JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS

doi:10.1017/S0022109023000121

The Design of a Central Counterparty

John Chi-Fong Kuong
Department of Finance of INSEAD
john.kuong@insead.edu

Vincent Maurin
Department of Finance of the Stockholm School of Economics, Swedish House of Finance
vincent.maurin@hhs.se (corresponding author)

Abstract

This article 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 to manage counterparty risk. 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.

I. Introduction

To decrease counterparty risks in over-the-counter (OTC) markets, regulators mandated the 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,

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 2022 BI Oslo, 2020 CMU Tepper, 2020 ICEF HSE, INSEAD, 2020 Stockholm School of Economics, 2021 UC3M, 2021 HKIMR, 2021 Vienna, the 2020 Asia-Pacific Corporate Finance Online Workshop, the 2021 Australasian Finance and Banking Conference, the 2022 CICF, the 2022 FT Webinar, the 2020 Gerzensee Workshop on Money, Payments, Banking and Finance, the 2020 Kelley Junior Finance Conference, the 2021 Microstructure Exchange, the 2021 NFN Young Scholars Webinar Series, 2021 OxFIT, the 2021 Paris December Finance Meeting, the 2021 SEC CFMR Conference, the 2021 SGF Conference, and the 2021 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.

[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 European Market Infrastructure Regulation (EMIR) regulation introduced similar requirements. See Spatt (2017) for an in-depth discussion of 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

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

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 prespecified "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 (ABN-AMRO,[3] Allianz, Barclays, and BlackRock (2020), CCP12[4] (2021)).

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 if investors are creditworthy. Investors' creditworthiness improves when their counterparty exerts due diligence to ascertain their financial soundness, which requires a monitoring effort.

Central clearing via a CCP can add value in two ways. First, the CCP can mutualize losses between investors (or members) due to idiosyncratic counterparty risk. The CCP then channels its own contributions and those of solvent members to members with defaulted counterparties, similar to a default fund in practice. Second, members can delegate counterparty monitoring to the CCP.[5] 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 occur only 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 a 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 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 expose their counterparties to "wrong-way

2013, the share of CCP-based repos increased from 42% to 71%, whereas bilateral repos declined from 50% to 19%.

[3]ABN-AMRO is a private bank.

[4]CCP12 is a global association of 37 members who operate more than 60 individual CCPs.

[5]A CCP can 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)).

risk."[6] 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 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 CCP defense against counterparty risks in practice, along with collateral requirements.

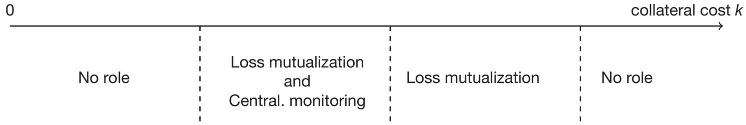
Figure 1 summarizes our results thus far by showing the possible roles of a CCP in the optimal contract against the cost of collateral. Central clearing can help

[6]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 (Basel Committee on Banking Supervision (BCBS) (2019))).

FIGURE 1

CCP Roles

Figure 1 shows the roles assumed by a CCP as a function of the collateral cost.



investors mutualize losses, and the CCP agent can also play an active role by monitoring investors.

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 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 CCPs' default waterfall 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 (London Clearing House (LCH) (2015))), CCP12 (2021)). We argue that small SITG observed in practice does not necessarily imply that CCPs lack incentives to manage risks, as incentives also come from their equity-like compensation.[7]

Our analysis delivers novel predictions about the size of SITG and hence the composition of the default waterfall. Empirically, the ratio of CCP capital to total prefunded resources (CCP capital plus members' collateral) varies substantially across CCPs (Paddrik and Zhang (2020)). We show that SITG relative to either total prefunded resources or CCP profit decreases when the number of members increases, due to the decline in the CCP's agency rent. This effect is compounded if larger CCPs have more bargaining power, and can thus resist 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. 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.

[7]Our novel implication resonates with McPartland (2021), who argues that no capital is needed for a 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.

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

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).[8] We analyze the tension between the mutualization of losses and the incentives to identify creditworthy counterparties, a version of the classic insurance versus incentive trade-off (Stiglitz (1974), Holmström (1979)).[9] 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 the consideration of the CCP as an agent rather than a mechanism designer. This feature allows us to endogenize the CCP's default waterfall (including SITG), the CCP compensation, and the optimal ownership structure of the CCP. 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 a CCP's default waterfall. Wang, Capponi, and Zhang (2022) also stress the need to align members' risk-management incentives and show that prefunded 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 a pecking order between types of collateral is absent in our analysis.

Instead, we endogenize another key element of the waterfall, CCP SITG capital, as part of a solution to the counterparty monitoring problem. Huang (2019) argues that for a given loss allocation, for-profit CCPs under-supply loss-absorbing capital to shift liabilities to surviving members.[10] 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 a CCP and predict that third-party CCPs dominate member-owned CCPs in large or opaque markets, due to endogenous economies of scale, as in Diamond (1984).

Our article focuses on CCPs' role in mitigating counterparty risks, which is most relevant to the default waterfall design. We thus abstract from other important

[8]Vuillemey (2020) provides an empirical analysis of counterparty risk hedging in a 19th-century CCP. Kuong (2021) argues that in the presence of collateral fire-sale externalities, a CCP can mitigate counterparty risk in repo markets by coordinating members' ex ante choices of margins and repo rates. See Vuillemey (2019) for related arguments and empirical analyses.

[9]Keoppl (2013) studies collusive moral hazard in central clearing and Palazzo (2016) argues that central clearing may foster peer monitoring for previously nonconnected investors.

[10]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 on CCP capital.

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

benefits of central clearing that have been discussed in the literature (see, e.g., 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 settlement and providing credit to clearing members.

The rest of the article 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).

1. 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.[11]

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

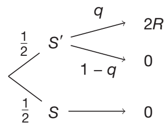
$$U_S = \frac{1}{2}\mathbb{E}[c(S')] + \frac{1}{2}\mathbb{E}[c(S) + (v-1)\min\{c(S), \hat{c}\}], \tag{1}$$

[11]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 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)).

FIGURE 2

Asset Payoff

Figure 2 shows the payoff tree for an S investor's asset with q the probability of success in state S'.



where $c(S)$ is consumption in state S, $v > 1$, 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 $(v-1)$ from consumption in state S until their consumption reaches $\hat{c}$. Investors' preferences exhibit risk aversion because their marginal utility differs across aggregate states.[12] Meanwhile, $(v-1)$ 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 $\{q, R, \hat{c}, v\} = \{0.8, 2, 1.8, 1.2\}$. 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 an 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

$$U = \frac{1}{2}[q(R-1.8)] + \frac{1}{2}q[1.8 + (v-1)1.8] = qR + \frac{1}{2} \times 0.2 \times 1.8 = 1.78. \tag{2}$$

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 that 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

[12]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 generated by a 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.

also relevant in OTC markets where a counterparty's overall risk exposure may be difficult to assess due to lack of transparency.

2. 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, an investor can use cash 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 and government bonds.[13]

3. Central Counterparty

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 refer to this cash contribution as CCP capital, with κ representing the (net) cost of capital. In addition to contributing capital, the CCP can monitor investors; its monitoring effort, however, 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. In Section III, we discuss the mapping to a cleared contract in practice. A contract specifies transfers, and, if necessary, a monitoring scheme: bilateral (counterparty) monitoring or centralized (CCP) monitoring. To streamline the exposition in the main text, we focus on contracts with monitoring. The optimal contract without monitoring is derived in the proof of Proposition 4, when we characterize conditions for monitoring to be optimal.

