Working Paper Series

No 44 / May 2017

Compressing over-the-counter markets

by
Marco D'Errico
Tarik Roukny

ESRB
European Systemic Risk Board
European System of Financial Supervision

Abstract

In this paper, we show both theoretically and empirically that the size of over-the-counter (OTC) markets can be reduced without affecting individual net positions. First, we find that the networked nature of these markets generates an *excess* of notional obligations between the aggregate gross amount and the minimum amount required to satisfy each individual net position. Second, we show conditions under which such excess can be removed. We refer to this netting operation as *compression* and identify feasibility and efficiency criteria, highlighting intermediation as the key element for excess levels. We show that a trade-off exists between the amount of notional that can be eliminated from the system and the conservation of original trading relationships. Third, we apply our framework to a unique and comprehensive transaction-level dataset on OTC derivatives including all firms based in the European Union. On average, we find that around 75% of market gross notional relates to excess. While around 50% can in general be removed via bilateral compression, more sophisticated multilateral compression approaches are substantially more efficient. In particular, we find that even the most conservative multilateral approach which satisfies relationship constraints can eliminate up to 98% of excess in the markets.
Keywords: OTC markets, compression, intermediation, derivatives, networks, optimization
JEL codes: C61, D53, D85, G01, G10, G12

Contents

1 Introduction **1**

2 Literature review **7**

3 The market **9**
- 3.1 Definition of excess 11
- 3.2 Existence condition 13
- 3.3 Excess decomposition 14

4 Compression **15**
- 4.1 Feasibility 16
- 4.2 Tolerances 16
- 4.3 Residual and redundant excess 17
- 4.4 Efficiency 18
- 4.5 Benchmark approaches 18
- 4.6 Feasibility and efficiency 20
- 4.7 Hybrid compression 27
- 4.8 Bilateral compression 28
- 4.9 Efficiency dominance among compression benchmarks 29

5 Empirical application **31**
- 5.1 Analysis strategy 31
- 5.2 Dataset description 32
- 5.3 General statistics 33
- 5.4 Quantifying excess and the efficiency of compression in original markets 34
- 5.5 Quantifying excess and the efficiency of compression in bilaterally compressed markets 36

6 Closing remarks **44**

A Proofs **50**

B A simple example with 3 market participants **62**

C Compression Algorithms **64**
- C.1 Non-Conservative Algorithm 64
- C.2 Conservative Algorithm 65
- C.3 Bilateral Algorithm 66

D Programming characterization and optimal algorithm **67**

1 Introduction

In contrast to centrally organized markets where quotes are available to all market participants and exchange rules are explicit, participants in over-the-counter (OTC) markets trade bilaterally and have to engage in search and bargaining processes. The decentralized nature of these markets makes them opaque as market information is often very limited for most agents. As a result of search frictions, dealers play the role of market makers and intermediate between buyers and sellers of a given asset (Duffie et al., 2005). Several OTC markets have an important role in the economy (Duffie, 2012) and, as in the case of OTC derivatives, can be large.[1] The size and lack of transparency of these markets have become an important concern for policy makers.[2]

In this paper, we show that the networked nature of decentralized markets where trading takes place over-the-counter generates *excess* of notional when trades are fungible and contingent. Formally, we define the excess of a market as the positive difference between the total outstanding gross notional of the market and the minimum aggregate amount required to satisfy every participants' net position. Intuitively, the excess of a market measures the amount of notional resulting from redundant trades, that is, trades that offset each other.

In turn, the existence of excess makes OTC markets *compressible*, i.e., the web of outstanding trades can be modified in order to remove redundant trades and, by doing so, reduce its excess. The main contribution of this paper is to provide a theoretical framework to understand and quantify the redundancy of trades leading to excess, propose methods to remove excess and provide an empirical quantification of the efficiency of each approach by applying the framework to a unique, transaction-level dataset on over-the-counter derivatives.

From an accounting perspective, the existence of a large excess in a market implies that an important gap exists between net and gross balance sheet based measures. Relying on one measure or the other thus leads to a distorted view of the market (Gros, 2010). Figure 1 illustrates the situation by mapping the network of obligations of an actual OTC market for Credit-Default-Swap (CDS) contracts. CDS buyers are on the left hand-side (green), sellers are on the right hand-side and dealers are in the middle (blue and purple). We observe two separate sets of obligations: customer-dealer obligations and dealer-to-dealer obligations. The first line below the figure retrieves the share of gross notional per set of market participants. The second line retrieves the average ratio between net and gross

[1] OTC derivatives markets amounted to \$ 553 trillion of outstanding gross notional at end of June 2015 (BIS, 2015).

[2] In September 2009, the G20 leaders committed to make OTC derivatives markets more transparent by mandating central clearing for certain derivative classes alongside mandatory reporting to trade repositories.

individual positions for participants in each set. While buyers and sellers have a combined gross share of less than 5%, their net position is equal to their gross position. In contrast, the set of dealers concentrates more than 95% of gross market share while, on average, only one fifth is covered by net positions. This characteristic shows that, on average, 80% of the notional flowing through the dealers is the result of offsetting trades.

In practice, some markets are already implementing mechanisms to reduce their excess levels. Indeed, firms engaging in certain derivatives markets eliminate some of the excess through the use of so-called *portfolio compression*. Portfolio compression is a post-trade netting technique through which market participants can modify or remove outstanding contracts and create new ones in order to reduce their overall market gross position without modifying their net positions.[3] In other words, compression aims at reducing counterparty risks of derivatives portfolios without changing their market risks. The methods we present in this paper follow the same principle.

Let us illustrate portfolio compression with the stylized example shown in Figure 2(a) of a market consisting of 4 institutions (i, j, k, l) selling and buying the same asset with different notional values: i has an obligation of notional value 5 to j, j has an obligation to k of notional value 10, k has obligations 20 and 10 towards k and l respectively.

The aggregate gross notional of the market is thus the sum of the individual contracts: $x = 5 + 10 + 20 + 10 = 45$. At the individual level, the gross notional position of i is equal to the sum of trades in which each i is involved: $5 + 20 = 25$. Instead, the net notional position of i is the difference between the amount due by i and the amount due to i: $5 - 20 = -15$. A way to compress the market is, for example, to remove the contract between i and j and accordingly reduce the obligations that both firms have with k by an amount of 5. The result is illustrated in Figure2(b). In such "compressed" market, the net position of each firm is the same as in Figure2(b) while the gross notional of the market is now given by: $x' = 5 + 15 + 10 = 30$. We have thus removed 15 units of notional from the market without modifying participants' net positions.

The above example represents a case of *multilateral* compression, i.e., several counterparties are involved and the exercise is run over the whole set of fungible

[3]Formally, the Markets in Financial Instrument Regulation (MiFIR, EU Regulation No 600/2014, Article 2 (47)) defines portfolio compression as follows: "Portfolio compression is a risk reduction service in which two or more counterparties wholly or partially terminate some or all of the derivatives submitted by those counterparties for inclusion in the portfolio compression and replace the terminated derivatives with another derivatives whose combined notional value is less than the combined notional value of the terminated derivatives".

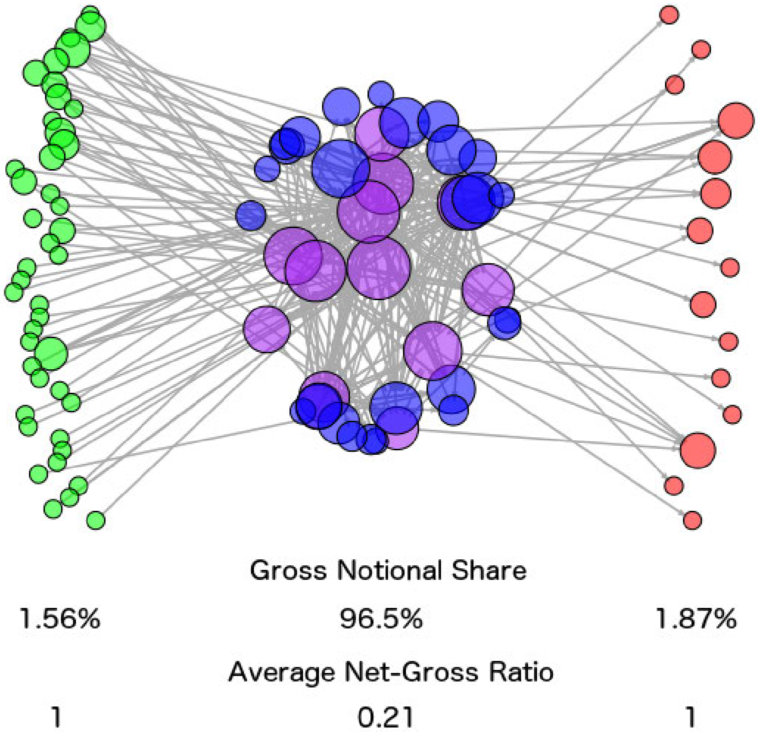
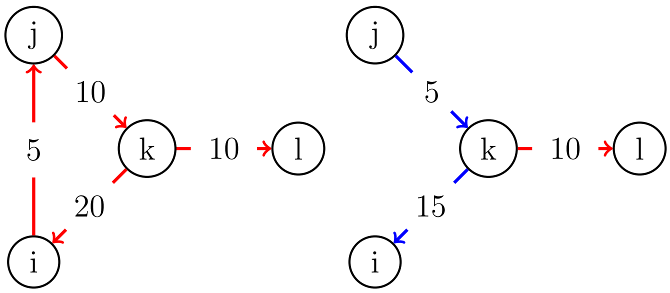


Figure 1: Network illustration of a real OTC derivative market, which maps all outstanding trades for credit default swap (CDS) contracts written on the same sovereign government reference entity for the month of April 2016. The data were collected under the EMIR reporting framework and thus contain all trades where at least one counterparty is legally based in the EU. Green nodes correspond to buyers. Red nodes correspond to sellers. Purple nodes are G16 dealers. Blue nodes are dealers not belonging to the G16 dealers set. The first line below the network report the share of gross outstanding notional based on individual positions for the segments: buyers, dealers, sellers. The second line reports the average net-to-gross ratio for each segment with the standard deviation in parenthesis for the dealer segment.



(a) Before Compression (b) After Compression

Figure 2: A graphical example of compression

trades outstanding between all counterparties.[4] Naturally, information disclosure is needed in order to run such process. Individual counterparties might not know the exact presence and amount of trades they are not directly involved in. In case institutions are reluctant to disclose their positions to other participants while still seeking to compress, a solution is to involve a third party (for instance, a dedicated service provider) that would be required to run the compression analysis. Such entity would recover portfolio information from each market participant seeking to compress their positions, reconstruct the web of trades, and propose a global compression procedure that satisfies every stakeholder.[5]

Despite portfolio compression being born out of the regulatory perimeter (Duffie et al., 2016), several regulatory bodies and post-crisis regulations have

[4]In the bilateral case where two institutions share several fungible trades that go in both directions, the exercise is much simpler as it merely consists of removing all bilateral contracts and creating a new contract between the two same institutions with a notional value equal to the net value of all original outstanding contracts.

[5]While the set of participants is theoretically heterogeneous (e.g., banks, insurances, funds, etc.), the list of existing service providers is limited. TriOptima, LMRKTS, Markit, Catalyst and SwapClear are among the most active compression service provider in OTC markets. Most compression operations are run on cleared and uncleared Interest Rate Swaps (IRS) and index and single-name Credit Default Swaps (CDS). Central Clearing Counterparties are increasingly involved in compression as well. Other instruments are also starting to be compressed: cross currency swaps, commodity swaps, FX forward, inflation swap. According to the International Swaps and Derivatives Association (ISDA), portfolio compression is responsible for a total of \$448.1 trillion of IRS derivatives elimination between 2003 and 2015 (ISDA, 2105). According to TriOptima, their portfolio compression service TriReduce has eliminated over \$861 trillions in notional until September 2016 (continuous updates are reported in `http://www.trioptima.com/services/triReduce.html`).

recently supported its adoption.[6]

As advertised by compression providers[7], the increasing interest for compression results from several benefits at the level of the individual market participant. Overall, we can distinguish between three major incentives for institutions to engage in compression. First, compression reduces counterparty risk. As contracts are removed and replaced by new contracts with lower notional amounts, counterparty risk deriving from the gross exposures to those trades is reduced. Second, compressing portfolios alleviates some regulatory constraints. For instance, banks' capital rules for derivatives are computed on the basis of gross exposures.[8] Hence, reducing the notional amounts of contracts can help reducing the corresponding capital needs of market participants. Third, by reducing the number of contracts, compression leads to a reduction of operational risks and an improvement of management, including trade count reduction, speed to auction in case of default, lower cash-flow needed to settle obligations, fewer reconciliations, lighter burden of settlement, lowered collateral and margin requirements, etc.

Despite the growing use of portfolio compression, limited policy and academic work has been devoted to understanding the determinants of excess, compression operations and the subsequent externalities. A more elaborated view of those aspects is indeed relevant to ensure a proper design and implementation of compression in OTC markets. Furthermore, compression affects market dynamics in ways that can distort the liquidity and risk analysis. In particular, contracts being terminated or created because of compression are subject to specific accounting and economic forces which should not be confused with other types of contract termination and creation processes. As monitoring markets in terms of both liquidity and counterparty risk is paramount to both micro and macro-prudential policy, the effects of compression must be explicitly accounted for. The current work seeks to fill in this gap by providing analytical and empirical insights at the individual and systemic level.

In this paper, we show that intermediation, determined by the existence of chains of fungible and outstanding trades, is *per se* a sufficient condition to observe

[6]For example, under the European Market Infrastructure Regulation (EMIR), institutions that trade more than 500 contracts with each other are required to seek to compress their trades at least twice a year. Article 14 of Commission Delegated Regulation (EU) No 149/2013 of 19 December 2012 supplementing Regulation (EU) No 648/2012 of the European Parliament and of the Council with regard to regulatory technical standards on indirect clearing arrangements, the clearing obligation, the public register, access to a trading venue, non-financial counterparties, and risk mitigation techniques for OTC derivatives contracts not cleared by a CCP (OJ L 52, 23.2.2013, p. 11- 'Commission Delegated Regulation on Clearing Thresholds' or 'RTS')

[7]See for example the advertising brochure by Swapclear: `http://www.swapclear.com/Images/lchswapcompression.pdf`

[8]For example capital requirements under the Basel framework are computed including gross derivatives exposures (BIS, 2016)

positive excess levels in markets. Dealers are thus at the heart of the generation of redundant trades. However, the share of excess that can be removed (i.e., redundant excess) is a function of potential constraints, which we refer to as *compression tolerances*, set by both individual participants and regulators. Hence, compression does not always remove the total amount of excess (i.e., there can be some residual excess after compression). We introduce a spectrum of benchmark compression tolerances settings and investigate their feasibility and efficiency. More precisely, we consider approaches that differ in the conservation of counterparties' trading relationships existing before compression. We show that a trade-off exists between the efficiency of compression and the levels of compression tolerance.

