DISCUSSION PAPER SERIES

DP17679

THE DESIGN OF A CENTRAL COUNTERPARTY

John Chi-Fong Kuong and Vincent Maurin

BANKING AND CORPORATE FINANCE

ISSN 0265-8003

THE DESIGN OF A CENTRAL COUNTERPARTY

John Chi-Fong Kuong and Vincent Maurin

Discussion Paper DP17679
Published 17 November 2022
Submitted 14 November 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

THE DESIGN OF A CENTRAL COUNTERPARTY

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 mutualizing losses among its members. This protection, however, weakens members' incentives for risk management. Delegating members' risk monitoring to the CCP alleviates this tension in large markets. To discipline the CCP at minimum cost, members offer the CCP a junior tranche and demand capital contribution. Our results endogenize key layers of the default waterfall and deliver novel predictions on its composition, collateral requirements, and CCP ownership structure.

JEL Classification: D86, G23, G28

Keywords: Collateral, Monitoring

John Chi-Fong Kuong - john.kuong@insead.edu
European Institute of Business Administration (INSEAD) and CEPR

Vincent Maurin - Vincent.Maurin@hhs.se
Stockholm School of Economics and Swedish House of Finance

The Design of a Central Counterparty

John Kuong and Vincent Maurin*

This version: November, 2022

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 mutualizing losses among its members. This protection, however, weakens members' incentives for risk management. Delegating members' risk monitoring to the CCP alleviates this tension in large markets. To discipline the CCP at minimum cost, members offer the CCP a junior tranche and demand capital contribution. Our results endogenize key layers of the default waterfall and deliver novel predictions on its composition, collateral requirements, and CCP ownership structure.

*Kuong, john.kuong@insead.edu, INSEAD and CEPR; Maurin (corresponding author), vincent.maurin@hhs.se, Stockholm School of Economics. We thank an anonymous referee, Jean-Edouard Colliard, Thierry Foucault (the editor), Thomas Gherig, David Murphy, Thomas Poulsen, Roberto Ricco, Anatoli Segura, Chester Spatt, Ernst-Ludwig von Thadden, Guillaume Vuillemey, Yan Xiong, Kostas Zachariadis, Haoxiang Zhu; and seminar and conference participants at BI Oslo, CMU Tepper, ICEF HSE, INSEAD, Stockholm School of Economics, UC3M, HKIMR, Vienna, the Asia-Pacific Corporate Finance Online Workshop, Australasian Finance and Banking Conference, CICF, the FT Webinar, the Gerzensee Workshop on Money, Payments, Banking and Finance, the Kelley Junior Finance Conference, the Microstructure Exchange, the NFN Young Scholars Webinar Series, OxFIT, Paris December Finance Meeting, the SEC CFMR Conference, the SGF Conference and Stern Salomon Microstructure Conference for useful comments. Kuong acknowledges the financial support from INSEAD and HKIMR. Maurin thanks the Swedish House of Finance for financial support.

I. Introduction

To tame counterparty risks in over-the-counter (OTC) markets, regulators mandated clearing of many OTC contracts via Central Counterparties (CCPs) following the 2008 global financial crisis.[1] For instance, the fraction of centrally cleared interest rate derivatives rose to 60% in 2018 from 15% in 2009 (FSB (2018)).[2] CCPs manage counterparty risks by setting collateral requirements, monitoring clearing member's financial soundness and mutualizing losses by maintaining a default fund. This mutualization process allocates losses imposed by a defaulting member to the CCP itself and to other members according to a pre-specified "default waterfall". Regulators view the waterfall design as critical to financial stability (Yellen (2013), FSB (2020)). However, CCPs and their members often disagree about the size and priority of their contributions to the default fund (CCP12 (2021), ABN-AMRO, Allianz, Barclays, and BlackRock (2020)).

In this paper, we propose a model to analyze these design aspects of CCPs. Risk-averse investors match in pairs to trade, subject to idiosyncratic counterparty default risk. Due to default risk, investors' transfers are credible only if they are sufficiently backed by cash collateral and investors are creditworthy. Investors' creditworthiness improves if their counterparty exerts due diligence to ascertain their financial soundness, which requires a monitoring effort.

Central clearing via a CCP can then add value in two ways. First, the CCP can mutualize losses between investors (or members) due to idiosyncratic counterparty risk. The CCP

[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%.

then channels contributions from itself and solvent members to members with defaulted counterparties, like in a default fund in practice. Second, members can delegate counterparty monitoring to the CCP.[3] Our key innovation is to model central clearing as a multilateral contracting problem in which investors collectively act as the *principal* and the CCP as an *agent*. Our analysis of the optimal contract sheds light on the loss mutualization mechanism and the design of CCPs' incentives.

With these basic ingredients, we achieve three main results. First, we compare central clearing with bilateral trading where transfers can only occur between paired investors. We find that central clearing dominates only when the cost of collateral is intermediate and market size is large. Second, under similar conditions, it is efficient to delegate all monitoring tasks to a CCP. Third, such a CCP holds a junior equity tranche in the default waterfall to align its incentives, and contributes capital at members' request. The equilibrium level of CCP capital is an outcome of bargaining between the CCP and its members, not necessarily a measure of CCP's incentive. Overall, our results have implications for the design of the default waterfall, the determinants of CCP capital, and the CCP ownership structure.

Our results arise due to two fundamental frictions. The first one is a moral hazard problem that limits the pledgeability of investors' future cash flows. As in Biais, Heider, and Hoerova (2016), investors would shirk for private benefits and default if they expect to make a large payment to other investors and the CCP. The shirking metaphor is meant to capture investors' under-investment in risk management or risk-shifting behavior that would

[3] A CCP could also use its capital to provide insurance. Our analysis shows, however, that collateral dominates capital as an insurance tool, unless capital is cheaper than collateral. This restrictive condition implies that CCPs' role as insurance providers is limited. Our finding resonates with the view expressed by regulators and CCPs that CCPs should primarily pool risk, not insure it (Coeuré (2015), LCH (2015)).

expose their counterparties to "wrong-way risk".[4] Investors can increase their payment capacity 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 requires a costly and unobservable effort. This is the second friction, which implies that monitoring needs to be incentivized.

To clearly analyze the effects of each friction in our model, we proceed in three steps. We first study the frictionless benchmark in which investors' asset is fully pledgeable. Then, we add the limited pledgeability friction, and finally, the friction of unobservable monitoring.

In the frictionless benchmark, investors use either collateral or loss mutualization via the CCP to mitigate counterparty risk. If collateral is cheap, investors pledge enough collateral to fully eliminate counterparty risk. Otherwise, investors do not pledge collateral and mutualize losses via the CCP. When few members are solvent, investors remain exposed to some counterparty risk when they mutualize losses.

When the limited pledgeability friction is introduced, the CCP instead needs collateral to implement loss mutualization. Limited pledgeability constrains investors' payment capacity. Under loss mutualization, investors expect to pay to the default fund when others default. Hence, they have to pledge collateral ex ante to expand their payment capacity.

Our first main result is that loss mutualization is useful, or, central clearing strictly dominates bilateral trading, only when the cost of collateral is intermediate. If collateral is expensive, using collateral to support loss mutualization is too costly and investors voluntarily remain exposed to bilateral counterparty risk. If instead collateral is cheap, full hedging

[4]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).

with collateral, which can be done bilaterally, is efficient. In addition, when there are more investors to mutualize losses, the benefits of central clearing increase.

When the second friction of unobservable monitoring is added, loss mutualization undermines investors' incentives to monitor the counterparty they matched with because it reduces their exposure to this counterparty. Delegating all monitoring tasks to a CCP resolves this tension but is costly for two reasons. First, investors' compensation to the CCP for its monitoring service must be backed by collateral due to limited pledgeability. Second, as monitoring is unobservable, investors must pay an agency rent over and above the CCP's effort costs.

When central clearing is optimal, investors delegate monitoring (tasks) to the CCP if collateral is cheap enough and the market is large. Cheaper collateral lowers the cost of hiring a CCP, and a large market favors centralizing monitoring due to economies of scale: The agency rent decreases in the number of investors monitored, as in Diamond (1984). CCPs' role as centralized monitors, our second main result, rationalizes member monitoring as a key CCPs' defense against counterparty risks in practice, along with collateral requirements.

Figure 1 summarizes our results so far by showing the possible roles of a CCP in the optimal contract against the cost of collateral. Central clearing can help investors mutualize losses and the CCP agent can also play an active role by monitoring investors.

[Insert Figure 1 approximately here]

The analysis of centralized monitoring delivers our third main result, which characterizes the compensation and capital contribution of the CCP. Investors pay the CCP only when no member defaults because such high-powered compensation minimizes the agency rent. 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) because the CCP loses it if any member defaults. Our results thus rationalize several key common features of the default waterfall of CCPs as observed in practice (see Section VI.B).

Our main implications regard CCP SITG, a topic of intense debate for practitioners and regulators. Large institutional investors who are clearing members often request more "meaningful" capital contribution from CCPs to align incentives for risk management (ABN-AMRO et al. (2020)). CCPs, meanwhile, resist these calls and argue that members should absorb the bulk of the losses caused by a defaulting member (LCH (2015), CCP12 (2021)). We argue that small SITG observed in practice needs not imply that CCPs lack incentives to manage risks, as incentives also come from their equity-like compensation.[5]

Our analysis gives novel predictions about the size of SITG and hence the composition of the default waterfall. Empirically, the ratio of CCP capital to total pre-funded resources—CCP capital plus members' collateral—varies substantially across CCPs (Paddrik and Zhang (2020)). We show that SITG relative to either total pre-funded resources or CCP profit decreases with the number of members, due to the decline in CCP's agency rent. This effect is compounded if larger CCPs have more bargaining power, thus resisting members' demand for capital contribution. This observation can explain the tension between members and CCPs about the desirable size of capital.

Finally, our analysis points to a new force shaping the optimal CCP ownership structure. Under centralized monitoring, the CCP is a third-party agent compensated by the members.

[5]Our novel implication resonates with McPartland (2021) who argues that no capital is needed for CCP's incentive purpose because executives of CCPs, who receive stocks and options in their compensation, will suffer tremendous personal losses when a member is in distress.

Under bilateral monitoring, however, the CCP merely channels transfers between members—an arrangement we interpret as a member-owned CCP. Hence, a large (small) market favors third-party (member-owned) CCPs.

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 et al. (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 (2022). Our analysis of member-owned CCPs thus broadly shares some of their conclusions. Our key innovation is to consider the CCP as an agent, rather than a mechanism designer. This feature allows us to endogenize the default waterfall of CCPs (including SITG), the CCP's compensation, and the optimal ownership structure of CCPs. 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, Capponi, and Zhang (2022) 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.

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

[7] Koeppl (2013) studies collusive moral hazard in central clearing and Palazzo (2016) argues that central clearing may foster peer monitoring for previously non-connected investors.

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.[8] We highlight that the CCP's capital contribution decision is an outcome of bargaining with its members while its incentives can be properly aligned with the junior tranche in the default waterfall. 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. Leitner (2011), Zawadowski (2013), and Acharya and Bisin (2014) argue that central clearing can reduce counterparty risk externalities. Koeppl, Monnet, and Temzelides (2012) show that a CCP can lower trading costs by deferring

[8] In a similar vein, Capponi and Cheng (2018) consider a CCP's trade-off between clearing volume and stability but focus on collateral requirements rather than CCP capital.

settlement and providing credit to clearing members.

The rest of the paper is organized as follows. Section II presents the model. Section III maps our general contracting approach to centrally cleared contracts in practice. In Section IV, we analyze the costs and benefits of central clearing by deriving the optimal contract when monitoring is observable. Section V analyzes the full problem when monitoring needs to be incentivized and compares bilateral monitoring with centralized monitoring. We gather practical implications of our model for CCP design in Section VI. Section VII concludes. All proofs are in Appendix A.

