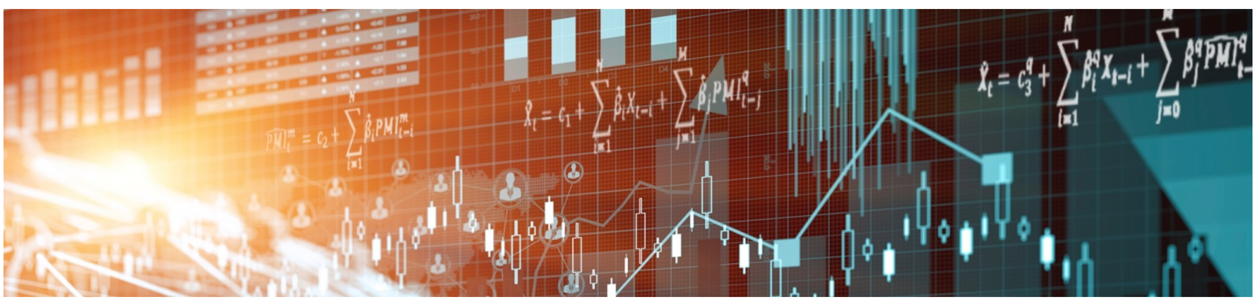
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2016-4

# To Share or Not to Share? Uncovered Losses in a Derivatives Clearinghouse



by Radoslav S. Raykov



www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2016-4

February 2016

# To Share or Not to Share? Uncovered Losses in a Derivatives Clearinghouse

by

**Radoslav S. Raykov**

Financial Stability Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
rraykov@bankofcanada.ca

ISSN 1701-9397 

## Acknowledgements

This research was performed at the Bank of Canada's Financial Stability Department. The author wishes to thank, in alphabetical order, Kartik Anand, Toni Anhert, Darrell Duffie, Itay Goldstein, Miguel Molico, and especially Hector Perez Saiz and Joshua Slive for extensively commenting on earlier versions of this working paper.

## Abstract

This paper studies how the allocation of residual losses affects trading and welfare in a central counterparty. I compare loss sharing under two loss-allocation mechanisms – variation margin haircutting and cash calls – and study the privately and socially optimal degree of loss sharing. For losses allocated using variation margin haircuts, I find that trading volume is sensitive to the degree of loss sharing and to the risk sensitivity of skin-in-the-game capital. By contrast, for cash calls, the degree of loss sharing does not affect trading volume but instead affects the chance that a cash call is honoured, which can constrain the recovery of funds. A welfare analysis characterizes the market outcome and compares it with the social optimum.



*JEL classification: G19, G21*
*Bank classification: Economic models; Payment clearing and settlement systems*



## Résumé

L'étude analyse l'incidence qu'exerce la répartition des pertes résiduelles sur le volume de transactions et le bien-être des participants d'une contrepartie centrale. Deux mécanismes de répartition des pertes sont comparés, soit l'application de décotes aux plus-values sur marge de variation et les appels de fonds. Le degré optimal de répartition des pertes est étudié du point de vue privé et social. Dans le cas des pertes réparties au moyen du premier mécanisme, on constate que le volume des transactions est influencé par le degré de répartition des pertes et par la sensibilité aux risques des « intérêts en jeu ». En revanche, pour ce qui est du second mécanisme – les appels de fonds –, le degré de répartition des pertes n'a aucune incidence sur le volume des transactions; il influe par contre sur la probabilité qu'un appel de marge soit honoré, ce qui peut restreindre le recouvrement des fonds. Une analyse du bien-être est réalisée : elle définit les résultats du marché et les compare aux résultats obtenus en situation d'optimum social.



*Classification JEL : G19, G21*
*Classification de la Banque : Modèles économiques; Systèmes de compensation et de règlement des paiements*

## Non-Technical Summary

This paper proposes a model for analyzing the effects of central counterparty loss-allocation tools – cash calls and variation margin gain haircuts (VMGH) – on the trading and welfare of clearinghouse participants. It is motivated by the increasing importance of centrally cleared markets, arising from the G-20 commitment to centrally clear standardized over-the-counter derivatives.

The paper quantifies, and generally confirms, the intuitive argument put forward by Singh (2015) that losses should be shared to the fullest extent possible. The analysis demonstrates that, in many cases, maximum loss sharing is indeed both privately and socially optimal. However, this paper also identifies an important special case where private and social welfare diverge and banks do not have the right incentives to trade under maximal loss sharing, creating a tradeoff between loss sharing and trading volume.

This special case is important for policy-makers concerned with the uninterrupted functioning of centrally cleared markets. It occurs when VMGH losses are shared fully across members and the CCP uses a risk-based, highly procylcial skin in the game. I show that overly procyclical skin in the game weakens incentives to trade when VMGH losses are shared fully because the CCP effectively subsidizes potential defaults by participants. This leads to a policy-making tradeoff between market activity and fairness in the allocation of losses. As a practical solution to avoid this tradeoff, I propose not using overly risk-sensitive skin in the game when losses are allocated using VMGH. (This consideration does not apply to cash calls.)

In contrast to VMGH, which affects trading volume ex ante, cash calls carry ex post performance risk that can undermine the CCP's ability to collect funds after a stress event, potentially threatening both its recovery and its post-recovery viability. The analysis shows that cash calls do not reduce trading before the onset of stress because clearing members retain the ability to not honor cash calls that they deem too intrusive. While this eliminates ex ante market reactions to the inclusion of cash calls in a recovery plan, it raises a different set of concerns. The model predicts that the banks' risk of non-performance to the cash call rises endogenously in response to higher cash call likelihoods, thereby placing a constraint on the total amount of funds that can be recovered. This can limit the success of the recovery effort, which suggests that CCP resolution authorities should be endowed with powers allowing them to resolve a CCP before recovery tools have become ineffective.

# 1 Introduction

The G-20 reforms committing to the central clearing of standardized over-the-counter derivatives have increased both the centrality and the importance of central counterparties (CCPs). Aimed at fostering transparency and standardization in derivatives markets, these commitments have also concentrated substantial new risks on CCPs by expanding the types of instruments subject to central clearing and by stimulating an increase in centrally cleared volumes (Duffie, 2014). This, in turn, has created concerns that clearinghouses may be becoming the next entities that are "too big" or "too important" to fail (*Financial Times*, 2014; Coeuré, 2015; Singh, 2015; Duffie, 2014). In reaction to these concerns, international regulatory bodies recently released recommendations for CCPs to develop recovery plans – contingency procedures that allow them to allocate uncovered losses to their members in the unlikely event that a default is so large that pre-funded clearinghouse resources are depleted (CPMI-IOSCO 2012, 2014). This paper studies the economic impact of two loss-allocation tools – variation margin gains haircutting (VMGH) and cash calls – that can be used to recover a CCP. The paper compares how these two loss-allocation tools affect the trading volume and welfare of clearinghouse participants and studies the privately optimal and socially optimal degree of loss sharing between survivors.

Loss-allocation tools are intrusive to CCP members because they expose participants to defaults by other members. Usually, loss allocation becomes relevant only after all pre-funded member resources are depleted and some of the CCP's own resources – often called "skin in the game" – are used up. Typically, a clearinghouse holds collateral in the form of margin deposits from participants in addition to a mutualized default fund with contributions from every member. Before loss allocation becomes necessary, then, the defaulter's margin deposit and default fund deposit, the CCP's skin in the game and the sum of all remaining members' default fund deposits must be depleted and insufficient to cover the realized default. (In most CCPs, skin in the game has to be exhausted before losses are mutualized.) Loss allocation can therefore be thought of as a form of tail risk mutualization – a very low-probability, yet potentially intrusive event. Since loss-allocation tools increase participants' exposures to the CCP, it is of interest to know how they affect member welfare and incentives to trade in centrally cleared instruments. Keeping in mind that activity in many core funding markets froze at the onset of the Great Recession, it is important to use loss-allocation tools that do not cause or exacerbate such disruptions. The uninterrupted functioning of OTC derivatives markets is important for financial stability because they improve the pricing of risk, add to liquidity, and help market participants better manage and diversify risks.

The paper focuses on losses distributed using variation margin gains haircuts (VMGH) and cash calls. VMGH are a form of gains withholding, whereby the clearinghouse withholds the mark-to-market gains originally due to members and uses them to address uncovered losses instead. Cash calls, by contrast, are contingent assessment powers that a clearinghouse may invoke when its resources become insufficient. I show that the exposures created by these two loss-allocation tools result in qualitatively different welfare and behavior despite looking similar on the surface. The difference stems mostly from the fact that exposure to VMGH withholdings is ex post irreducible, while a cash call can be defaulted on even if the default is costly to the participant. I am particularly interested in the optimal degree of loss sharing in the context of private and social welfare.

The paper quantifies and broadly confirms the intuitive argument put forward by Singh (2015) that losses should be shared to the broadest extent possible. The analysis demonstrates that maximum loss sharing is indeed both privately and socially optimal in many cases. However, it also identifies an important special case where private and social welfare diverge and banks do not have the right incentives to trade under maximal loss sharing, creating a tradeoff between loss sharing and trading volume.