Finally, we provide an empirical assessment of our framework. Using a unique and granular dataset comprising all CDS contracts traded by institutions based in the European Union (EU) between 2014 and 2016, we are able to quantify the levels of excess exhibited by real OTC markets as well the efficiency of the different compression approaches. We find that the vast majority of markets defined as the set of transactions written on the same reference entity with the same maturity, exhibit levels of excess accounting for 75% or more of the total gross notional. While around 50% can in general be removed via bilateral compression, multilateral compression approaches can remove almost all the excess. In particular, we find that even the most conservative approach which satisfies pre-existing relationship constrains can eliminate up to 98% of excess in the markets. Nevertheless, we find that the efficiency conservative multilateral compression is impaired if market participants first seek to bilaterally compress their positions. This sensitivity is dampened when the constraints on compression are relaxed on the intra-dealer segment of markets.

Despite the application of our framework on derivatives markets, our findings and methods can similarly be applied to other OTC markets. Indeed, as long as a market exhibits fungibility, contingency and intermediation, our framework identifies levels of excess and offers ways of reducing it. Hence markets such as credit or bond markets are other potential candidates for such exercise.

The rest of the paper is organized as follows. We provide an overview of the relevant literature for this work in Section 2. In Section 3, we introduce the general setting for our analysis describing a model of an OTC market and the formal definition of excess. Section 4 provides the core of the paper. It describes compression as a network operation over the market; discusses the issues of compression tolerances; proposes benchmark cases; analyses the feasibility and efficiency of each approach. In Section 5, we report the results of our empirical analysis of excess and compression efficiency in real OTC derivatives markets. Last, we conclude and discuss avenues for further research as well as comments on some operational and regulatory aspects of compression in Section 6. The appendices provide proofs of the propositions and lemmas as well as the analytical details for the algorithms used in the paper.

2 Literature review

The study of the structure of OTC markets has gained attention in the last decade prompted by both their role in the 2008 financial crisis and the increased data availability for these previously opaque markets (Duffie, 2012; Abad et al., 2016).

Despite the use of different datasets and the focus on different instruments, many contributions show similar findings. First, there are typically two types of market participants: dealers and customers. Customers enter the market either to buy or sell a particular product while dealers act as intermediaries by concomitantly buying and selling and tend to keep balanced positions (i.e., they seek to have a relatively low net balance of contracts bought and sold with respect to their gross amount). In particular, this feature has been well documented for derivatives markets by Shachar (2012); Benos et al. (2013); Peltonen et al. (2014); D'Errico et al. (2016); Abad et al. (2016); Ali et al. (2016). To a larger extent, this same feature is in line with the core-periphery structure of OTC credit markets reported by Craig and Von Peter (2014); in't Veld and van Lelyveld (2014); Fricke and Lux (2015). In general, Craig and Von Peter (2014) show that institutions acting as dealers are typically large banks.

Atkeson et al. (2015) propose a parsimonious theoretical model which generates the above described feature for derivatives markets: they show that banks entering an OTC derivatives market for incentives stemming from intermediation profit need to be larger to bear the entry cost while not benefiting from a long or short position in the market. D'Errico et al. (2016) empirically observe that in the global OTC CDS market these intermediaries form a very tight structure which entails closed intermediation (exposure) chains. The authors report that this structure occurs on almost all reference entities and suggest its relation to "hot potato" trading, a feature of OTC derivatives observed and modeled by Burnham (1991); Flood (1994); Lyons (1995, 1997).

Furthermore, these markets are characterized by large concentration of notional within the intra-dealer segment. For the CDS market, Atkeson et al. (2013) report that, in the US, on average, about 95% of OTC derivatives gross notional held on banks' balance sheet is concentrated in the top five banks; D'Errico et al. (2016) show that between 70% and 80% of the notional in CDS markets is in the intra-dealer market across all reference entities. Abad et al. (2016) report similar levels for IRS markets and FX markets in the EU market.

Overall, this large amount of intra-financial exposures relates to a deeper role of financial intermediation, as detailed by Allen and Santomero (1997), who find that certain derivative markets have mainly become "markets for intermediaries rather than individuals or firms". Furthermore, the authors note that standard intermediation theories could not explain the large surge in intermediation as merely a result of reduced transaction costs and informational asymmetries.

In contrast to research efforts cited above to better understand OTC markets, limited attention has been devoted to market compression in the literature. This reflects both the novelty of this financial innovation and the recent adoption by market participants due to the contemporary regulatory changes (e.g., the Basel III leverage ratio framework which accounts for derivative gross exposures).

On the theoretical side, O'Kane (2014) stands as the main contribution. The author analyses, by means of simulations, the performances of different compression algorithms on a synthetic network where all banks are connected. The benchmark algorithm is in the spirit of the approach followed by compression service providers to the author's claim and is based on a depth-first search algorithm. The author shows that, if performed optimally, compression mitigates counterparty risk and suggests compression be encouraged by regulators. On the empirical side, Benos et al. (2013) use CDS transaction data from the UK to show that reduction breaks in dealers' gross positions are due to compression, suggesting the monthly frequency of compression cycles. The Bank for International Settlements highlighted the role of compression in the global reduction of the gross notional size of OTC derivatives markets (Schrimpf, 2015; Ehlers and Eren, 2016).

Compression is a particular method to net exposures. Other works have looked at ways of netting liabilities via the introduction of Central Clearing Counterparties (CCPs). Duffie and Zhu (2011) provide the ground work of this approach. The authors show that, while central clearing helps reduce exposures at the asset class level, differentiating asset classes in clearing design can rip of the benefits of netting. Interestingly, the authors also suggest that compression reduces the needs and benefits of central clearing. Cont and Kokholm (2014) build on the model developed by Duffie and Zhu (2011) and explore the effect of heterogeneity across asset classes. The authors show that a more risk sensitive approach to asset classes can alleviate the need to concentrate all netting activities in one single CCP. Our paper shows the extent to which exposure reduction is feasible without the introduction of new market players (e.g., CCP).

In spirit, this paper is also related to works on the *gridlock problem* in payments system as discussed by Flannery (1996). In particular, Rotemberg (2011) analyzes the issues of minimal settlements and models a system of payments interconnected via due payments. The author identifies conditions under which the market can be cleared with minimal endowments of liquid assets. This approach is relevant as compression can also be seen as a procedure that seeks to reduce the conditional payments without affecting the expected net flow from each market participant. Importantly, the author shows that in the absence of closed chains of intermediation, solvency necessarily implies the settlements of all obligations.

From a policy perspective, our work relates to ongoing debates on the adequacy of the regulatory framework. In particular, the way net and gross positions information are currently used under different accounting rules is subject to concerns as they do not allow to fully capture the risks associated (see Blundell-Wignall and Atkinson, 2010). In particular, Gros (2010) shows that under different legislation (i.e., US and Europe), the same financial institutions can exhibit very different profiles. Compression, by affecting the gross levels without changing the net levels, can therefore have an effect on the accounting approach followed by policy makers and other market analysts.

Finally, our work relates to the growing stream of works highlighting the important relationship between interconnectedness and systemic risk in financial markets (see Allen and Babus, 2009; Yellen, 2013). These works explore the role of interdependencies on the propagation of distress at different levels: link formation (Babus, 2016; Gofman, 2016), default cascades (Allen and Gale, 2000; Elliott et al., 2014; Acemoglu et al., 2015) and regulatory oversight (Roukny et al., 2016). This paper contributes to this literature by showing how post-trade practices can affect the network profile of a financial market. Intuitively, compression affects counterparty risk which has held a central role in the unfolding of the 2007-2009 financial crises together with OTC derivatives markets (Haldane, 2009; European Central Bank, 2009).

However, in this work, we do not build explicit links between compression and systemic risk. The focus we take is rather on providing a first comprehensive framework to understand the mechanics underlying the possibilities to reduce gross notional: future work will be dedicated to the effects of compression on systemic risk. Nevertheless, the results of this work provide way to understand how counterparty risk reduction, collateral demand and capital requirements can be modified in post-trade situations which play a role in financial stability.

3 The market

We consider an Over-The-Counter (OTC) market made of n market participants (institutions) denoted by the set $N = \{1, 2, ..., n\}$. These institutions trade contracts with each other and establish a series of bilateral obligations. While we keep the contract type very general, we assume that these obligations are *fungible*, that is, the traded contracts have the same payoff structure from the market participants' perspective and can thus be algebraically summed. The whole set of outstanding obligations in the market constitutes the financial network. Formally, we have the following definition:

Definition (Financial Network). *The network or graph G is the pair (N, E) where*

N is a set of institutions present in the market and E is a set of directed outstanding fungible obligations (i.e., edges) between two institutions in the market. An outstanding obligation is represented by e_{ij} whose value corresponds to the notional value of the obligation and the directionality departs from the seller i to the buyer j with $i, j \in N$.

From the financial network, we infer two measurements of an individual's position in the OTC market: the *gross position* and the *net position.* On the one hand, the gross position of an institution i is the sum of all obligations' notional value involving this institution on any side of the trade (i.e., buyer and seller).

Definition (Gross position)**.** *The gross position of i is given by:*

$$v_i^{gross} = \sum_j e_{ij} + \sum_j e_{ji} = \sum_j (e_{ij} + e_{ji})$$

On the other hand, the net position of an institution i is the difference between the sum of the notional values all i's obligations' towards other nodes in the network and the sum of the notional values of the obligations from other nodes in the network to i:

Definition (Net position)**.** *The net position of i is given by:*

$$v_i^{net} = \sum_j e_{ij} - \sum_j e_{ji} = \sum_j (e_{ij} - e_{ji})$$

We also define the *total gross notional* of the market as the sum of the notional amounts of all trades:

Definition (Total gross notional)**.** *The total gross notional of a market $G = (N, E)$ is given by:*

$$x = \sum_i \sum_j e_{ij}$$

We further classify market participants according to their activity in the market. A market can contain two types of institutions: customers and dealers. Customers only enter the market to buy or sell a given contract and are active on one side of each trade. In contrast, dealers also intermediate between other market participants and, thus, act both as buyers and sellers of the same contract type. We use the following indicator to identify dealers in the market[9]:

[9]Note that this definition would consider two participants selling and buying from each as dealers (i.e., $e_{ij}.e_{ji} > 0$) which can be misleading. However, our framework accounts for those cases in the same way and the same result hold. Hence, we employ in the first part of we will employ the notion of dealer in a general way. In the second part, which analyses empirical data, we will define dealers as intermediaries between different market participants (i.e., after having bilaterally netted all contracts). While, the results remain the same, this more elaborated definition will help the interpretation of the empirical findings.

Definition (Dealer indicator). *Given a market $G = (N, E)$, let $\delta()$ indicate whether a market participant is a dealer in the market in the following way:*

$$\delta(i) = \begin{cases} 1 & \text{if } \sum e_{ij} . \sum e_{ji} > 0 \qquad \textbf{(dealer)} \\ 0 & \text{otherwise} \qquad\qquad\quad \textbf{(customer)} \end{cases}$$

In a sense, we generalize the modeling approach of Atkeson et al. (2015) with regard to market participant types. Note that, as a result, only three types of trading relationships can exist in the market: dealer-customer, dealer-dealer and customer-customer.

3.1 Definition of excess

We now elaborate on the concept of excess and the condition for markets to exhibit positive levels.

Let us start by introducing a post-trade mathematical operator that acts upon an extant market in order to modify the set of outstanding liabilities. Such operation can be subject to different types of constraints. Here we focus on the concept of *net-equivalence*. In our framework, an operation on a networked market is net-equivalent if, despite exhibiting a different set of edges, the resulting market keeps the net position of each institution equal to its original value (i.e., before the operation). Formally, we have:

Definition (Net-Equivalent Operation). *Given a market $G = (N, E)$ an operation $\Omega()$ such that $G' = \Omega(G) : (N, E) \to (N', E')$ is net-equivalent if*

$$N = N'$$

and

$$v_i^{net} = v_i^{\prime net} \qquad \forall i \in N$$

where v_i^{net} and $v_i^{\prime net}$ are the net positions of i in G and G' respectively.

Notice that the networks G and G' differ by the configuration of their obligations which could be due to changes in the notional value of existing trades or creation and removal of trades. Furthermore, the aggregate gross notional of each net-equivalent market does not need to be equal.

We now show that, given an original market, it is possible to compute the minimum level of gross notional that can be obtained from a net-equivalent market.

Proposition 1. *Given a market $G = (N, E)$, if a net-equivalent operator Ω on G is such that:*

$$G' = \Omega(G) = \min_{x'}(\Omega(G) : (N, E) \to (N', E'))$$

then

$$x' = \frac{1}{2}\sum_{i=1}^{n} |v_i^{net}| = \sum_{i:\ v_i^{net}>0}^{n} v_i^{net} \tag{1}$$

Proof. See Appendix ■

In fact, as the market we defined is a closed system (i.e., both sides of all the trades are market participants, $\forall e_{i,j} \in E, i \in N$ and $j \in N$), the sum of all net positions must be equal to zero ($\sum_i v_i^{net} = 0$). Nevertheless, looking only at the institutions with a positive net position (i.e., institutions for which total selling outbalances total buying), we obtain the total out-flow of the market. This total out-flow is necessarily equal, in absolute values, to the total in-flow obtained from all the institutions with a negative net notional. The out-flow is also equal to half the absolute sum of all net notional positions as the sum of all positive and all negative net positions are equal. If the total amount of notional in the market is smaller then the total out-flow, there will be no configuration of trades such that the resulting market is net-equivalent because there will exist at least one market participant with $\sum_j (e'_{ij} - e'_{ji}) < v_i^{net}$. Hence, in order to be net-equivalent, the resulting market's gross notional must be at least equal to the total out-flow. Note that there can exist several G' but they all share the same level of gross notional (i.e., $v'^{gross} = \frac{1}{2}\sum_{i=1}^{n} |v_i^{net}|$).

We can now formally define the excess of a market. In fact, if, for a given market, there exists a net-equivalent operation that reduces the aggregated gross notional, we conclude that the original market exhibits trades that can be removed or modified without affecting the net positions of any market participant.

Given the previous result, we can quantify the total level of excess in a market as the difference between the aggregate gross notional of a given market and the aggregate gross notional of the net-equivalent market with the minimum market aggregate gross notional. Formally, we define and quantify the excess in a market as follows:

Definition (Excess)**.** *The excess in the market is defined as*

$$\Delta(G) = x - x' \tag{2}$$

$$= \left(\sum_{i=1}^{n}\sum_{j=1}^{n} e_{ij} - \frac{1}{2}\sum_{i=1}^{n} |v_i^{net}|\right) \tag{3}$$

$$= \left(\sum_{i=1}^{n}\sum_{j=1}^{n} e_{ij} - \sum_{i:\ v_i^{net}>0}^{n} v_i^{net}\right)$$

Note that Equation 2 and Equation 3 are equivalent as long as the market under study is a closed-system. The excess in the market is thus the amount of notional generated by trades that offset each other: it corresponds to the amount of notional that can be removed without affecting the net position. Note that, at this stage, we are not accounting for the potential positive value of some offsetting outstanding contracts for market participants or market regulators. We elaborate on that aspect in Section 4.2.