II. A Model of Central Clearing

A. The Framework

There are two dates $t \in \{0, 1\}$. At date 1, there are two equiprobable aggregate states of the world, A and B. We denote S a generic element of $\{A, B\}$ and S' the other element. The economy is populated by investors and a CCP agent. 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. Each S-investor is endowed with one unit of a nontradable 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 2. The success or failure of the asset is independent across S-investors, conditional on the realization of state S'. Both state S and investors' asset success or failure are observable.[9]

[9]Having N pairs of investors, rather than a continuum, allows us to vary the size of the CCP. As we shall show, the size of the CCP is a key determinant of various economic forces in our model, namely, the benefits of loss mutualization, the investors' incentives to monitor each other bilaterally, and the economies of scale

[Insert Figure 2 approximately here]

Investors from a different group can gain from transferring consumption across states because an S-investor prefers to consume in aggregate state S in which only S'-investors have positive asset payoffs. Specifically, an S-investor's utility function is:

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

where $c(S)$ is consumption in state S, $\nu > 1$, and $\hat{c} > 0$. Variable $c(S)$ is random due to the randomness in S'-investors' asset cash flows. In words, S-investors derive extra marginal utility $(\nu - 1)$ from consumption in state S until their consumption reaches $\hat{c}$. Investors' preferences exhibit risk aversion because their marginal utility differs across aggregate states.[10] Meanwhile, $(\nu - 1)$ captures the gains from hedging and $\hat{c}$ the hedging demand.

To show how gains from trades can be achieved, consider a numerical example with $\{q, R, \hat{c}, \nu\} = \{0.8, 2, 1.8, 1.2\}$. In autarky, each investor's utility is given by the expected cash flow of its asset, $qR = 1.6$. Suppose each S-investor matches with a S'-investor and promises to pay $\hat{c} = 1.8$ when its asset succeeds in state S' while the S'-investor promises 1.8 in state S. Expected gains from trade 0.18 are realized because each investor now enjoys

$$(2) \qquad U = \frac{1}{2}\left[q(R - 1.8)\right] + \frac{1}{2}q\left[1.8 + (\nu - 1)1.8\right] = qR + \frac{1}{2} \times 0.2 \times 1.8 = 1.78.$$

in centralized monitoring by the CCP. Furthermore, comparative statics analyses with respect to N deliver novel empirical predictions regarding CCP capital contribution and ownership structure. Empirically, the number of members varies greatly across CCPs (Domanski, Gambacorta, and Picillo (2015)).

[10] Investors have state-dependent preferences, which can be microfounded with standard preferences and liquidity shocks as in Holmström and Tirole (1998). It can also be viewed as a reduced form for state-varying marginal utilities due to difference in endowments or beliefs. To identify robust principles for clearing, we do not specify a particular hedging instrument. Our model can be easily adapted to accommodate one-sided hedging needs as in the Credit Default Swaps (CDS) market.

Trading is constrained by the fact that the asset's cash flow is not fully pledgeable. An investor can privately shirk at date 0 and enjoy a private benefit $\tilde{B} = q\left(R - \frac{\tilde{\beta}}{2}\right)$ per unit of asset, which causes asset failure. Parameter $\tilde{\beta} \in \{0, \beta\}$ represents the asset pledgeability, that is, the maximum amount an investor can credibly pay in expectation out of the asset's cash flow. The limited pledgeability friction captures investors' concerns for counterparties taking excessive risks or shirking proper risk management effort (see footnote 4).

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 an important defense against counterparty risks (see Section VI.A). It is also relevant in OTC markets where a counterparty's overall risk exposure may be difficult to assess due to lack of transparency.

Collateral

At date 0, any fraction of an investor's asset can be converted into cash at an exogenous rate of 1. 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 from the other group, cash collateral can protect against counterparty default. Third, as we will show, cash collateral expands investors' 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 (net) 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.[11]

CCP The CCP agent is risk-neutral and competitive, and has no hedging need. Its utility function is given by $U_C = c_0 + c_1$. The CCP agent has a large endowment E of asset with per-unit payoff $\kappa + 1$ at date 1 where $\kappa > 0$. Its asset is nonpledgeable but each unit can be liquidated for a unit of cash at date 0. The CCP can thus contribute cash to help satisfy investors' hedging needs. To distinguish from investors' collateral, we call this cash contribution CCP capital, with κ representing the (net) cost of capital. Besides contributing capital, the CCP can also monitor investors but its monitoring effort is as costly as the investors' and it is also unobservable.

B. Contracting

At date 0, each S-investor matches with an S'-investor, called her counterparty. In practice, these investors would sign a bilateral contract which can then be novated to and cleared by a CCP. In addition to bilateral payments, a cleared contract implicitly specifies contingent transfers among all investor pairs and the CCP.

In the model, we consider a general multilateral contract between all investors and the CCP. We discuss the mapping to a cleared contract in practice in Section III. A contract specifies transfers, and if necessary, a monitoring scheme: bilateral (counterparty) monitoring or centralized (CCP) monitoring. To streamline the exposition, we focus on contracts with

[11] 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).

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.[12] Given their preferences, S-investors should receive payments and S'-investors should pay only in states S when the former have high marginal utility of consumption. The model is symmetric across S, so we focus on contracts with symmetric transfers. We then drop the reference to S and label investors with their ex-post role: *payer* or *receiver*. Then, a general contract specifies an investor's transfer contingent on three factors: the aggregate state summarized by the number of defaulting payers d, the investor's ex-post role, and, within each pair, the payer's asset outcome $o \in \{s, f\}$ where s stands for success and f for failure (with asset payoff or pledgeability being zero). This last feature implies that a receiver transfer can depend on the outcome of her matched payer. CCP's date-1 transfer is indexed by d only. Formally, a contract is 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 nonnegative 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).*

Definition 1 illustrates what we mean by symmetric contracts. For a given combination

[12] 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.

(o,d), a S-investor receives the same transfer $r_o(d)$ in state S as a S'-investor in state S'. The same applies to payments. Transfers $r_s(N)$, $p_s(N)$ and $r_f(0)$, $p_f(0)$ are set to 0 as they are not well-defined: there cannot be N (0) defaulting payers if one payer succeeds (fails).

We now formally define the investors' problem.

Investors' Problem

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

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

$$\text{(5)} \qquad \forall\, d, \quad p_f(d) \le x,$$

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

$$(\text{PC}_{CCP}) \qquad \mathbb{E}[\pi(d)] \ge (\kappa+1)e + \mathbb{1}_{cm}2\psi,$$

$$(\text{LP}) \qquad \mathbb{E}_s[p_o(d)] - \mathbb{E}_f[p_o(d)] \le (1-x)\beta,$$

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

$$\text{If } \mathbb{1}_{cm} = 0, \quad \frac{\psi}{q(1-\alpha)} \le \frac{1}{2}\Big(\mathbb{E}_s[r_o(d)] - \mathbb{E}_f[r_o(d)]\Big)$$

$$(\text{MIC}_{bm}) \qquad + \frac{\nu-1}{2}\Big(\mathbb{E}_s[\min\{r_o(d),\hat{c}\}] - \mathbb{E}_f[\min\{r_o(d),\hat{c}\}]\Big),$$

where expectation $\mathbb{E}[\cdot]$ is over d, the number of defaulting payers, and $\mathbb{E}_{o'}[\cdot] \equiv \mathbb{E}[\cdot|o=o']$.

We note that, as the contract can implement any date-0 investment decisions and date-1 consumption profiles, the solution to the investors' problem is also the constrained-efficient allocation chosen by a social planner who maximizes the investors' expected utility.[13] We discuss the elements of the Investors' Problem below.

[13]Specifically, x and e determine the amount of collateral and CCP capital at date 0. The date-1 consumption for receivers, payers with succeeded assets, payers with failed assets, and the CCP agent are respectively $r_o(d)$, $x+(1-x)2R-p_s(d)$, $x-p_f(d)$, and $\pi(d)$.

Investor's expected utility is given by equation (3). To obtain equation (3) from equation (1), we substitute for the expected payment using an expected version of equation (6), $\mathbb{E}[p_o(d)] = \mathbb{E}[r_o(d)] + \mathbb{E}[\pi(d)] - x - e$, which we derive in the Appendix. The second term of equation (3) captures gains from trades; the third to last terms represent costs from using collateral, monitoring, and CCP compensation.

Resource constraints are represented in equation (4) and equation (5) at individual payer's level and equation (6) at aggregate level. 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}).

Investors' Limited Pledgeability constraint (LP) is the first key constraint. Recall that an investor can shirk at date 0 to enjoy private benefit $B = q\left(R - \frac{\beta}{2}\right)$ per unit of asset while causing the asset to fail. Hence, an investor would not shirk if and only if

(7)

$$\frac{1}{2}\Big\{q\big(x + (1-x)2R - \mathbb{E}_s[p_o(d)]\big) + (1-q)\big(x - \mathbb{E}_f[p_o(d)]\big)\Big\} \geq \frac{1}{2}\big(x - \mathbb{E}_f[p_o(d)]\big) + B(1-x),$$

which says that a non-shirking payer's expected consumption, hence, utility, is weakly higher than a shirking payer's expected consumption plus the private benefits from shirking. Such incentive constraint can be rewritten as the limited pledgeability constraint (LP), implying 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.[14]

[14]We impose only that equation (LP) holds under the expectation that other investors behave. That is, we abstract from coordination problem whereby investors would shirk because they expect others to shirk.

The second key constraints are the Monitoring Incentive Constraints, equation (MIC_{cm}) or (MIC_{bm}), imposed because monitoring efforts are unobservable. Under the centralized monitoring scheme, equation (MIC_{cm}) ensures that the CCP prefers monitoring everyone to no one (we verify later this is the relevant deviation). Under the bilateral monitoring scheme, equation (MIC_{bm}) guarantees that each investor monitors her counterparty.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 $[q(1-\alpha)]^{-1}$.

[Insert Figure 3 approximately here]

Figure 3 provides a numerical example of a contract and the transfers involved at date 1 with $N = 2$ pairs of investors. In the left panel, no payer defaults ($d = 0$) and receivers consume their desired amount $\hat{c} = 1.8 = r_s(0)$. In the right panel, payer $P1$ defaults ($d = 1$) and pays with all his collateral holding, $p_f(1) = x = 0.3$. Despite $P1$'s default, both receivers' consumption remains unchanged ($r_s(1) = r_f(1) = 1.8$) because loss mutualization is at play. The comparison across panels reveals how the loss is mutualized. The payment shortfall of $P1$ is $p_s(0) - p_f(1) = 1.5$. Part of the loss is then absorbed by the CCP whose compensation drops from $\pi(0) = 0.8$ to $\pi(1) = 0$. The residual loss ($1.5 - 0.8 = 0.7$) is borne by the surviving payer $P2$ whose payment increases from $p_s(0) = 1.8$ to $p_s(1) = 2.5$. The aggregate budget constraint (6) can be used to directly pin down the surviving payer $P2$'s contribution in state $d = 1$. By plugging in the ex ante collateral and capital contribution ($x = 0.3$, $e = 0.1$), receivers' consumption ($r_s(1) = r_f(1) = 1.8$), defaulter's payment $p_f(1) = 0.3$, and the CCP's compensation $\pi(1) = 0$, $p_s(1) = 2.5$ is obtained. More generally, this example shows that the CCP enables loss mutualization by aggregating payers' resources and redistributing to receivers, as described in aggregate budget constraint (6).