This special case is important for policy-makers interested in the continuous functioning of centrally cleared markets. It occurs when VMGH losses are equally shared with high probability and the CCP uses a risk-based, highly procylcial skin in the game that correlates positively with the amount of default risk. I show that, in such circumstances, procyclical skin in the game weakens incentives to trade because the CCP effectively subsidizes potential participant defaults in times of high risk. Without the right incentives to both trade in high volumes and share losses in full, participants find it optimal to reduce their trading positions. This outlines an important tradeoff between market stability and fairness in the allocation of losses. If VMGH losses are equally shared with high probability, there is the potential that market activity may decline even before the onset of stress, thereby reducing market liquidity. If, on the other hand, the social planner places more weight on avoiding a market freeze, he or she may have to tolerate some inequality in the ex post distribution of losses. To avoid this tradeoff altogether, one solution is to simply not use overly risk-sensitive skin in the game. This does not preclude a CCP from using some measure of member default risk to size its own default fund contribution, as long as skin in the game does not increase too rapidly with default risk.

In contrast to VMGH, which affects trading volume ex ante, cash calls carry ex post performance risk that can undermine a CCP's ability to collect funds after a stress event, potentially threatening both its recovery and its post-recovery viability. The analysis shows that cash calls do not reduce trading before the onset of stress because clearing members retain the ability to default on cash calls that they deem too intrusive. While this eliminates ex ante market reactions to including cash calls in a recovery plan, it raises a different set of concerns. The model predicts that the banks' risk of non-performance to cash calls rises endogenously when cash calls become more likely, thereby constraining the total amount of funds that can be recovered. This can limit the success of the recovery effort, which suggests that CCP resolution authorities should be endowed with powers allowing them to resolve a CCP before recovery tools become ineffective.

# 2 Related Literature

The notion of stabilizing a clearinghouse through loss allocation is not new, but was it not embedded in regulation until the G-20 countries adopted the Principles for Financial Market Infrastructures – a set of comprehensive risk-management standards requiring CCPs to, among other things, develop dedicated recovery plans, including plans for loss allocation (CPMI-IOSCO 2012, 2014). The efforts of regulators and central banks to provide guidance with respect to recovery planning are ongoing, which explains the relative dearth of academic literature from which to draw upon. One contribution of this paper is that it provides a tractable analysis of recovery tools and draws several policy recommendations from it.

Conceptual considerations for CCP recovery and resolution have been outlined by Duffie

(2014a; 2014b) and Singh (2015), while Gibson (2013) provides a good non-technical overview of recovery tools. One of the few rigorous studies to focus on loss allocation is that of Heath, Kelly, and Manning (2015), who are interested in how CCPs could transmit financial stress through loss allocation and examine liquidity and solvency risk under different clearing configurations. By contrast, this paper focuses on the effects of loss allocation on CCP participants exposed to different loss-allocation tools. Does loss allocation affect trading patterns when losses are distributed using variation margin haircuts, compared to cash calls? Is there a natural limit beyond which cash calls become ineffective while stabilizing a CCP? What are the implications of cash calls and variation margin haircuts for the welfare of CCP participants and for social welfare? The main contribution of the paper is that it sheds light on these previously unanswered questions. The paper also helps define a natural boundary between CCP recovery and CCP resolution by studying the effectiveness of cash calls.

The analysis draws partly on the generic CCP model developed by Santos and Scheinkman (2001), which I modify to allow for loss allocation. Unlike the Santos and Scheinkman model, however, this is not a model of adverse selection or moral hazard because the results I obtain do not depend on hidden action or self-selection. For example, unobserved risk-taking by banks is not necessary in order to obtain the result that high-probability VMGH losses dampen trading in the presence of too much skin in the game (Proposition 1). Similarly, the banks' observable exposures to loss-allocation tools are entirely sufficient to generate the finding that banks become less likely to respond to high-probability cash calls (this result is, in fact, only strengthened by moral hazard). Thus, asymmetric information is not necessary to obtain the main results in the paper; risk mutualization is sufficient.[1] This increases the model's robustness and shows that mutualized risk exposures within a CCP can have significant effects on participant trading and welfare even in the absence of adverse selection and moral hazard.

The rest of the paper is structured as follows. Section 3 sets up the model. Section 4 presents the results on variation margin haircutting, and Section 5 the results on cash calls; both result sections compare the effects of loss sharing on participant and social welfare.

[1]For a more advanced treatment of moral hazard in CCPs, see, for example, Koeppl (2013) and Koeppl and Monnet (2012).

Section 6 summarizes the findings and the policy recommendations.

# 3 Model

I build a partial equilibrium model consisting of two groups of risk-averse banks, trading a stylized derivative contract through a competitive central counterparty, as in Santos and Scheinkman (2001). The CCP becomes the buyer for every seller and the seller for every buyer, thus replacing the original contractual obligations between the two banks with obligations involving only the CCP in a process called *novation*. After novation, each bank contractually faces only the CCP, which takes on the responsibility to make good on the trade regardless of the original counterparty's performance; in that respect, the CCP is similar to an insurance provider. The risk of having to replace a trade not honored by a bank is commonly known as replacement cost risk and is funded by pre-pledged resources (margin deposit and default fund) held by the CCP, supplemented by a layer of CCP's own equity (skin in the game). Losses in excess of pre-funded resources are redistributed among survivors so that the CCP breaks even in the long run.

## 3.1 Economic Environment

The model features two groups of risk-averse banks, represented by Bank 1 and Bank 2, with identical concave utility functions[2] $u(\cdot)$. Each group consists of a continuum of banks with total measure 1. Banks in each group experience idiosyncratic random shocks to their consumption good endowments, which motivates them to try to smooth out consumption by trading in a stylized derivative contract. The contract is handled by a competitive CCP, which interposes itself between buyers and sellers, collects collateral (margin) and allocates uncovered losses to survivors after a default exceeding its pre-funded resources, including skin in the game. The clearinghouse breaks even over the long run, consistent with free market entry.[3]

[2]Most results do not depend of the choice of utility function; where they do, this is clearly indicated.

[3]Free entry is not an unrealistic assumption for the central clearing market. Several authors (Murphy, 2012; Zhu, 2011) point to the introduction of several new CCPs over the past decade as evidence of increasing competition between clearinghouses. Santos and Scheinkman (2001) also study CCP competition.

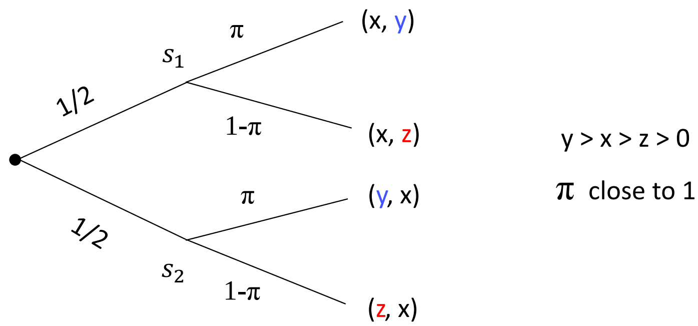



Figure 1: Stochastic endowments for Bank 1 and Bank 2

A derivative trade typically features uncertainty about two kinds of events: (1) Who will gain from the trade upon maturity? and (2) Will the payor will honor its obligations? The model reflects both uncertainties in a stylized way by assuming that the banks face an endowment process similar to that in Allen and Gale (2000) or Santos and Scheinkman (2001) (see Figure 1). Under this process, Bank 1 and Bank 2 face equal chance of ending up as the payee or as the payor once the true state of nature is known. This is reflected by the leftmost node in Figure 1. The payee, once determined, also faces additional uncertainty whether the payor will default to the CCP, which happens with probability $1 - \pi$, as reflected by the remaining two nodes; the default probability is common knowledge. In the event of default, the residual loss not covered by margin is distributed among survivors using loss-allocation tools. The outcomes $y > x > z$ at the right of the figure represent the randomly realized endowments of a consumption good that emulate asset returns. If the payor receives the good endowment $y$, which occurs with probability $\pi$ close to 1, the contract is honored and the payor has to transfer to the payee the contracted amount $\theta$ ($\theta \leq y$) using the CCP. However, if the payor receives a bad realization $z$, which occurs with a small probability $(1 - \pi)$, it is unable to honor the contract, and the CCP keeps only the collateralized fraction of the position $\Phi\theta$, where $\Phi$ is the margin requirement expressed in percent ($\Phi \in [0, 1]$). For the CCP to pay out its obligation on the remaining leg of the transaction, it allocates the uncovered loss to surviving members.

Thus, the trade normalizes to a generic derivative contract in which the bank holding a long position is entitled to receive one unit of the consumption good in state $s_1$ and must deliver one unit of the consumption good in state $s_2$. By convention, position size $\theta_i$ is positive for long positions and negative for short positions. Banks can default on the contract if the disutility of default is smaller than the disutility from getting the idiosyncratic shock $z$. Similar to Diamond (1984) and Dubey, Genakoplos and Shubik (2005), I assume that default is costly and is associated with a disutility proportional to the defaulted amount. Specifically, if a bank must deliver $\ell$ dollars on a position but delivers a smaller amount, $D$, it sustains a utility penalty equal to

$$\lambda \max\{\ell - D, 0\} \tag{1}$$

subtracted directly from the utility obtained in the state triggering the default. The literature interprets the exogenous parameter $\lambda > 0$ as the economy-wide bankruptcy code, or the marginal disutility from a dollar defaulted. In the context of a central counterparty, $\lambda$ can be broadly interpreted as the disutility from losing CCP membership status, including the inability to have transactions cleared and settled, reputation loss, and fees or other penalties imposed by the CCP on a non-conforming member.