3.2 Existence condition

Not all markets exhibit notional excess. As mentioned above, the existence of excess is due to the existence of a difference between net and gross positions of (some) individual positions. In the following, we identify a necessary and sufficient condition for excess to emerge in a market: the existence of intermediation. In fact, for excess to exist in the market, we need at least one institution to have its gross position larger than its net position. As we show below, such case only exists if the institution is selling and buying the same type of contract at the same time (even if done at different levels of notional), that is, if the institution is a dealer. From a network perspective, this situation is present when there exists at least two edges where the same institution is found at each ends. More formally, we define intermediation as follows:

Definition (Intermediation)**.** *A market $G = (N, E)$ exhibits intermediation i.i.f.*

$$\exists i \in N \quad s.t. \quad \delta(i) = 1$$

At the market level, we thus have the following result:

Lemma 1. *Given a market $G = (N, E)$, if*

$$\sum_{i \in N} \delta(i) > 0$$

then

$$\Delta(G) > 0$$

In fact, if there is no intermediation, net positions are equal to gross positions as every participant is active only on the buy or sell side (i.e., only customers in the market). As a result, markets with no intermediation do not exhibit notional excess. This result provides a global market view on the effect of intermediation in distorting gross and net measurements. It generalizes measurements at the individual level as shown in the entry-exit model of Atkeson et al. (2015). This result also explicitly shows why the existence of notional excess is intrinsic to OTC markets: the presence of dealer institutions is the source of notional excess in those markets. We conclude that the two main types of market organizations (i.e., over-the-counter and centralized exchange-traded markets) have different levels of notional excess,

Corollary 1. *Centralised exchange-traded markets exhibit no excess.*

Centralized exchange-traded market markets can indeed be framed as bipartite networks consisting of customers exclusively interacting with each other on the buy and sell spectrum and thus $v_i^{net} = v_i^{gross}, \quad \forall i \in N$.

Corollary 2. *In the presence of dealers, over-the-counter markets always exhibit positive notional excess.*

Even if some OTC markets exhibit customer-customer trading relationships, those interactions do not contribute to notional excess. It is the activity of dealers that generates notional excess both in the intra-dealer segment and in the dealer-customer segment. Several studies have stated the prevalent role of dealers in over-the-counter markets (Duffie et al., 2005) and others have shown the high levels of notional concentration in the dealers segment of OTC markets (Atkeson et al., 2013; Abad et al., 2016; D'Errico et al., 2016) as illustrated in Figure 1. We also document these feature in the empirical section of this paper.

Finally, note the special case of bilaterally netted positions. It often happens that two institutions having an outstanding trade decide to terminate this trade by creating an offsetting trade (i.e., contract of similar characteristics in the opposite direction). Such a situation also generates excess as trades are accounted for in the gross position while they do not contribute to the net position of each counterparty. While those mechanisms cannot be framed as intermediation, the formal network definition still applies (i.e., both institutions are active on the buy and sell side) and the related results are unchanged (i.e., existence of notional excess).

3.3 Excess decomposition

We now explore the decomposition of excess with respects to two segments of the market: the intra-dealer market and the customer market.

The intra-dealer (sub-)market only contains obligations between dealers while the customer (sub-)market contains obligations where at least one counterparty is a customer. Formally we have:

Definition (Intra-dealer and customer market)**.** *The set of contracts E can be segmented in two subsets E^D and E^C such that*

$$\delta(i).\delta(j) = 1 \quad \forall e_{ij} \in E^D$$

$$\delta(i).\delta(j) = 0 \quad \forall e_{ij} \in E^C$$

Where E^D is the intra-dealer market and E^C is the customer market and $E^D + E^C = E$.

In general, the excess is not additive: quantifying the excess of each segment separately does not lead to excess of the entire market. Special cases of excess additivity are presented in the following result:

Proposition 2 (Additivity of excess)**.** *Given a market $G = (N, E)$, and the two markets $G^1 = (N, E^1)$ and $G^2 = (N, E^2)$ obtained from the partition $\{E^1, E^2\}$ of E, then:*

$$\Delta(G) \geq \Delta(G^1) + \Delta(G^2)$$

which implies that:

$$\Delta(N, E) \geq \Delta(N, E^D) + \Delta(N, E^C)$$

In particular, we have additivity, $\Delta(N, E) = \Delta(N, E^D) + \Delta(N, E^C)$ if

1. $\sum_h^{dealer}(e_{dh} - e_{hd}) = 0, \quad \forall d \in D$, *or*

2. $\sum_{c^+}^{customer^+} e_{dc^+} - \sum_{c^-}^{customer^-} e_{c^-d} = 0, \quad \forall d \in D$

Proof. See Appendix. ■

The above results state that if all dealers have a zero net position w.r.t. to all their outstanding trades with (1) their dealer counterparties or (2) their customer counterparties, then the excess can be decomposed between the intra-dealer excess and the dealer-customer excess. In general, we have $\Delta(E) \geq \Delta(E^D)+\Delta(E^C)$. The insights from this results will become useful when we consider applying different methods of excess reduction for the different segments of the market.

4 Compression

Building on the framework introduced in the previous section, we now focus on ways to reduce the excess of markets, that is, we investigate the extent to which the excess of OTC markets can be *compressed*. In particular, we adopt an analogous concept as that of *portfolio compression* already in place in some derivatives markets. Portfolio compression is a technique that aims at terminating outstanding trades and creating new ones in order to reduce gross individual positions without affecting net positions.

In our framework, compression is an operation over the market's underlying network of outstanding trades that effectively reduces the excess of notional. Formally, we have the following definition of *compression* in OTC markets:

Definition (Compression)**.** *Given a market $G = (N, E)$ and a market $G' = (N, E') := c(N, E)$ is compressed w.r.t. to G if and only if*

$$v_i^{'net} = v_i^{net} \text{ and } v_i^{'gross} \leq v_i^{gross} \quad \text{for all } i \in N$$

with at least one strict inequality and where $c()$ *is a net-equivalent network operator.*

Compression, at the market level, is thus an operation on the network of outstanding trades (i.e., $c(N, E)$) that reconfigures the set of edges ($(N, E') := c(N, E)$) while (i) keeping all net positions constant (i.e. net-equivalence) and (ii) reducing the individual gross notional of at least one node. By construction, this latter property leads to a reduction of gross notional at the market level (i.e, $x' < x$). As a result, compression on a market always reduces the excess. The above definition is a canonical definition of compression. Several refinements can be added to the compression operator. We discuss these aspects in Section 4.2.

4.1 Feasibility

As, by definition, compression acts upon market excess, a direct consequence of Lemma 1 is that compression can only take place if there is intermediation in the market:

Corollary 3 (Necessary condition for compression)**.** *Compression can only take place if there is intermediation in the market.*

Similar to the excess conditions, such result informs us that centralized exchange-traded markets are not candidate for compression. Note that the intermediation condition is necessary but not sufficient as additional factors can be accounted for to determine the sufficiency of compression. Those factors, called *compression tolerances*, can limit the capacity to compress the excess of a market.

4.2 Tolerances

In realistic settings, designing a compression operator also includes factors such as individual preferences or regulatory restrictions. For instance, at the individual level, market participants might not be willing to compress certain trades; at the regulatory level, policy makers might refuse that new trades be created between specific counterparties in the market. We call these additional constraints *compression tolerances*, as they define the extent to which modifications can be applied to the set of portfolios during the compression exercise both in terms of change in currently existing contracts and creation of new ones with new counterparties. Compression tolerances thus determine the degrees of freedom for a compression operation to take place.

Formally, compression tolerances form a set of constraints at the bilateral level of each potential edge in the networked market.

Definition (Compression tolerances)**.** *A compression operator* $c()$ *s.t.* $G' = (N, E') := c(N, E)$ *satisfies the set of compression tolerances* $\Gamma = \{(a_{ij}, b_{ij})|a, b \in \mathbb{R}, i, j \in N\}$ *if*

$$a_{ij} \leq e'_{ij} \leq b_{ij} \quad \forall i, j \in N$$

with $0 \leq a_{ij} \leq e_{ij}$, $e_{ij} \leq b_{ij}$ $\forall (i, j) \in N$.

For each potential contract between two counterparties in the resulting compressed market, there exist a lower (i.e., a_{ij}) and upper bound (i.e., b_{ij}). Those constraints are tolerances and hence cannot force an expected value for the resulting obligation, that is why lower bound (resp. upper bound) cannot be higher (resp. lower) than the original obligation notional, i.e., $a_{ij} \leq e_{ij}$ (resp. $e_{ij} \leq b_{ij}$).

The levels of compression tolerances affect how much excess can be removed from compression: there is a potential opportunity cost in the efficiency of compression resulting from how participants' portfolios can be modified[10].

Finally, note that compression tolerances on a bilateral obligation (i, j) are set from the combination of both participants i and j constraints, as they must satisfy each participant's individual sets of constraints (both on the asset and the liability side).

4.3 Residual and redundant excess

The set of all individual compression tolerances determines the trades that can be deemed redundant and thus modified. Hence, the total excess of a market as in Definition 3.1 can be divided in two levels: *redundant* excess and *residual* excess. The former is the excess that can be compressed while the latter is the excess that remains after compression. The determination of those levels is conditional upon (1) the underlying network of outstanding fungible contracts and (2) the set of compression tolerances set by the market participants or the market regulator. Formally, we have:

Definition (Residual and redundant excess)**.** *A compression operator* $c()$ *s.t.* $G' = (N, E') := c(N, E)$ *satisfying the set of compression tolerances* $\Gamma = \{(a_{ij}, b_{ij})|a, b \in \mathbb{R}, i, j \in N\}$ *generates:*

- $\Delta_{res}(G) = \Delta(G')$ ***(residual excess)***

[10] In the context of clients to a compression service provider, compression tolerances determine how much the compression participants clients are willing not to alter their original positions. In derivatives markets, service providers such as TripOptima refer to these constraints as *risk tolerances*. As they can reduce the efficiency of a compression exercise, bargaining can also take place between the service provider and its clients in order to modify those constraints. *Dress rehearsals* are steps in the compression exercise where the service provider informs all the clients on a candidate compression solution and seeks their confirmation. Several iterations can be needed before an optimal solution satisfying all participants is reached.

- $\Delta_{red}(G) = \Delta(G) - \Delta(G')$ ***(redundant excess)***

We have the following relationship: $\Delta(G) = \Delta_{res}(G) + \Delta_{red}(G)$

4.4 Efficiency

Given a market, there exist many possible compression operations. In order to compare them, we associate each compression operator $c_k(N, E)$ with its redundant excess. We can thus assess the *efficiency* of different compression operations using the associated levels of excess reduction.

Definition (Efficiency of Compression). *A compression operator over a network G, $c_s(N, E)$ is more efficient than another compression operator, $c_t(N, E)$ if*

$$c_s(N, E) \succ c_t(N, E) \Leftrightarrow \Delta^s_{red}(G) > \Delta^t_{red}(G)$$

From this definition it appears that a compression operator that yields a complete reduction of the overall excess achieves the highest level of efficiency (i.e., $\Delta_{res}(G) = 0$).

The definition can be re-expressed in relative terms by introducing a *compression ratio*, i.e.:

$$c_s(N, E) \succ c_t(N, E) \Leftrightarrow \rho_s > \rho_t$$

Where $\rho_s = \frac{\Delta^s_{red}(G)}{\Delta(G)}$ and $\rho_t = \frac{\Delta^t_{red}(G)}{\Delta(G)}$ are the compression ratios of c_s and c_t respectively, i.e., the fraction of notional obligation eliminated via the compression operation. The ratio provides a natural way to compare different compression operators when applied to networks where obligations are of a dissimilar type (e.g. expressed in different currencies or with different underlying in case of a derivative).

4.5 Benchmark approaches

In practice, compression tolerances are set to cover a wide range of heterogeneous preferences from market participants and regulators. As a result, the space of possible compression tolerance combinations is infinite. Nevertheless, in the following, we study specific compression benchmarks as ways to define the conditions and maximum levels of compression that can be achieved according to some standardized set of preferences. As such, we consider the two following case:

1. $(a_{ij}, b_{ij}) = (0, e_{ij}) \;\; \forall i, j \in N$
2. $(a_{ij}, b_{ij}) = (0, +\infty) \;\; \forall i, j \in N$

Those two benchmarks incorporate different preferences with regards to previously existing trading relationships. As such we call these approaches *conservative* and *non-conservative*. Intuitively, the non-conservative case has the highest levels of compression tolerance: it discards all counterparty constraints. The approach is deemed non-conservative with respect to the original web of contracts in the market. In the conservative case: compression tolerances are such that $e'_{ij} \leq e_{ij}$ for all links. The compression tolerances are such that all original dependencies can be reduced or removed but no new relationships can be created. It is conservative with respect to the original trading relationships of the market. Below, we formalize those two approaches.

4.5.1 Non-conservative compression

In the non-conservative compression approach: the resulting set of new trades E' is not determined in any way by the previous configuration of trades E.

Definition (Non-Conservative Compression). $c(N, E)$ *is a non-conservative compression operator i.f.f.* $c()$ *is a compression operator that satisfies the compression tolerances set* Γ*:*

$$a_{ij} = 0 \quad \textit{and} \quad b_{ij} = +\infty, \quad \forall (a_{ij}, b_{ij}) \in \Gamma,$$

In practice, such benchmark approach is unlikely to be the default modus operandi. However, it is conceptually useful to study as it sets up the bar for the most compression tolerant case.

4.5.2 Conservative compression

The second compression approach is defined as conservative. A compression operation is *conservative* if the set of new trades resulting from the compression is strictly obtained from the reduction in notional values of previously existing trades. Trades can be removed (i.e., complete reduction of notional) but no new trade can be introduced. Formally, we have:

Definition (Conservative Compression). $c(N, E)$ *is a conservative compression operator i.f.f.* $c()$ *is a compression operator that satisfies the compression tolerances set* Γ*:*

$$a_{ij} = 0 \quad \textit{and} \quad b_{ij} = e_{ij}, \quad \forall (a_{ij}, b_{ij}) \in \Gamma, e_{ij} \in E$$

The resulting graph $G' = (N, E')$ *is a 'sub-graph' of the original graph* $G = (N, E)$*.*

Such benchmark approach is arguably close to the way most compression takes place in derivatives markets (O'Kane, 2014).

We provide a simple example of a market consisting of 3 market participants in the Appendix B

4.6 Feasibility and efficiency

For each compression approach, we identify the conditions under which compression can take place and the efficiency of each approach. We conclude by proving the existence of a trade-off between the approaches and the level of efficiency.