C. Assumptions

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

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 equation (LP), each payer can pay at most 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)**.** $\psi \leq \bar{\psi} \equiv \min\left\{\frac{(1-q)(\nu-1)}{\nu(2-\beta\alpha q)(1-\alpha q)}, \frac{1}{2}\right\}\beta q(1-\alpha)\left(1-\frac{\hat{c}}{2}\right)$

The first term in the minimum argument 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 term in the argument plays a technical role.

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

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 $d \leq N-1$. Assumption 3 simplifies our analysis in that the only aggregate risk receivers must bear is that of all payers' joint default.[15]

[15]We derive the optimal contract when Assumption 3 fails in Internet Appendix C for the case $N = 3$. Then, 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.

Assumption 4 (Capital Cost)**.** $\kappa \geq k$

Assumption 4 states that CCP capital is not cheaper than investors' collateral, that is, the CCP does not possess a superior technology to convert date-1 resources into cash. We will show that CCP capital plays a role in the optimal contract despite this cost disadvantage.[16]

III. Optimal contract properties

We first provide a result that characterizes the set of relevant contracts for our analysis. This characterization also allows us to map our multilateral contracts to centrally 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*

$$p_s(d) = \underbrace{r_s - x - e}_{\textit{Bilateral transfer}} + \underbrace{\frac{d}{N-d}(r_f - 2x - e)}_{\textit{Loss Mutualization transfer}} + \underbrace{\frac{N}{N-d}\pi(d)}_{\textit{CCP compensation}}. \tag{8}$$

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 mentioned above, receivers are risk-averse and thus wish to minimize

[16]In Appendix B, we consider the case $\kappa < k$ and show that CCP capital can play another role as insurance against counterparty risks, similar to collateral. Our analysis shows, however, that CCP capital will never fully substitute for investors' collateral due to the limited pledgeability friction.

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 pre-committed 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. The bilateral component of the surviving payer's transfer in equation (8) is the transfer from a payer to the receiver he matched with. The second term captures loss-mutualization transfers across investor pairs that compensate receivers whose payer default. Without loss mutualization, each receiver with a defaulting payer could consume at most $r_f(N) = 2x + e$, which is the collateral plus the CCP capital per investor pair. When some (other) payers survive, they can transfer resources to the receiver whose consumption r_f can increase above $r_f(N)$. This loss-mutualization transfer corresponds to investors' contributions to a default fund in practice. Finally, the third-term captures surviving payer's contribution to the CCP compensation.

IV. Clearing with Observable Monitoring

In this section, we provide two benchmarks in which monitoring does not need to be incentivized. In section IV.A we characterize the frictionless case where monitoring is redundant. In section IV.B we analyze the case with limited pledgeability and observable monitoring effort. We also provide conditions for (observable) monitoring to be optimal in Section IV.C.

A. Frictionless Benchmark

In the frictionless benchmark, there is no benefits from shirking ($\tilde{B} = 0$) hence the asset is fully pledgeable ($\tilde{\beta} = 2R$). Monitoring is thus redundant because its only role is to increase asset pledgeability.

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

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

The results are intuitive. If collateral is cheap enough, investors fully secure payments to meet their hedging needs in all states of the world. If not, they rely on the CCP to mutualize losses to deal with counterparty risks. Loss mutualization is said complete because a receiver's transfer is not affected by the default of her counterparty ($r_f = r_s = \hat{c}$) as long as one other payer survives. Loss mutualization does not involve collateral whose only role here is to hedge the joint default state (probability $(1 - q)^N$). The CCP does not pledge capital because it is more expensive than collateral as an insurance tool (Assumption 4). As the CCP does not monitor investors, it receives no compensation.[17] Central clearing via the CCP is needed, however, to implement loss mutualization.

[17]When $\kappa = k$ and $k \leq \underline{k}_N$, the preference for collateral over CCP capital is weak, and capital can replace collateral to provide full hedging. See also Appendix B for a detailed discussion of the case $\kappa < k$.

B. Limited pledgeability

In this section, we add back the limited pledgeability friction but monitoring remains observable. Specifically, we solve the Investor's Problem without equation (MIC_{bm}) or equation(MIC_{cm}).

The limited pledgeability friction gives collateral a new function: satisfying 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 nondefaulting payer can credibly pay $x + (1 - x)\beta$ in expectation (substituting $p_f(d) = x$ in equation (LP)). Also the receiver can use her own collateral x for consumption, thereby reducing her hedging needs to $\hat{c} - x$. Together, a nondefaulting payer's payment capacity in excess of her receiver's needs is

$$x + (1 - x)\beta - (\hat{c} - x). \tag{9}$$

Without collateral, this excess payment capacity is negative because $\beta < \hat{c}$, that is, an investor's payment capacity β already falls short of her counterparty's hedging needs $\hat{c}$. Hence, without collateral, central clearing cannot improve upon a simple contract that features no loss mutualization.

Equation (9) shows that pledging collateral increases excess payment capacity at the investor-pair level because $\beta < 2$ (from Assumption 1). Therefore, collateral is needed to support payments for loss mutualization or to compensate the CCP.

We begin the analysis with the choice of monitoring scheme.

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

Lemma 1 stems from the additional collateral cost of monitoring by CCP when asset pledgeability is limited. To support the CCP's compensation at $t = 1$, investors have to pledge collateral ex ante. Since the CCP has no intrinsic technological advantage as a monitor and collateral is costly, bilateral monitoring is superior. As we will show in Section V, when monitoring is not observable, the above conclusion can be overturned.

Proposition 3 (Optimal clearing with observable monitoring)**.** *There exists a threshold for collateral cost* $\bar{k} \equiv \frac{1}{2}(\nu - 1)(2 - q\beta) > \underline{k}_N$ *such that the contract solving the Investors' Problem without equation* (MIC_{bm}) *and equation* (MIC_{cm}) *is as follows:*

1. *For* $k \leq \underline{k}_N$, *it is the full-hedging contract of the frictionless case (Proposition 2).*

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*

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

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 with loss mutualization when collateral is costly enough.

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 mutualize losses (Case 2). The intuition is that investors must be able to pay to the default fund with loss mutualization, and their 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 satisfy their hedging needs ($r_s < \hat{c}$), and they are fully exposed to counterparty risk ($r_f = 0$). This threshold measures the total hedging value of collateral as counterparty risk insurance and as a tool to increase pledgeability. If $k > \bar{k}$, hedging and thus loss mutualization are too costly.

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. Notably, the contracts in Case 1 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 1), 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.

These observations imply that clearing benefits are hump-shaped in the cost of collateral.

Corollary 1 (Essentiality of CCP)**.** *A CCP is essential*

for $k \in [\underline{k}_N, \bar{k}]$. This region strictly expands with N.

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 market size becomes larger. 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.

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

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).*

The effect of central clearing on collateral needs depends on the contract investors would choose if they could only trade bilaterally. As collateral is bilateral traders' only defense against counterparty risk, they fully hedge when collateral is cheap. Central clearing then economizes on collateral because it relies on loss mutualization to mitigate counterparty risk. When collateral is expensive, however, it serves only to support incentive-compatible transfers. Large transfers in a CCP due to loss mutualization, relative to bilateral trading, increase collateral needs in this case.

C. Optimal Monitoring

To conclude this section, we provide conditions for monitoring to be optimal.

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 $\hat{k}^m \in [\underline{k}_N, \bar{k})$.*

Intuitively, as monitoring substitutes for collateral in expanding asset pledgeability, it is optimal when collateral cost is high. A lower bound for $\hat{k}^m$ is $\underline{k}_N$ because monitoring is suboptimal when the contract is fully collateralized ($k < \underline{k}_N$). 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 $k \in [\hat{k}^m, \bar{k}]$ to show how the incentive problem in monitoring affects the contract design and the role of the CCP.[18]

V. 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.

Monitoring incentive matters only if the *OM-contract* is not incentive-compatible with bilateral monitoring. The following lemma describes the parameter region for such a case.

Lemma 2. *When monitoring is optimal, the OM-contract violates equation* (MIC$_{bm}$) *when* $k \in (\hat{k}^m, \bar{k})$ *and* $N > N^*$*, where* N^* *is the largest value of* N *such that*

$$\frac{\psi}{q(1-\alpha)} \leq \nu(1-q)^{N-1}\left(\frac{\hat{c}}{2} - x^{OM}\right), \tag{11}$$

with x^{OM} *given by equation* (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 and the monitoring cost is low enough (Assumption 2) to induce monitoring. In the intermediate case $k \in (\underline{k}_N, \bar{k})$ with loss mutualization, investors retain only partial exposure to counterparty risk. This exposure and thus investors' incentives to monitor are captured by the right-hand side of

[18] In Internet Appendix D.B we extend Corollary 1, which characterizes the region in which a CCP is essential, to account for the optimal monitoring choice. The lower bound of the essential region changes relative to the case in which monitoring is imposed, but the result that this region expands with N remains.

equation (11), which decreases with N for two reasons. First, the only state in which losses are not mutualized, that is, the joint-default state, becomes less likely as N increases. Second, as the amount of collateral x^{OM} increases with N, the "loss given joint default" $\hat{c} - 2x^{OM}$ is also reduced.

To focus on the interesting case where monitoring is optimal but the monitoring in *OM-contract* is not incentive compatible, we impose parametric restrictions in Assumption 5.

Assumption 5. *$k \in [\hat{k}^m, \bar{k}]$ and $N > N^*$.*

The rest of Section V proceeds as follows. We derive the optimal contract under bilateral monitoring in Section V.A and under centralized monitoring in Section V.B. We compare the two schemes to show when the CCP emerges as the efficient monitor in Section V.C. Section V.D discusses the equilibrium level of CCP capital.

A. 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. We use the superscript $*$ for the equilibrium variables of the optimal contract with unobservable monitoring.

Proposition 5 (Optimal contract under bilateral monitoring)**.** *Let $\bar{k}^{bm} = \frac{1-q}{1-q+\nu q}\bar{k}$. 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 the following analysis of monitoring by the CCP.

B. Centralized Monitoring by the CCP

In this section, we analyze clearing with centralized monitoring. As 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 service is costly, however, because it increases investors' liability and thus requires additional collateral (Lemma 1). Investors minimize this cost by optimally designing the CCP compensation $\pi(d)$ and its capital contribution e.

Proposition 6 (Centralized monitoring contract)**.** *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*

$$\pi^*(0) = \frac{2\psi}{q^N(1-\alpha^N)}, \quad \pi^*(d) = 0 \text{ for } d > 0, \quad \text{and} \tag{12}$$

$$e^* = \underline{e} \equiv \frac{1}{(\kappa+1)} \frac{2\psi\alpha^N}{(1-\alpha^N)}. \tag{13}$$

Proposition 6 shows first that, relative to the *OM-contract*, investors post additional collateral $x^* - x^{OM}$ to support the CCP's compensation. However, as monitoring and risk sharing are now separated, investors can continue to mutualize loss completely ($r_o^* = r_o^{OM}$), unlike with bilateral monitoring.

Proposition 6 delivers two new insights for the CCP compensation and capital contribution. 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 paid when no payer defaults ($\pi^*(d) > 0$ only if $d = 0$), which is the state most indicative of CCP monitoring efforts.

The optimal compensation is then the minimum value of $\pi(0)$ that satisfies equation (MIC_{cm}). As such compensation scheme implies that the CCP loses all of its promised compensation when one or more payer default, it effectively holds a junior tranche and absorbs losses right after the defaulters' pre-committed resources (i.e., collateral) have been exhausted.[19]

[19] In practice, for-profit CCPs also collect noncontingent 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 does not depend on bargaining power as it is used to efficiently sustain the CCP's monitoring incentives.

The second insight is that the monitoring role of a CCP provides a rationale for CCP capital. In the *OM-contract*, the CCP does not pledge capital because it is too costly to be used for hedging counterparty risk. 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, equation (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.