To maintain viability in the long run, the CCP faces a break-even condition requiring expected pay-ins to equal expected payouts. This requires the CCP to redistribute the expected uncovered loss across its surviving clearing members after it has exhausted both the defaulter's margin deposit and the CCP's own skin in the game. To obtain cleaner results for loss allocation with recovery tools (as opposed to risk mutualization using the default fund), I normalize the default fund size to zero so all residual losses are allocated using recovery tools; a number of other papers explore default fund sizing and loss allocation (for example, see Elliott, 2013, Nahai-Williamson et al., 2013, and Haene and Sturm, 2009).

Since the contracted amount is $\theta$, and the probability of default on the contract is $1 - \pi$, the expected uncovered loss is $L = (1 - \pi)(1 - \Phi)\theta$, where $(1 - \Phi)$ is the uncollateralized fraction of the position $\theta$. I will refer to the expression $\kappa \equiv (1 - \pi)(1 - \Phi)$ as the *uncovered default risk* from the trade. After a default loss has been realized, the CCP first uses its skin in the game, $s$, to absorb part of the loss before mutualizing losses between its members. Therefore, for loss allocation to be applicable, $\kappa\theta > s$, where $s$ is an exogenous skin-in-the-

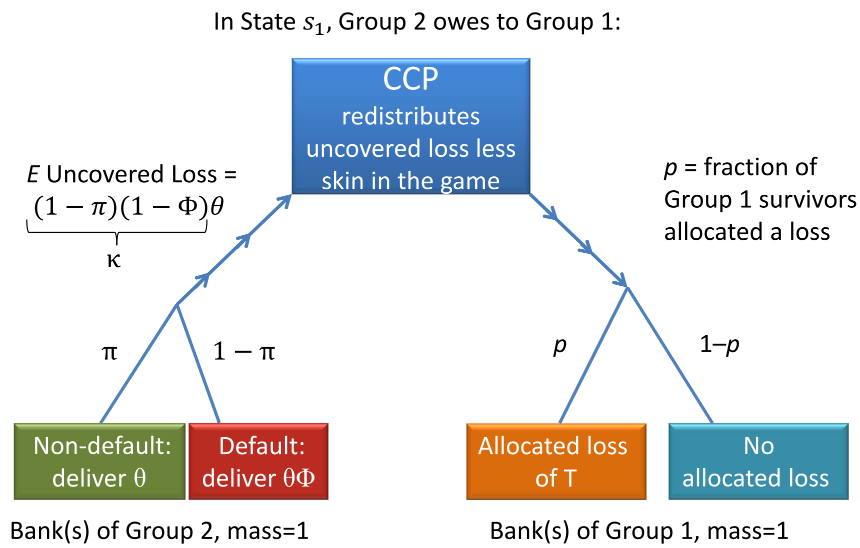



Figure 2: Generation and redistribution of uncovered losses using VMGH.

game capital.[4] Figure 2 illustrates the flow of the uncovered loss from defaulters to the CCP and following loss redistribution across members. The situation illustrated corresponds to state $s_1$, where Group 2 banks owe a payment $\theta$ to Group 1, and the CCP imposes a loss $T$ on the surviving bank (or group of banks).[5] The flowchart for state $s_2$ is symmetric (the flow of obligations operates in the opposite direction).

For the CCP to break even, the losses allocated to members must equal the aggregate expected loss from default minus any loss absorbed by the CCP's skin in the game, $s$. Let $p$ denote the probability that a survivor bank is allocated a loss of $T$. Then the break-even condition requires that

$$\mathbb{E}[\text{Loss allocated to members}] = \mathbb{E}\,[\text{Uncovered default loss}] - [\text{Skin in the game}], \text{ or}$$

$$pT = \kappa\theta - s. \tag{2}$$

Since the banks in each group form a continuum, the probability $p$ that any individual bank is assigned a loss can also be interpreted as the exact *fraction* of banks assigned a loss. This, in turn, allows one to interpret $p$ as the *degree of loss sharing*, which can exogenously

[4]The capital structure of the CCP, including skin-in-the-game capital, is taken as exogenous since the positive past profits required to generate capital require a short-run model to emulate. By contrast, in this long-run model, the CCP realizes zero economic profit.

[5]To preserve the isomorphism between the individual bank and the continuum of banks in each group, the model does not allocate losses to more than one group of banks at a time.

be varied by the CCP. I will switch between the two interpretations where this facilitates exposition. I do not impose any particular assumptions on skin in the game. For example, $s$ can be a flat amount, as is common in some CCPs, or it can be risk-based (procyclical) so that the CCP has more skin in the game in riskier periods. In the latter case, $s$ could be a function of risk – broadly defined – such as the default probability on a trade, or of the degree of loss sharing. For the purpose of this study, whenever I analyze risk-based skin in the game, I define it as a function of loss sharing ($s = s(p)$) since, all else equal, a higher default probability increases the chance $p$ of an allocated loss. In any case, adding more skin in the game reduces the expected loss $pT$ to be shared by survivors, and so changes the tradeoff between loss size and loss probability $p$.

The break-even condition implies that broader loss sharing helps reduce allocated losses per member, but at the expense of more banks participating in the loss-sharing process, implying a higher probability that any individual survivor bank is allocated a loss. This tradeoff between loss size $T$ and loss probability $p$ is illustrated in Figure 3 (holding skin in the game and trading position constant).

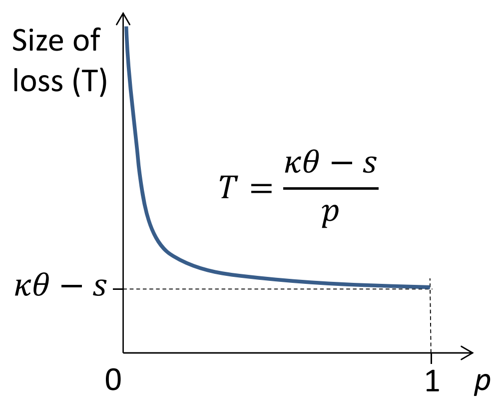



Figure 3: The tradeoff between loss size and loss probability.

This tradeoff is at the core of the analysis that follows, and many of the results can be more easily explained by referring to Figure 3. When selecting an optimal trading position $\theta$, banks realize that they are subject to this $(p, T)$ constraint, and their choices of trading position size $\theta$ will vary as the quantitative terms of the $(p, T)$ tradeoff are changed by other parameters, such as skin in the game. The next section describes the banks' optimal choice of trading volume $\theta$.

## 3.2 The Banks' Optimization Problem

Before entering the trade, banks need to determine the size of their position $\theta$ (number of contracts traded) by weighing the marginal utility benefit of trading against its marginal utility cost. Since the two representative banks are symmetric, it suffices to consider Bank 1's optimization problem. Bank 1 maximizes its expected utility from trading by choosing an optimal trading position $\theta$, subject to the CCP's margin requirement $\Phi$ and the bank's risk exposure to losses $T$ from recovery tools, occurring with probability $p$. The loss size $T$ is determined by the zero-profit condition, $T = \frac{\kappa\theta - s}{p}$, given a loss-sharing probability $p$, which is determined exogenously by the CCP. Because of the symmetry of the endowment process, Bank 1 will go long ($\theta_1 > 0$) and Bank 2 will go short ($\theta_2 < 0$), so that, by symmetry of their utility functions, $\theta_1 = -\theta_2$, and the market clears. Equilibrium is attained when each bank chooses its position $\theta_i$ optimally and all applicable resource constraints (break-even condition and collateral constraint) are satisfied.

Bank 1's utility is determined by the realized endowment ($x, y$ or $z$) and by the CCP-imposed loss $T$, if any. Thus a representative Group 1 bank maximizes its expected utility

$$U = \frac{1}{2}\Big[(1-p)\overbrace{u(x+\theta)}^{\text{No loss}} + p\,\overbrace{u(x+\theta-T)}^{\text{Honor loss of T}}\Big] + \\ + \frac{\pi}{2}\overbrace{u(y-\theta)}^{\text{Deliver fully}} + \frac{1-\pi}{2}\Big[\overbrace{u(z-\theta\Phi) - \lambda(1-\Phi)\theta}^{\text{Default on trade}}\Big], \quad (3)$$

subject to the zero-profit condition $T = \frac{\kappa\theta - s}{p}$ and the collateral constraint $z \geq \theta\Phi$. The first two terms of $U$ refer to the state where Bank 1 is due to receive $\theta$ from Bank 2 but faces a chance $p$ of having a loss $T$ allocated and receiving $\theta - T$, instead of $\theta$. The remaining two terms refer to the opposite state, where Bank 1 owes $\theta$ to Bank 2, and stochastically defaults or delivers on the trade with probabilities $(1-\pi)$ or $\pi$, respectively.

