,m}$

Combining the results above, we obtain

$$\underline{k}^{ess,m} = \min\left\{\nu_C - 1, \hat{k}^m, (\nu-1)(1-\alpha q)^N\right\}, \quad \text{with} \quad \hat{k}^m \quad \text{given by} \quad (\text{D.1})$$

The third argument of the min strictly decreases with N. We also show that $\hat{k}^m$ strictly decreases with N when it is given by (D.1). For this, define $g : (y,k) \mapsto \frac{k+y(\nu-1)}{2+2y-\beta q}$ and apply the Implicit Function Theorem to equation (D.1). We obtain

$$\frac{\partial k}{\partial N} = -\frac{\frac{\partial g}{\partial y}\frac{\partial \bar{y}}{\partial N}}{\frac{\partial g}{\partial k}}$$

with $y = -(1-q)^N$. As $\frac{\partial g}{\partial k} > 0$ and $\frac{\partial \bar{y}}{\partial N} > 0$, the derivative is negative if and only if

$$0 < \frac{\partial g}{\partial y} \quad \Leftrightarrow \quad 0 < \frac{(\nu-1)(2-\beta q) - 2k}{[2+2y-\beta q]^2} = \frac{2(\bar{k}-k)}{[2+2y-\beta q]^2}$$

The last inequality holds because by Proposition 4, $\hat{k}^m$ lies below $\bar{k}$.

When $\underline{k}^{ess,m}$ is equal to the second or third argument, it strictly decreases with N. Besides, for N large enough, $\underline{k}^{ess,m}$ is equal to the second or third argument as $\lim_{N\to\infty}(\nu-1)(1-\alpha q)^N = 0$. This proves that the result in Corollary 1 is robust when investors can choose whether to monitor.

References

ABN-AMRO, Allianze, B. B., 2020, "A path forward for CCP Resilience, Recovery, and Resolution," Discussion paper, ABN-AMRO.

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, forthcoming, "Transparency and collateral: central versus bilateral clearing," Theoretical Economics, p. to appear.

Armakolla, A., and B. Bianchi, 2017, "The European central counterparty (CCP) ecosystem," in Bank for International Settlements (ed.), Data needs and Statistics compilation for macroprudential analysis . , vol. 46, Bank for International Settlements.

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.

Cerasi, V., and S. Daltung, 2000, "The optimal size of a bank: Costs and benefits of diversification," European Economic Review, 44, 1701 – 1726.

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.

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.

FSB, 2020, "Guidance on Financial Resources to Support CCP Resolution and on the Treatment of CCP Equity in Resolution," Discussion paper, Financial Stability Board.

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.

Huang, W., and H. Zhu, 2021, "CCP Auction Design," Discussion paper, Bank of International Settlements.

ICE, 2019, "Central Clearing: Proven, Transparent, Regulated Means of Reducing Systemic Risk," Discussion paper, ICE.

ISDA, 2019, "CCP Best Practices," .

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

Koeppl, T., and C. Monnet, 2010, "The emergence and future of central counterparties," Discussion paper.

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

Laux, C., 2001, "Limited-liability and incentive contracting with multiple projects," Rand Journal of Economics, 32, 514–526.

LCH, 2015, "CCP Conundrums," Discussion paper, LCH.

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, 2021, "The Economics of Central Clearing," Annual Review of Financial Economics.

MRAC, 2021, "DCO Capital and Skin in the Game," Discussion paper, CCP Risk and Governance Subcommittee, Market Risk Advisory Committee of the U.S. Commodity Futures Trading Commission.

OCC, 2020, "Optimizing Incentives, Resilience and Stability in Central Counterparty Clearing," Discussion paper, OCC.

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

Pirrong, C., 2011, "The economics of central clearing: theory and practice," Discussion paper, ISDA.

Santos, T., and J. A. Scheinkman, 2001, "Competition among exchanges," The Quarterly Journal of Economics, 116, 1027–1061.

Spatt, C. S., 2017, "Designing Reliable Clearing and Settlement Systems for Enhancing the Transparency and Stability of Derivatives Markets," Discussion paper, Carnegie Mellon University.

Stiglitz, J. E., 1974, "Incentives and Risk Sharing in Sharecropping," The Review of Economic Studies, 41, 219–255.

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

Vuillemey, G., 2020, "The Value of Central Clearing," The Journal of Finance, 75, 2021–2053.

Wang, J. J., A. Capponi, and H. Zhang, forthcoming, "Central Counterparty and Collateral Requirements," Management Science, p. to appear.

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.