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.[20] These economies of scale can be seen in the reduction of total CCP capital contribution (Ne^* decreases with N). As we discuss in Section V.C, 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. In addition, even if resource constraints bind for low-d states, a single monitor, hence a single CCP, remains optimal.*

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

C. 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 4 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.

[Insert Figure 4 approximately here]

In both panels, centralized monitoring is 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 is better dealt with collateral. If collateral is very expensive, bilateral monitoring (blue region) is more efficient than centralized monitoring: although loss mutualization is reduced, it requires less collateral (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 nonempty 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.

D. Bargaining over CCP Capital

Our model follows the principal-agent literature in assuming that the principal (investors) has all the bargaining power. As a result, investors require CCP to pledge capital as a way to recoup the CCP's rent from monitoring. In this section, we show that the CCP would never pledge capital if it had the bargaining power, that is, if it could make a take-it-or-leave-it offer to all the investors collectively. The novel takeaway is that CCP capital contribution is determined by the relative bargaining power between members and the CCP. As we discuss later, this finding echoes the ongoing debate between members and CCPs about the suitable amount of capital contribution (see Section VI.C).

Proposition 8. *The CCP would not pledge capital if it had the bargaining power.*

We recall the result in Proposition 6 that when investors have bargaining power, they require the CCP to pledge capital only when it monitors in order to recoup the agency rent from monitoring. It follows that a CCP with the bargaining power would prefer not to contribute capital as doing so would lower its profit.

We note that both investors' and the CCP's optimal choice of CCP capital are Pareto efficient. As we show in the proof of Proposition 8, however, these capital levels, may not maximize total welfare, the objective of a utilitarian planner. The reason is that utility is not transferable and investors request costly capital in order to capture the CCP rent.

VI. Implications for CCP Design

In this section, we relate our results to practical questions about CCP design and derive associated empirical predictions.

A. CCP Roles and Determining Factors

Our results rationalize potential roles of a CCP and qualitatively assess their relevance. First, a CCP can play the role as a risk pooler. By ex-ante arranging a loss mutualizing scheme, a CCP pools idiosyncratic member default risks. Second, as discussed below, the CCP can monitor investors. In Appendix B, we analyze yet a third role of CCPs as insurance providers. A CCP can use its capital as insurance against members' defaults, but this is efficient only when the CCP is small and has a lower cost of capital than that of members' collateral. In our view, these conditions are very restrictive and hence the CCP's role as an insurance provider is very limited.

The monitoring role of CCPs is the novel emphasis of our paper. Monitoring mitigates counterparty risk and is a valuable substitute of costly collateral. CCPs can emerge as efficient monitors due to endogenous economies of scale. In practice, adequate monitoring of members is indeed often cited by many CCPs as their first line of defense against counterparty risks.

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 (Pirrong 2011). 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.

B. 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, we rationalize the commonly observed defaulter-pay principle because seizing the pledged collateral of defaulting members efficiently discourages ex-ante 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 the relatively junior position of CCP in the default waterfall. A CCP's incentives to monitor its members is best preserved when it holds an equity tranche, which would be wiped out when members default. This default waterfall structure is indeed very common among CCPs in practice (Duffie (2015)).

C. The Determinants of CCP Capital

Our analysis can shed light on the intense debate about the size of CCP capital, the so-called skin-in-the-game (SITG). SITG is in general small as a fraction of total pre-funded resources, which some commentators take as evidence that SITG is either unimportant or insufficient. We argue that SITG is a consequence of bargaining between members and the CCP and is related to the rent paid to the CCP for its monitoring role. It needs not be large as incentives come in the form of the equity tranche held by the CCP or by its executives.[21] The view that SITG is an outcome of bargaining is acknowledged by market participants.[22]

While SITG is on average small as a fraction of total pre-funded resources, there is substantial heterogeneity across asset classes and jurisdictions of the CCPs.[23] Our model can generate such variations provided that asset pledgeability β differs across assets (e.g. constructing a portfolio with "wrong-way" risk is easier for some assets than others) and jurisdictions (e.g. some courts enforces contracts better than others). In addition, we predict that market size (number of members) is also a determinant of such ratio.

[21]CCPs in practice make executive compensation contingent on the actual usage of SITG to induce risk management effort. 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 nondefaulting clearing members' default fund contributions" (OCC (2020)). LCH, another CCP, states that besides SITG, "LCH has further strengthened this incentive structure by linking management compensation directly to usage of the SITG layer." (LCH (2015))

[22]In 2020, a group of twenty major institutional investors and investment banks has collectively issued a discussion paper (ABN-AMRO et al. (2020)) to request more substantial capital contribution from CCPs. The International Swap and Derivative Association 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." (ISDA (2019)).

[23]The ratio of CCP capital to total funded resources varies from 1.6% in Interest rate CCPs to 9.1% in Commodity CCPs and from 0.1% in CCPs in South America to 12.2% in Asia (Paddrik and Zhang (2020)).

Empirical prediction 1. *The ratio of CCP capital to total pre-funded resources* $\frac{e^*}{x^*+e^*}$ *strictly increases with* β *and decreases with* N.

The proof of these results is in Appendix A.M. As pledgeability improves, less collateral is required which increases the ratio of capital to collateral. The second result is driven by the reduction in the CCP's agency rent from monitoring when the number of members increases. The amount of CCP capital that investors can request thus decreases.

In practice, a larger CCP would have more bargaining power vis-à-vis its members and could thus further reduce its capital contribution, as shown in Section V.D. This effect would reinforce our result.

Another key metrics considered by market participants is the ratio of SITG capital to CCP realized profit, whose model equivalent is the CCP profit when no member defaults.[24]

Empirical prediction 2. *The ratio of CCP capital to realized profit* $\frac{e^*}{\pi^*(0)-2\psi}$ *strictly decreases with* N.

As observed above, CCP capital decreases with N. In addition, as CCP compensation is concentrated in the state with no member default, the realized profit increases with N.

D. CCP Ownership Structure

The discussion of default waterfall and CCP capital would be incomplete without considering the CCP's ownership structure. In a member-owned CCP, the line between CCP capital and members' collateral is blurry (McPartland and Lewis (2017)). In contrast, as a third-party

[24]There is substantial variation in this ratio. The European Association of CCP Clearing Houses reports an average ratio of 1.6 for EU and UK CCPs, and our own calculations based on regulatory reports from ESMA for 16 CCPs show this number can vary from 0.3 to 9.51. The data is self-collected from the various CCPs' disclosure and reports in 2020.

CCP contributes its own capital and retains profit from clearing, the seniority of CCP's claims vis-à-vis members in the default waterfall is relevant.

Our analysis of monitoring schemes relates to ownership structure. Under bilateral monitoring, the CCP purely mutualizes losses and does not receive compensation. Thus, this scheme resembles member-owned CCP. Under centralized monitoring, the CCP contributes capital ex ante and receives an equity-like compensation, resembling a third-party agent. Therefore, Proposition 7 yields the following prediction.

Empirical prediction 3. *A third-party CCP is preferable to a member-owned CCP when the number of clearing members is large.*

E. Collateral Requirement in Cleared Contracts

An important concern raised by market participants about central clearing is that it can substantially increase collateral requirements. Corollary 2 shows this needs not be the case.

Empirical prediction 4. *Bilateral contracts require more (less) collateral than cleared contracts when collateral is cheap (expensive).*

While indeed central clearing requires collateral to perform loss mutualization, bilateral contracts also rely on collateral because there is no other way to mitigate counterparty risks.

VII. 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.

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 allow 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.

References

ABN-AMRO; Allianz; Barclays; and BlackRock. "A Path Forward for CCP Resilience, Recovery, and Resolution." ABN-AMRO (2020).

Acharya, V., and A. Bisin. "Counterparty Risk Externality: Centralized versus Over-The-Counter Markets." *Journal of Economic Theory*, 149 (2014), 153–182.

Antinolfi, G.; F. Carapella; and F. Carli. "Transparency and Collateral: Central versus Bilateral Clearing." *Theoretical Economics*, 17 (2022), 185–217.

Armakolla, A., and B. Bianchi. "The European central counterparty (CCP) ecosystem." In *Data Needs and Statistics Compilation for Macroprudential Analysis*, Vol. 46, B. f. I. Settlements, ed. Bank for International Settlements (2017).

BCBS. "Internal models method for counterparty credit risk." (2019).

Biais, B.; F. Heider; and M. Hoerova. "Clearing, Counterparty Risk, and Aggregate Risk." *IMF Economic Review*, 60 (2012), 193–222.

Biais, B.; F. Heider; and M. Hoerova. "Risk-sharing or Risk-taking? Counterparty Risk, Incentives, and Margins." *The Journal of Finance*, 71 (2016), 1669–1698.

Board, T. F. S. E. "Financial Stability Report." Bank of England (2010).

Capponi, A., and W. A. Cheng. "Clearinghouse Margin Requirements." *Operations Research*, 66 (2018), 1542–1558.

CCP12. "CCP12 responses to ESMA's seven Consultation Papers on CCP Recovery and Resolution Regime (CCPRRR)." (2021).

Cerasi, V., and S. Daltung. "The Optimal Size of a Bank: Costs and Benefits of Diversification." *European Economic Review*, 44 (2000), 1701–1726.

Coeuré, B. "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. "Financial Intermediation and Delegated Monitoring." *The Review of Economic Studies*, 51 (1984), 393–414.

Domanski, D.; L. Gambacorta; and C. Picillo. "Central Clearing: Trends and Current Issues." BIS (2015).

Duffie, D. In *Making Failure Feasible.* "Resolution of Failing Central Counterparties.", K. E. Scott; T. H. Jackson; and J. B. Taylor, eds. Hoover Institution, Stanford University (2015), chapter 4.

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

ESMA. "CCPs' Membership Criteria and Due Diligence." (2020).

FSB. "Incentives to Centrally Clear Over-The-Counter (OTC) Derivatives." BIS (2018).

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

Hart, O., and J. Moore. "The Governance of Exchanges: Members' Cooperatives versus Outside Ownership." *Oxford Review of Economic Policy*, 12 (1996), 53–69.

Holmström, B. "Moral Hazard and Observability." *The Bell journal of economics*, 74–91.

Holmström, B., and J. Tirole. "Private and Public Supply of Liquidity." *Journal of Political Economy*, 106 (1998), 1–40.

Huang, W. "Central counterparty Capitalization and Misaligned Incentives." *BIS Working Paper 767*.

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

ISDA. "CCP Best Practices." (2019).

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

Koeppl, T., and C. Monnet. "The Emergence and Future of Central Counterparties." Federal Reserve Bank of Philadelphia (2010).

Koeppl, T.; C. Monnet; and T. Temzelides. "Optimal Clearing Arrangements for Financial Trades." *Journal of Financial Economics*, 103 (2012), 189 – 203.

Laux, C. "Limited-liability and Incentive Contracting with Multiple Projects." *Rand Journal of Economics*, 32 (2001), 514–526.

LCH. "CCP Conundrums." LCH (2015).

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

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

McPartland, J. "Challenging Skin in the Game." *WFE focus*.

McPartland, J., and R. Lewis. "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. "The Economics of Central Clearing." *Annual Review of Financial Economics*, 13 (2021), 153–178.

OCC. "Optimizing Incentives, Resilience and Stability in Central Counterparty Clearing." OCC (2020).

Paddrik, M. E., and S. Zhang. "Central Counterparty Default Waterfalls and Systemic Loss." *OFR WP*, 20–04.

Palazzo, F. "Peer Monitoring via Loss Mutualization." *Bank of Italy Temi di Discussione (Working Paper) No*, 1088.