The Lagrangian $\mathscr{L}$ for the optimal choice of $\theta$ depends on the size of the allocated loss

$T$, which is a function $T(\theta, p)$ of both trading volume $\theta$ and the loss probability $p$:

$$2\mathscr{L} = (1-p)u(x+\theta) + pu\big((x+\theta - T(\theta,p)\big) + \pi u(y-\theta) + \\ + (1-\pi)\big[u(z-\theta\Phi) - \lambda\theta(1-\Phi)\big] - 2\gamma(\theta\Phi - z) \quad (4)$$

The Lagrangian is strictly concave in $\theta$, with first-order condition of the form $G(\theta, p) = 0$ given by

$$(1-p)u'(x+\theta) + pu'\big(x+\theta - T(\theta,p)\big)\left[1 - \frac{\partial T}{\partial \theta}\right] + \\ - \pi u'(y-\theta) - (1-\pi)\big[\Phi u'(z-\theta\Phi) + \lambda(1-\Phi)\big] - 2\gamma\Phi = 0. \quad (5)$$

The first-order condition means that utility is maximized when the marginal utility benefit from trading, captured by the first two (positive) terms, equals the marginal utility cost, captured by the remaining (negative) terms. The first term is simply the marginal utility benefit $u'(x+\theta)$ from receiving the gain $\theta$, weighted by the probability $(1-p)$; it reflects the "speculative" benefit of scoring a gain. By contrast, the second term reflects the marginal benefit of a larger position $\theta$ in helping offset a potential loss $T$, thus providing a benefit similar to self-insurance. On the cost side, utility costs of trading come from two sources: the two states where Bank 1 has to pay the CCP (with or without default) and the shadow cost of collateral, $\gamma\Phi$. The balance between marginal utility costs and benefits determines the optimal trading position $\theta$ and links it to the remaining model parameters. It is used to derive the VMGH results in the next section.

# 4 Results on Variation Margin Gains Haircutting

As shown in Figure 3, sharing losses with low probability among just a few participants results in very large losses per member. On the other hand, sharing smaller losses among more banks (with $p$ near 1) implies that any individual bank has a higher chance of being allocated a loss, which, all else equal, is also undesirable for a risk-averter. When determining their optimal trading positions, banks therefore need to take this tradeoff into account. The terms of the tradeoff, however, can be altered by the variables that the banks have under their control – for example, the trading position $\theta$, which acts as a vertical shifter of the $T$ curve in

Figure 3. This suggests that banks could endogenously reduce their exposures to loss sharing by adjusting their trading positions. This section shows that when one of the two elements of the exposure – either the size or the probability of the loss – becomes excessive, banks can react by flattening out their positions. The model shows that excessively large losses, even when allocated with a very small probability, lead to reduced trading. For smaller, higher-probability losses, the outcome depends on the behavior of other loss-absorbent resources, such as the skin in the game, and particularly, on whether skin in the game is risk-based (procyclical) or not. This happens because risk-based skin in the game changes the terms of the tradeoff between the loss size and loss probability, and hence, alters the optimum trading position size. Analyzing the extent to which VMGH discourages trading provides a better understanding of how banks can react to recovery tool exposures and is formalized by the following proposition.

**Proposition 1.** *(Effect of variation margin gain haircutting on trading)*

*(a) If skin in the game is sufficiently procyclical, as measured by the size of $s'(p)$, then trading volume $\theta$ peaks at a strictly interior loss-sharing value $p_M < 1$.*

*(b) If skin in the game is a flat amount ($s'(p) = 0$) or zero, trading volume $\theta$ peaks at the maximal loss-sharing ($p = 1$).*[6]

*(c) Therefore the degree of procyclicality of skin in the game $s'(p)$ acts as a horizontal shifter of the trade volume peak.*

*Proof.* See the appendix.

The intuition behind Proposition 1 is that overly procyclical skin in the game weakens incentives to fully share VMGH losses and thus creates a tradeoff between trading and loss sharing in which higher loss probabilities only come at the cost of reduced trading.

To understand this result better, recall that to maximize utility, the marginal utility benefit of trading must equal the marginal utility cost. In this model, there are two benefits to trading: a speculative benefit and an insurance benefit. The speculative benefit from

[6]This result is proven for log and constant absolute risk aversion utility; with constant relative risk aversion utility, the analysis is intractable.

trade is that it provides a good chance to score gains and improve final wealth, i.e., this is the normal motive why market participants trade. In the context of loss allocation, however, trading also has an added insurance benefit: potential gains make it easier to handle an allocated loss. As $p$ approaches unity, this self-insurance motive becomes dominant because losses become virtually certain. In this context, higher skin in the game actually discourages members from self-insuring by trading because skin in the game absorbs some of the default loss before it is passed on to participants; with less incentive to self-insure, members also have less incentive to trade because, in addition to benefits, trading also involves costs – the shadow cost of pledged collateral plus the marginal utility costs of those states where the bank has to pay. Thus, at high loss probabilities $p$ near 1, procyclical skin in the game actually has a disincentivizing effect on trade.[7]

The result in Proposition 1 outlines an important tradeoff between fairness in the allocation of losses and the market stability. Losses that are fairly allocated ex post here result in lower transaction volumes that could, in times of stress, challenge market liquidity and financial stability (Brunnermeier and Pedersen, 2009). To maximize market activity in this state, the policy-maker has to inevitably sacrifice ex post fairness and be satisfied with the weaker, ex ante, fairness notion. This naturally leads to the question how private and social welfare compare when losses are allocated using VMGH.

A natural measure of the banks' private utility are their equilibrium utility levels, which are a function of the degree of loss sharing. How the utility depends on loss sharing is shown in Proposition 2 below.

**Proposition 2.** *(Individual welfare effects of variation margin gain haircutting)*
*Given position size $\theta$, bank welfare is maximized at the point of full loss sharing ($p = 1$).*

*Proof.* See the appendix.

Proposition 2 shows that private welfare from sharing a given loss is maximized when the

[7] To see this on a technical level, consider the first-order condition (5). In it, the (expected) marginal speculative benefit of trading is $(1-p)\ u'(x+\theta)$, and the marginal insurance benefit is $p\ u'(x+\theta-T)$. Clearly, when $p \to 1$, the insurance benefit term dominates. Now suppose $u'(x+\theta-T)$ is balanced against the utility costs of trading, given by the remaining terms in equation (5) (the costs of having to pay plus the costs of collateral $\gamma\Phi$) and consider an exogenous increase in skin in the game $s$. Skin in the game reduces the loss $T$, thereby increasing wealth and lowering the marginal utility. To restore the equilibrium, trading volume $\theta$ must go *down* in order to bring marginal utility back up. This explains the drop in $\theta$ near $p = 1$.

loss is shared fully since broadest loss sharing maximally reduces the individual loss per member.[8] It is important to note, however, that Proposition 2 does *not* imply that trading volume $\theta$ remains fixed. Indeed, Proposition 1(a) already reveals that it does not: while it may be optimal, full loss sharing can also create disincentives to trade. Thus, a policy-maker who wants to both maintain market volume and share VMGH losses fully will need to decide which of these two aspects is more important. This decision can be informed by formalizing the policy-maker's tradeoff by a social welfare function.

If one uses trading volume as a rough proxy for the welfare of the rest of the economy that derives utility from the existence of banking services, one can define a social welfare function as the sum of the banks' private utilities and the reduced-form utility of non-banks. I assume the latter to be roughly proportional to the activity of the banking sector, resulting in the social welfare function

$$SWF(p) = \int_{[0,1]} U_i(p)di + \theta(p) = U(p) + \theta(p). \tag{6}$$

Since we have graphs for each of the two components of the social welfare function, it is easy to visualize the SWF (see Figures 4 and 5). The results from this maximization are shown in the corollary below.

**Corollary.** *(VMGH Welfare Analysis – Private vs Social Optimum).*

(a) *If skin in the game is highly procyclical (a sufficient condition for this, with log utility, is $s'(1) > \kappa y - s(1)$), then the socially optimal level of loss sharing $p_S$ lies in the interval $(p_M, 1]$.*

(b) *If skin in the game is flat or zero, resulting in a trading volume peak at $p = 1$, then the socially optimal and privately optimal levels of loss sharing coincide at unity, so $p_S = p^* = 1$.*

Figures 4 and 5 imply that, consistent with intuition, in many cases it is both privately and socially optimal to share uncovered losses as fully as possible. This conforms to the

[8] This intuitive result is listed as a separate theorem mostly because of the contrast it provides to the welfare analysis of cash calls in Section 5, which shows that individual welfare is invariant to the degree of loss sharing under cash calls.