[13]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).

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

With monitoring, all investors have the same asset pledgeability β and thus, a single contract is offered to all investors.[14] 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. The 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 $d \in \{0, \ldots, N\}$ 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 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 (o, d), an S-investor receives the same transfer $r_o(d)$ in state S as an S'-investor in state S'. The same applies to payments. Transfers $r_s(N), p_s(N), r_f(0)$, and $p_f(0)$ 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

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

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

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

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

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

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

[14]As we show in Proposition A1 in Appendix A, 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.

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

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

$$\text{(MICbm)} \qquad \text{If } 1_{\text{cm}} = 0, \quad \frac{\psi}{q(1-\alpha)} \leq \frac{1}{2}\left(\mathbb{E}_s[r_o(d)] - \mathbb{E}_f[r_o(d)]\right) + \frac{v-1}{2}\left(\mathbb{E}_s[\min\{r_o(d), \hat{c}\}] - \mathbb{E}_f[\min\{r_o(d), \hat{c}\}]\right),$$

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.[15] We discuss the elements of the investors' problem below.

An 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 Appendix A. The second term of equation (3) captures gains from trades; the third, fourth, and fifth terms represent costs from using collateral, monitoring, and CCP compensation.

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

$$\text{(7)} \qquad \frac{1}{2}\left\{q(x + (1-x)2R - \mathbb{E}_s[p_o(d)]) + (1-q)\left(x - \mathbb{E}_f[p_o(d)]\right)\right\} \geq \frac{1}{2}\left(x - \mathbb{E}_f[p_o(d)]\right) + B(1-x),$$

which says that a nonshirking payer's expected consumption (and utility) is weakly higher than a shirking payer's expected consumption and 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.[16]

---

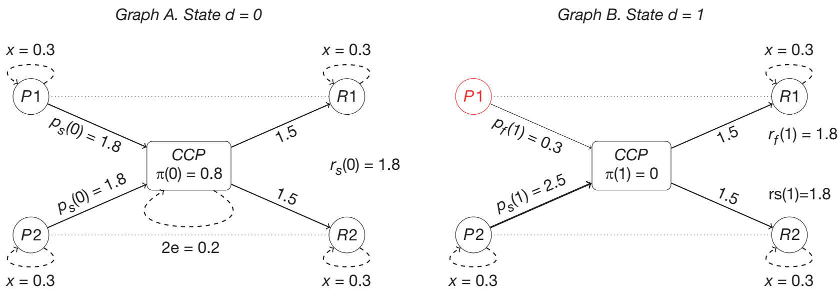
[15]Specifically, x and e determine the amount of collateral and CCP capital at date 0. The date-1 consumption for receivers, payers with succeeded assets, payers with failed assets, and the CCP agent are, respectively, $r_o(d)$, $x + (1-x)2R - p_s(d)$, $x - p_f(d)$, and $\pi(d)$.

[16]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.

FIGURE 3

Example – Contract with $N = 2$ and $\hat{c} = 1.8$

In Figure 3, date-1 transfers are represented in states $d = 0$ (Graph A) and $d = 1$ (Graph B). 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.



The second key constraints are the monitoring incentive constraints, equation (MIC_{cm}) or equation (MIC_{bm}), imposed because monitoring efforts are unobservable. Under the centralized monitoring scheme, equation (MIC_{bm}) 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 states 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 $[q(1-\alpha)]^{-1}$.

Figure 3 provides a numerical example of a contract and the transfers involved at date 1 with $N = 2$ pairs of investors. In Graph A, no payer defaults ($d = 0$) and receivers consume their desired amount $\hat{c} = 1.8 = r_s(0)$. In Graph B, 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 $\left(r_s(1) = r_f(1) = 1.8\right)$ because loss mutualization is at play. The comparison across graphs 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 $\left(r_s(1) = r_f(1) = 1.8\right)$, defaulter's payment $p_f(1) = 0.3$, and the CCP's compensation $\pi(1) = 0$, $p_s(1) = 2.5$ are obtained. More generally, this example shows that the CCP enables loss mutualization by aggregating payers' resources and redistributing them 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 from her counterparty.

Assumption 2 (monitoring cost). $\psi \le \overline{\psi} \equiv \min\left\{\frac{(1-q)(v-1)}{v(2-\beta\alpha q)(1-\alpha q)}, \frac{1}{2}\right\}\beta q(1-\alpha)\left(1-\frac{\hat{c}}{2}\right)$

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). $N \le \frac{2R}{\hat{c}}$

Assumption 3 ensures that the hedging demand $N\hat{c}$ of all receivers can be satisfied even if only one payer's asset succeeds, as the asset pays out $2R$ in this case. This implies that the resource constraint (4) is slack for all $d \le N-1$. Assumption 3 simplifies our analysis in that the only aggregate risk receivers must bear is that of all payers' joint default.[17]

Assumption 4 (capital cost). $\kappa \ge k$

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.[18]

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, $r_f(d) = r_f \le r_s$ if at least one (other) payer survives ($d < N$), and $r_f(N) = 2x + e \le r_f$ if all payers default.

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

[17]In Appendix IA.1 of the Supplementary Material, we derive the optimal contract when Assumption 3 fails 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.

[18]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.

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

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 the variability of their transfers. Yet, transfers may be state-contingent for two reasons. First, receivers are exposed to the aggregate risk of a joint payer default. In this state of the world, by budget constraint (6), their transfer $r_f(N)$ cannot exceed precommitted resources $2x+e$, as no payer survives. Second, investors may optimally retain some counterparty risk exposure ($r_s > r_f$) to satisfy the bilateral monitoring constraint (MIC_{bm}). For payers now, it is optimal to set $p_f(d) = x$ because a larger payment in default relaxes investors' limited pledgeability constraint, equation (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 with whom he matched. The second term captures loss-mutualization transfers across investor pairs that compensate receivers whose payer defaults. 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 the surviving payer's contribution to the CCP compensation.

IV. Clearing with Observable Monitoring

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

A. Frictionless Benchmark

In the frictionless benchmark, there are 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 $k \leq \underline{k}_N \equiv (v-1)(1-q)^N$, a full-hedging contract with $r_s = r_f = \hat{c}$ and $(x,e) = \left(\frac{\hat{c}}{2}, 0\right)$,

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

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 to be 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.[19] Central clearing via the CCP is needed, however, to implement loss mutualization.

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}). We refer to the optimal contract in this case as the observable monitoring (OM) contract.

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

[19]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 $\kappa < k$.

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 $\bar{k} \equiv \frac{1}{2}(v-1)(2-q\beta) > \underline{k}_N$ such that the OM contract is as follows:

1) For $k \leq \underline{k}_N$, it is the full-hedging contract of the frictionless case (Proposition 2).
2) For $k \in \left[\underline{k}_N, \bar{k}\right]$, there is complete loss mutualization: $r_s^{\mathrm{OM}} = r_f^{\mathrm{OM}} = \hat{c}$, $e^{\mathrm{OM}} = 0$ and

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

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

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 too costly.

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 be expanded only by pledging collateral.

When collateral is needed for central clearing, as Proposition 3 shows, 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 cases 1 and 3 can be implemented bilaterally. In both cases, CCP capital is too expensive to be used for insurance. Additionally, loss mutualization is not used for different reasons. When collateral is cheap (case 1), the payer's transfer is fully backed by collateral ($p_o^{\mathrm{OM}} = x$) and the receiver is fully hedged ($r_o^{\mathrm{OM}} = 2x = \hat{c}$), which leaves no counterparty risk to mutualize. When collateral is expensive (case 3), loss mutualization, which requires collateral, is too costly.