Pirrong, C. "The Economics of Central Clearing: Theory and Practice." ISDA (2011).

Santos, T., and J. A. Scheinkman. "Competition Among Exchanges." *The Quarterly Journal of Economics*, 116 (2001), 1027–1061.

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

Stiglitz, J. E. "Incentives and Risk Sharing in Sharecropping." *The Review of Economic Studies*, 41 (1974), 219–255. ISSN 00346527, 1467937X.

Vuillemey, G. “The Value of Central Clearing.” *The Journal of Finance*, 75 (2020), 2021–2053.

Wang, J. J.; A. Capponi; and H. Zhang. “A Theory of Collateral Requirements for Central Counterparties.” *Management Science*, (2022), published online.

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

Zawadowski, A. “Entangled Financial Systems.” *The Review of Financial Studies*, 26 (2013), 1291–1323.

Figure 1: CCP roles

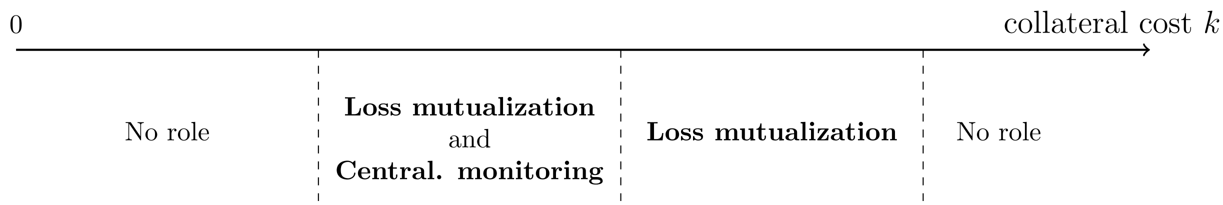


Figure 2: Payoff from an S-investor's asset

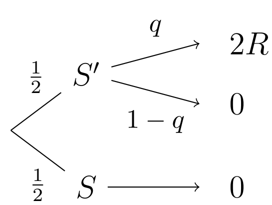


Figure 3: Example - Contract with N=2 and $\hat{c} = 1.8$

Date-1 transfers are represented in states $d = 0$ and $d = 1$. Values for investors' collateral and CCP capital are $(x, e) = (0.3, 0.1)$. Transfers are $\{\pi(0), p_s(0), r_s(0)\} = \{0.4, 1.8, 1.8\}$ for $d = 0$, $\{\pi(2), p_f(2), r_f(2)\} = \{0, 0.3, 0.7\}$ for $d = 2$ (not depicted) and $\{\pi(1), p_f(1), p_s(1), r_f(1), r_s(1)\} = \{0, 0.3, 2.5, 1.8, 1.8\}$ for $d = 1$. Dotted lines are bilateral investor links. Label P (R) is for payer (receiver). A red circle indicates a default.

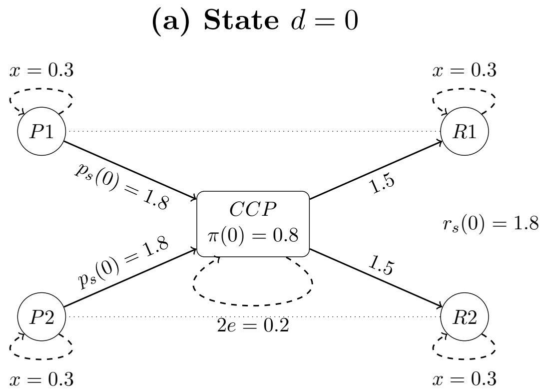
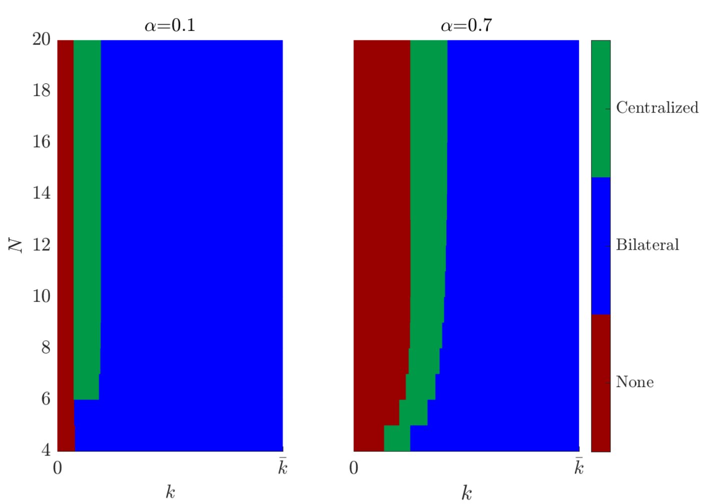


Figure 4: Optimal Monitoring with $\hat{c} = 0.8$, $\beta = 0.4$, $v = 2$, $q = 0.7$, $\kappa = 0.9$, $\psi = 5.6 \times 10^{-3}$.



Appendix

A. Proofs

A. Derivation of Equation (3)

We first derive equation (II.B). 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 equation (6), we thus obtain

(A.1)
$$\mathbb{E}_s[p_o(d)] = \sum_{d=0}^{N-1} (1-q)^d q^{N-1-d} \binom{N-1}{d} \left[r_s(d) + \frac{d}{N-d}(r_f(d) - p_f(d)) - \frac{N}{N-d}(x+e-\pi(d)) \right],$$

$$= \mathbb{E}_s[r_o(d)] + \sum_{d=1}^{N-1} (1-q)^d q^{N-1-d} \binom{N-1}{d-1} (r_f(d) - p_f(d)) - (x+e) \sum_{d=0}^{N-1} (1-q)^d q^{N-1-d} \binom{N}{d}$$

(A.2)
$$+ \sum_{d=0}^{N-1} (1-q)^d q^{N-1-d} \binom{N}{d} \pi(d),$$

(A.3)
$$= \mathbb{E}_s[r_o(d)] + \frac{1-q}{q} \sum_{l=0}^{N-2} (1-q)^l q^{N-1-l} \binom{N-1}{l} (r_f(l+1) - p_f(l+1)),$$

$$- \frac{(x+e)}{q} \left[1 - (1-q)^N\right] + \frac{1}{q} \left[\mathbb{E}[\pi(d)] - (1-q)^N \pi(N)\right]$$

(A.4)
$$= \mathbb{E}_s[r_o(d)] + \frac{1-q}{q}(\mathbb{E}_f[r_o(d)] - \mathbb{E}_f[p_o(d)]) - \frac{x+e}{q} + \frac{\mathbb{E}[\pi(d)]}{q},$$

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

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

(A.5)
$$U = \frac{1}{2}\Big(q(1-x)2R + x - \mathbb{E}[p_o(d)]\Big) + \frac{1}{2}\Big(\mathbb{E}[r_o(d)] + (\nu-1)\mathbb{E}\big[\min\{r_o(d), \hat{c}\}\big]\Big) - (1 - \mathbb{1}_{cm})\psi.$$

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

(A.6)
$$U = qR + \frac{1}{2}x - qRx + \frac{1}{2}(x+e) - \frac{1}{2}\mathbb{E}[\pi(d)] + \frac{\nu-1}{2}\mathbb{E}\big[\min\{r_o(d), \hat{c}\}\big] - (1 - \mathbb{1}_{cm})\psi,$$

which is equivalent to equation (3).

B. 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 in equation (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 the inequality in equation (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

$$\text{(A.7)} \qquad \Delta r_f(N) \leq \Delta p_f(N), \qquad \Delta\mathbb{E}_s[r_o(d)] \geq \nu\Delta\mathbb{E}_f[r_o(d)], \qquad \Delta\mathbb{E}_s[r_o(d)] \leq \Delta\mathbb{E}_s[p_o(d)].$$

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 the objective function in equation (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

$$\text{(A.8)} \qquad (\Delta r_f(d'), \Delta r_f(d)) = \left(\epsilon, -\frac{f(d')}{f(d)}\nu\epsilon\right),$$

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

$$\text{(A.9)} \qquad \Delta\mathbb{E}_s[p_o(d)] = -\frac{1-q}{q}(\nu - 1)f(d')\epsilon.$$

The perturbation strictly increases the objective function in equation (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 in equation (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:

$$\text{(A.10)} \qquad \Delta r_f < 0, \qquad \Delta r_s = -\frac{1-q-(1-q)^N}{q}\Delta r_f, \qquad \text{such that} \quad r_f + \Delta r_f = r_s + \Delta r_s.$$

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 equation (II.B). 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 the inequality in equation (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 $2x + e > \hat{c}$.

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

$$\text{(A.11)} \qquad xk + \frac{1}{2}e\kappa > \frac{\hat{c}}{2}k + \frac{1}{2}e(\kappa - k) > \frac{\hat{c}}{2}k,$$

where the last inequality follows from Assumption 4. The last expression is the cost for the the full-hedging contract. Hence, the candidate contract cannot be optimal.

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

$$\text{(A.12)} \qquad \mathbb{E}[r] = q\beta - \kappa e - 2\psi.$$

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 equation (LP) and equation (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.

C. 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

$$\max_{x,e,r_s,r_f} \frac{\nu-1}{2}\left[q\min\{r_s,\hat{c}\}+(1-q)\Big(\left[1-(1-q)^{N-1}\right]\min\{r_f,\hat{c}\}+(1-q)^{N-1}(2x+e)\Big)\right]$$
$$-x(qR-1)-\frac{1}{2}e\kappa. \tag{A.13}$$

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:

$$U'(e) = \frac{1}{2}(\nu-1)(1-q)^N - \frac{1}{2}\kappa, \tag{A.14}$$
$$U'(x) = (\nu-1)(1-q)^N - k. \tag{A.15}$$

First, equation (A.14) and equation (A.15) imply that $U'(x) \geq 2U(e)$ with a strict inequality if $\kappa > k$. Hence, if $e > 0$, a perturbation $(\Delta x, \Delta e) = (1/2e, -e)$ increases investors' utility, which means $e = 0$ is optimal. Furthermore, equation (A.15) shows that setting $r_f(N) = 2x$ equal to $\hat{c}$ is optimal if and only if $k \leq (\nu-1)(1-q)^N$.This concludes the proof.

D. 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, equation (PC_{CCP}), and equation (II.B), we obtain

$$q\Big(\mathbb{E}_s[p_o(d)] - \mathbb{E}_f[p_o(d)]\Big) = \mathbb{E}_s[r_o(d)] - 2x + \kappa e \tag{A.16}$$
$$= qr_s + (1-q)\left[1-(1-q)^{N-1}\right]r_f - \Big[1-(1-q)^N\Big](2x+e) + \kappa e. \tag{A.17}$$

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

$$qr_s + (1-q)\left[1-(1-q)^{N-1}\right]r_f \leq q\beta + \Big(2 - q\beta - 2(1-q)^N\Big)x - \left[\kappa + 1\left\{1-(1-q)^N\right\}\right]e. \tag{A.18}$$

Investors thus solve the problem described in equation (A.13) under constraint (A.18).

Step 2. Analysis

We first show that the optimal level of CCP capital is $e^{OM} = 0$. We showed in Proposition 2 that $e = 0$ is optimal in the absence of the limited pledgeability friction. In the presence of constraint (LP), equation (A.18) shows that increasing e tightens this constraint. Hence, setting $e^{OM} = 0$ remains optimal.

We now argue we can consider two different cases for the analysis: Either $r_s = r_f = \hat{c}$ or constraint (A.18) binds. This observation follows from Proposition 2 where we showed $r_s = r_f = \hat{c}$ is optimal in the absence of constraint (A.18). Besides, the relative weight on these two variables is the same in the objective function in equation (A.13) and in constraint (A.18).