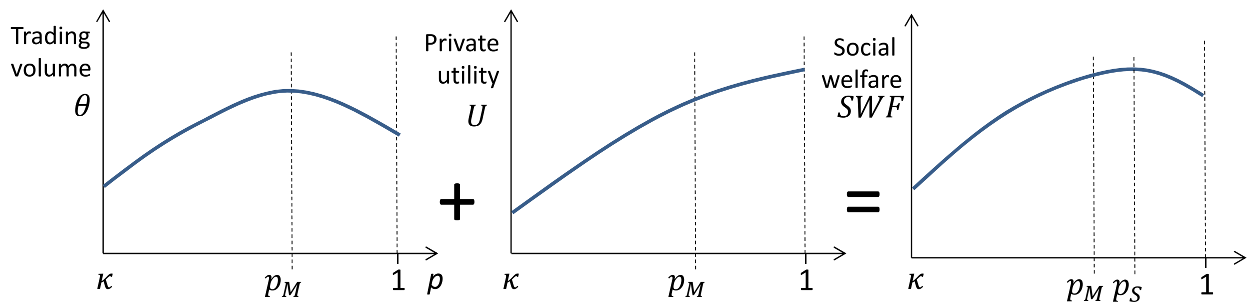



Figure 4: Example of social and private welfare for highly procyclical skin in the game.

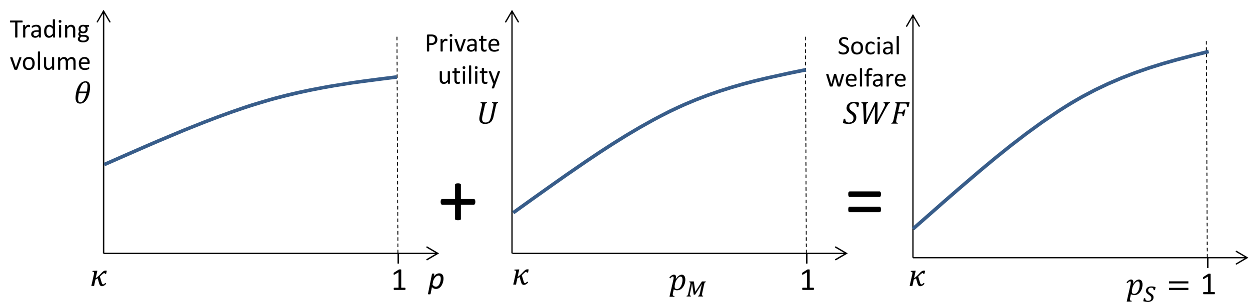



Figure 5: Social and private welfare for flat or zero skin in the game.

heuristic argument put forward by Singh (2015) that reducing loss sizes should generally pay off, even if it implies a higher loss probability. The more interesting case is where the private and social optimum diverge – that is, the case with highly procyclical skin in the game. Even though, from a private standpoint, each bank would prefer to have losses shared fully, full loss sharing does not maximize trading and is therefore socially suboptimal.

As discussed, partial loss sharing has implications for ex post fairness. Although ex ante, all members are equal, ex post, if $p < 1$, some members will have suffered losses while others will not have. Such allocations raise some practical concerns about fairness since many CCPs have rules "hard-coding" the symmetric and proportionate treatment of members in their rulebooks. Rather than trying to achieve a social optimum with partial loss sharing, avoiding overly procyclical skin in the game appears as a more attractive practical solution because the tradeoff cannot be optimized so that both private and social incentives are fully aligned. As an added benefit, this solution is is fair both ex ante and ex post.

# 5 Results on Cash Calls

A fundamental difference between VMGH and cash calls is the risk that a cash call may not be honored. When using VMGH, a CCP retains variation margin payments originally due to be transferred to members with position gains; gaining members are therefore unable to avoid the withholding. By contrast, cash calls introduce the *performance risk* that they may not be paid out to the CCP as due (or within the required time frame for replenishment). A 2014 incident at the KRX CCP (affiliated with Korea Exchange) provides a recent example.[9] After a series of incorrect trades submitted by HanMag Securities resulted in default fund assessments, KRX was forced to extend its time frame for replenishment for months after the stress event had occurred due to the slow rate at which members were making payments.

Non-performance on a cash call is not costless for a bank in terms of market reputation and compliance with CCP rules and can lead to a loss of access to centrally cleared markets. To reflect this cost, each bank in the model compares the marginal utility of avoiding the loss against the marginal disutility $\lambda_1$ caused by reputation and non-compliance costs. The net effect of these actions on member utility is $u(x+\theta)-\lambda_1$: if a member defaults, it avoids the financial loss $T$ but suffers a utility loss $\lambda_1$, similar to the reputation loss associated with defaulting on a trade. To reflect the ability of banks to endogenously adjust performance risk in response to losses, I assume that member banks also optimize over their performance probability $q$ to a cash call, as illustrated in Figure 6.

In this optimization, the bank maximizes with respect to $q$ the Largangian

$$\begin{aligned}\mathscr{L} = \frac{1}{2}\Big[(1-p)\overbrace{u(x+\theta)}^{\text{No loss}} + pq\overbrace{u(x+\theta-T)}^{\text{Honor loss of T}} + p(1-q)\overbrace{[u(x+\theta)-\lambda_1]}^{\text{Not honor loss}}\Big]+ \\ +\frac{\pi}{2}\underbrace{u(y-\theta)}_{\text{Deliver}} + \frac{1-\pi}{2}\Big[\underbrace{u(z-\theta\Phi)-\lambda(1-\Phi)\theta}_{\text{Default on trade}}\Big] - \gamma(\theta\Phi - z), \quad (7)\end{aligned}$$

resulting in the first-order condition for $q$

$$J \equiv p\Big[u\Big(x+\theta-T(p,q,\theta)\Big) - u(x+\theta) + \lambda_1 + qu'\Big(x+\theta-T(p,q,\theta)\Big)\frac{-\partial T(p,q,\theta)}{\partial q}\Big] = 0 \quad (8)$$

[9]See Vaghela (2014). I am indebted to Darrell Duffie for pointing out this example.

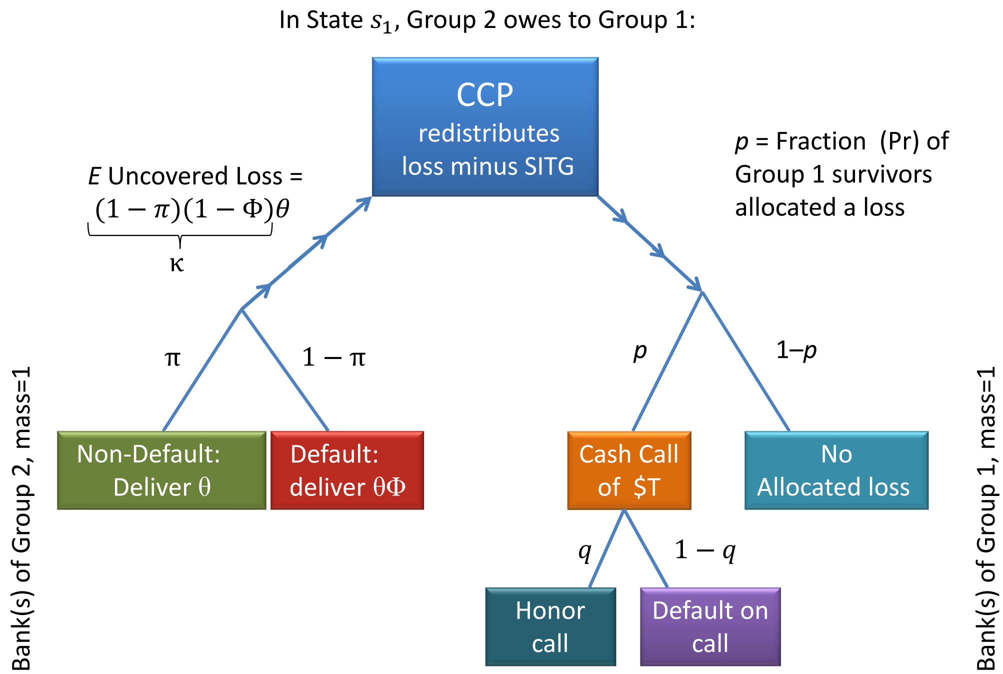



Figure 6: Generation and redistribution of uncovered losses using cash calls.

Now there are two endogenously determined variables that can interact: the level of trading $\theta$ and the performance probability $q$. For the outcome to be an equilibrium, the first-order condition for $q$ must hold jointly with the first-order condition for $\theta$,

$$(1-pq)u'(x+\theta) + pqu'\big(x+\theta-T(\theta,p,q)\big)\left[1-\frac{\partial T}{\partial \theta}\right] + \\ -\pi u'(y-\theta) - (1-\pi)\big[\Phi u'(z-\theta\Phi) + \lambda(1-\Phi)\big] - 2\gamma\Phi = 0. \quad (9)$$

This system of equations fully describes the equilibrium outcome.

When trading volume and the performance probability to cash calls are determined jointly, a bank has two channels for optimizing its behavior in response to the risk exposures caused by CCP-imposed losses. All else held equal, the bank can either decrease its trading volume down to reduce its future exposure or it can increase its probability of not honoring the cash call. Proposition 3 shows that, in equilibrium, banks find it optimal to take the latter route: managing the exposure by increasing performance risk. The benefit of doing so is that banks retain the flexibility to react to a future loss only if it occurs, without affecting trading in normal periods in response to a future event that may not materialize.