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

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. Additionally, 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 (high).

The effect of central clearing on collateral needs depends on the contract that 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 $k \in \left[\hat{k}^m, \bar{k}\right]$ to show how the incentive problem in monitoring affects the contract design and the role of the CCP.[20]

[20]In Appendix IA.2 of the Supplementary Material, 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.

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

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 incentives matter only if the OM contract is not incentive-compatible with bilateral monitoring. The following lemma describes the corresponding parameter region.

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

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

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 $k \in \left(\underline{k}_N, \bar{k}\right)$ 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 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" $\hat{c} - 2x^{\text{OM}}$ is also reduced.

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

Assumption 5. $k \in \left[\hat{k}^m, \bar{k}\right]$ 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. Finally, 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 $\bar{k}^{\text{bm}} = \frac{1-q}{1-q+vq}\bar{k}$. The optimal contract with incentive-compatible bilateral monitoring is as follows:

1) If $k \le \bar{k}^{\text{bm}}$, 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, $x^* > x^{\text{OM}}$.

2) If $k \in \left[\bar{k}^{\text{bm}}, \bar{k}\right]$, 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, $x^* < x^{\text{OM}}$.

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 $\left(k < \bar{k}^{\text{bm}}\right)$. This is intuitive because a larger transfer to receivers requires more collateral to increase investors' excess payment capacity.

The main takeaway 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 $x^* > x^{\text{OM}}$. The CCP breaks even; its compensation and capital contribution are given by

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

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

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

Proposition 6 delivers two new insights for the CCP compensation and capital contribution. Regarding compensation, the CCP should get paid only 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

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

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$_{\text{cm}}$). As this compensation scheme implies that the CCP loses all of its promised compensation when one or more payers default, the CCP effectively holds a junior tranche and absorbs losses immediately after the defaulters' pre-committed resources (i.e., collateral) have been exhausted.[21]

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 that the CCP earns from monitoring. Indeed, equation (PC$_{\text{CCP}}$) 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.[22] 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, that is, $d = 1$, then $d = 2$, and so on. Additionally, even if resource constraints bind for low-d states, a single monitor, hence a single CCP, remains optimal.

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.

In both graphs, 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

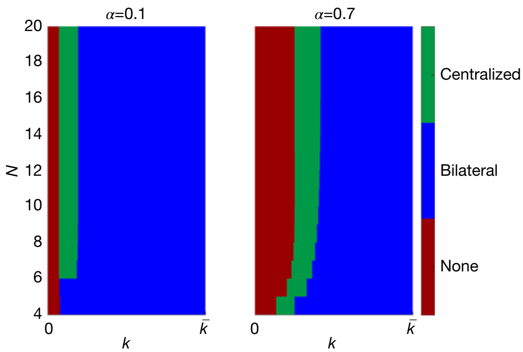
[21] In practice, for-profit CCPs also collect noncontingent fees from members. In our model, if, instead, the CCP has bargaining power, it would charge such fees to extract members' benefits from central clearing (formal results are available from the authors). In contrast, the high-powered compensation described in Proposition 6 does not depend on bargaining power, as it is used to efficiently sustain the CCP's monitoring incentives.

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

FIGURE 4

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

Figure 4 shows the parameter region in which centralized monitoring, bilateral monitoring, or no monitoring is optimal. The collateral cost k is on the x-axis and the number of pair of investors N is on the y-axis. The parameter alpha captures the severity of the monitoring friction.



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 be optimal only 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 $k \in \left[\hat{k}^{\mathrm{cm}}, \bar{k}^{\mathrm{cm}}\right]$, where $\hat{k}^{\mathrm{cm}} > \hat{k}^{\mathrm{m}}$ and $\bar{k}^{\mathrm{cm}} < \bar{k}$. This region is nonempty, as $\hat{k}^{\mathrm{cm}} < \bar{k}^{\mathrm{cm}}$ 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 $\hat{k}^{\mathrm{cm}}$ and $\bar{k}^{\mathrm{cm}}$ both increase with α in the proof.

D. Bargaining over CCP Capital

Our model follows the principal-agent literature in assuming that the principal (here the 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,

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

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 of 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 of 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 article. 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 conduct on-site 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 are thus paramount and, as we discuss below, have implications for the loss allocation process. Therefore, the two key roles of a CCP in our article are intertwined.

B. Default Waterfall Design

Our analysis of the loss mutualization role of CCPs explains important features of the loss allocation process, also known as the CCP's default waterfall. 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 are 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 sheds light on the intense debate about the size of CCP capital, the so-called skin in the game (SITG). SITG is, in general, a small fraction of total prefunded 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.[23] The view that SITG is an outcome of bargaining is acknowledged by market participants.[24]

While SITG is on average a small fraction of total prefunded resources, there is substantial heterogeneity across asset classes and jurisdictions of the CCPs.[25] 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). Additionally, we predict that market size (number of members) is also a determinant of such ratio.

[23]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)).

[24]In 2020, a group of 20 major institutional investors and investment banks 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 judgment 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" (International Swaps and Derivatives Association (ISDA) (2019))).

[25]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 prefunded resources $\frac{e^*}{x^*+e^*}$ strictly increases with β and decreases with N.

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 metric 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.[26]

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. Additionally, 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 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 is not necessarily the case.

[26]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 that this number can vary from 0.3 to 9.51. The data is self-collected from the various CCPs' disclosure and reports in 2020.

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 article, 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. We show that when the market is large 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 SITG capital.

To understand the basic determinants of the default waterfall, we assumed that 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.

Appendix A. Proofs

A.1. Derivation of equation (3)

We first derive the expected version of equation (6), given by

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

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

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

$$+ \sum_{d=0}^{N-1} (1-q)^d q^{N-1-d} \binom{N}{d} \pi(d), \tag{A-3}$$

$$\text{(A-4)} \qquad = \mathbb{E}_s[r_o(d)] + \frac{1-q}{q}\sum_{l=0}^{N-2}(1-q)^l q^{N-1-l}\binom{N-1}{l}\Big(r_f(l+1) - p_f(l+1)\Big), \\ -\frac{(x+e)}{q}\Big[1-(1-q)^N\Big] + \frac{1}{q}\Big[\mathbb{E}[\pi(d)] - (1-q)^N \pi(N)\Big]$$

$$\text{(A-5)} \qquad = \mathbb{E}_s[r_o(d)] + \frac{1-q}{q}\left(\mathbb{E}_f[r_o(d)] - \mathbb{E}_f[p_o(d)]\right) - \frac{x+e}{q} + \frac{\mathbb{E}[\pi(d)]}{q},$$

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

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

$$\text{(A-6)} \qquad U = \frac{1}{2}(q(1-x)2R + x - \mathbb{E}[p_o(d)]) + \frac{1}{2}(\mathbb{E}[r_o(d)] + (v-1)\mathbb{E}[\min\{r_o(d), \hat{c}\}]) \\ -(1-\mathbb{1}_{cm})\psi.$$

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

$$\text{(A-7)} \qquad U = qR + \frac{1}{2}x - qRx + \frac{1}{2}(x+e) - \frac{1}{2}\mathbb{E}[\pi(d)] + \frac{v-1}{2}\mathbb{E}[\min\{r_o(d), \hat{c}\}] \\ -(1-\mathbb{1}_{cm})\psi,$$

which is equivalent to equation (3).

A.2. Proof of Proposition 1

We prove the results in several steps. Step 1 proves that resource constraint (4) 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 contracts with $2x + e \le \hat{c}$ without loss of generality. Finally, in Step 5, we prove that $r_f > r_f(N)$. For some arguments in this proof, we will refer to certain contracts introduced later in the main text.