Suppose first that $r_s = r_f = \hat{c}$ and $k \leq \underline{k}_N = (\nu - 1)(1 - q)^N$. Then, increasing x until $r_f(N) = 2x$ equal $\hat{c}$ is optimal because it increases investors' utility as shown by condition (A.15) in the proof of Proposition 2 and it relaxes constraint (A.18). Hence, in this case, the optimal OM-contract is the full-hedging contract derived in Proposition 2.

Suppose now $k > \underline{k}_N = (\nu - 1)(1 - q)^N$. We want to find conditions such that $r_s = r_f = \hat{c}$ is optimal and $r_f(N) = 2x < \hat{c}$. In this case, it must be that the inequality in equation(A.18) binds. Otherwise decreasing x while maintaining (r_s, r_f) constant strictly increases investors' utility because $k > \underline{k}_N = (\nu-1)(1-q)^N$. A contract with the conjectured properties is optimal if decreasing x when constraint (A.18) binds decreases the objective function. We have in this case

$$U'(x) = \frac{\nu - 1}{2} \frac{\partial \mathbb{E}[r_o(d)]}{\partial x}_{|e=0,(A.18)\text{ binds}} - k = \frac{\nu - 1}{2}(2 - q\beta) - k \equiv \bar{k} - k. \tag{A.19}$$

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

We are left to describe the case $k \leq \bar{k}$ in which$r_s, r_f < \hat{c}$. In this case, it is also optimal to set x to 0 since the marginal benefit of collateral is given by equation (A.19). 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.18), that is,

$$r_s + \frac{1 - q}{q}\left[1 - (1 - q)^{N-1}\right] r_f = \beta. \tag{A.20}$$

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.

E. 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.B. We verify that the OM-contracts of Proposition 3 satisfy Definition 2 only in Cases 1 and 3.

For Case 1, 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 2 has $r_f^{OM}(d) = \hat{c} > p_f^{OM}(d) + x^{OM}$ for all $d < N$, and thus this contract violates Definition 2. It follows that the upper bound for the essential CCP region is given by $\bar{k}$ and the lower bound is $\underline{k}_N$. This concludes the proof.

F. Proof of Corollary 2

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

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 (10) 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 $\bar{k}$.

For the second part of the result, observe that $\underline{k}_N = (\nu - 1)(1-q)^N$ strictly decreases with N. Hence, by Proposition 3, when $k \in [\underline{k}_N, \bar{k}_1]$, the multilateral contract features full loss mutualization with $x^{OM} < \frac{\hat{c}}{2}$ while the optimal bilateral contract features full hedging, that is $x_1^{OM} = \frac{\hat{c}}{2}$. This proves the result.

G. Proof of Proposition 4

We first derive the optimal contract without monitoring in Section G.1 and then derive the optimal monitoring decision in Section G.2.

G.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 in equation (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. *Without monitoring, there are two thresholds of collateral cost*

$$\underline{k}_N^{\not{m}} = (\nu - 1)(1-\alpha q)^N, \tag{A.21}$$

$$\bar{k}^{\not{m}} = \frac{1}{2}(\nu - 1)(2 - \alpha q \beta), \tag{A.22}$$

such that

1. *if $k \leq \underline{k}_N^{\not{m}}$, the contract in Case 1 of Proposition 3 is optimal,*
2. *if $k \in [\underline{k}_N^{\not{m}}, \bar{k}^{\not{m}}]$, a complete LM contract is optimal with $r_s^{OM,\not{m}} = r_f^{OM,\not{m}} = \hat{c}$ and*

$$x^{OM,\not{m}} \equiv \frac{\left[1-(1-\alpha q)^N\right]\hat{c} - \beta q}{2\left[1-(1-\alpha q)^N\right] - \beta\alpha q} \in \left(0, \frac{\hat{c}}{2}\right), \tag{A.23}$$

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

G.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, the contract consecutive to the full-hedging contract is the full-loss-mutualization contract both with and without monitoring. For each case of Proposition 3 or Proposition A.1, 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 than the predecessor contract, which follows directly from Proposition 3.

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 1 when both contracts are the same and thus require the same amount of collateral.

The third claim is that the thresholds between consecutive contracts are strictly higher under no monitoring. The comparison between $\bar{k}$ and $\bar{k}^{\not{m}}$ on the one hand and $\underline{k}_N$ and $\underline{k}_N^{\not{m}}$ on the other hand shows immediately that this is the case because $\alpha < 1$.

These three claims together imply that the benefit from monitoring is strictly increasing with k except when $k \leq \underline{k}_N$ where it is constant, negative, and equal to $-\psi$ because then 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 $\hat{k}^m > \underline{k}_N$ for $\psi > 0$. For the degenerate case $\psi = 0$, any value in $[0, \underline{k}_N]$ is admissible.

Since $\underline{k}_N < \bar{k}$ 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 1 or 2. In general, investors' utility is given by

$$U^{\not{m}}_{|k=\bar{k}} = qR + \Big[(\nu - 1) - \bar{k}\Big]\frac{\hat{c}}{2} + \max\Big\{0, \bar{k} - \underline{k}_N^{\not{m}}\Big\}\left(\frac{\hat{c}}{2} - x^{OM,\not{m}}\right) \tag{A.24}$$

$$= qR + q\beta\frac{\nu - 1}{2}\frac{\hat{c}}{2} + \frac{\nu - 1}{2}\beta\alpha q\left(1 - \frac{\hat{c}}{2}\right)\max\left\{0, \frac{2 - q\beta - 2(1-\alpha q)^N}{2\big[1 - (1-\alpha q)^N\big] - \beta\alpha q}\right\}. \tag{A.25}$$

The second term of $U^{\not{m}}_{|k=\bar{k}}$ in the second line 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,

$$U^{\not{m}}_{|k=\bar{k}} \leq qR + q\beta\frac{\nu - 1}{2}\frac{\hat{c}}{2} + \frac{\nu - 1}{2}\left(1 - \frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2 - q\beta)}{2 - q\alpha\beta}. \tag{A.26}$$

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

$$0 \leq U_{k=\bar{k}} - \left\{ qR + q\beta \frac{\nu-1}{2}\frac{\hat{c}}{2} + \frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta} \right\} \tag{A.27}$$

$$\leq qR + \frac{\nu-1}{2}q\beta - \psi - \left\{ qR + q\beta \frac{\nu-1}{2}\frac{\hat{c}}{2} + \frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta} \right\} \tag{A.28}$$

$$\leq \frac{\nu-1}{2}q\beta\left(1-\frac{\hat{c}}{2}\right) - \frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta} - \psi \tag{A.29}$$

$$\leq \frac{\beta q(1-\alpha)(\nu-1)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right) - \psi. \tag{A.30}$$

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.

H. 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 5, 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 (10). Plugging these variables into the bilateral monitoring constraint (MIC_{bm}), we obtain condition (11).

I. Proof of Proposition 5

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

$$\begin{aligned}\frac{\psi}{1-\alpha} \leq\ & \frac{1}{2}\Big[r_s - r_f + (1-q)^{N-1}\left(r_f - (2x+e)\right)\Big] \\ & + \frac{\nu-1}{2}\Big[\min\{r_s,\hat{c}\} - \left(\left[1-(1-q)^{N-1}\right]\min\{r_f,\hat{c}\} + (1-q^{N-1})((2x+e))\right)\Big].\end{aligned} \tag{A.31}$$

The optimal contract under bilateral monitoring solves problem (A.13) under limited pledgeability constraint (A.18) and constraint (A.31) 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. Equations (LP) and (MIC_{bm}) bind

Under Assumption 5, constraint (A.31) binds because the *OM-contract* in Proposition 3 violates equation (A.31). 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.31). Hence, the marginal effect on investors' utility from this perturbation is given by $-U'(x)$ in equation (A.15), which is positive because $k > \bar{k}_N$ by Assumption 5.

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 5 when constraint (A.31) is not imposed. Hence, it is still optimal under additional constraint (A.31) because increasing r_s relaxes this constraint. It is also optimal to increase r_f until equation (A.31) binds. Under Assumption 5, 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 equation (A.18) and equation (A.31), we have (for given x and e).

$$qr_s + (1-q)\big[r_f - (1-q)^{N-1}(r_f - 2x - e)\big] = (2-q\beta)x + q\beta - \kappa e \tag{A.32}$$

$$r_s - v\big[r_f - (1-q)^{N-1}(r_f - 2x - e)\big] = \frac{2\psi}{q(1-\alpha)} - (\nu - 1)\hat{c}. \tag{A.33}$$

Hence, we obtain

(A.34)

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

We can plug this relationship into the expression for investors' utility in equation (A.13). 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

$$k \leq \frac{\nu-1}{2}\frac{1-q}{1-q+\nu q}(2-q\beta) = \frac{1-q}{1-q+\nu q}\bar{k} \equiv \bar{k}^{bm} < \bar{k}. \tag{A.35}$$

Step 3. Optimal distortion

CCP capital e tightens monitoring constraint (A.31). 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 equation (A.32) and equation (A.33), we obtain

$$\hat{c}\Big[1-q-(1-q)^N+q-\nu q(1-q)^{N-1}\Big] - q\beta + \frac{2\psi}{1-\alpha} = \Big(2-2(1-q)^{N-1}\big[\nu q + 1 - q\big] - \beta q\Big)x. \tag{A.36}$$

Hence,

$$x^* = \frac{\hat{c}\Big(1-(1-q)^{N-1}\big[\nu q+1-q\big]\Big) - q\beta + \frac{2\psi}{1-\alpha}}{2-2(1-q)^{N-1}\big[\nu q+1-q\big] - \beta q} > x^{OM}. \tag{A.37}$$

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 equation (A.32) and equation (A.33) to solve for r_f^* and x^* setting again $e^* = 0$. We obtain

$$x^* = \frac{\hat{c} - q\beta - \frac{2\psi(1-q)}{q\nu(1-\alpha)}}{2-q\beta} < x^{OM}, \tag{A.38}$$

$$r_f^* = \frac{\hat{c} - 2(1-q)^{N-1}x^* - \frac{2\psi}{\nu q(1-\alpha)}}{1-(1-q)^{N-1}}. \tag{A.39}$$

This concludes the proof.

J. 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

$$IC(d) = 1 - \frac{\mathbb{P}[d|\text{ shirk}]}{\mathbb{P}[d|\text{ effort}]}, \tag{A.40}$$

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,

$$\mathbb{P}[d|\text{ shirk}] = \sum_{d_m=0}^{d} \binom{n_m}{d_m}\binom{N-n_m}{d-d_m}(1-q)^{d_m} q^{n_m-d_m}(1-\alpha q)^{d-d_m}(\alpha q)^{N-n_m-d+d_m}. \tag{A.41}$$

After some manipulation, we obtain

$$\frac{\mathbb{P}[d|\text{ shirk}]}{\mathbb{P}[d|\text{ effort}]} = \frac{\sum_{d_m=0}^{d}\binom{n_m}{d_m}\binom{N-n_m}{d-d_m}\left[\frac{1-\alpha q}{\alpha(1-q)}\right]^{d-d_m}}{\binom{N}{d}} = \sum_{d_m=0}^{d} w_{n_m}(d_m)\left[\frac{1-\alpha q}{\alpha(1-q)}\right]^{d-d_m}, \tag{A.42}$$

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 equation (MIC_{cm}) holds as an equality. It is defined by

$$Nq^N\underline{\pi}(0) - 2N\psi = \max_{n_m \in [|0,..N-1|]}\left\{Nq^N\alpha^{N-n_m}\underline{\pi}(0) - 2n_m\psi\right\}, \tag{A.43}$$

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

$$g'(y) \propto 1 - \alpha^y + y\alpha^y log(\alpha) \geq 1 - \alpha(1 - log(\alpha)). \tag{A.44}$$

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 equation (A.43), we find that $\underline{\pi}(0)$ is given by equation (12). With $\underline{\pi}(0)$, $\underline{e}$ given by equation (13) is the amount of capital such that equation (PC_{CCP}) binds.