**Proposition 3.** *(Effect of cash calls on performance and trading)*

(a) *When losses are allocated using cash calls, any increase in loss sharing is fully offset by the lower chance that the cash call will be honored. The point elasticity of the performance probability $q^*$ with respect to loss sharing $p$ is* $\epsilon_{q^*,p} = -1$ *so that the overall probability that a cash call is assigned to an individual bank and honored remains* $pq^* = const.$

(b) *Trading volume is irresponsive to the degree of loss sharing when losses are allocated using cash calls. Specifically, the point elasticity of trading volume $\theta^*$ with respect to loss sharing $p$ is* $\epsilon_{\theta^*,p} = 0$.

*Proof:* See the appendix.

Proposition 3 means that ex ante trading volume $\theta$ is perfectly inelastic with respect to loss sharing because any increase in the loss probability is offset by an equally large, endogenous increase in performance risk. Concretely, Proposition 3(a) implies that the joint probability $pq^*$ that a cash call is assigned and honored remains constant in equilibrium and therefore so does the expected loss $pq^*T$. This implies that a bank's exposure to losses imposed using cash calls no longer varies with the degree of loss sharing and, hence, there is no need to adjust trading activity to reduce exposure.

The response of trading activity to cash calls, therefore, is qualitatively different compared with VMGH. Ex ante market activity does not respond to the possibility of future cash calls because of the clearing members' ability to default to cash calls ex post, after they know that a stress event has materialized. Even when the CCP increases the individual loss size to account for potential defaults, as reflected by the model, that comes at the cost of reduced post-recovery viability.

A corollary of this finding extends to the analysis of bank welfare. Since equilibrium utility $U^*$ is a function of $pq^*$, which remains constant, equilibrium expected utility remains unaffected by exogenous changes to the loss-sharing probability, as illustrated below and in Figure 7.

**Corollary.** *The fact that, in equilibrium, $pq^* = const.$, also implies that the bank's equilibrium utility, $U^*(pq^*)$, is independent of loss sharing $p$. The same is true of the social welfare $SWF(pq^*)$.*

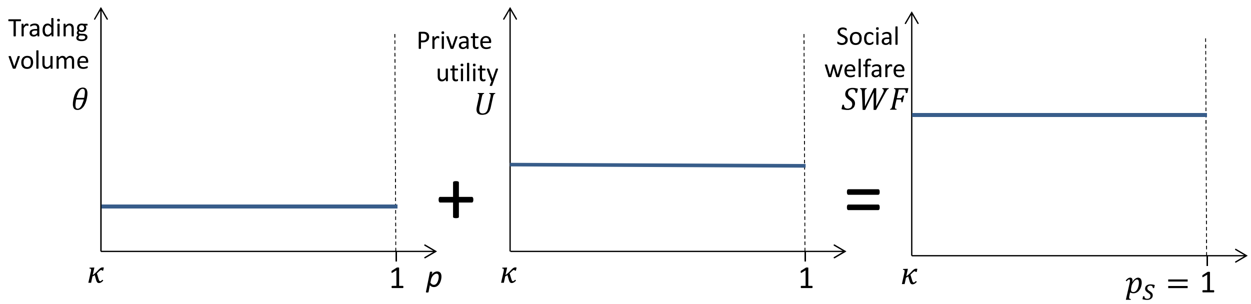



Figure 7: Social and private welfare for cash calls.

The probability of an honored cash call, $pq^*$, which is constant in equilibrium, indicates an important constraint faced by the CCP when allocating losses. The constraint comes from the fact that the expected amount collected from cash calls in equilibrium, $pq^* \cdot T(pq^*)$, remains constant, no matter how broadly CCP members share uncovered losses, because increasing the loss probability also boosts non-performance risk percent-for-percent. As illustrated by the 2014 incident with the KRX CCP, a slow response or a non-response to a cash call is entirely realistic, even when clearing members are not exposed to systemic risk.

This could be labeled as the CCP's *recovery constraint* because it outlines the endogenous limits of the recovery process. It reflects the expectation that members are likely to become unresponsive to additional cash calls after a certain sum has been collected. This consideration helps better define the boundary between CCP recovery and resolution. It seems reasonable to argue that resolution should be the preferred course of action when the recovery constraint is about to be reached since additional cash calls do not help stabilize the CCP any further. From a practical standpoint, this also suggests that CCP resolution regimes would benefit from sufficient powers allowing a resolution authority to initiate resolution before the CCP exhausts all recovery tools, especially when further loss allocation threatens to be ineffective. The lessons from VMGH loss allocation also indicate that a good resolution regime should be able to pre-empt situations where the recovery plan threatens the viability and normal trade patterns of participants; CCP-induced participant defaults could be just as detrimental to financial stability as the failure of a systemically important clearinghouse.

# 6 Conclusion

This paper proposes a model for analyzing the effects of cash calls and variation margin haircuts on the trading and welfare of clearinghouse participants. It is motivated by the increasing importance of centrally cleared trading, arising from the G-20 commitment to centrally clear standardized OTC derivatives. The paper studies how exposures to two main CCP loss-allocation tools – cash calls and variation margin haircuts – affect participant welfare and incentives to trade in centrally cleared instruments. It also explores the optimal amount of loss sharing and compares the privately and socially optimal outcomes.

The paper quantifies, and generally confirms, the intuitive argument put forward by Singh (2015) that losses should be shared to the fullest extent possible. The analysis demonstrates that, in many cases, maximum loss sharing is indeed both privately and socially optimal. However, the paper also identifies an important special case where private and social welfare diverge, and banks do not have the right incentives to share losses fully, creating a tradeoff between loss sharing and trading.

This special case is important for policy-makers who are concerned with the continuous functioning of centrally cleared markets. It occurs when VMGH losses are shared with high probability and the CCP uses a risk-based, highly procylcial skin-in-the-game capital that correlates positively with the amount of default risk. I show that, in this case, overly procyclical skin in the game weakens incentives to fully share VMGH losses and creates a tradeoff between trading and loss sharing. This, in turn, leads to tradeoff between market stability and fairness in the allocation of losses. If VMGH losses are shared in full, there is the potential that market activity may decline even before the onset of stress. If, on the other hand, the social planner places higher weight on market stability than on fairness, he or she may need to tolerate some inequality in the ex post distribution of losses. To avoid this tradeoff altogether, one solution is to simply not use overly risk-sensitive skin in the game. This does not preclude a CCP from using some measure of member default risk to size its own default fund contribution, as long as skin in the game does not increase too rapidly with default risk.

In contrast to VMGH, which affects trading volume ex ante, cash calls carry ex post performance risk that can undermine the CCP's ability to collect funds after a stress event, potentially threatening both its recovery and its post-recovery viability. The analysis shows that cash calls do not reduce trading before the onset of stress because clearing members retain the ability to default on cash calls that they deem too intrusive. While this eliminates ex ante market reactions to the inclusion of cash calls in a recovery plan, it raises a different set of concerns. The model predicts that the banks' risk of non-performance to cash calls rises endogenously when cash calls become more likely, thereby constraining the total amount of funds that can be recovered. This can limit the success of the recovery effort, which suggests that CCP resolution authorities should be endowed with powers allowing them to resolve a CCP before recovery tools become ineffective.

# 7 Appendix: Proofs

**Proposition 1**. (*Effect of variation margin gain haircutting on trading*)

(a) *If skin in the game is sufficiently procyclical, as measured by the size of $s'(p)$, then trading volume $\theta$ peaks at a strictly interior loss-sharing value $p_M < 1$.*

(b) *If skin in the game is a flat amount ($s'(p) = 0$) or zero, trading volume $\theta$ peaks at the maximal loss sharing ($p = 1$) for log and CARA utility.*

(c) *Therefore the degree of procyclicality of skin in the game $s'(p)$ acts as a horizontal shifter of the trade volume peak.*

*Proof.* (a) The proof is based on the idea that the slope of the continuous function $\theta(p)$ is positive at the left corner of $p$ and can be either positive or negative at the right corner $p = 1$, depending on the procyclical properties of the skin-in-the-game amount $s$. When $\theta$'s slope is positive at the right corner $p = 1$, the trading volume maximum occurs there; when negative, the maximum occurs at a strictly interior point $p_M$.