Step 1. Resource constraint (4) 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 that limited pledgeability constraint (LP) still holds. Consider then a joint increase in $r_f(N)$ and $\mathbb{E}_s[r_o(d)]$ such that

$$\text{(A-8)} \qquad \Delta r_f(N) \le \Delta p_f(N), \quad \Delta\mathbb{E}_s[r_o(d)] \ge v\Delta\mathbb{E}_f[r_o(d)], \quad \text{and} \quad \Delta\mathbb{E}_s[r_o(d)] \le \Delta\mathbb{E}_s[p_o(d)].$$

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 that there are two states (d, d') such that $r_s(d) > r_s(d')$. We argue that the following perturbation weakly increases investors' utility: a decrease in $r_s(d)$ and $p_s(d)$ and an increase in $r_s(d')$ 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 limited pledgeability 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. $r_f(d) = r_f \le \min\{r_s, \hat{c}\}$ for all $d < N$.

We first show that setting $r_f(d) = r_f$ for all $d < N$ is optimal. Suppose instead that 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 $r_f(d) > r_f(d') \ge \hat{c}$ or if $r_f(d') < r_f(d) \le \hat{c}$. Hence, we are left to analyze the case in which $r_f(d') < \hat{c} < r_f(d)$. For $\varepsilon > 0$ small enough, consider the following perturbation:

$$\text{(A-9)} \qquad \left(\Delta r_f(d'), \Delta r_f(d)\right) = \left(\varepsilon, -\frac{f(d')}{f(d)} v\varepsilon\right),$$

with $f(d)$ the probability that d payers default among $N-1$. The perturbation is designed such that the right-hand side of bilateral monitoring 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 $\Delta p_s(d') = \frac{1-q}{q} \Delta r_f(d')$. Limited pledgeability constraint (LP) still holds after the perturbation as the expected payment $\mathbb{E}_s[p_o(d)]$ increases by

$$\text{(A-10)} \qquad \Delta \mathbb{E}_s[p_o(d)] = -\frac{1-q}{q}(v-1) f(d')\varepsilon.$$

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

We then show that $r_f \le \min\{r_s, \hat{c}\}$ is optimal. The result $r_f \le \hat{c}$ 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 bilateral monitoring constraint (MIC_{bm}).

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

$$\text{(A-11)} \qquad \Delta r_f < 0, \qquad \Delta r_s = -\frac{1-q-(1-q)^N}{q} \Delta r_f, \qquad \text{such that } r_f + \Delta r_f = r_s + \Delta r_s.$$

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 (A-1). This implies that limited pledgeability constraint (LP) still holds. Hence, the perturbation is feasible under limited pledgeability constraint (LP) and bilateral monitoring constraint (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

$$xk + \frac{1}{2}e\kappa > \frac{\hat{c}}{2}k + \frac{1}{2}e(\kappa - k) > \frac{\hat{c}}{2}k, \tag{A-12}$$

where the last inequality follows from Assumption 4. The last expression is the cost for 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 (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 $\frac{v-1}{2}(2 - q\beta) \geq k$, 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 $\frac{v-1}{2}(2 - q\beta) > k$, it is optimal to decrease x until it reaches 0 so that

$$\mathbb{E}[r] = q\beta - \kappa e - 2\psi. \tag{A-13}$$

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 limited pledgeability constraint (LP) binds. The argument when $\frac{v-1}{2}(2 - q\beta) > k$ is similar to that above. Suppose then that $\frac{v-1}{2}(2 - q\beta) \leq k$. This implies that x should be increased until $r_s = \hat{c}$. Increasing r_f, however, entails an additional cost because the bilateral 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 that $r_s = \hat{c}$ and limited pledgeability constraint (LP) and bilateral monitoring constraint (MIC_{bm}) hold as equality. This contract is the contract considered in case 2 of Proposition 5, and, as we show there, it satisfies $r_f \geq 2x + e$ under Assumption 2. This concludes the proof. □

A.3. Proof of Proposition 2

Using Proposition 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

$$\text{(A-14)} \quad \max_{x,e,r_s,r_f} \frac{v-1}{2}\left[q\min\{r_s,\hat{c}\} + (1-q)\left(\left[1-(1-q)^{N-1}\right]\min\{r_f,\hat{c}\} + (1-q)^{N-1}(2x+e)\right)\right] \\ -x(qR-1) - \frac{1}{2}e\kappa.$$

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 the following variables:

$$\text{(A-15)} \qquad U'(e) = \frac{1}{2}(v-1)(1-q)^N - \frac{1}{2}\kappa, \quad \text{and}$$

$$\text{(A-16)} \qquad U'(x) = (v-1)(1-q)^N - k.$$

First, equations (A-15) and (A-16) 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-16) shows that setting $r_f(N) = 2x$ equal to $\hat{c}$ is optimal if and only if $k \leq (v-1)(1-q)^N$. This concludes the proof. □

A.4. Proof of Proposition 3

Step 1. Limited Pledgeability Constraint.

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

$$\text{(A-17)} \quad q\left(\mathbb{E}_s[p_o(d)] - \mathbb{E}_f[p_o(d)]\right) = \mathbb{E}_s[r_o(d)] - 2x + \kappa e$$

$$\text{(A-18)} \qquad = qr_s + (1-q)\left[1-(1-q)^{N-1}\right]r_f - \left[1-(1-q)^N\right](2x+e) + \kappa e.$$

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

$$\text{(A-19)} \quad qr_s + (1-q)\left[1-(1-q)^{N-1}\right]r_f \leq q\beta + \left(2 - q\beta - 2(1-q)^N\right)x - \left[\kappa + 1\left\{1-(1-q)^N\right\}\right]e.$$

Investors thus solve the problem described in equation (A-14) under constraint (A-19).

Step 2. Analysis.

We first show that the optimal level of CCP capital is $e^{\mathrm{OM}}=0$. We showed in Proposition 2 that $e=0$ is optimal in the absence of the pledgeability friction. In the presence of limited pledgeability constraint (LP), equation (A-19) shows that increasing e tightens this constraint. Hence, setting $e^{\mathrm{OM}}=0$ remains optimal.

We now argue that we can consider two different cases for the analysis: Either $r_s=r_f=\hat{c}$ or constraint (A-19) binds. This observation follows from Proposition 2, where we showed that $r_s=r_f=\hat{c}$ is optimal in the absence of constraint (A-19). Additionally, the relative weight on these two variables is the same in the objective function in equation (A-14) and in constraint (A-19).

Suppose first that $r_s=r_f=\hat{c}$ and $k\leq\underline{k}_N=(v-1)(1-q)^N$. Then, increasing x until $r_f(N)=2x$ equals $\hat{c}$ is optimal because it increases investors' utility as shown by condition (A-16) in the proof of Proposition 2, and it relaxes constraint (A-19). Hence, in this case, the optimal OM contract is the full-hedging contract derived in Proposition 2.