Step 2. Optimal Contract

Observe first that the expected compensation to the CCP is a fixed cost. Hence, under Assumption 5, 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 equation (12). When e is increased over $\underline{e}$, however, the inequality in equation(PC_{CCP}) is tight, and any increase in CCP capital requires an increase in expected compensation by a factor $\kappa + 1$. Using equation (A.18) for constraint (LP), we obtain the following result

$$U'(e)_{|r_s^*=r_f^*=\hat{c},\,(\text{LP})binds} = \frac{\partial U}{\partial e} + \frac{\partial U}{\partial x}\frac{\partial x}{\partial e} \tag{A.45}$$

$$= \begin{cases} \frac{\nu-1}{2}(1-q)^N - [(\nu-1)(1-q)^N - k]\frac{1-(1-q)^N}{2-2(1-q)^N-\beta q} & \text{if} \quad e \leq \underline{e}, \\ [\underline{k}_N - k]\frac{\kappa+(1-q)^N}{2-q\beta-2(1-q)^N} & \text{if} \quad e > \underline{e}. \end{cases} \tag{A.46}$$

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

$$\hat{c}\left[1-(1-q)^N\right] + (1-q)^N(2x^* + e^*) + \mathbb{E}[\pi^*] = q\beta + (2-q\beta)x^* + e^*. \tag{A.47}$$

We obtain

$$x^* = \frac{\hat{c}\left[1-(1-q)^N\right] - \beta q}{2\left[1-(1-q)^N\right] - \beta q} + \frac{\left(\kappa + 1 - \left[1-(1-q)^N\right]\right)e^* + 2\psi}{2\left[1-(1-q)^N\right] - \beta q}, \tag{A.48}$$

$$= x^{OM} + \frac{2\psi}{(\kappa+1)(1-\alpha^N)}\frac{\kappa + 1 - \alpha^N\left[1-(1-q)^N\right]}{2\left[1-(1-q)^N\right] - \beta q}. \tag{A.49}$$

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

$$\psi \leq \frac{1-\alpha^N}{2 - \frac{\beta q \alpha^N}{\kappa+1}}\beta q\left(1 - \frac{\hat{c}}{2}\right). \tag{A.50}$$

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.

K. 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 5 implies $k \leq \bar{k}^{\not{m}}$, and the lower bound of the region $\underline{k}_N^{\not{m}}$ converges to 0 as N grows large. Using Proposition 6 and A.1, we derive the following expressions for investors' utility:

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

$$- \frac{1}{2}\Big[(\nu_{cm}-1) - (\nu-1)(1-q)^{N-1}\Big]e^* - \psi \tag{A.51}$$

$$U^{OM,\cancel{m}} = qR + \Big[\nu - 1 - k\Big]\frac{\hat{c}}{2} + \Big[k - (\nu-1)(1-\alpha q)^N\Big]\left(\frac{\hat{c}}{2} - x^{OM,\cancel{m}}\right). \tag{A.52}$$

From Proposition 6 and A.1 again, we have

$$\frac{\hat{c}}{2} - x^{*,cm} = \frac{\beta q\left(1-\frac{\hat{c}}{2}\right)}{2\big[1-(1-q)^N\big] - \beta q} - \frac{2\psi}{(\kappa+1)(1-\alpha^N)}\frac{\kappa + 1 - \alpha^N\big[1-(1-q)^N\big]}{2\big[1-(1-q)^N\big] - \beta q} \tag{A.53}$$

$$\frac{\hat{c}}{2} - x^{OM,\cancel{m}} = \frac{\beta\alpha q}{2\big[1-(1-\alpha q)^N\big] - \beta\alpha q}\left(1-\frac{\hat{c}}{2}\right), \tag{A.54}$$

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, $U^{*,cm} \geq U^{OM,\cancel{m}}$ if and only if

$$\frac{k}{2-\beta q}\left[\beta q\left(1-\frac{\hat{c}}{2}\right) - 2\psi\right] - \psi \geq \frac{k}{2-\beta\alpha q}\beta\alpha q\left(1-\frac{\hat{c}}{2}\right). \tag{A.55}$$

Under Assumption 2, we have

$$\psi \leq \frac{\beta q(1-\alpha)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right). \tag{A.56}$$

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

$$\hat{k}^{cm} = \frac{2-\beta q}{\frac{\beta q(1-\alpha)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right) - \psi}\frac{\psi}{2}. \tag{A.57}$$

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

$$U^* = qR + \Big[\nu - 1 - k\Big]\frac{\hat{c}}{2} + \Big[k - (\nu-1)(1-q)^N\Big]\left(\frac{\hat{c}}{2} - x^*\right) - \psi. \tag{A.58}$$

Using equation (A.51) and equation (A.58), centralized monitoring dominates Case 1 of bilateral monitoring if and only if

(A.59)

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

Using the expression for the collateral requirement in equation (A.37), we obtain

$$x^* - x^{OM} = \frac{2\psi}{\left[1-\alpha\right]\left[2(1-(1-q)^N) - \beta q\right]} - \frac{\nu q(1-q)^{N-1}}{2(1-(1-q)^N) - \beta q}(\hat{c} - 2x^*) \tag{A.60}$$

$$= \frac{2\psi}{\left[1-\alpha\right]\left[2(1-(1-q)^N) - \beta q\right]} - \frac{\nu q(1-q)^{N-1}}{2(1-(1-q)^N) - \beta q}\frac{\beta q(2-\hat{c}) - \frac{4\psi}{1-\alpha}}{2\left[1-(1-q)^{N-1}(\nu q + 1 - q)\right] - \beta q}. \tag{A.61}$$

We thus obtain the following condition

$$\frac{1}{2}\left(\kappa - (\nu-1)(1-q)^N\right)e^* \leq \left[k - (\nu-1)(1-q)^N\right](x^* - x^{*,cm}) \tag{A.62}$$

$$\frac{1}{2}\left(\kappa - (\nu-1)(1-q)^N\right)e^* \leq \frac{k - (\nu-1)(1-q)^N}{2(1-(1-q)^N) - \beta q}\left[\frac{2\psi}{1-\alpha} - \frac{2\psi}{1-\alpha^N} - \frac{\nu q(1-q)^{N-1}\left(\beta q(2-\hat{c}) - \frac{4\psi}{1-\alpha}\right)}{2\left[1-(1-q)^{N-1}(\nu q + 1 - q)\right] - \beta q}\right]. \tag{A.63}$$

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

$$\left(k-(\nu-1)(1-q)^N\right)\left(x^{cm,*} - x^{OM}\right) + \frac{1}{2}\left(\kappa-(\nu-1)(1-q)^N\right)e^* \leq \left[\frac{\nu-1}{2}(2-q\beta) - k\right]\left(x^{OM} - x^*\right). \tag{A.64}$$

Using equation (10) for x^{OM} and equation (A.38) for x^*, we obtain

$$x^{OM} - x^* = \frac{2\psi(1-q)}{\nu q(1-\alpha)(2-q\beta)} - \frac{\beta q(2-\hat{c})(1-q)^N}{\left[2-q\beta\right]\left[2(1-(1-q)^N) - \beta q\right]}. \tag{A.65}$$

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

$$\frac{\frac{\nu-1}{2}(2-q\beta) - k}{2-q\beta}\frac{2\psi(1-q)}{\nu q(1-\alpha)} \geq \frac{k}{2-\beta q}2\psi. \tag{A.66}$$

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

$$\bar{k}^{cm} \equiv \frac{1-q}{1-q+\nu q(1-\alpha)}\bar{k} < \bar{k}. \tag{A.67}$$

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

$$0 = \frac{1-q}{1-q+\nu q(1-\alpha)} \frac{\nu-1}{2}(2-q\beta) - \frac{2-\beta q}{\frac{\beta q(1-\alpha)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right)-\psi}\frac{\psi}{2} \tag{A.68}$$

$$0 = (1-q)(\nu-1)\frac{\beta q(1-\alpha)}{2-\beta\alpha q} - (1-q)(\nu-1)\psi - \psi\big[1-q+\nu q(1-\alpha)\big] \tag{A.69}$$

$$\psi = \frac{\beta q(1-q)(1-\alpha)(\nu-1)}{v(2-\beta\alpha q)(1-\alpha q)}\left(1-\frac{\hat{c}}{2}\right). \tag{A.70}$$

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.

L. Proof of Proposition 8

We first prove that a CCP would never pledge capital if it had the bargaining power. We then prove the additional result mentioned in the text that an utilitarian planner maximizing total surplus may choose a lower level of capital than investors.

The 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 5 the net value of CCP capital to investors is negative when its cost is $\kappa+1$. 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 $\kappa+1$ 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 $e \in [0, e^*]$,

$$U(e) = qR + \frac{\nu-1-k}{2}\hat{c} - \frac{1}{2}\mathbb{E}[\pi^*] + \Big[k-(\nu-1)(1-q)^N\Big]\left(\frac{\hat{c}}{2}-x-\frac{e}{2}\right) + \frac{1}{2}\big[1+k\big]e. \tag{A.71}$$

where x is a function of e given implicitly by equation (A.47) 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 $N(\kappa+1)e^*$. Hence, for a planner maximizing total surplus, the allocation with $e = e^*$ dominates if and only if

$$0 \leq 2NU(e^*) - (2NU(e=0) + N(\kappa+1)e^*_C) \tag{A.72}$$

$$\Leftrightarrow \quad 0 \leq \Big[k-(\nu-1)(1-q)^N\Big]\left(x(e=0)-x^*-\frac{e^*}{2}\right) - \Big[\kappa-k\Big]\frac{e^*_C}{2} \tag{A.73}$$

$$\Leftrightarrow \quad 0 \leq \left\{\frac{k-(\nu-1)(1-q)^N}{\bar{k}-(\nu-1)(1-q)^N}\beta q(\nu-1) - \Big[\kappa-k\Big]\right\}\frac{e^*_C}{2}. \tag{A.74}$$

When κ 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 $e = e^*$.

M. Proof of Empirical Prediction 1

CCP capital is given by equation (13) and collateral x^* in a third-party CCP is given by equation (A.49) in the proof of Proposition 6.

First, we prove the comparative statics with respect to β. Proposition 6 shows e^* is independent from β. Hence, it is enough to show that x^* is decreasing with β. To prove this result, use the implicit function theorem on equation (A.47) to obtain

$$\frac{\partial x^*}{\partial \beta} = -\frac{q(1-x^*)}{2-q\beta-2(1-q)^N} < 0. \tag{A.75}$$

The inequality follows because $x^* < 1$ when the contract is not fully collateralized and the denominator is strictly positive for all $N \geq 2$ $\beta < 2$ because $\beta < 2$ by Assumption 1.