4.6.1 Non-conservative compression

With non-conservative compression operators, the set of trades prior to compression does not matter for determining the new set of trades, only the net and gross positions of each individual does. We can thus generalize the Corollary 3 as follows:

Proposition 3. *Given a market $G(N, E)$ and compression $c^n()$ satisfying a non-conservative compression tolerance set Γ:*

$$\Delta_{red}^{c^n}(G) > 0 \quad \Leftrightarrow \quad \sum_{i \in N} \delta(i) > 0$$

Furthermore, once non-conservative compression is possible, we can analyze the efficiency of such compression operation. The efficiency criterion is solely based on the amount of excess notional that is successfully removed after compression is applied. Given the role of intermediation in generating excess, removing chains of intermediation present in the network directly reduces the excess. Recall from Lemma 1 that if all intermediation chains are broken, the market exhibits zero excess. Moreover, the resulting market is composed of two kinds of participants: selling customers on one side and buying customers on the other side. No institution combines both activities anymore, that is, non-conservative compression either removes dealers from the market (if their net position is zero) or makes them buying customers (resp., selling customers) if their net position is negative (resp., positive). Such market is thus necessarily characterized by a directed bipartite underlying network structure:

Definition (Directed Bipartite Graph)**.** *A graph G=(N,E) is bipartite if the set of nodes can be decomposed into 2 subsets N^{out} and N^{in} where each set is strictly composed of only one kind of node: respectively, nodes with only outgoing edges and nodes with only incoming edges. The edges are characterized as follows: e_{ij} with $i \in N^{out}$ and $j \in N^{in}$. Also, a bipartite graph has no dealers:*

$$\sum_{i \in N} \delta(i) = 0$$

Note that any compression operator that transforms a market with intermediation into a market that is bipartite is necessarily non-conservative. More importantly, any compression operation leading to a bipartite structure is also a perfectly efficient compression as all the excess becomes redundant:

Proposition 4. *Given a market $G = (N, E)$, there exists a set of non-conservative compression operators C such that*

$$C = \{c^n | \Delta^{c^n}_{res}(G) = 0\} \neq \emptyset$$

Moreover, let $G' = c^n(G)|c^n \in C$, then G' is bi-partite.

Proof. See Appendix ■

The proof of existence stems from the following algorithm: from the original network, compute all the net positions then empty the network and generate edges such that the gross and net positions are equal at the end. As net and gross positions are equal, the resulting market has a bipartite underlying architecture: there is no intermediation.

Corollary 4. *Given a market $G(N, E)$ a compression operation $c(N, E)$:*

$$\Delta^{c}_{res}(G) = 0$$

if

$$\sum_{i \in N'} \delta(i) = 0$$

Hence, by generating a method that removes all intermediation in the market, while keeping the net positions constant, all the (redundant) excess is removed. Such a method can be formalized under an algorithmic framework. Obviously, there exist many ways to devise algorithm that conduct intermediation removal and there exist multiple solutions that achieve a similar level of efficiency. For illustrative purposes, we provide a simple algorithm for such type of compression in the Appendix.

Remark: a more realistic approach to non-conservative compression

In a realistic setting, exposure limits exist, either set by individuals or by regulators. In the non-conservative case, this implies a cap on the upper bound of each compression tolerance (i.e., b_{ij}). In the following we consider the case of a non-conservative compression with a common exposure limit set to any bilateral relationship in the market (e.g., set by the regulator). We have:

$$(a_{ij}, b_{ij}) = (0, \lambda) \quad \text{with} \quad \lambda > max\{e_{ij}\} \quad \lambda \in \mathbb{R}^+.$$

The value of λ will affect the efficiency of the non-conservative case. Nevertheless, it is possible to determine the value beyond which the previous results on the efficiency of non-conservative compression still hold (i.e., achieving full compression).

Proposition 5. *Given a market $G = (N, E)$, if compression tolerances $\Gamma = \{(a_{ij}, b_{ij}) | a, b \in \mathbb{R}, i, j \in N\}$ are set such that:*

$$(a_{ij}, b_{ij}) = (0, \lambda) \quad \textit{with} \quad \lambda > \max e_{ij} \quad \lambda \in \mathbb{R}^+$$

then,

$$C = \{c | G' = c(G) : \Delta^c_{res}(G) = 0\} \neq \emptyset \quad \Leftrightarrow \quad \lambda \geq \frac{|v^{net}_{i^*}|}{|N^{-1.sign(v^{net}_{i^*})}|}$$

where $i^* \in N$ s.t. $|v^{net}_{i^*}| = max\{|v^{net}_i| \ \forall i \in N\}$

Proof. See Appendix ■

More generally we see that, a solution with 0 residual excess is possible for any compression tolerance set Γ that satisfies the following conditions:

$$a_{ij} = 0, \quad b_{ij} \geq \frac{|v^{net}_{i^*}|}{|N^{-1.sign(v^{net}_{i^*})}|} \qquad \forall (a_{ij}, b_{ij}) \in \Gamma$$

A regulator can thus identify conditions under which all the excess can be removed from the system under regulatory constraints on the exposure limit.

4.6.2 Conservative compression

In the conservative case, an operator can only reduce or remove existing trades. As we noted before, only non-conservative compression can be applied to general chains of intermediation as the breaking of intermediation chains generates new ties. Nevertheless, when chains of intermediation are closed, we show that compression can be used without requiring the creation of new ties. Let us formalize the concept of closed intermediation chains:

Definition (Directed Closed Chain of Intermediation)**.** *A directed closed chain of intermediation is a set of edges $K = (N, E)$ arranged in a chain of intermediation such that the first and last node are the same and no other node appears twice in the set:*

$$E = \{e_{1,2}, ..., e_{i,i+1}, ..., e_{n,1}\}$$

Hence

$$\prod_{i,j} e_{ij} > 0$$

This structure constitutes the necessary and sufficient condition for conservative compression to be applicable to a market:

Proposition 6. *Given a market $G(N, E)$ and a compression operator c^c satisfying a conservative compression tolerance set Γ:*

$$\Delta^{c^c}_{red}(G) > 0 \quad \Leftrightarrow \quad \exists E^* \subset E \quad s.t. \prod_{e^* \in E^*} e^* > 0$$

Proof. See Appendix ■

Next, we show that the most efficient conservative compression (i.e., compression that removes the highest level of excess) on a single directed closed chain consists of removing the contract with the lowest notional value in the chain.

Lemma 2. *Given a directed closed chain $K = (N, E)$, consider the set of compression operations C satisfying a conservative compression tolerance set Γ such that*

$$C = \min_{x'}(c(N, E) : (N, E) \to (N', E'))$$

then

$$e'_{ij} = e_{ij} - min_e\{E\}. \quad \forall e' \in E'$$

and

$$\Delta^{c}_{res}(K) = \Delta(K) - \Phi(E),$$

where $\Phi(E) = |E| min_{e \in E}\{E\}$.

Proof. See Appendix. ■

On a directed chain, withdrawing the smallest trade removes the maximum redundant excess without having to change the directionality of other trades. To keep balances equal, when the trade is removed, its notional value is subtracted from all other trades in the chain resulting in an excess reduction equal to the value of the removed trade times the initial number of trades in the closed chain of intermediation.

Given a market of several closed chains of intermediation, a conservative compression algorithm would thus aim at breaking chains by removing the contract with the smallest notional value. Breaking a closed chain of intermediation (i.e., the set of edges $e^{chain}_{ij} \in E^{chain}$ such that $\prod e^{chain}_{ij} > 0$) results in a reduction of excess by:

$$\Delta_{res}(G) = \Delta(G) - \Phi(E^{chain}).$$

At the end of the algorithm, the resulting compressed market does not contain directed closed chains anymore: it is a Directed Acyclic Graph (DAG).[11]

Definition (Directed Acyclic Graph)**.** *A Directed Acyclic Graph is a graph that does not contain any directed cycle (i.e. closed directed chains).*

[11] In the Graph Theory literature, closed chains of intermediation are also called *cycles*.

Corollary 5. *A market resulting from a conservative compression is a directed acyclic graph.*

The fact that conservative compression cannot take place if there is no closed chain of intermediation also yields a result on the efficiency limitation of such compression class of operators. We show that, in general, the residual excess of a conservative compression is positive (i.e., not all the excess can be removed). However, there is a specific configuration of closed chains that allows complete removal of excess. Consider the following type of chain.

Definition (Balanced chain)**.** *A balanced chain is a chain of intermediation $K = (N, E)$ which has the two following features:*

1. $|\{e|e = min_e\{E\}\}| \geq \frac{|E|}{2}$
2. *if* $\exists e_i \in E|e_i > \min_e\{E\}$ *then* $\{e_{i-1}, e_{i+1}\} = \{\min_e\{E\}, \min_e\{E\}\}$

The first property of such chain is that more than half of edges have the same value and this value is the minimum value of all the set of edges. The second property states that, for any edge that has a value higher the the minimum value, the edges preceding and succeeding it in the sequence of edges in the chain (i.e., e_{i-1} and e_{i+1}) have the minimum value. A chain in which all edges have the same value is thus a special case of a balanced chain.

We now show that conservative compression can remove all the excess only when all closed chains of intermediation in the market are balanced.

Proposition 7. *Given a market $G(N, E)$ and a compression operator $c()$ satisfying a conservative compression tolerance set Γ:*

$$\Delta^{c^c}_{res}(G) = 0$$

i.i.f. all chains in E are closed and balanced.

Proof. See Appendix. ■

From this result, we also obtain the following corollary:

Corollary 6. *If there is at least one closed chain of intermediation that is not balanced in $G = (N, E)$, then:*

$$\Delta^{c^c}_{res}(G) > 0.$$

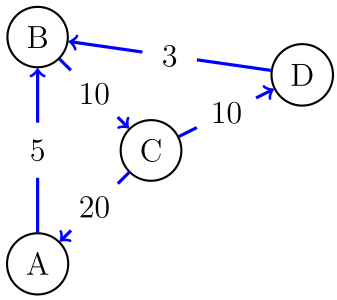


Figure 3: Example of market with entangled chains

The intuition behind this result is that, in order to completely remove the excess, there must be no more intermediation in the resulting market. But since a conservative compression cannot compress an intermediation chain that is not (i) closed nor (ii) not balanced, there will be a level of excess that cannot be removed from conservative compression, in the general case. Analyzing further the efficiency of such approach is less straightforward than the non-conservative case. In fact, the network structure of the market plays an important role that is not merely captures in the excess values in part because the number of closed chains of intermediation will affect the efficiency of a conservative compression. In contrast with the non-conservative case, it is not possible to establish general expressions for the expected residual and redundant excess under a conservative approach. Next we establish conditions under which such formulation is feasible and, then, we propose an algorithmic method to determine the conservative residual and redundant excess amounts for any given network structure.

Special case

In order to reach a directed acyclic graph any algorithm would need to identify and break all closed chains of intermediation. Nevertheless, the sequences of chains to be compressed can affect the results. In fact, if two chains share edges, compressing one chain modifies the value of the contracts also present in the other one. There can be different values of residual excess depending on which closed chain is compressed first.

Formally, we identify such case as a case of *entangled chains of intermediation.*

Definition (Entangled Chains)**.** *Two chains of intermediation, $K_1 = (N_1, E_1)$ and $K_2 = (N_2, E_2)$, are entangled if they share at least one edge:*

$$E_1 \cap E_2 \neq \emptyset$$

An illustration of entangled chains is provided in Figure 3 where the edge BC is share by two chains of intermediation (i.e., ABC and BCD).

As such, we formulate the following feature on a graph:

Definition. *(Chain Ordering Proof). A market is chain ordering proof w.r.t. to the conservative compression if the ordering of entangled chains by Φ does not affect the efficiency of compression.*

If the configuration of entangled chains is such that, according to the initial ordering of excess reduction resulting from a compression on each chain, the optimal sequence is not affected by the effects of compression on other entangled chains, the market is said to be chain ordering proof. Under the above Definition, the optimal conservative compression yields a Directed Acyclic Graph (DAG) where the excess is given by the following expression:

Proposition 8. *Given a market $G = (N, E)$. If there are no entangled chains, we have:*

$$\Delta_{res}(G) = \Delta(G) - \sum_{K_i \in \Pi} \Phi(E_{K_i})$$

In the presence of entangled chains, if $G = (N, E)$ is chain-ordering proof, we have

$$\Delta_{res}(G) < \Delta(G) - \sum_{K_i \in \Pi} \Phi(E_{K_i})$$

Where Π is the set of all chains of intermediation in G.

Proof. See Appendix ■

For illustrative purpose, we present an algorithm that always reaches a global solution under the chain ordering proof assumption in the Appendix.

Generalization

In practice, many markets can exhibit entangled chains with an ordering effect. When the chain ordering proof assumption does not hold, the sequence of chains upon which conservative compression is applied will affect the efficiency of the compression. In order to guarantee a global solution, we characterize conservative compression as a linear programming problem and apply the network simplex algorithm to determine the most efficient compression procedure. Details regarding the program characterization and the network simplex algorithm are provided in the Appendix.[12]

[12]For further information on algorithmic solutions for linear programming problems and the network simplex, see (Ahuja et al., 1993)

4.7 Hybrid compression

In more realistic settings, compression tolerances can be subject to the strategical role of specific trading relationships. In the following, we consider a hybrid model that results from 2 main assumptions of market participants' preferences:

Assumption 1. *Dealers prefer to keep their intermediation role with customers*

Assumption 2. *Intra-dealer trades can be switched at negligible cost.*

The first assumption states that dealers value their interaction with customers and will reject compression exercises that remove such contracts. In the case of a balanced intermediation chain (i.e., where the intermediary has 0 net position), the intermediary(ies) can be removed from the solution and a sole contract would be created between the two end-customers. The assumption here is that dealers prefer to stick with the original situation and will set low compression tolerances on their customer contracts.

The second assumption posits that the intra-dealer networks forms a well-connected club where the interactions are so frequent overall that the instance of a specific trade does not signal a strong preference towards a specific dealer counterparty. As a result, switching counterparties in the intra-dealer network as a result of compression has negligible costs in comparison with the overall benefits of compression. The assumption thus results in high compression tolerances on the contracts between dealers.

In our framework, these two assumptions lead to a segmentation of the market into the two subsets defined by Definition 3.3: the intra-dealer market, i.e., E^D, and the customer market, i.e., E^C. For each we have a different set of compression tolerances. We have the following formal definition:

Definition (Hybrid compression)**.** *$c(N, E)$ is a hybrid compression operator i.f.f. $c()$ is a compression operator that satisfies the compression tolerances set Γ:*

$$a_{ij} = 0 \quad and \quad b_{ij} = e_{ij}, \quad \forall (a_{ij}, b_{ij}) \in \Gamma, e_{ij} \in E^C$$

$$a_{ij} = 0 \quad and \quad b_{ij} = +\infty, \quad \forall (a_{ij}, b_{ij}) \in \Gamma, e_{ij} \in E^D$$

Where E^C and E^D are the customer market and the intra-dealer market, respectively, with $E^C + E^D = E$.