The slope of $\theta$ with respect to $p$, in turn, can be inferred by use of the implicit function theorem, applied to the first-order condition for $\theta$. Recall that Bank 1's Lagrangian is

$$2\mathscr{L} = (1-p)u(x+\theta) + pu\big((x+\theta - T(\theta,p)\big) + \pi u(y-\theta) + \\ + (1-\pi)\big[u(z-\theta\Phi) - \lambda\theta(1-\Phi)\big] - 2\gamma(\theta\Phi - z). \quad (10)$$

The first-order condition for $\theta$, denoted as $G(\theta, p) = 0$, is

$$(1-p)u'(x+\theta) + pu'\big(x+\theta - T(\theta,p)\big)\left[1 - \frac{\partial T}{\partial \theta}\right] + \\ - \pi u'(y-\theta) - (1-\pi)\big[\Phi u'(z-\theta\Phi) + \lambda(1-\Phi)\big] - 2\gamma\Phi = 0. \quad (11)$$

According to the implicit function theorem, the sign of the slope of $\theta$ with respect to $p$ is given by

$$\operatorname{sgn}\left\{\frac{d\theta}{dp}\right\} = \operatorname{sgn}\left\{-\frac{\partial G/\partial p}{\partial G/\partial \theta}\right\} = \operatorname{sgn}\{\partial G/\partial p\} \quad (12)$$

since, by concavity of the Lagrangian in $\theta$, it already follows that $\partial G/\partial\theta < 0$. The expression of interest is therefore

$$2\frac{\partial G}{\partial p} = -u'(x+\theta) + u'\big(x+\theta-T(\theta,p)\big)\left[1-\frac{\partial T}{\partial\theta}\right] + \\ + p\left[u''\big(x+\theta-T(\theta,p)\big)\frac{-\partial T}{\partial p}\left(1-\frac{\partial T}{\partial\theta}\right) + u'\big(x+\theta-T(\theta,p)\big)\frac{-\partial^2 T}{\partial\theta\partial p}\right], \quad (13)$$

where the amount of allocated loss $T$ is determined by the CCP break-even condition $pT = \kappa\theta - s(p)$ so that $T = [\kappa\theta - s(p)]/p$, where $s(p)$ is skin in the game. Skin in the game is allowed to be potentially procyclical; that is, to increase with the loss probability $p$ if desired. This case corresponds to $s'(p) > 0$; otherwise, for flat-amount skin in the game, $s'(p) = 0$.

Next, I show that $\partial G/\partial p > 0$ near $p = 0$ and that the sign of $dG/dp$ near $p = 1$ depends on the extent of procyclicality of skin in the game $s(p)$ – in other words, on the size of the derivative $s'(1)$.

First, observe that the model has no economic content if the probability of a loss $p$ allocated to an individual bank is smaller than the uncovered default risk $\kappa = (1-\pi)(1-\Phi)$ on the other side of the same trade (otherwise, the expected default losses accruing to the CCP would exceed total redistributed losses allocated to survivors, violating the break-even condition). Therefore, it must be the case that $p \geq \kappa$.[10] Using the derivatives of $T$

$$\frac{\partial T}{\partial\theta} = \frac{\kappa}{p}, \quad \frac{\partial T}{\partial p} = \frac{-\kappa\theta + s(p) - ps'(p)}{p^2}, \quad \frac{\partial^2 T}{d\theta\partial p} = -\frac{\kappa}{p^2},$$

I first evaluate the sign of $\partial G/\partial p$ at the left corner $p = \kappa$ and, after cancellation of terms, obtain

$$2\frac{\partial G}{\partial p}\bigg|_{p=\kappa} = u'(x+\theta-T) - u'(x+\theta) > 0. \quad (14)$$

Thus $\partial G/\partial p$, and hence the slope of $\theta(p)$, is positive at the left corner $p = \kappa$.

Next I evaluate $\partial G/\partial p$ at the right corner, $p = 1$, and obtain

$$2\frac{\partial G}{\partial p}\bigg|_{p=1} = \underbrace{u'(x+\theta-T) - u'(x+\theta)}_{>0}\underbrace{-u''(x+\theta-T)[1-\kappa]\big(-\kappa\theta + s(1) - s'(1)\big)}_{<0} \quad (15)$$

The sign of this expression depends on the balance between the two braced terms. Hence, when $s'(1)$ is sufficiently large – that is, when skin in the game is sufficiently procyclical

[10] Observe that the requirement $p \geq \kappa$ also implies that the haircut $T$ cannot exceed the realized gain $\theta$, which is a realistic feature of VMGH.

– the negative term dominates and $\frac{\partial G}{\partial p}\Big|_{p=1} < 0$, implying that the slope $d\theta/dp < 0$ at $p = 1$. Further, one can obtain the critical value of $s'(1)$ above which this happens by setting equation (15) to zero and solving for $s'(1)$. (The concrete critical value of $s'(1)$ depends on the particular utility function chosen.) Doing so for the log utility function using the shorthand notation for the residual loss uncovered by skin in the game, $R \equiv \kappa\theta - s(1)$, implies the critical value

$$s'(1) = \frac{[k\theta - s(1)]\big(x + (1-\kappa)\theta + s(1)\big)}{(1-\kappa)(x+\theta)} = \frac{R(x+\theta-R)}{(x+\theta)(1-\kappa)} = R - \frac{R^2}{x+\theta} + \varepsilon < \\ < R = \kappa\theta - s(1) < \kappa y - s(1), \quad (16)$$

where I used the fact that $\theta \leq y$ and $\kappa \equiv (1-\pi)(1-\Phi)$ is near zero.[11] This provides a sufficient condition, $s'(1) > \kappa y - s(1)$, that guarantees the result in part (a).

Collecting the results that $\theta'(\kappa) > 0$ and $\theta'(1) < 0$, and recalling that the function $\theta(p)$ is differentiable and hence continuous, it follows that there exists (at least one) interior point $p_M \in (0, 1)$ where $\theta'(p) = 0$. Since $\theta(p)$ is continuous over the compact set $[\kappa, 1]$, at least one $p_M$ corresponds to a maximum. Hence trading volume $\theta$ peaks at an interior value of $p$, as claimed. This proves part (a). $\square$

(b) It is not so when there is a flat-amount skin in the game (including zero). First consider the case of a flat amount $s$, implying that $s'(p) = 0$. Then

$$2\frac{\partial G}{\partial p}\bigg|_{p=1} = u'(x+\theta-T) - u'(x+\theta) - [1-\kappa]u''\big(x+\theta-T\big)[s-\kappa\theta]. \quad (17)$$

This result is demonstrated separately for the log and the CARA utility functions.

For the log utility $u(w) = \ln(w)$, this expression translates to

$$2\frac{\partial G}{\partial p}\bigg|_{p=1} = \frac{1}{x+\theta-T} - \frac{1}{x+\theta} + \frac{(1-\kappa)(s-\kappa\theta)}{(x+\theta-T)^2}, \quad (18)$$

whose sign is determined by the numerator

$$(x+\theta)(x+\theta-T) - (x+\theta-T)^2 - (1-\kappa)(x+\theta)(\kappa\theta - s) \quad = \quad T[\kappa x + s] > 0. \quad (19)$$

Hence, with a flat or zero skin in the game $s \geq 0$, the expression $\partial G/\partial p$ and the slope of $\theta(p)$ at the right corner $p = 1$ are both strictly positive, implying that trading volume is maximized at full loss sharing $p = 1$, as claimed.

[11] This follows from the near-zero default probability $\pi$ and from $\Phi < 1$.

For CARA utility, the expression $2\partial G/\partial p$ takes the form

$$2\frac{\partial G}{\partial p}\bigg|_{p=1} = re^{-r(x+\theta-T)} - re^{-r(x+\theta)} - (1-\kappa)Tr^2e^{-r(x+\theta-T)} \tag{20}$$

Dividing by $re^{-r(x+\theta)}$, and rearranging terms, one only needs to verify that

$$e^{rT} > [1 - rT(1-\kappa)]^{-1}. \tag{21}$$

To do this, take the log of both sides and obtain $rT > -\ln[1 - rT(1-\kappa)]$. Putting $\xi \equiv rT$ and taking logs, it remains only to show that $\xi > -\ln[1-\xi(1-\kappa)]$. Since empirically, $r$ is near-zero (see Cohen and Einav, 2007), so is $\xi$, which allows one to use the log approximation near unity $\ln(\omega) = -1 + \omega$ resulting in

$$\xi > (1-\kappa)\xi, \tag{22}$$

which always holds because $\kappa > 0$. This establishes that for CARA utility, $\frac{\partial G}{\partial p}\Big|_{p=1} > 0$.

(c) The result follows directly from (a) and (b). □

**Proposition 2.** (*Individual welfare effects of variation margin gain haircutting*) *Given position size $\theta$, bank welfare is maximized at the point of full loss sharing ($p = 1$).*

*Proof.* (a) To demonstrate this, it is sufficient to show that the bank's Largangian, which is concave in $p$, has a positive slope at $p = 1$. Concretely, I will prove that the slope

$$2\frac{d\mathscr{L}}{dp} = -u(x+\theta) + u\big(x+\theta-T(p,\theta)\big) + pu'\big(x+\theta-T(p,\theta)\big)\frac{-\partial T}{\partial p} > 0. \tag{23}$$

Evaluating this expression at $p = 1$, results in

$$\underbrace{-u(x+\theta) + u\big(x+\theta-T(1,\theta)\big)}_{>-Tu'(x+\theta-T)} + u'\big(x+\theta-T(1,\theta)\big)[\kappa\theta - s(1) + s'(1)], \tag{24}$$

so it is sufficient to prove that

$$-Tu'(x+\theta-T) + u'(x+\theta-T)[\kappa\theta - s(1) + s'(1)] \geq 0. \tag{25}$$