Next, we prove the comparative statics with respect to N. We have

$$r_{xe}(N) \equiv \frac{2Nx^*}{Ne^*} = \frac{2x^*}{e^*} = 2\frac{\kappa + (1-q)^N + \frac{\hat{c}\left[1-(1-q)^N\right]-\beta q+2\psi}{e^*(N)}}{2\left[1-(1-q)^N\right]-\beta q}. \tag{A.76}$$

Taking the first-order derivative with respect to N, we obtain

$$r'_{xe}(N) = -2\frac{\frac{\partial e^*}{\partial N}}{(e^*)^2}\frac{\hat{c}\left[1-(1-q)^N\right]-\beta q+2\psi}{2\left[1-(1-q)^N\right]-\beta q} \tag{A.77}$$

$$-2\log(1-q)(1-q)^N\frac{\left(\frac{\hat{c}}{e^*}-1\right)(2-\beta q)-2\left(\frac{\hat{c}-\beta q+2\psi}{e^*}+\kappa\right)}{\left(2\left[1-(1-q)^N\right]-\beta q\right)^2}. \tag{A.78}$$

The term in the first line is positive because $\frac{\partial e^*}{\partial N} < 0$, that is, CCP capital per investor is decreasing in N. Hence to show that $r'_{xe}(N) > 0$, it is enough to show that the numerator of the second term, call it A, is positive. Indeed $-\log(1-q)(1-q)^N > 0$. We have

$$e^*A = (\hat{c}-e^*)(2-\beta q) - 2(\hat{c}-\beta q+2\psi+\kappa e^*) \tag{A.79}$$

$$= \beta q(2-\hat{c}) - 4\psi - 2(\kappa+1)e^* + \beta q e^* \tag{A.80}$$

$$= \beta q(2-\hat{c}) - \frac{4\psi}{1-\alpha^N} + \beta q\frac{2\psi\alpha^N}{(\kappa+1)(1-\alpha^N)} = \beta q(2-\hat{c}) - \frac{2\psi}{1-\alpha^N}\left[2-\frac{\beta q\alpha^N}{\kappa+1}\right]. \tag{A.81}$$

Assumption 2 implies this expression is positive. Hence r_{xe} is increasing with N which implies $\frac{e^*}{x^*+e^*}$ is decreasing with N. This concludes the proof.

B. Cheap CCP Capital

In this section, we analyze the case in which CCP capital is cheaper than investors' collateral, that is, $\kappa < k$. To clearly highlight the new role of capital in this case, we consider the version of the model with observable monitoring from Section IV. We show that capital can substitute for collateral as an insurance tool if $\kappa < k$, but tapping into CCP capital nevertheless requires investors' collateral due to the limited pledgeability problem.

The following result extends Proposition 3 for any value of CCP capital κ.

Proposition B.1. *There exists a threshold for collateral cost*

$$\tilde{\underline{k}}_N \equiv (\nu - 1)(1-q)^N + \frac{1}{2}\frac{2 - q\beta - 2(1-q)^N}{\kappa + (1-q)^N} \max\left\{(\nu - 1)(1-q)^N - \kappa, 0\right\}, \tag{B.1}$$

such that the optimal contract in Case 2 and 3 of Proposition 3 is identical substituting $\underline{k}_N$ with $\tilde{\underline{k}}_N$. For $k \leq \tilde{\underline{k}}_N$, the optimal contract features full hedging with $r_s^{OM} = r_f^{OM} = \hat{c}$ and

1. *$(e^{OM}, x^{OM}) = \left(0, \frac{\hat{c}}{2}\right)$ if $k < \kappa$*
2. *$(e^{OM}, x^{OM}) = \left(\frac{q\beta(2-\hat{c})}{2(\kappa+1)-q\beta}, \frac{\hat{c}-e^{OM}}{2}\right)$ otherwise.*

The proof of this result is in Internet Appendix D. When CCP capital is cheap, the OM contract of Proposition 3 changes only in the full-hedging case. When CCP capital is cheap, it can play a role similar to investors' collateral in hedging counterparty risks. When the cost of collateral or capital is lower than the value of hedging the joint default state, $(\nu - 1)(1-q)^N$, investors use the cheapest of the two resources to hedge. Case 2 of Proposition B.1 shows, however, that collateral is always part of the optimal contract even when it is more expensive than CCP capital. This asymmetry arises because of the limited pledgeability problem. Collateral relaxes investors' limited pledgeability constraint (LP). To use CCP capital, however, investors must compensate the CCP at date 1, which adds to the liability of investors, thereby exacerbating the pledgeability problem. This effect is reflected in condition $k < \tilde{\underline{k}}_N$ for Case 2: no CCP capital is used if collateral is too expensive because compensating the CCP for its capital contribution requires investors' collateral.

Overall, our robustness analysis strikes a cautious note about the role of CCP capital as insurance. CCP capital can be used as insurance only if is cheaper than collateral, which we view as a restrictive condition. This condition is not even sufficient as CCP capital comes with a shadow cost of collateral.

Internet Appendix

C. Contract with binding resource constraint

We relax Assumption 3 to analyze the situation in which the resource constraint (4) may bind. We 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

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

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 equation (10). 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 $N = 3$ when equation (A3n) holds. 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 C.1. *Let $N = 3$, $\psi = 0$ and $\kappa > k$. Under Assumption (A3n), there exists a threshold*

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

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*

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

The proof is in Internet Appendix D. Case 2 of Proposition C.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.

D. Additional Proofs

A. Proof of Proposition B.1

The first step of the proof is identical to that of Proposition 3, that is, we can rewrite the limited pledgeability constraint as equation (A.18).

We first show two results about CCP capital e. First, CCP capital may be used only if $k < \kappa$. 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.18). Second, if CCP capital is used, it must be that equation (A.18) 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.18) still holds and the objective function increases because $k < \kappa$ 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.18) binds. This observation follows from Proposition 2 where we showed $r_s = r_f = \hat{c}$ is optimal in the absence of constraint (A.18). Besides, the relative weight on these two variables is the same in the objective function in equation (A.13) and in constraint (A.18).

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. $k \leq (\nu - 1)(1 - q)^N$

Increasing x until $r_f(N) = 2x + e = \hat{c}$ is then optimal because condition (A.15) shows that investors' utility increases with x and because increasing x relaxes constraint (A.18). If in addition $k < \kappa$, 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 case is thus identical to Case 1 of Proposition 3.

If instead $k > \kappa$, CCP capital should be used and, as shown above, constraint (A.18) 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 equation (A.18) binds. This corresponds to Case **??** of Proposition 3.

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

Then, it is optimal to decrease x until constraint (A.18) binds because $U'(x) < 0$. Equation (A.14) shows that increasing e until $r_f(N) = 2x + e = \hat{c}$ can still be optimal if $\kappa \leq (\nu - 1)(1 - q)^N$. To determine the sufficient condition, we need to account for the effect of e on constraint (A.18) when computing the total derivative of the objective function with respect to e. Maintaining r_s and r_f constant in equation (A.18), we have

$$\frac{\partial x}{\partial e}_{|r_f = r_s = \hat{c},\ \text{(A.18) binds}} = \frac{\kappa + (1 - q)^N}{2 - q\beta - 2(1 - q)^N}. \tag{D.1}$$

We thus obtain

(D.2)

$$U'(e)_{|r_f = r_s = \hat{c},\ \text{(A.18) binds}} = \frac{\partial U}{\partial e} + \frac{\partial U}{\partial x}\frac{\partial x}{\partial e}_{|r_f = r_s = \hat{c},\ \text{(A.18) binds}}$$

$$= \frac{1}{2}\left[(\nu - 1)(1 - q)^N - \kappa\right] + \left[(\nu - 1)(1 - q)^N - k\right]\frac{\kappa + (1 - q)^N}{2 - q\beta - 2(1 - q)^N}. \tag{D.3}$$

This term is positive if and only if $k \leq \underline{\tilde{k}}_N$ with $\underline{\tilde{k}}_N$ defined in equation (B.1). 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 equation (A.18) 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 (B.1) does not hold while still assuming $r_s = r_f = \hat{c}$. Then the analysis above shows that setting $e = 0$ is optimal. Since $k > \tilde{\underline{k}}_N$ and thus $k > (\nu - 1)(1-q)^N$, the collateral amount x is pinned down by saturating constraint (A.18) with $e = 0$. The analysis is then similar to that of Case 2 of Proposition 3 and thus the same contract obtains. The threshold with the full-hedging region changes from $\underline{k}_N$ to $\tilde{\underline{k}}_N$ but threshold $\bar{k}$ over which $r_s = r_s = \hat{c}$ is no longer optimal remains the same.

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

For the same reaons, the analysis of the case $k \geq \bar{k}$ is similar to that of Case 3 of Proposition 3. This concludes the proof.

B. Proof of Corollary 1 with optimal monitoring

We extend Corollary 1 taking into account investor's optimal monitoring choice analyzed in Section A.G. We show that the comparative statics with respect to N 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.G, 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 $\underline{k}_N^m < \bar{k}$ of the essential CCP region.

By Proposition 3 and A.1, we have $\underline{k}_N^m \geq \underline{k}_N$ because the region with full hedging in which a CCP is not essential is larger without monitoring. Define $\hat{k}^m$ as the threshold such that investors are indifferent between the complete loss mutualization contract with monitoring and the full hedging contract without monitoring. This threshold solves

$$0 = U_{k=\hat{k}^m} - U^{\not{m}}_{|k=\hat{k}^m} \tag{D.4}$$

$$= qR + \left[\nu - 1 - \hat{k}^m\right]\frac{\hat{c}}{2} - (\hat{k}^m - \underline{k}_N)\left(\frac{\hat{c}}{2} - x^{OM}\right) - \psi - \left\{qR + \left[\nu - 1 - \hat{k}^m\right]\frac{\hat{c}}{2}\right\} \tag{D.5}$$

$$= \beta q \left(1 - \frac{\hat{c}}{2}\right) \frac{\hat{k}^m - \underline{k}_N}{2\left[1 - (1-q)^N\right] - \beta q} - \psi. \tag{D.6}$$

Two cases are then possible. Either $\hat{k}^m \leq \underline{k}_N^{\not{m}}$ which implies $\underline{k}_N^m = \hat{k}^m$ or $\hat{k}^m > \underline{k}_N^{\not{m}}$ and $\underline{k}_N^m = \underline{k}_N^{\not{m}}$.We thus have

$$\underline{k}_N^m = \min\{\hat{k}^m, \underline{k}_N^{\not{m}}\}, \tag{D.7}$$

with $\hat{k}^m$ defined implicitly by equation (D.6) and $\underline{k}_N^{\not{m}} = (\nu - 1)(1 - \alpha q)^N$.

The second argument of the min in equation (D.7) strictly decreases with N by definition. We are thus left to show that $\hat{k}^m$ strictly decreases with N as well. 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.6). We obtain

$$\frac{\partial k}{\partial N} = -\frac{\frac{\partial g}{\partial y}\frac{\partial \bar{y}}{\partial N}}{\frac{\partial g}{\partial k}}, \tag{D.8}$$

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

$$0 < \frac{\partial g}{\partial y} \quad \Leftrightarrow \quad 0 < \frac{(\nu-1)(2-\beta q)-2k}{[2+2y-\beta q]^2} = \frac{2(\bar{k}-k)}{[2+2y-\beta q]^2}. \tag{D.9}$$

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

C. Proof of Proposition C.1

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 equation (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

$$\mathbb{E}[r_o(d)] = \left[q^3 + 3(1-q)q^2\right]\hat{c} + 3q(1-q)^2\bar{r}_3(2,x) + (1-q)^3 2x = x(2-q\beta) + q\beta. \tag{D.10}$$

Solving for x in equation (D.10), we find x_{IC} as given by equation (C.3). 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 equation (3) with respect to x, we obtain

$$U'(x) = \begin{cases} (\nu-1)\Big[(1-q)^3 + q(1-q)^2(3-R)\Big] - k & \text{if} \quad x \in [x_{IC}, \hat{x}_2(3)], \\ \underline{k}_2 - k & \text{if} \quad x \in [\hat{x}_2(3), \frac{\hat{c}}{2}]. \end{cases} \tag{D.11}$$

To obtain the derivative $\frac{\partial \mathbb{E}[r_o(d)]}{\partial x}$ for the first expression, we use the middle term of equation (D.10). 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 $k \in [\underline{k}_3(2), \bar{k}]$