The hybrid compression approach sets high compression tolerance in the intra-dealer sub-network and low compression tolerance for contracts involving customers. Hence, it is a combination of a non-conservative approach in the intra-dealer network and a conservative approach in the customer network.

Corollary 7. *The feasibility conditions of the hybrid model are*

- *non-conservative condition for* E^D
- *conservative condition for* E^C

In a market following the definitions of dealers and customers provided in Section 3, we thus see that compression will only take place in the intra-dealer network because no closed chains of intermediation will be present in the customer network. This situation is similar to the conservative case. Nevertheless, the compression on the intra-dealer network is now non-conservative. As a result, the intra-dealer network will form a bi-partite graph with 0 residual intra-dealer excess.

Proposition 9. *Given a market* $G = (N, E)$*, if*

$$\Delta(N, E) = \Delta(N, E^D) + \Delta(N, E^C)$$

then, a compression operator $c^h()$ *satisfying a hybrid compression tolerance set* Γ *leads to*

$$\Delta_{res}^{c^h}(N, E) = \Delta(N, E^C)$$

As a result, we see that, in case the excess is additive, it is straightforward to obtain the efficiency of the hybrid compression. When it is not, a specific algorithm must be implemented to obtain the exact level (see Appendix).

4.8 Bilateral compression

Finally, we look at a last benchmark: bilateral compression. In this case, market participants do not share information about their portfolio, that is, there is no centralization mechanism to identify compression opportunities beyond pairs of counterparties. Hence, compression is limited to the bilateral sets of contracts that exists between each pair of counterparty. Formalising this compression approach allows us to assess the added-value of a third party agent (i.e., a compression service provider follow either of the multilateral approaches presented above) in comparison with the efficiency that market participants can reach when they can compress their portfolio only given the information they hold (i.e., what they sell and what they buy). In our framework, bilateral compression is defined as follows:

Definition (Bilateral compression)**.** $c(N, E)$ *is a bilateral compression operator i.f.f.* $c()$ *is a compression operator that satisfies the compression tolerances set* Γ*:*

$$a_{ij} = b_{ij} = \max\{e_{ij} - e_{ji}, 0\}, \quad \forall (a_{ij}, b_{ij}) \in \Gamma, e_{ij} \in E.$$

For each pair of market participants i and j, we collect their bilateral exposures, e_{ij} and e_{ji} and keep the largest exposure reduced by the lowest exposure, thereby eliminating a cycle of length two. Hence, if we assume $e_{ij} > e_{ji}$, we have: $e'_{ij} = e_{ij} - e_{ji}$ and $e'_{ji} = 0$ after bilateral compression.

In terms of feasibility, the mere existence of excess is not enough for bilateral compression to be applicable. In particular, we need at least two exposures for the same pair of counterparties going with opposite signs. Formally, we have the following results:

Proposition 10. *Given a market $G(N, E)$ and a compression operator c^b satisfying a bilateral compression tolerance set Γ:*

$$\Delta^{c^b}_{red}(G) > 0 \quad \Leftrightarrow \quad \exists i, j \in N \quad s.t. \quad e_{ij}.e_{ji} > 0 \quad where \quad e_{ij}, e_{ji} \in E$$

Proof. See Appendix ■

The efficiency of bilateral compression is straightforward. It corresponds to the effect of netting out each pair of bilateral exposures. We thus obtain the following efficiency results:

Proposition 11. *Given a market $G = (N, E)$ and a compression operator $c^b()$ satisfying a bilateral compression tolerance set Γ leads to*

$$\Delta^{c^b}_{res}(G) = \Delta(G) - \sum_{i,j \in N} \min\{e_{ij}, e_{ji}\} \quad where \quad e_{ij}, e_{ji} \in E.$$

Proof. See Appendix. ■

Technically, bilateral compression results in the removal of all chains of intermediation of length 1. Hence, a bilaterally compressed market exhibit a maximum of one and only directed exposure between each pair of market participants.

4.9 Efficiency dominance among compression benchmarks

We close the theoretical discussion on compression by investigating the differences in efficiency between the four benchmark approaches we have introduced, namely, conservative, non-conservative, hybrid and bilateral. We are thus interested in ranking the capacity of each approach to remove excess from a given generic market. For each approach, we consider the maximum amount of excess that can be removed, that is, we consider globally optimal solutions given compression tolerances and the net-equivalent condition.

Proposition 12. *Given a market $G = (N, E)$ and the set of compression operators $\{c^c(), c^n(), c^h(), c^b()\}$ such that:*

- $c^c()$ *satisfies a conservative compression tolerance set* Γ^c *such that* $\Delta^{c^c}_{red}(G)$ *is maximized,*
- $c^n()$ *satisfies a non-conservative compression tolerance set* Γ^n *such that* $\Delta^{c^n}_{red}(G)$ *is maximized,*
- $c^h()$ *satisfies a hybrid compression tolerance set* Γ^h *such that* $\Delta^{c^h}_{red}(G)$ *is maximized,*
- $c^b()$ *satisfies a bilateral compression tolerance set* Γ^b *such that* $\Delta^{c^b}_{red}(G)$ *is maximized,*

the following weak dominance holds:

$$\Delta^{c^b}_{red}(G) \leq \Delta^{c^c}_{red}(G) \leq \Delta^{c^h}_{red}(G) \leq \Delta^{c^n}_{red}(G) = \Delta(G).$$

Proof. See Appendix. ■

This result shows a clear dominance sequence from the least to the most efficient compression operator in reducing market excess. First, we see that the non-conservative compression is the most efficient. This stems from the fact that a global non-conservative solution always removes all the excess from a market (see Proposition 4). The second most efficient compression operator is represented by hybrid compression, followed by the conservative. The least efficient approach is the bilateral compression. This lack of efficiency is due to the fact that bilateral compression cannot remove excess resulting from chains of length higher than two. The proof of this proposition derives from an analysis of the in compression tolerance sets of each approach. In fact, it can be shown that the bilateral compression tolerance set is a subset of the conservative set which in turn is a subset of the hybrid set which is also a subset of the non-conservative set. This nested structure for the compression tolerances sets ensures that any globally optimal solution of a superset is at least as efficient as the globally optimal solution of any subset. Additional analysis on the relative efficiencies of each approach (e.g., strong dominance, quantities, etc.) needs to include further information about the underlying set of edges E (i.e., network characteristics). However, having established the dominance sequence, we proceed next with a quantitative analysis based on real market data.

5 Empirical application

5.1 Analysis strategy

In this Section, we apply the theoretical framework developed in this paper to a unique, transaction-level dataset for Credit-Default-Swaps (CDS) derivatives[13]. The dataset covers all CDS transactions in which at least one counterparty is legally based in the European Union from October 2014 to April 2016.[14]

Our analysis will be split in two main steps. First, in the next subsection, we analyze the original markets as obtained from the procedure. For each market, we compute the (i) dealer-customer network characteristics, (ii) excess statistics and (iii) efficiency of three compression approaches: bilateral, conservative and hybrid compression. We do not report results from non-conservative compression as an optimal solution always leads to zero residual excess (see Proposition 4). Bilateral compression is the result of a bilateral netting between all pairs of counterparties in the market. In the case of the conservative and hybrid compressions, the results are not trivial and require more sophisticated algorithmic approaches that ensure globally optimal solutions. We use a linear programming framework to design such solutions.[15] For each market G, we implement each compression algorithm and compute its efficiency (i.e, redundant excess) as a fraction of the total level of excess:

- Bilateral: $\rho_b = \frac{\Delta^b_{red}(G)}{\Delta(G)}$;
- Conservative : $\rho_c = \frac{\Delta^c_{red}(G)}{\Delta(G)}$;
- Hybrid : $\rho_h = \frac{\Delta^h_{red}(G)}{\Delta(G)}$.

In the previous Section, we have established theoretically the following efficiency dominance ordering: bilateral, conservative and hybrid. Comparing the efficiency, quantitative results from the three approaches allows us to assess 1) the effect of synchronized multilateral compression (i.e., conservative and hybrid cases) versus asynchronised bilateral compression (i.e., bilateral case)[16] and 2) the

[13]CDS contracts are the most used types of credit derivatives. A CDS offers protection to the buyer of the contract against the default of an underlying reference. The seller thus assumes a transfer of credit risk from the buyer. CDS contracts played an important role during the 2007-2009 financial crisis. For more information, see (Stulz, 2010).

[14]For more details on the dataset, the general cleaning procedure and other statistics, see (Abad et al., 2016)

[15]All algorithms used to solve these problems are described in the Appendix.

[16]The synchronization aspect stems from the fact that both the conservative and hybrid approaches assume coordination among market participants (i.e., they all agree to compress the submitted observed trades at the same time). This condition is not necessary in the bilateral compression.

effect of relaxing compression tolerances from bilateral to conservative to hybrid approaches.

Second, we analyze markets after bilateral compression. In derivatives market, participants, specially dealers, offset some positions by simply writing a symmetric contract in the opposite direction with the same counterparty. Analyzing the bilaterally compressed market thus allows us to quantify excess and compression efficiency beyond the redundancy incurred by this specific behavior. As such, we similarly analyze (i) dealer-customer network characteristics, (ii) excess statistics and (iii) efficiency of two compression approaches: conservative and hybrid compression. Doing so, we are thus able to quantify the marginal gains from syncronized multilateral compression once all market participants have maximally compressed their positions using only their local information. More broadly, comparing results from applying multilateral compression on the original market and on the bilaterally compressed market, we are able to quantify the potential losses in excess reduction due to the sequence of bilateral-then-multilateral compression.

5.2 Dataset description

There are multiple selection options to aggregate the data and build the network of obligations. In the following, we expose our strategy which focuses on the most *conservative* approach (i.e., we gather trades with minimal assumptions on their fungibility).

Each bilateral transaction reports the identity of the two counterparties, the underlying reference entity, the maturity of the contract, the currency and its notional amount. For a given reference entity there can be several identifiers (e.g., government bonds with different maturities). At each point in time, we select the most traded reference identifier (i.e., ISIN) associated to the reference with the most traded maturity (by year). In practice, participants to a compression process may combine a larger variety of contracts. For sake of simplicity and consistency, we do not consider such cases in the following. At the participant level, we select participants using their Legal Entity Identifier (i.e., LEI), that is, the entity reporting the transaction. In practice, financial groups may decide to submit trades coming from different legal entities of the same group. We do not consider such case in the following.

We consider 19 mid-month snapshots from October 2014 to April 2016. Overall, our sample comprises 7300 reference entities. The vast majority of the notional, however, is concentrated in a much lower number of entities. This allows us to focus on a restricted sample of entities to illustrate our framework. We opt to retain 100 reference entities which we find to be a good compromise between the amount of notional traded and clarity of analysis. Our restricted sample comprises 43 sovereign entities (including the largest EU and G20 sovereign entities), 27 financials (including the largest banking groups) and 30 non-financials entities (including large industrial and manufacturing groups).

5.3 General statistics

An overview of the main statistics of the original markets is reported in Table 1. In particular, the total notional of the selected 100 entities varies between 380Bn Euros and 480Bn Euros retaining roughly $30 - 34\%$ of the original total gross notional. The average number of counterparties across the 100 entities is stable and varies between 45 and 58 individual counterparties.

Time	Gross notional of 100 top ref. (E+11 euros)	Share of gross notional among all ref.	Avg num. of counterparties
Oct-14	3.88	0.358	54
Nov-14	4.16	0.349	55
Dec-14	4.4	0.357	58
Jan-15	4.73	0.361	57
Feb-15	4.67	0.355	57
Mar-15	4.35	0.351	51
Apr-15	3.87	0.338	46
May-15	3.91	0.337	45
Jun-15	3.86	0.343	47
Jul-15	3.9	0.347	50
Aug-15	3.9	0.344	52
Sep-15	3.94	0.350	53
Oct-15	4.08	0.349	55
Nov-15	4.18	0.351	55
Dec-15	4.24	0.348	55
Jan-16	4.39	0.351	55
Feb-16	4.33	0.348	56
Mar-16	3.94	0.350	49
Apr-16	4.37	0.352	49

Table 1: General statistics of the dataset.

Table 2 provides further statistics on the market segments (i.e, intra-dealer and customer markets). Note that we identify dealers as intermediaries beyond bilateral interactions. Indeed, from the formal definition of dealer in Definition 3, two market participants buying and selling from each other would be identified as dealers which does not properly reflect the role of dealers in derivatives markets.

As such, by convention, we set market participants as dealer if they appear as intermediary in the bilaterally compressed market. Similarly, buying customers and selling customers are determined using the bilaterally compressed market. This convention does not affect the theoretical results and provides a more grounded interpretation of the empirical results, in particular for the hybrid compression.

We compute the average number of dealers, customers on the buy side and customers on the sell side across all entities in the different snapshots. We observe stability of these numbers across time: per reference entity, there are on average 18 to 19 dealers, 12 to 17 customers buying a CDS contract, 14 to 21 customers selling a CDS contract. The average number of contracts per reference entity varies more through time but remains between 140 and 170 contracts. Taken as a whole, markets are quite sparse with an average density of contracts around 0.10. This means that, on average, only 10% of all possible bilateral contracts between all market participants are actually present. Interestingly, this measure is almost three times bigger when we only consider the intra-dealer market. We thus see that the bulk of the activity in those market revolves around intra-dealer trades. The amount of intra-dealer notional also highlights the level of activity concentration around dealers: it averages around 80% of the total notional. Finally, the last column of Table 2 confirms the very low frequency of customer-customer trades: on average, less then 0.2% of all contracts are written without a dealer on either side of the trade.

5.4 Quantifying excess and the efficiency of compression in original markets

After this general analysis, we focus on quantitative assessments of the measures introduced in this paper. We start by measuring the level of excess present in the original markets as a function of the total gross notional (i.e., $\epsilon(G) = \frac{\Delta(G)}{x}$). Table 3 reports the statistics of excess levels of six snapshots equally spread between October 2014 and April 16 including minimum, maximum, mean, standard deviation and quartiles, computed across all reference entities in our sample. Results on the means and medians are stable over time and often higher than 0.75. We thus see that, in general, around three quarters of the gross notional in the most traded CDS markets (at least by EU institutions) is in excess vis-a-vis market participants' net position. For example, in October 2014, this means that around 298Bn Euros correspond to notional excess. At the extremes, we note a high degree of variability: the minimum levels of excess oscillate around 0.45 while the maximum was 0.90. Nevertheless, this result shows that all markets exhibits at least 50% of excess in notional.