Suppose now that $k>\underline{k}_N=(v-1)(1-q)^N$. 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-19) binds. Otherwise, decreasing x while maintaining (r_s,r_f) constant strictly increases investors' utility because $k>\underline{k}_N=(v-1)(1-q)^N$. A contract with the conjectured properties is optimal if decreasing x when constraint (A-19) binds decreases the objective function. We have in this case

$$\text{(A-20)}\qquad U'(x)=\frac{v-1}{2}\frac{\partial\mathbb{E}[r_o(d)]}{\partial x}_{|e=0,(\mathrm{A19})\,\mathrm{binds}}-k=\frac{v-1}{2}(2-q\beta)-k\equiv\bar{k}-k.$$

The conjecture is thus optimal if $k\in\left[\underline{k}_N,\bar{k}\right]$. 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-20). The optimal contract is then characterized by $e^{\mathrm{OM}}=0$, $x^{\mathrm{OM}}=0$. The values of r_s and r_f are pinned down by the binding constraint (A-19); that is,

$$\text{(A-21)}\qquad r_s+\frac{1-q}{q}\left[1-(1-q)^{N-1}\right]r_f=\beta.$$

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

A.5. Proof of Corollary 1

We prove the result here in the case where monitoring is imposed. The proof for the case where investors can choose whether to monitor is in Appendix IA.2 of the Supplementary Material. 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 $r_f^{\mathrm{OM}}(d)=\hat{c}>p_f^{\mathrm{OM}}(d)+x^{\mathrm{OM}}$ 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. □

A.6. Proof of Corollary 2

The optimal bilateral contract is obtained from Proposition 3 for the case with monitoring and Proposition A1 without monitoring, 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 $\underline{k}_N = (v-1)(1-q)^N$ strictly decreases with N. Hence, by Proposition 3, when $k \in \left[\underline{k}_N, \underline{k}_1\right]$, the multilateral contract features full loss mutualization with $x^{\mathrm{OM}} < \frac{\hat{c}}{2}$, while the optimal bilateral contract features full hedging; that is $x_1^{\mathrm{OM}} = \frac{\hat{c}}{2}$. This proves the result.

A.7. Proof of Proposition 4

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

A.7.1. Optimal Contract Without Monitoring

We first establish that a single (pooling) contract is offered, although investors may have different ex post types. Without monitoring, each investor has pledgeability β with probability α or 0 with probability $1-\alpha$. With unobservable types, a menu of contracts could be used to screen investors. In our environment, however, screening is not possible due to a failure of the Mirrless–Spence sorting condition. The investor type changes the asset pledgeability but investors' utility 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, independent of the investor type.

It 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 A1. Without monitoring, there are two thresholds of collateral cost

$$\underline{k}_N^{\not m} = (v-1)(1-\alpha q)^N, \quad \text{and} \tag{A-22}$$

$$\bar{k}^{\not m} = \frac{1}{2}(v-1)(2-\alpha q \beta), \tag{A-23}$$

such that

1) if $k \leq \underline{k}_N^{\not m}$, then the contract in case 1 of Proposition 3 is optimal,

2) if $k \in \left[\underline{k}_N^{\not m}, \bar{k}^{\not m}\right]$, then a complete LM contract is optimal, with $r_s^{\mathrm{OM},\not m} = r_f^{\mathrm{OM},\not m} = \hat{c}$, and

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

$$\text{(A-24)} \qquad x^{\mathrm{OM},\not{m}} \equiv \frac{\left[1-(1-\alpha q)^N\right]\hat{c}-\beta q}{2\left[1-(1-\alpha q)^N\right]-\beta\alpha q} \in \left(0, \frac{\hat{c}}{2}\right), \quad \text{and}$$

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

A.7.2. Optimal Monitoring Decision

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

Step 1. Threshold condition.

The argument relies on three claims.

The first claim is that for a given monitoring choice, the difference in investor's utility across consecutive contracts is strictly increasing with k. A contract is consecutive to a reference contract if it is the next optimal contract when increasing k. For example, 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 A1, 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 Proposition A1 shows that the desired inequality holds strictly in all cases except in case 3, 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 A1, the optimal contract without monitoring is given by case 1 or 2. In general, investors' utility is given by

$$\text{(A-25)} \qquad U^{\not{m}}_{|k=\bar{k}} = qR + \left[(v-1)-\bar{k}\right]\frac{\hat{c}}{2} + \max\left\{0, \bar{k} - \underline{k}_N^{\not{m}}\right\}\left(\frac{\hat{c}}{2} - x^{\mathrm{OM},\not{m}}\right)$$

$$\text{(A-26)} \qquad = qR + q\beta\frac{v-1}{2}\frac{\hat{c}}{2} + \frac{v-1}{2}\beta\alpha q\left(1-\frac{\hat{c}}{2}\right)\max\left\{0, \frac{2-q\beta-2(1-\alpha q)^N}{2\left[1-(1-\alpha q)^N\right]-\beta\alpha q}\right\}.$$

The last term of $U^{\bar{m}}_{|k=\bar{k}}$ in the second line of (A-26) is increasing in N. Hence, an upper bound for $U^{\bar{m}}_{|k=\bar{k}}$ is obtained by letting $N\to\infty$, that is,

$$\text{(A-27)}\qquad U^{\bar{m}}_{|k=\bar{k}} \le qR+q\beta\frac{\nu-1}{2}\frac{\hat{c}}{2}+\frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta}.$$

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

$$\text{(A-28)}\qquad 0\le U_{k=\bar{k}}-\left\{qR+q\beta\frac{\nu-1}{2}\frac{\hat{c}}{2}+\frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta}\right\}$$

$$\text{(A-29)}\qquad \le qR+\frac{\nu-1}{2}q\beta-\psi-\left\{qR+q\beta\frac{\nu-1}{2}\frac{\hat{c}}{2}+\frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta}\right\}$$

$$\text{(A-30)}\qquad \le\frac{\nu-1}{2}q\beta\left(1-\frac{\hat{c}}{2}\right)-\frac{\nu-1}{2}\left(1-\frac{\hat{c}}{2}\right)\frac{q\alpha\beta(2-q\beta)}{2-q\alpha\beta}-\psi$$

$$\text{(A-31)}\qquad \le\frac{\beta q(1-\alpha)(\nu-1)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right)-\psi.$$

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 lies strictly below $\bar{k}$. This concludes the proof. □

A.8. *Proof of Lemma 2*

Suppose first that $k\in\left[\hat{k}^m,\underline{k}_N\right]$. In this case, by Proposition 3, the OM contract is given by case 3, with $r_s^{\text{OM}}=r_f^{\text{OM}}=r_f^{\text{OM}}(N)$. This implies that the bilateral monitoring constraint (MIC_{bm}) is violated. Suppose now that $k\ge\underline{k}_N$. Under Assumption 5, the OM contract is given by case 2 of Proposition 3, with $r_s^{\text{OM}}=r_f^{\text{OM}}=\hat{c}$, $e^{\text{OM}}=0$, and x^{OM} given by equation (10). Plugging these variables into the bilateral monitoring constraint (MIC_{bm}), we obtain condition (11).

A.9. *Proof of Proposition 5*

We first rewrite the bilateral monitoring constraint (MIC_{bm}) using the results from Proposition 1, as follows:

$$\text{(A-32)}\qquad \frac{\psi}{1-\alpha}\le\frac{1}{2}\left[r_s-r_f+(1-q)^{N-1}\left(r_f-(2x+e)\right)\right]$$
$$+\frac{\nu-1}{2}\left[\min\{r_s,\hat{c}\}-\left(\left[1-(1-q)^{N-1}\right]\min\{r_f,\hat{c}\}+\left(1-q^{N-1}\right)((2x+e))\right)\right].$$

The optimal contract under bilateral monitoring solves problem (A-14) under constraints (A-19) and (A-32), which correspond, respectively, to limited pledgeability constraint (LP) and bilateral monitoring constraint (MIC_{bm}) in the investor's problem. In Step 1, we show that constraints (A-19) and (A-32) bind. In Step 2, we derive the threshold $\bar{k}^{\text{bm}}$. Finally, in Step 3, we characterize the optimal distortion to the OM contract of Proposition 3.