Substituting in this equation $T(1,\theta) = \kappa\theta - s(1)$, after cancellation of terms, it remains only to prove that

$$u'(x+\theta-T)s'(1) \geq 0, \tag{26}$$

which is always true since skin in the game is either flat or procyclical ($s'(p) \geq 0$). Therefore, $\frac{d\mathscr{L}}{dp}\Big|_{p=1} > 0$, which implies that the bank's Lagrangian, which is a concave function of $p$, peaks strictly at the welfare maximizer $p = 1$. $\square$

**Proposition 3.** *(Effect of cash calls on performance and trading)*

(a) *When losses are allocated using cash calls, any increase in loss sharing is fully offset by the lower chance that the cash call will be honored. The point elasticity of the performance probability $q^*$ with respect to loss sharing $p$ is $\epsilon_{q^*,p} = -1$ so that the overall probability that a cash call is assigned to an individual bank and honored remains $pq^* = const$.*

(b) *Trading volume is irresponsive to the degree of loss sharing when losses are allocated using cash calls. Specifically, the point elasticity of trading volume $\theta^*$ with respect to loss sharing $p$ is $\epsilon_{\theta^*,p} = 0$.*

*Proof.* (a) In the presence of performance risk, define the *effective* loss-sharing probability as $pq$ (the probability that a cash call is not only assigned but also honored). Likewise define skin in the game as *procyclical* if and only if $s'(pq) > 0$.[12]

To prove the assertion, I will first demonstrate that an exogenous increase in loss sharing $p$ reduces the optimal performance probability $q^*$ for any level of trading $\theta$. Let the first-order condition for $q^*$ be denoted as

$$J \equiv p\Big[u(x+\theta-T) - u(x+\theta) + \lambda_1 + qu'\big(x+\theta-T\big)\frac{-\partial T}{\partial q}\Big] = 0 \tag{27}$$

The equation $J(q^*, p) = 0$ provides an implicit link between the optimal choice of $q^*$ and the parameter $p$ through the implicit function theorem, according to which

$$\frac{dq^*}{dp} = -\frac{\partial J/\partial p}{\partial J/\partial q}. \tag{28}$$

The components of this expression, written out individually, are

$$\frac{\partial J}{\partial p} = -\frac{\partial T}{\partial p}Tu''(x+\theta-T) < 0 \tag{29}$$

[12]Results are qualitatively similar if skin in the game is defined as a function of $p$ only. The elasticity in this case is still negative but not exactly equal to 1.

and

$$\frac{\partial J}{\partial q} = -\frac{\partial T}{\partial q} T u''(x + \theta - T) < 0, \tag{30}$$

where

$$\frac{\partial T}{\partial p} = \frac{-pqs'(pq) - \kappa\theta + s(pq)}{p^2 q} \quad \text{and} \quad \frac{\partial T}{\partial q} = \frac{-pqs'(pq) - \kappa\theta + s(pq)}{pq^2} \tag{31}$$

resulting in

$$\frac{dq^*}{dp} = -\frac{\partial T/\partial p}{\partial T/\partial q} < 0, \tag{32}$$

which implies that a higher probability of allocating a loss endogenously reduces the bank's likelihood of delivering a cash call. It is of interest to find out the size of this effect, which depends on the relative sizes of the derivatives $\partial T/\partial p$ and $\partial T/\partial q$. The break-even condition $pqT = \kappa\theta - s(pq)$ implies that the ratio of the two derivatives of $T$ is

$$\frac{dq^*}{dp} = -\frac{\big(-pqs'(pq) - \kappa\theta + s(pq)\big)/p^2 q}{\big(-pqs'(pq) - \kappa\theta + s(pq)\big)/pq^2} = -\frac{pq^2}{p^2 q} = -\frac{q}{p}. \tag{33}$$

This helps compute the point elasticity of $q^*$ with respect to $p$ as

$$\epsilon_{q^*,p} = \frac{dq^*}{dp}\frac{p}{q^*} = -\frac{q^* p}{pq^*} = -1, \tag{34}$$

therefore proving part (a). (This result continues to hold if in the above calculations one sets skin in the game $s$ to zero or to a constant amount, which would imply $s' = 0$).

(b) The negative unit elasticity of $q^*$ with respect to $p$ implies that $\%\Delta q^* = -\%\Delta p$, and therefore, near $q^*$,

$$\%\Delta(pq^*) = \%\Delta p + \%\Delta q^* = 0 \tag{35}$$

so that percentage changes in $p$ and $q$ offset each other exactly. Hence, the quantity $pq^*$ remains constant regardless of exogenous changes in $p$. Next, I consider how the optimally chosen trading volume $\theta^*$ responds to changes in $p$ when $q^*$ is endogenous. In the first-order condition for optimal volume, $\theta^*$ is now defined as an implicit function of $pq$ in the same way as it was a function of $p$ before I introduced the possibility of default on a cash call. The implicit function theorem likewise guarantees that $\theta^*(pq)$ is differentiable and hence continuous in its argument $pq$. However, it was just proved that when $q$ is chosen optimally (that is, $q = q^*$), changes in $p$ and $q$ offset each other so that $pq^* = const$. This immediately implies that $\theta^*(const.) = const.$, so the trading volume is irresponsive to $p$. By definition, then, the elasticity $\epsilon_{\theta^*,p} = 0$, thereby proving part (b). $\square$

# References


[1] Allen, Franklin, and Gale, Douglas. (2000). "Financial Contagion." *Journal of Political Economy* 108, 1–33.

[2] Brunnermeier, Markus, and Pedersen, Lasse. (2009). "Market Liquidity and Funding Liquidity." *Review of Financial Studies* 22, 2201–2238.

[3] Coeuré, Benoît. (2015). "Ensuring an Adequate Loss-absorbing Capacity of Central Counterparties." In *BIS Central Bankers Speeches*. Bank of International Settlements: Basel, Switzerland.

[4] Cohen, Alma, and Einav, Liran. (2007). "Estimating Risk Preferences from Deductible Choice." *American Economic Review* 97, 745–788.

[5] CPMI-IOSCO. (2012). Principles for Financial Market Infrastructures. Bank of International Settlements: Basel, Switzerland.

[6] CPMI-IOSCO. (2014). Recovery of Financial Market Infrastructures. Bank of International Settlements: Basel, Switzerland.

[7] Diamond, Douglas. (1984). "Financial intermediation and delegated monitoring." *Review of Economic Studies* 51, 393–414.

[8] Dubey, Pradeep; Geanakoplos, John; and Shubik, Martin. (2005). "Default and punishment in general equilibrium." *Econometrica* 73, 1–37.

[9] Duffie, Darrell. (2014a). "Financial Market Infrastructure: Too Important to Fail." In *Across the Great Divide: New Perspectives on the Financial Crisis*, Martin Neil Baily and John B. Taylor, editors. Hoover Institution Press: Stanford, California.

[10] Duffie, Darrell. (2014b). "Resolution of Failing Central Counterparties." Graduate School of Business, Stanford University, Working Paper No. 19890.

[11] Elliott, David. (2013). "Central counterparty loss-allocation rules." Bank of England Financial Stability Paper No. 20.

[12] *Financial Times.* (2014). "'Too Big to Fail' Worries Reach Clearing Houses." 2 December.

[13] Gibson, Matt. (2013). "Recovery and Resolution of Central Counterparties." *Reserve Bank of Australia Bulletin*, December Quarter 2013.

[14] Gibson, Rajna, and Murawski, Karsten. (2013). "Margining in derivatives markets and the stability of the banking sector." *Journal of Banking & Finance* 37, 1119–1132.

[15] Haene, Philipp, and Sturm, Andy. (2009). "Optimal Central Counterparty Risk Management." Swiss National Bank Working Paper 2009-7.

[16] Heath, Alexandra; Kelly, Gerard; and Manning, Mark. (2014). "Central Counterparty Loss Allocation and Transmission of Financial Stress." Reserve Bank of Australia Research Discussion Paper RDP-2015-02.

[17] Koeppl, Thorsten. (2013). "The Limits of Central Counterparty Clearing: Collusive Moral Hazard and Market Liquidity." Working paper, Queen's University.

[18] Koeppl, Thorsten, and Monnet, Cyril. (2012). "Central Counterparty Clearing and Systemic Risk Insurance in OTC Derivatives Markets," *Revue d'Économie Financière*.

[19] Murphy, David. (2012). "The systemic risks of OTC derivatives central clearing." *Journal of Risk Management in Financial Institutions* 5 (3), 319–334.

[20] Nahai-Williamson, Paul; Ota, Tomohiro; Vital, Mathieu, and Wetherilt, Anne. (2013). "Central Counterparties and their Financial Resources – A Numerical Approach." Bank of England Financial Stability Paper No. 19.

[21] Santos, Tano, and Scheinkman, Jose. (2001). "Competition Among Exchanges." *Quarterly Journal of Economics* 116, 1027–1061.

[22] Singh, Manmohan. (2015). "Limiting taxpayer "puts": an example from central counterparties." *Journal of Financial Market Infrastructures* 3(3), 1–15.

[23] Vaghela, Viren. (2014). "Korea trading error prompts global clearing house review." *Risk Magazine*, 4 March.

[24] Zhu, Siyi. (2011). "Is there a race to the bottom in central counterparties competition? Evidence from LCH.Clearnet SA, EMCF and EuroCCP." DNB Occasional Studies 9 (6).