We now move to the efficiency of different compression approaches reported in Table 4. After having implemented the compression algorithms on each market

Time	Avg num. dealers	Avg num. customers buying	Avg num. customers selling	Avg num. contracts	Avg. share intra-dealer notional	Avg. density	Avg. intra-dealer density	Avg. intra-customer density
Oct-14	18.10	15.57	19.88	153.72	0.812	0.105	0.332	0.0010
Nov-14	18.51	16.10	20.70	162.26	0.831	0.109	0.345	0.0006
Dec-14	19.22	17.12	21.44	171.13	0.829	0.109	0.339	0.0005
Jan-15	19.34	17.00	21.10	171.25	0.827	0.106	0.334	0.0006
Feb-15	19.16	16.80	21.06	168.45	0.826	0.106	0.335	0.0004
Mar-15	18.46	15.16	17.44	154.42	0.832	0.110	0.339	0.0007
Apr-15	18.03	12.68	14.87	143.71	0.829	0.110	0.344	0.0005
May-15	18.16	12.49	14.56	143.66	0.827	0.108	0.336	0.0008
Jun-15	18.36	13.20	14.12	142.01	0.828	0.106	0.323	0.001
Jul-15	18.51	14.32	14.36	143.68	0.813	0.101	0.314	0.0009
Aug-15	18.78	14.59	16.79	149.12	0.821	0.101	0.308	0.0011
Sep-15	18.95	15.58	17.29	151.36	0.804	0.098	0.302	0.0018
Oct-15	19.28	16.06	17.62	155.52	0.815	0.099	0.297	0.0013
Nov-15	19.31	16.94	18.76	158.10	0.810	0.099	0.293	0.0017
Dec-15	19.50	16.71	18.57	158.78	0.821	0.098	0.292	0.0012
Jan-16	19.52	16.88	18.13	159.58	0.822	0.098	0.291	0.0013
Feb-16	19.42	17.02	18.08	159.80	0.813	0.098	0.291	0.0012
Mar-16	18.24	14.17	16.64	144.11	0.790	0.096	0.301	0.0018
Apr-16	18.62	13.71	16.68	146.42	0.811	0.098	0.301	0.0019

Table 2: Dealers/customers statistics of original markets.

Total Excess	Oct-14	Jan-15	Apr-15	Jul-15	Oct-15	Jan-16	Apr-16
min	0.529	0.513	0.475	0.420	0.533	0.403	0.532
max	0.904	0.914	0.895	0.901	0.903	0.890	0.869
mean	0.769	0.777	0.766	0.757	0.751	0.728	0.734
stdev	0.077	0.082	0.085	0.090	0.082	0.096	0.080
first quart.	0.719	0.733	0.712	0.703	0.693	0.660	0.678
median	0.781	0.791	0.783	0.769	0.758	0.741	0.749
third quart.	0.826	0.847	0.832	0.822	0.808	0.802	0.796

Table 3: Excess statistics of original markets

(i.e., for all time snapshots, we run the algorithms on each of the 100 different markets considered separately), we compute the efficiency of the compression as defined in Section 5.1. As expected (and not reported in the table), for the non-conservative compression, the amount of excess removed is equal in every part to the results of Table 3.[17]

Analysing the means and medians, we observe that the bilateral compression already manages to remove 50% of excess on average. Nevertheless both multilateral compression approaches (i.e., conservative and hybrid) perform much better by removing around 85% and 90% of the excess for the conservative and hybrid approach respectively. Those levels are even larger than the maximum efficiency achievable by bilateral compression which oscillates around 75%. In comparison with the bilateral efficiency, the conservative and hybrid approaches perform similarly on the extremes: with minima around 55% and 62% and maxima around 98% and 99%, respectively. Multilateral compression approaches appear almost fully efficient. In particular, results from the conservative compression show that, even under the severe constrain that contracts can only be removed and no new relationship can be introduced, the vast majority of market's excess can be reduced, at least by half, at most leaving 1.2% of excess.

5.5 Quantifying excess and the efficiency of compression in bilaterally compressed markets

As we have seen, bilateral excess, on average, accounts for half the excess of the original markets. In order to understand excess and compression beyond bilateral offsetting, we follow up by analyzing deeper the bilaterally compressed markets.

[17] Note that Table 3 also provides us with the upper efficiency limit of any compression approach and that the current compression exercise does not represent the amount of compression achieved in the market, rather, it is the amount of compression that is still achievable given the current state of outstanding trades

Bilateral (ρ_b)	Oct-14	Jan-15	Apr-15	Jul-15	Oct-15	Jan-16	Apr-16
min	0.278	0.281	0.286	0.277	0.276	0.276	0.260
max	0.779	0.791	0.759	0.777	0.717	0.711	0.746
mean	0.528	0.536	0.524	0.522	0.513	0.512	0.543
stdev	0.101	0.106	0.103	0.105	0.107	0.109	0.108
first quart.	0.464	0.460	0.469	0.452	0.448	0.444	0.448
median	0.526	0.542	0.535	0.530	0.517	0.528	0.555
third quart.	0.583	0.597	0.590	0.600	0.596	0.597	0.623

Conservative (ρ_c)	Oct-14	Jan-15	Apr-15	Jul-15	Oct-15	Jan-16	Apr-16
min	0.558	0.547	0.545	0.507	0.491	0.528	0.574
max	0.985	0.982	0.973	0.967	0.968	0.979	0.969
mean	0.836	0.857	0.848	0.843	0.828	0.827	0.834
stdev	0.091	0.087	0.090	0.091	0.104	0.106	0.090
first quart.	0.781	0.816	0.810	0.800	0.777	0.773	0.788
median	0.852	0.880	0.868	0.858	0.849	0.847	0.860
third quart.	0.906	0.925	0.913	0.915	0.902	0.907	0.904

Hybrid (ρ_h)	Oct-14	Jan-15	Apr-15	Jul-15	Oct-15	Jan-16	Apr-16
min	0.589	0.626	0.636	0.653	0.574	0.619	0.676
max	0.990	0.994	0.988	0.990	0.994	0.989	0.990
mean	0.878	0.898	0.894	0.893	0.881	0.882	0.898
stdev	0.079	0.072	0.074	0.073	0.085	0.080	0.069
first quart.	0.821	0.859	0.862	0.865	0.831	0.836	0.863
median	0.894	0.916	0.918	0.912	0.901	0.908	0.911
third quart.	0.935	0.952	0.947	0.951	0.948	0.945	0.947

Table 4: Statistics of compression efficiency of original markets

We obtain dealer-customer network characteristics reported in Table 5. While the participant-based statistics are equal to Table 2, there is a reduction in all contract-related statistics except the intra-customer density which remains the same: the average number of contracts is reduced by 25 percentage points. while the intra-dealer share of notional is only affected by 5 percentage points. Hence, we see that, despite the density reduction, concentration remains onto the intra-dealer activity after bilateral compression.[18]

In terms of excess, Table 6 complements the results from the bilateral compression efficiency and reports statistics similar to Table 3.[19] At the extremes, we note a high degree of variability: for example, in mid-January 2016, the minimum level of excess was 0.261 while the maximum was 0.809. Nevertheless, results on the means and medians are stable over time and alway higher than 0.5. We thus see that, in general, around half of the gross notional of bilaterally compressed market remains in excess vis-a-vis market participants' net position. Note that the gross notional used here is the total notional left after bilateral compression on the original market.

Table 7 reports the results related to the efficiency of conservative and hybrid compression applied to the already bilaterally compressed market. On the extremes, both the conservative and the hybrid compression perform with various degrees of efficiency: the minimum amount of excess reduction via conservative compression (resp. hybrid compression) oscillates around 15% (resp. 35%) while the maximum amount of excess oscillates around 90% (resp. 97%). This shows that compression can perform very efficiently and very poorly with both approaches. However, the fact that conservative compression reaches 90% of excess removal shows the possibility of having very efficient compression despite restrictive compression tolerances. The mean and the median of both approaches are stable over time: both around 60% for the conservative compression and 75% for the hybrid compression. Overall, we find that on average each compression algorithm is able to remove more than half of the excess from the market, the hybrid compression allowing for greater performances as a result of constraints alleviation (i.e., relaxation of intra-dealer compression tolerances).

To summarize the difference of results in efficiency when applying conservative and hybrid compression on either the original market or the the bilaterally compressed market, we show in Figure 4, box plots reporting the efficiency ratio with

[18]Note that the average intra-customer density is equal to Table 2. In theory, we should have doubled the value as the density of the bilaterally netted intra-customer networks should be seen as the density of a undirected graph. We kept the previous definition to highlight the fact that the intra-customers contracts were not affected by the bilateral compression and avoid a misinterpretation of density increase.

[19]The relationship between the bilateral compression efficiency, ρ_b and the relative excesses in the original market, ϵ^o, and the bilaterally compressed market, ϵ^b, is given by $\rho_b = (1 - \frac{1-\epsilon^o}{1-\epsilon^b})\frac{1}{\epsilon^o}$. This expression directly follows from the definition of each parameter.

Time	Avg num. dealers	Avg num. customers buying	Avg num. customers selling	Avg num. contracts	Avg. share intra-dealer notional	Avg. density	Avg. intra-dealer density	Avg. intra-customer density
Oct-14	18.10	15.57	19.88	115.87	0.767	0.075	0.221	0.0010
Nov-14	18.51	16.10	20.70	121.61	0.779	0.077	0.227	0.0006
Dec-14	19.22	17.12	21.44	128.24	0.777	0.076	0.221	0.0005
Jan-15	19.34	17.00	21.10	127.94	0.778	0.075	0.219	0.0006
Feb-15	19.16	16.80	21.06	126.31	0.782	0.075	0.221	0.0004
Mar-15	18.46	15.16	17.44	114.74	0.786	0.078	0.225	0.0007
Apr-15	18.03	12.68	14.87	106.28	0.786	0.079	0.229	0.0005
May-15	18.16	12.49	14.56	106.18	0.782	0.078	0.224	0.0008
Jun-15	18.36	13.20	14.12	105.20	0.783	0.076	0.216	0.0010
Jul-15	18.51	14.32	14.36	107.05	0.766	0.072	0.211	0.0009
Aug-15	18.78	14.59	16.79	111.49	0.776	0.073	0.208	0.0011
Sep-15	18.95	15.58	17.29	114.12	0.755	0.071	0.204	0.0018
Oct-15	19.28	16.06	17.62	117.22	0.766	0.072	0.200	0.0013
Nov-15	19.31	16.94	18.76	120.52	0.762	0.072	0.198	0.0017
Dec-15	19.50	16.71	18.57	120.76	0.772	0.072	0.198	0.0012
Jan-16	19.52	16.88	18.13	121.06	0.774	0.072	0.198	0.0013
Feb-16	19.42	17.02	18.08	121.05	0.763	0.071	0.197	0.0012
Mar-16	18.24	14.17	16.64	108.03	0.739	0.070	0.205	0.0018
Apr-16	18.62	13.71	16.68	109.29	0.759	0.071	0.204	0.0019

Table 5: Dealers/customers statistics after bilateral compression.

Total Excess	Oct-14	Jan-15	Apr-15	Jul-15	Oct-15	Jan-16	Apr-16
min	0.422	0.423	0.290	0.257	0.366	0.261	0.293
max	0.811	0.811	0.798	0.809	0.820	0.809	0.781
mean	0.614	0.621	0.614	0.602	0.597	0.570	0.558
stdev	0.087	0.087	0.091	0.095	0.097	0.112	0.098
first quart.	0.562	0.558	0.562	0.544	0.531	0.489	0.503
median	0.617	0.618	0.614	0.613	0.594	0.569	0.566
third quart.	0.670	0.684	0.674	0.663	0.654	0.653	0.635

Table 6: Excess statistics after bilateral compression

Conservative (ρ_c)	Oct-14	Jan-15	Apr-15	Jul-15	Oct-15	Jan-16	Apr-16
min	0.160	0.203	0.140	0.163	0.165	0.119	0.098
max	0.894	0.927	0.923	0.878	0.912	0.911	0.878
mean	0.568	0.622	0.599	0.592	0.555	0.552	0.525
stdev	0.166	0.160	0.164	0.158	0.175	0.183	0.172
first quart.	0.456	0.505	0.512	0.489	0.435	0.437	0.409
median	0.562	0.636	0.594	0.591	0.537	0.550	0.546
third quart.	0.685	0.729	0.728	0.705	0.680	0.687	0.643

Hybrid (ρ_h)	Oct-14	Jan-15	Apr-15	Jul-15	Oct-15	Jan-16	Apr-16
min	0.370	0.460	0.377	0.281	0.259	0.281	0.135
max	0.971	0.973	0.968	0.963	0.977	0.974	0.981
mean	0.724	0.763	0.760	0.755	0.738	0.735	0.752
stdev	0.149	0.130	0.130	0.130	0.146	0.140	0.148
first quart.	0.623	0.691	0.678	0.674	0.626	0.642	0.679
median	0.735	0.785	0.781	0.778	0.775	0.756	0.784
third quart.	0.846	0.866	0.859	0.866	0.849	0.851	0.845

Table 7: Statistics of compression efficiency after bilateral compression

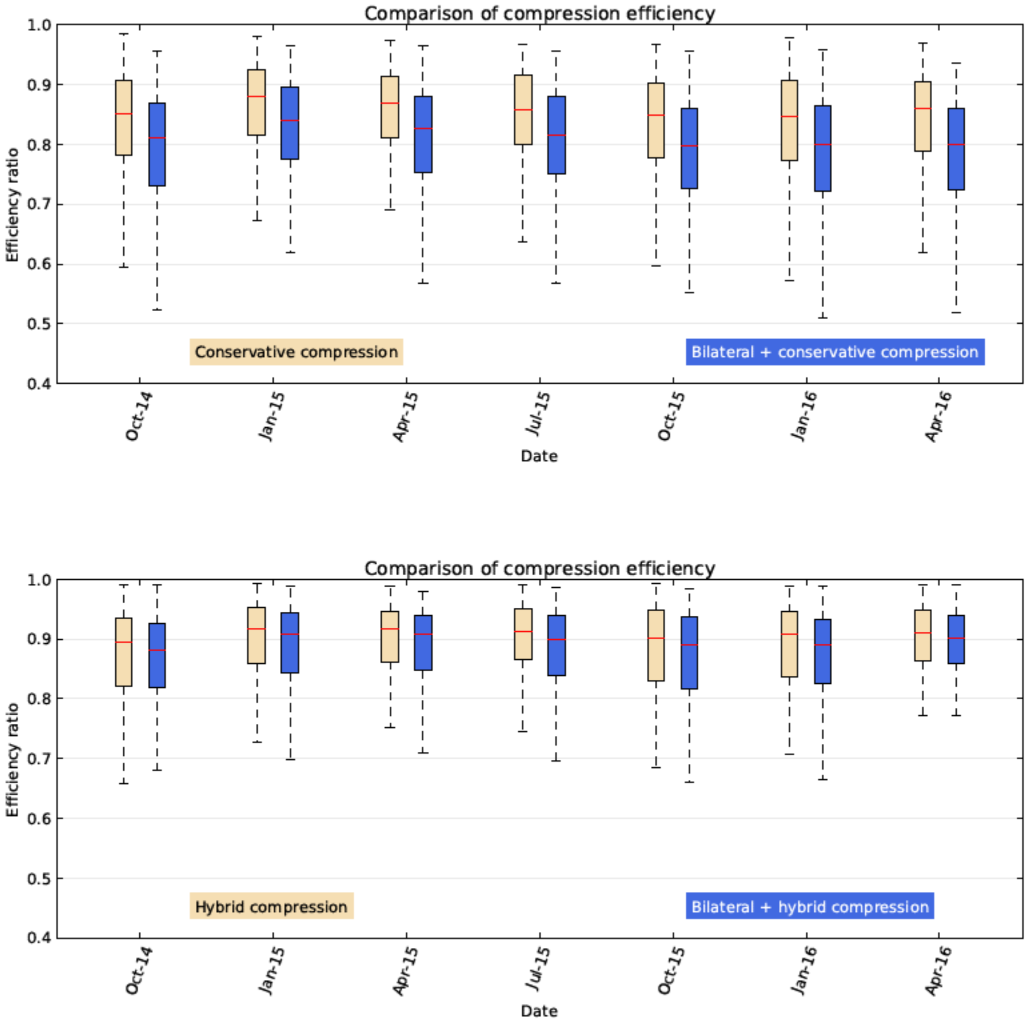


Figure 4: Time comparison of the efficiency between the application of conservative and hybrid compression on the original markets (upper panel) and application of bilateral compression on the original markets followed by conservative and hybrid compression on the bilaterally compressed markets (lower panel).

respect to the original level of notional: the beige-color related results are similar to the ones characterized in Table 4 for the conservative and hybrid approach while the blue-related results are obtained from adding the absolute bilateral results characterized in Table 4 to the absolute excess reduction for the conservative and hybrid approach as in Table 7 then dividing by the notional of the original markets.