Step 1. Equations (A-19) and (A-32) bind.

Under Assumption 5, constraint (A-32) binds because the OM contract in Proposition 3 violates equation (A-32). The constraint (A-19) must also bind. If it does not, decrease x while keeping r_s and r_f constant. This change relaxes constraint (A-32). Hence, the marginal effect on investors' utility from this perturbation is given by $-U'(x)$ in equation (A-16), which is positive because $k > \bar{k}_N$ by Assumption 5.

Step 2. Threshold $\bar{k}^{\text{bm}}$ 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-32) is not imposed. Hence, it is still optimal under constraint (A-32) because increasing r_s relaxes this constraint. It is also optimal to increase r_f until equation (A-32) 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 $r_f \in \left[\underline{r}_f, \hat{c}\right]$. From equations (A-19) and (A-32), we have (for given x and e).

$$\text{(A-33)} \qquad qr_s + (1-q)\left[r_f - (1-q)^{N-1}\left(r_f - 2x - e\right)\right] = (2-q\beta)x + q\beta - \kappa e \quad \text{and}$$

$$\text{(A-34)} \qquad r_s - v\left[r_f - (1-q)^{N-1}\left(r_f - 2x - e\right)\right] = \frac{2\psi}{q(1-\alpha)} - (v-1)\hat{c}.$$

Hence, we obtain

$$\text{(A-35)} \qquad (1-q)\left[r_f - (1-q)^{N-1}\left(r_f - 2x - e\right)\right] = \frac{(1-q)[(2-q\beta)x + q\beta - \kappa e] - q(1-q)\left[\frac{2\psi}{1-\alpha} - (v-1)\hat{c}\right]}{qv + (1-q)}.$$

We can plug this relationship into the expression for investors' utility in equation (A-14). 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

$$\text{(A-36)} \qquad k \leq \frac{v-1}{2}\frac{1-q}{1-q+vq}(2-q\beta) = \frac{1-q}{1-q+vq}\bar{k} \equiv \bar{k}^{\text{bm}} < \bar{k}.$$

Step 3. Optimal distortion.

CCP capital e tightens (A-32). 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 $\bar{k}^{\text{bm}}$.

Case i. $k \leq \bar{k}^{\text{bm}}$

In this case, $r_f^* = \hat{c}$. Setting $e^* = 0$ and solving for x using equations (A-33) and (A-34), we obtain

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

(A-37)
$$\hat{c}\left[1-q-(1-q)^{N}+q-vq(1-q)^{N-1}\right]-q\beta+\frac{2\psi}{1-\alpha} = \left(2-2(1-q)^{N-1}[vq+1-q]-\beta q\right)x.$$

Hence,

(A-38)
$$x^{*}=\frac{\hat{c}\left(1-(1-q)^{N-1}[vq+1-q]\right)-q\beta+\frac{2\psi}{1-\alpha}}{2-2(1-q)^{N-1}[vq+1-q]-\beta q}>x^{\mathrm{OM}}.$$

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

Case ii. $k\geq\bar{k}^{\mathrm{bm}}$

In this case, $r_s^{*}=\hat{c}$. We then use equations (A-33) and (A-34) to solve for r_f^{*} and x^{*} setting again $e^{*}=0$. We obtain

(A-39)
$$x^{*}=\frac{\hat{c}-q\beta-\frac{2\psi(1-q)}{qv(1-\alpha)}}{2-q\beta}<x^{\mathrm{OM}}\quad\text{and}$$

(A-40)
$$r_f^{*}=\frac{\hat{c}-2(1-q)^{N-1}x^{*}-\frac{2\psi}{vq(1-\alpha)}}{1-(1-q)^{N-1}}.$$

This concludes the proof. □

A.10. Proof of Proposition 6

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

Step 1. CCP compensation schedule.

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

(A-41)
$$\mathrm{IC}(d)=1-\frac{\mathbb{P}[d|\mathrm{shirk}]}{\mathbb{P}[d|\mathrm{effort}]},$$

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

(A-42)
$$\mathbb{P}[d|\mathrm{shirk}]=\sum_{d_m=0}^{d}\binom{n_m}{d_m}\binom{N-n_m}{d-d_m}(1-q)^{d_m}q^{n_m-d_m}(1-\alpha q)^{d-d_m}(\alpha q)^{N-n_m-d+d_m}.$$

After some manipulation, we obtain

$$\text{(A-43)} \qquad \frac{\mathbb{P}[d|\text{shirk}]}{\mathbb{P}[d|\text{effort}]} = \frac{\sum_{d_m=0}^{d} \binom{n_m}{d_m}\binom{N-n_m}{d-d_m}\left[\frac{1-\alpha q}{\alpha(1-q)}\right]^{d-d_m}}{\binom{N}{d}}$$

$$= \sum_{d_m=0}^{d} w_{n_m}(d_m)\left[\frac{1-\alpha q}{\alpha(1-q)}\right]^{d-d_m},$$

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

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

$$\text{(A-44)} \qquad Nq^N\underline{\pi}(0)-2N\psi = \max_{n_m\in[|0,..N-1|]}\left\{Nq^N\alpha^{N-n_m}\underline{\pi}(0)-2n_m\psi\right\},$$

where on the right-hand side, n_m is the number of investor pairs that 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 $g: y \to y\left(1-e^{y\log(\alpha)}\right)^{-1}$ is increasing with y for $y\geq 1$. We have

$$\text{(A-45)} \qquad g'(y) \propto 1-\alpha^y + y\alpha^y\log(\alpha) \geq 1-\alpha(1-\log(\alpha)).$$

The inequality obtains because $y\geq 1$ and $\alpha \leq 1$. We thus have $g'(y)\geq 0$ because $\alpha \mapsto \alpha(1-\log(\alpha))$ is increasing and $\lim_{\alpha\to 1}\alpha(1-\log(\alpha))=1$. Setting $n_m=0$ on the right-hand side of equation (A-44), 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^*.

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 limited pledgeability 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-19), which is equivalent to limited pledgeability constraint (LP), we obtain the following result:

$$\text{(A-46)} \qquad U'(e)_{|r_s^*=r_f^*=\hat{c},\text{(LP)binds}} = \frac{\partial U}{\partial e}+\frac{\partial U}{\partial x}\frac{\partial x}{\partial e},$$

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

$$
\text{(A-47)} \quad = \begin{cases} \dfrac{v-1}{2}(1-q)^N - \left[(v-1)(1-q)^N - k\right]\dfrac{1-(1-q)^N}{2-2(1-q)^N-\beta q} & \text{if } e \le \underline{e}, \\ \dfrac{\kappa+(1-q)^N}{2-q\beta-2(1-q)^N} & \text{if } e > \underline{e}. \end{cases}
$$

Since $k > \underline{k}_N$, $U'(e) \geq 0$ if and only if $e \le \underline{e}$. 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

$$
\text{(A-48)} \quad \hat{c}\left[1-(1-q)^N\right] + (1-q)^N(2x^*+e^*) + \mathbb{E}[\pi^*] = q\beta + (2-q\beta)x^* + e^*.
$$

We obtain

$$
\text{(A-49)} \quad x^* = \frac{\hat{c}\left[1-(1-q)^N\right]-\beta q}{2\left[1-(1-q)^N\right]-\beta q} + \frac{\left(\kappa+1-\left[1-(1-q)^N\right]\right)e^*+2\psi}{2\left[1-(1-q)^N\right]-\beta q},
$$

$$
\text{(A-50)} \quad = x^{\mathrm{OM}} + \frac{2\psi}{(\kappa+1)(1-\alpha^N)} \frac{\kappa+1-\alpha^N\left[1-(1-q)^N\right]}{2\left[1-(1-q)^N\right]-\beta q}.
$$