From Figure 4, we observe that applying a multilateral compression algorithm on the original market is always more efficient than through the sequence of bilateral plus multilateral compression. Nevertheless, the difference for hybrid compression is very small (i.e., about a 1 percentage point improvement in the median) while it is more apparent in the conservative compression (i.e., up to 7 percentage points) showing a marked sensitivity of the conservative efficiency vis-a-vis prior bilateral compression step. Overall, the Figure also highlights, again, the generally high efficiency of the two multilateral approaches.

Finally, in order to better appreciate the levels of compression achievable in the CDS market, we "zoom-in" into the top 5 reference entities by notional across all time snapshots and investigate how much notional value can be eliminated via compression. The top five reference entities are all large sovereigns. For each market, let e_{ijk} be the notional contract between i and j on the k-th reference entity and $x_k = \sum_k e_{ijk}$ be the total gross notional outstanding on reference entity k. Let $w_k = \frac{x_k}{\sum_k x_k}$ be the relative gross notional for entity k vis-a-vis the total notional of the 5 markets aggregated. Consider the relative excess ratio $\epsilon(G) = \frac{\Delta(G)}{x}$, we compute the following ratio for each compression approach:

$$\begin{aligned} \text{Non-conservative:} \quad & \epsilon_n^k(G) = \epsilon^k(G), \\ \text{Hybrid:} \quad & \epsilon_h^k(G) = \rho_h^k \times \epsilon(G), \\ \text{Conservative:} \quad & \epsilon_c^k(G) = \rho_c^k \times \epsilon(G). \end{aligned}$$

Finally, we compute the weighted average for each of these ratios as follows:

$$\begin{aligned} \text{Non-conservative:} \quad & \epsilon_n = \sum_{k=1}^{5} \left(w_k \, \epsilon^k(G) \right), \\ \text{Hybrid:} \quad & \epsilon_h = \sum_{k=1}^{5} \left(w_k \, \rho_h^k \times \epsilon(G) \right), \\ \text{Conservative:} \quad & \epsilon_c = \sum_{k=1}^{5} \left(w_k \, \rho_c^k \times \epsilon(G) \right). \end{aligned}$$

These ratios can be easily interpreted as the fraction of notional that can be eliminated over all the top five entities, when taken individually. Results for these weighted averages are reported in Figure 5. The circled series highlighted by the light blue shade represent the weighted non-conservative compression ratio ϵ_n (which coincides with the weighted level of excess); the triangle markers represent the weighted hybrid compression ratio ϵ_h; the squared markers represent the conservative compression ratio ϵ_c. From the figure, we observe again large levels of excess across time, i.e. between 60% and 70%. In addition, the conservative

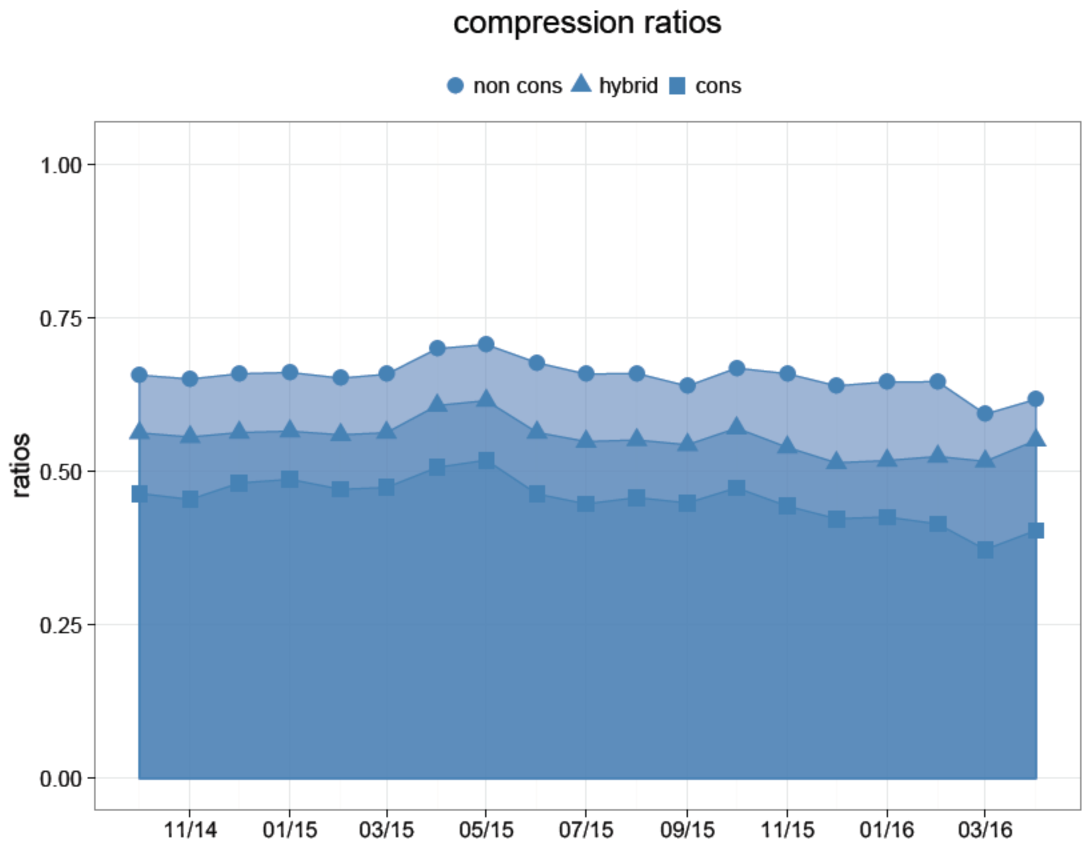


Figure 5: Weighted averages of the non-conservative, hybrid and conservative compression ratios for the top-5 reference entities in our sample.

compression ratio ranges between 40% and 50% of total notional across time. This implies that, in the case of the top 5 largest single-name CDS markets and even under very conservative assumptions, almost half of the notional can be eliminated while keeping both individual net positions and counterparty relationships.

6 Closing remarks

In this work, we show that Over-The-Counter (OTC) markets with contingent and fungible trades generate gross volumes that can far exceed the amounts satisfying every market participants' net position. We call this difference *excess*. In particular, we show that the activity of dealers acting as intermediaries between customers but also between other dealers is the factor determining the level of excess in a market. In turn, we show ways to removed this excess while keeping net positions satisfied through a netting operation called *compression*. To the best of our knowledge, this work is the first to propose a comprehensive framework to analyze the *mechanics* of compression (i.e., feasibility and efficiency conditions). Using this framework, we find the existence of a trade-off between the fraction of excess that can be eliminated from the market via compression and the degree to which the operation is limited by specific constraints, such as pre-existing trading exposures. Furthermore, using a unique granular dataset on bilateral exposures from Credit Default Swaps (CDS) contracts, we quantify the levels of excess and the efficiency of different compression procedures in real OTC markets. On average we find that around 75% of market sizes relates to excess of notional. While around 50% can in general be removed via bilateral compression, multilateral compression approaches can remove almost all the excess. In particular, we find that even the most conservative approach which satisfies pre-existing relationship constrains can eliminate up to 98% of excess in the markets.

The framework introduced in this paper accommodates several novelties. A natural extension is to allow compression over pools of multiple asset classes. Indeed, OTC market participants are often involved in several markets (D'Errico et al., 2016) and the intricacies of these layered positions can generate efficiency trade-offs (Duffie and Zhu, 2011; Cont and Kokholm, 2014). Such extension would require the introduction of explicit fungibility relationships between the different classes of obligations and means to identifying the cost and benefits of compression.

Along these lines, as OTC trades are increasingly called for standardization and mandatory central clearing, the role of Central Clearing Counterparties (CCPs) in determining excess and efficiency of compression is hence becoming a paramount matter. While we do not explicitly tackle this aspect, a simple intuition can be drawn from the results. In fact, CCPs in our framework would be peculiar nodes: they would have a strictly balanced position. In case the market contains only one CCP and all trades are cleared, the excess in the market would simply be equal to half of the aggregate gross notional. In fact, each bilateral trade would be novated to the CCP resulting in a double counting of the related notional. Nevertheless, in the presence of several CCPs and the co-existence of cleared and non-cleared trades, the situation becomes non-trivial as the determination of excess and compression efficiency will depend on the level of both cleared and uncleared trades. Such analysis is straightforward using our framework. Note that the effect on the efficiency of CCPs in presence of compression is also non-trivial as discussed by Duffie and Zhu (2011).

The role that compression can play for systemic risk mitigation is major. Our results show the way compression, by reducing counterparty gross exposures, may limit the effects of contagion in times of distress. Interestingly, at the onset of the Great Financial Crisis, Lehman, which was believed to be counterparty to around $5 trillion of CDS contracts (Haldane, 2009) was reportedly subject to a compression exercise run in collaboration with policymakers[20]. However, the attempt was unsuccessful due to technical limitations and timing constraints. Though it is difficult to gauge the exact impact of a successful procedure, our findings suggest that, as a major dealer, Lehman's portfolio might have been largely compressed potentially curbing the systemic effects of its default.

In general, the relationship between compression and financial stability can be explored in several ways. Here, we list three major avenues. The first relates to the elimination (or mitigation) of the effects of chains of obligations, which have been identified as a source of uncertainty and instability by Roukny et al. (2016) and giving rise to frictions such as payments gridlocks (Rotemberg, 2011). Given these premises, compression provides an additional mechanism to mitigate these effects. Second, compression naturally leads to a reshaping of the underlying web of exposures. Compression solutions could, in some cases, concentrate exposures into specific market segments. While this effect can be both intended or unintended, it does involve non-negligible changes in counterparty risk, levels of collateral exchange, etc. Third, compression impacts capital management by its effect leverages. For example, OTC derivative exposures are accounted for in the Basel III leverage ratio. Banks' capital must therefore include those gross exposures. As compression reduces gross exposures, banks lower the amount of capital necessary to cover the same positions. There can be two views on the matter. On the one hand, reducing capital requirements frees up "unused" leverage, reduces inventory costs for dealers and thus increases liquidity. On the other hand, reducing capital requirements affects the loss absorption capacity of market participants potentially making the stability of the market as a whole worse off.

Finally, our framework does not pertain uniquely to contingent claims arising from financial institutions. It can also find applications for non-financial firms, thereby allowing to quantify the levels of excess and potential compression in the real economy. Indeed, as long as a market exhibits outstanding fungible positions and intermediaries, compression may reduce the total amount of outstanding debt due and liquidity needs (Kiyotaki et al., 1997).

[20]See the Bloomberg article by B. Ivry, C. Harper and M. Pittman, "Missing Lehman Lesson of Shakeout Means Too Big Banks May Fail" September 8, 2009.

References

Abad, J., Aldasoro, I., Aymanns, C., D'Errico, M., Fache Rousova, L., Hoffmann, P., Langfield, S., Neychev, M., and Roukny, T. (2016). Shedding light on dark markets: first insights from the new EU-wide OTC derivative dataset. *ESRB Occasional Paper Series No 11-16*.

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608.

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network flows: theory, algorithms, and applications*. Prentice Hall.

Ali, R., Vause, N., and Zikes, F. (2016). Systemic risk in derivatives markets: a pilot study using CDS data. *Bank of England Financial Stability Paper*, (38).

Allen, F. and Babus, A. (2009). Networks in finance. *The Network Challenge: Strategy, Profit, and Risk in an Interlinked World*.

Allen, F. and Gale, D. (2000). Financial contagion. *Journal of Political Economy*, 108(1):1–33.

Allen, F. and Santomero, A. M. (1997). The theory of financial intermediation. *Journal of Banking & Finance*, 21(11):1461–1485.

Atkeson, A. G., Eisfeldt, A. L., and Weill, P.-O. (2013). The market for OTC derivatives. Technical report, National Bureau of Economic Research.

Atkeson, A. G., Eisfeldt, A. L., and Weill, P.-O. (2015). Entry and exit in OTC derivatives markets. *Econometrica*, 83(6):2231–2292.

Babus, A. (2016). The formation of financial networks. *The RAND Journal of Economics*, 47(2):239–272.

Benos, E., Wetherilt, A., and Zikes, F. (2013). The structure and dynamics of the UK credit default swap market. *Bank of England Financial Stability Paper*, (25).

BIS (2015). OTC derivatives statistics at end-June 2015. Technical report, Basel Committee on Banking Supervision - Bank of International Settlements.

BIS (2016). Minimum capital requirements for market risk. Technical report, Basel Committee on Banking Supervision - Bank of International Settlements.

Blundell-Wignall, A. and Atkinson, P. (2010). Thinking beyond Basel III. *OECD Journal: Financial Market Trends*, 2010(1):9–33.

Burnham, J. (1991). Current structure and recent developments in foreign exchange markets. *editor: S. Khoury, Recent Developments in International Banking and Finance*, pages 123—-153.

Cont, R. and Kokholm, T. (2014). Central clearing of OTC derivatives: bilateral vs multilateral netting. *Statistics & Risk Modeling*, 31(1):3–22.

Craig, B. and Von Peter, G. (2014). Interbank tiering and money center banks. *Journal of Financial Intermediation*, 23(3):322–347.