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

$$
\text{(A-51)} \quad \psi \le \frac{1-\alpha^N}{2-\frac{\beta q \alpha^N}{\kappa+1}}\beta q\left(1-\frac{\hat{c}}{2}\right).
$$

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

A.11. *Proof of Proposition 7*

We first compare centralized monitoring to no monitoring. To avoid confusion, we add a superscript $^{\mathrm{cm}}$ to variables for the optimal centralized monitoring contract. For large N, Proposition A1 shows that the OM contract without monitoring is given by case 2. This is because the condition $k \le \bar{k}$ in Assumption 5 implies $k \le \bar{k}^{\not m}$, and the lower bound of the region $\underline{k}_N^{\not m}$ converges to 0 as N grows large. Using Propositions 6 and A1, we derive the following expressions for investors' utility:

$$
\text{(A-52)} \quad U^{*,\mathrm{cm}} = qR + [v-1-k]\frac{\hat{c}}{2} + \left[k-(v-1)(1-q)^N\right]\left(\frac{\hat{c}}{2}-x^{*,\mathrm{cm}}\right) - \frac{1}{2}\left[(v_{\mathrm{cm}}-1)-(v-1)(1-q)^{N-1}\right]e^* - \psi \quad \text{and}
$$

(A-53) $$U^{\mathrm{OM},\not{m}} = qR + [v-1-k]\frac{\hat{c}}{2} + \left[k-(v-1)(1-\alpha q)^N\right]\left(\frac{\hat{c}}{2} - x^{\mathrm{OM},\not{m}}\right).$$

From Propositions 6 and A1 again, we have

(A-54) $$\frac{\hat{c}}{2} - x^{*,\mathrm{cm}} = \frac{\beta q\left(1-\frac{\hat{c}}{2}\right)}{2\left[1-(1-q)^N\right]-\beta q} - \frac{2\psi}{(\kappa+1)(1-\alpha^N)}\,\frac{\kappa+1-\alpha^N\left[1-(1-q)^N\right]}{2\left[1-(1-q)^N\right]-\beta q}, \quad \text{and}$$

(A-55) $$\frac{\hat{c}}{2} - x^{\mathrm{OM},\not{m}} = \frac{\beta\alpha q}{2\left[1-(1-\alpha q)^N\right]-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right),$$

When $N \to \infty$, e^* converges to 0 at an exponential rate by Proposition 6. The second term of $\frac{\hat{c}}{2} - x^{\mathrm{cm},*}$ above also converges at an exponential rate as $N \to \infty$. In the limit, centralized monitoring dominates no monitoring, that is, $U^{*,\mathrm{cm}} \geq U^{\mathrm{OM},\not{m}}$ if and only if

(A-56) $$\frac{k}{2-\beta q}\left[\beta q\left(1-\frac{\hat{c}}{2}\right) - 2\psi\right] - \psi \geq \frac{k}{2-\beta\alpha q}\beta\alpha q\left(1-\frac{\hat{c}}{2}\right).$$

Under Assumption 2, we have

(A-57) $$\psi \leq \frac{\beta q(1-\alpha)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right).$$

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

(A-58) $$\hat{k}^{\mathrm{cm}} = \frac{2-\beta q}{\frac{\beta q(1-\alpha)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right)-\psi}\,\frac{\psi}{2}.$$

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

(A-59) $$U^* = qR + [v-1-k]\frac{\hat{c}}{2} + \left[k-(v-1)(1-q)^N\right]\left(\frac{\hat{c}}{2} - x^*\right) - \psi.$$

Using equations (A-52) and (A-59), centralized monitoring dominates case 5 of bilateral monitoring if and only if

(A-60) $$\begin{aligned}&\left(k-(v-1)(1-q)^N\right)\left(x^{\mathrm{cm},*} - x^{\mathrm{OM}}\right) + \frac{1}{2}\left(\kappa-(v-1)(1-q)^N\right)e^* \\ &\leq \left(k-(v-1)(1-q)^N\right)\left(x^* - x^{\mathrm{OM}}\right).\end{aligned}$$

Using the expression for the collateral requirement in equation (A-38), we obtain

(A-61) $$x^* - x^{\mathrm{OM}} = \frac{2\psi}{[1-\alpha]\left[2\left(1-(1-q)^N\right)-\beta q\right]} - \frac{vq(1-q)^{N-1}}{2\left(1-(1-q)^N\right)-\beta q}(\hat{c}-2x^*)$$

(A-62)
$$= \frac{2\psi}{[1-\alpha]\left[2\left(1-(1-q)^N\right)-\beta q\right]}$$
$$-\frac{vq(1-q)^{N-1}}{2\left(1-(1-q)^N\right)-\beta q} \frac{\beta q(2-\hat{c})-\frac{4\psi}{1-\alpha}}{2\left[1-(1-q)^{N-1}(vq+1-q)\right]-\beta q}.$$

We thus obtain the following condition

(A-63)
$$\frac{1}{2}\left(\kappa-(v-1)(1-q)^N\right)e^* \leq \left[k-(v-1)(1-q)^N\right](x^*-x^{*,\mathrm{cm}})$$

$$\frac{1}{2}\left(\kappa-(v-1)(1-q)^N\right)e^* \leq \frac{k-(v-1)(1-q)^N}{2\left(1-(1-q)^N\right)-\beta q}\left[\frac{2\psi}{1-\alpha}-\frac{2\psi}{1-\alpha^N}\right.$$

(A-64)
$$\left.-\frac{vq(1-q)^{N-1}\left(\beta q(2-\hat{c})-\frac{4\psi}{1-\alpha}\right)}{2\left[1-(1-q)^{N-1}(vq+1-q)\right]-\beta q}\right].$$

Observe that the terms that 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

(A-65)
$$\left(k-(v-1)(1-q)^N\right)\left(x^{\mathrm{cm},*}-x^{\mathrm{OM}}\right)+\frac{1}{2}\left(\kappa-(v-1)(1-q)^N\right)e^*$$
$$\leq \left[\frac{v-1}{2}(2-q\beta)-k\right]\left(x^{\mathrm{OM}}-x^*\right).$$

Using equation (10) for x^{OM} and equation (A-39) for x^*, we obtain

(A-66)
$$x^{\mathrm{OM}}-x^* = \frac{2\psi(1-q)}{vq(1-\alpha)(2-q\beta)}-\frac{\beta q(2-\hat{c})(1-q)^N}{[2-q\beta]\left[2\left(1-(1-q)^N\right)-\beta q\right]}.$$

We observe again that the terms that 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

(A-67)
$$\frac{\frac{v-1}{2}(2-q\beta)-k}{2-q\beta}\frac{2\psi(1-q)}{vq(1-\alpha)} \geq \frac{k}{2-\beta q}2\psi.$$

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

(A-68)
$$\bar{k}^{\mathrm{cm}} \equiv \frac{1-q}{1-q+vq(1-\alpha)}\bar{k}<\bar{k}.$$

Finally, we are left to derive the maximum value of the monitoring cost ψ such that the interval $\left[\hat{k}^{\text{cm}}, \bar{k}^{\text{cm}}\right]$ is nonempty. Observe that $\bar{k}^{\text{cm}}$ is independent of ψ, while $\hat{k}^{\text{cm}}$ is strictly increasing with ψ. Solving for the value of ψ such that $\hat{k}^{\text{cm}} = \bar{k}^{\text{cm}}$, we get

$$0 = \frac{1-q}{1-q+vq(1-\alpha)}\frac{v-1}{2}(2-q\beta) - \frac{2-\beta q}{\frac{\beta q(1-\alpha)}{2-\beta\alpha q}\left(1-\frac{\hat{c}}{2}\right)-\psi}\frac{\psi}{2}, \tag{A-69}$$

$$0 = (1-q)(v-1)\frac{\beta q(1-\alpha)}{2-\beta\alpha q} - (1-q)(v-1)\psi - \psi[1-q+vq(1-\alpha)], \tag{A-70}$$

$$\psi = \frac{\beta q(1-q)(1-\alpha)(v-1)}{v(2-\beta\alpha q)(1-\alpha q)}\left(1-\frac{\hat{c}}{2}\right). \tag{A-71}$$

This is the first argument of the minimum in the expression for the upper bound on ψ given by Assumption 2. Hence, for any $\psi < \bar{\psi}$, the interval $\left[\hat{k}^{\text{cm}}, \bar{k}\right]$ is nonempty.

A.12. 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 a 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 Proposition 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 that 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^* being 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 $e \in [0, e^*]$,

$$U(e) = qR + \frac{v-1-k}{2}\hat{c} - \frac{1}{2}\mathbb{E}[\pi^*] + \left[k-(v-1)(1-q)^N\right]\left(\frac{\hat{c}}{2} - x - \frac{e}{2}\right) + \frac{1}{2}[1+k]e, \tag{A-72}$$

where x is a function of e given implicitly by equation (A-48), replacing e^* with $e \in [0, e^*]$. When $e = e^*$ the CCP breaks even, while when $e = 0$, 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

$$0 \le 2NU(e^*) - \left(2NU(e=0) + N(\kappa+1)e^*_C\right), \tag{A-73}$$

$$\Leftrightarrow\ 0 \le \left[k-(v-1)(1-q)^N\right]\left(x(e=0) - x^* - \frac{e^*}{2}\right) - [\kappa - k]\frac{e^*_C}{2}, \tag{A-74}$$

https://doi.org/10.1017/S0022109023000121 Published online by Cambridge University Press

$$\text{(A-75)} \qquad \Leftrightarrow \quad 0 \leq \left\{ \frac{k-(v-1)(1-q)^N}{\bar{k}-(v-1)(1-q)^N}\beta q(v-1) - [\kappa - k] \right\} \frac{e_C^*}{2}.$$

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

A.13. 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-50) in the proof of Proposition 6.

First, we prove the comparative statics with respect to β. Proposition 6 shows that 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-48) to obtain

$$\text{(A-76)} \qquad \frac{\partial x^*}{\partial \beta} = -\frac{q(1-x^*)}{2-q\beta-2(1-q)^N} < 0.$$

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

$$\text{(A-77)} \qquad r_{\mathrm{xe}}(N) \equiv \frac{2Nx^*}{Ne^*} = \frac{2x^*}{e^*} = 2\frac{\kappa+(1-q)^N+\frac{\hat{c}\left[1-(1-q)^N\right]-\beta q+2\psi}{e^*(N)}}{2\left[1-(1-q)^N\right]-\beta q}.$$

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

$$\text{(A-78)} \qquad r'_{\mathrm{xe}}(N) = -2\frac{\frac{\partial e^*}{\partial N}}{(e^*)^2}\frac{\hat{c}\left[1-(1-q)^N\right]-\beta q+2\psi}{2\left[1-(1-q)^N\right]-\beta q}$$

$$\text{(A-79)} \qquad -2\log(1-q)(1-q)^N\frac{\left(\frac{\hat{c}}{e^*}-1\right)(2-\beta q)-2\left(\frac{\hat{c}-\beta q+2\psi}{e^*}+\kappa\right)}{\left(2\left[1-(1-q)^N\right]-\beta q\right)^2}.$$

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 $r'_{\mathrm{xe}}(N) > 0$, 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

$$\text{(A-80)} \qquad e^*A = (\hat{c}-e^*)(2-\beta q) - 2(\hat{c}-\beta q+2\psi+\kappa e^*)$$

$$\text{(A-81)} \qquad = \beta q(2-\hat{c}) - 4\psi - 2(\kappa+1)e^* + \beta q e^*$$

$$\text{(A-82)} \qquad = \beta q(2-\hat{c}) - \frac{4\psi}{1-\alpha^N} + \beta q\frac{2\psi\alpha^N}{(\kappa+1)(1-\alpha^N)} = \beta q(2-\hat{c}) - \frac{2\psi}{1-\alpha^N}\left[2-\frac{\beta q\alpha^N}{\kappa+1}\right].$$

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

Appendix 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 B1. There exists a threshold for collateral cost,

$$\text{(B-1)} \quad \underline{k}_N \equiv (v-1)(1-q)^N + \frac{1}{2}\frac{2-q\beta-2(1-q)^N}{\kappa+(1-q)^N}\max\left\{(v-1)(1-q)^N-\kappa,0\right\},$$

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

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

The proof of this result is in Appendix IA.2 of the Supplementary Material. When CCP capital is cheap, the OM contract of Proposition 3 changes only in the full-hedging case. CCP capital can then 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, $(v-1)(1-q)^N$, investors use the cheapest of the two resources to hedge. Case 2 of Proposition B1 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 $k < \underline{k}_N$ 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 it 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.

Supplementary Material

To view supplementary material for this article, please visit http://doi.org/10.1017/S0022109023000121.

References

ABN-AMRO; Allianz; Barclays; and BlackRock. "A Path Forward for CCP Resilience, Recovery, and Resolution." Technical Report, 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, Bank for International Settlements, ed. Basel: Bank for International Settlements (2017).

Basel Committee on Banking Supervision (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." *Journal of Finance*, 71 (2016), 1669–1698.

Board, T. F. S. E. "Financial Stability Report." Technical 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)." Technical Report (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 (2015).

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

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

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

Duffie, D., and H. Zhu. "Does a Central Clearing Counterparty Reduce Counterparty Risk?" *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." Technical Report, BIS (2018).

FSB. "Guidance on Financial Resources to Support CCP Resolution and on the Treatment of CCP Equity in Resolution." Technical Report, 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." *Bell Journal of Economics*, 10 (1979), 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 (2019).

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

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

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

Koeppl, T., and C. Monnet. "The Emergence and Future of Central Counterparties." Working Papers 10–30, 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.

Kuong, J. C.-F. "Self-Fulfilling Fire Sales: Fragility of Collateralized Short-Term Debt Markets." *Review of Financial Studies*, 34 (2021), 2910–2948.

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

London Clearing House (LCH). "CCP Conundrums." Technical Report, 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." *Review of Financial Studies*, 29 (2015), 1747–1779.

McPartland, J. "Challenging Skin in the Game." WFE Focus (2021).

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 (2017), 1–13

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." Technical Report, OCC (2020).

Paddrik, M. E., and S. Zhang. "Central Counterparty Default Waterfalls and Systemic Loss." OFR WP 20-04 (2020).

Palazzo, F. "Peer Monitoring via Loss Mutualization." Bank of Italy Temi di Discussione (Working Paper) No. 1088 (2016).

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

Santos, T., and J. A. Scheinkman. "Competition Among Exchanges." *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." Technical Report, Carnegie Mellon University (2017).

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

Vuillemey, G. "Mitigating Fire Sales with Contracts: Theory and Evidence" (2019).

Vuillemey, G. "The Value of Central Clearing." *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*, 68 (2022), 6355–7064.

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

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