D'Errico, M., Battiston, S., Peltonen, T., and Scheicher, M. (2016). How does risk flow in the Credit Default Swap market? *European Systemic Risk Board Working Paper Series No 33*.

Duffie, D. (2012). *Dark markets: Asset pricing and information transmission in over-the-counter markets*. Princeton University Press.

Duffie, D. et al. (2016). Financial regulatory reform after the crisis: An assessment. In *ECB Forum on Central Banking on "The Future of the International Monetary and Financial Architecture*, pages 27–29.

Duffie, D., Gârleanu, N., and Pedersen, L. H. (2005). Over-the-Counter Markets. *Econometrica*, 73(6):1815–1847.

Duffie, D. and Zhu, H. (2011). Does a central clearing counterparty reduce counterparty risk? *Review of Asset Pricing Studies*, 1(1):74–95.

Ehlers, T. and Eren, E. (2016). The changing shape of interest rate derivatives markets. *BIS Quarterly Review*.

Elliott, M., Golub, B., and Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, 104(10):3115–3153.

European Central Bank (2009). Credit Default Swaps and Counterparty Risk.

Flannery, M. J. (1996). Financial crises, payment system problems, and discount window lending. *Journal of Money, Credit and Banking*, 28(4):804–824.

Flood, M. D. (1994). Market structure and inefficiency in the foreign exchange market. *Journal of International Money and Finance*, 13(2):131–158.

Fricke, D. and Lux, T. (2015). Core–periphery structure in the overnight money market: evidence from the e-mid trading platform. *Computational Economics*, 45(3):359–395.

Gofman, M. (2016). Efficiency and stability of a financial architecture with too interconnected to fail institutions. *Journal of Financial Economics*, forthcoming.

Gros, D. (2010). Too interconnected to fail= too big to fail: What's in a leverage ratio? *CEPS Commentary*, 28.

Haldane, A. G. (2009). Rethinking the financial network. Speech delivered at the Financial Student Association, Amsterdam, The Netherlands.

in't Veld, D. and van Lelyveld, I. (2014). Finding the core: Network structure in interbank markets. *Journal of Banking & Finance*, 49:27–40.

ISDA (2105). The impact of compression on the interest rate derivatives market. Technical report, International Swaps and Derivatives Association - Research Note.

Kiyotaki, N., Moore, J., et al. (1997). Credit chains. *Journal of Political Economy*, 105(21):211–248.

Lyons, R. K. (1995). Tests of microstructural hypotheses in the foreign exchange market. *Journal of Financial Economics*, 39(2):321–351.

Lyons, R. K. (1997). A simultaneous trade model of the foreign exchange hot potato. *Journal of International Economics*, 42(3-4):275–298.

O'Kane, D. (2014). Optimizing the compression cycle: algorithms for multilateral netting in OTC derivatives markets. *Available at SSRN 2273802*.

Peltonen, T. A., Scheicher, M., and Vuillemey, G. (2014). The network structure of the CDS market and its determinants. *Journal of Financial Stability*, 13:118–133.

Rotemberg, J. J. (2011). Minimal settlement assets in economies with interconnected financial obligations. *Journal of Money, Credit and Banking*, 43(1):81–108.

Roukny, T., Battiston, S., and Stiglitz, J. E. (2016). Interconnectedness as a source of uncertainty in systemic risk. *Journal of Financial Stability (in press).*

Schrimpf, A. (2015). Outstanding OTC derivatives positions dwindle as compression gains further traction. *BIS Quarterly Review*, pages 24–5.

Shachar, O. (2012). Exposing the exposed: Intermediation capacity in the credit default swap market. *Federal Researve Bank of New York Working Paper*.

Stulz, R. M. (2010). Credit default swaps and the credit crisis. *The Journal of Economic Perspectives*, 24(1):73–92.

Yellen, J. (2013). Interconnectedness and systemic risk: Lessons from the financial crisis and policy implications. Speech presented at The American Economic Association/American Finance Association joint luncheon, San Diego, CA.

A Proofs

A.1 Proposition 1

Proof. The proof consists of two steps.

1. First, we show that given a market $G = (N, E)$, we can always find a net-equivalent market G' with total notional of x' as in Equation 1.

   Consider the partition of N into the following disjoint subsets: $N^+ = \{i | v_i^{net} > 0\}, N^- = \{i | v_i^{net} < 0\}$ and $N^0 = \{i | v_i^{net} = 0\}$ (such that $N = N^+ \bigcup N^- \bigcup N^0$). Let $B \in N \times N$ be a new set of edges (each with weight b_{ij}) such that:

   - $\forall b_{ij}$ s.t. $(i, j) \in B$, $i \in N^+, j \in N^-$;
   - $\sum_j b_{ij} = v_i^{net}, \quad \forall i \in N^+$;
   - $\sum_i b_{ij} = v_j^{net}, \quad \forall j \in N^-$.

   The total notional of the market $G' = (N, B)$ is thus given by:

   $$x' = \sum_i \sum_j b_{ij} = \sum_{i \in V^+} v_i^{net} = \sum_{i \in V^-} |v_i^{net}|.$$

   As edges in B only link two nodes within N (i.e., the system is closed), the sum of all net position is equal to 0: $\sum_i v_i^{net} = 0$. Hence, we have: $\sum_{i \in V^+} v_i^{net} + \sum_{j \in V^-} v_j^{net} = 0$. We see that, in absolute terms, the sum of net positions of each set (V^+ and V^-) are equal: $|\sum_{i \in V^+} v_i^{net}| = |\sum_{j \in V^-} v_j^{net}|$. As all elements in each part have the same sign by construction, we obtain: $\sum_{i \in V^+} |v_i^{net}| = \sum_{j \in V^-} |v_j^{net}|$. As a result, we have: $\sum_{i \in V^+} v_i^{net} = \frac{1}{2} |\sum_{i \in V} v_i^{net}|$.

2. Second, we show that x' is the minimum total notional attainable from a net-equivalent operation over $G = (N, E)$. We proceed by contradiction. Consider $G' = (N, B)$ as defined above and assume there exists a $G^* = (N, B^*)$ defined as a net-equivalent market to G' such that $x^* < x'$. At the margin, such result can only be obtained by a reduction of some weight in B: $\exists b^*_{ij} < b_{ij}$. If $x^* < x'$, then there exists at least one node for which this reduction is not compensated and thus $\exists v_i^{*net} < v_i^{net}$. This violates the net-equivalent condition. Hence, $x' = \sum_{i:\ v_i^N > 0}^n v_i^N$ is the minimum net equivalent notional.

■

A.2 Lemma 1

Proof. By definition, $\delta(i) = 1 \Leftrightarrow \sum_j e_{ij} \cdot \sum_j e_{ji} > 0$: a dealer has thus both outgoing and incoming edges. Then it holds that:

$$\delta(i) = 1 \quad \Rightarrow \quad v_i^{gross} > |v_i^{net}| \quad \Leftrightarrow \quad \sum_j e_{ij} + \sum_j e_{ji} > \left| \sum_j e_{ij} - \sum_j e_{ji} \right|.$$

In contrast, for a customer $\sum_j e_{ij} \cdot \sum_j e_{ji} = 0$ and thus $\delta(i) = 0$. Then it holds that:

$$\delta(i) = 0 \quad \Rightarrow v_i^{gross} = |v_i^{net}| \quad \Leftrightarrow \quad \sum_j e_{ij} + \sum_j e_{ji} = \left| \sum_j e_{ij} - \sum_j e_{ji} \right|.$$

The equality is simply proven by the fact that if i is a customer selling (resp. buying) in the market, then $\sum_j e_{ji} = 0$ (resp. $\sum_j e_{ij} = 0$) and thus both ends of the above equation are equal.

If $G = (N, E)$ has $\sum_{i \in N} \delta(i) = 0$, then all market participants are customers, and we thus have: $v_i^{gross} = |v_i^{net}| \; \forall i \in N$. As a result, the excess is given by

$$\Delta(G) = x - \frac{1}{2} \sum_i \left| v_i^{net} \right| = x - \frac{1}{2} \sum_i |v_i^{gross}|.$$

As in the proof of Proposition 1, the market we consider is closed (i.e., all edges relate to participants in N) and thus: $\sum_i |v_i^{gross}| = 2x$. We thus have no excess in such market: $\Delta(G) = 0$.

If $G = (N, E)$ has $\sum_{i \in N} \delta(i) > 0$, then some market participants have $v_i^{gross} > |v_i^{net}|$. As a result, the excess is given by:

$$\begin{aligned}\Delta(G) &= x - \frac{1}{2} \sum_i |v_i^{net}| = \frac{1}{2} \sum_i |v_i^{gross}| - \frac{1}{2} \sum_i |v_i^{net}| = \\ &= \sum_i |v_i^{gross}| - \sum_i |v_i^{net}| > 0\end{aligned}$$

■

A.3 Proposition 2

Proof. For sake of clarity, in the following we only focus the notation on the set of edges for the computation of excess. In general, let us decompose the set of edges E in two subsets A and B such that $E = A \cup B$ and $\sum_{ij} e_{ij} = \sum_{ij} a_{ij} + \sum_{ij} b_{ij}$. We want to verify if

$$\Delta(E) = \Delta(A) + \Delta(B)$$

We decompose each part according to the definition of excess:

$$\begin{aligned}
\sum_{ij} e_{ij} - 0.5 \sum_i |\sum_j (e_{ij} - e_{ji})| &= \sum_{ij} a_{ij} - 0.5 \sum_i |\sum_j (a_{ij} - a_{ji})| + \sum_{ij} b_{ij} + \\
&\quad - 0.5 \sum_i |\sum_j (b_{ij} - b_{ji})| = \\
-0.5 \sum_i |\sum_j (e_{ij} - e_{ji})| &= -0.5 \sum_i |\sum_j (a_{ij} - a_{ji})| + \\
&\quad - 0.5 \sum_i |\sum_j (b_{ij} - b_{ji})| \\
\sum_i |\sum_j (e_{ij} - e_{ji})| &= \sum_i |\sum_j (a_{ij} - a_{ji})| + \sum_i |\sum_j (b_{ij} - b_{ji})| \\
\sum_i |\sum_j (e_{ij} - e_{ji})| &= \sum_i |\sum_j (a_{ij} - a_{ji})| + \sum_i |\sum_j (b_{ij} - b_{ji})| \\
\sum_i |\sum_j (a_{ij} + b_{ij} - a_{ji} - b_{ji})| &= \sum_i |\sum_j (a_{ij} - a_{ji})| + \sum_i |\sum_j (b_{ij} - b_{ji})| \\
\sum_i |\sum_j (a_{ij} - a_{ji}) + \sum_j (b_{ij} - b_{ji})| &= \sum_i \Big(|\sum_j (a_{ij} - a_{ji})| + |\sum_j (b_{ij} - b_{ji})| \Big)
\end{aligned}$$

This later relationship is not true in general due to the convexity of the absolute value function. Using Jensen's inequality we thus have the following relationship:

$$\Delta(E) \geq \Delta(A) + \Delta(B)$$

We now identify specific cases under our framework in which the relationship holds. Let us decompose the original additivity expression:

$$\begin{aligned}
\Delta(E) &= \Delta(E^D) + \Delta(E^C) \\
\sum_i |\sum_j (e_{ij} - e_{ji})| &= \sum_i |\sum_j (e^D_{ij} - e^D_{ji})| + \sum_i |\sum_j (e^C_{ij} - e^C_{ji})|
\end{aligned}$$

We can decompose each part in the context of a dealer-customer network.

1) For the whole network we have

$$\sum_i |\sum_j (e_{ij} - e_{ji})| = \sum_d^{dealer} |\sum_j (e_{dj} - e_{jd})| + \sum_c^{customer} |\sum_j (e_{cj} - e_{jc})|$$

$$= \sum_d^{dealer} |\sum_j (e_{dj} - e_{jd})| + \sum_{c^+}^{customer^+} |\sum_j (e_{c^+j} - e_{jc^+})| +$$

$$+ \sum_{c^-}^{customer^-} |\sum_j (e_{c^-j} - e_{jc^-})| =$$

$$= \sum_d^{dealer} |\sum_j (e_{dj} - e_{jd})| + \sum_{c^+}^{customer^+} |\sum_j (e_{c^+j})| +$$

$$+ \sum_{c^-}^{customer^-} |\sum_j (-e_{jc^-})| =$$

$$= \sum_d^{dealer} |\sum_j (e_{dj} - e_{jd})| + \sum_{c^+}^{customer^+} \sum_d^{dealer} e_{c^+d} + \sum_{c^-}^{customer^-} \sum_d^{dealer} e_{dc^-}$$

2) For the dealer network we have

$$\sum_i |\sum_j (e^D_{ij} - e^D_{ji})| = \sum_d^{dealer} |\sum_h^{dealer} (e^D_{dh} - e^D_{hd})|$$

3) For the customer network we have

$$\sum_i |\sum_j (e^C_{ij} - e^C_{ji})| = \sum_d^{dealer} |\sum_j (e^C_{dj} - e^C_{jd})| +$$

$$+ \sum_{c^+}^{customer^+} |\sum_j (e^C_{c^+j} - e^C_{jc^+})| \sum_{c^-}^{customer^-} |\sum_j (e^C_{c^-j} - e^C_{jc^-})|$$

$$= \sum_d^{dealer} |\sum_j (e^C_{dj} - e^C_{jd})| + \sum_{c^+}^{customer^+} \sum_d^{dealer} e^C_{c^+d} + \sum_{c^-}^{customer^-} \sum_d^{dealer} e^C_{dc^-}$$

Combining equations, we obtain:

$$\sum_d^{dealer} |\sum_j^{n} (e_{dj} - e_{jd})| = \sum_d^{dealer} |\sum_h^{dealer} (e^D_{dh} - e^D_{hd})| + \sum_d^{dealer} |\sum_c^{customer} (e^C_{dc} - e^C_{cd})|$$

We continue decomposing the different elements.

1) For the whole network:

$$\sum_{d}^{dealer} |\sum_{j}^{n}(e_{dj} - e_{jd})| = \sum_{d}^{dealer} |\sum_{h}^{dealer}(e_{dh} - e_{hd}) + \sum_{c^+}^{customer^+}(e_{dc^+} - e_{c^+d}) + \sum_{c^-}^{customer^-}(e_{dc^-} - e_{c^-d})|$$

$$= \sum_{d}^{dealer} |\sum_{h}^{dealer}(e_{dh} - e_{hd}) + \sum_{c^+}^{customer^+} e_{dc^+} - \sum_{c^-}^{customer^-} e_{c^-d}|$$

2) for the dealer and customer networks:

$$\sum_{d}^{dealer} |\sum_{h}^{dealer}(e_{dh}^{D} - e_{hd}^{D})| + \sum_{d}^{dealer} |\sum_{c}^{customer}(e_{dc}^{C} - e_{cd}^{C})| =$$

$$\sum_{d}^{dealer} |\sum_{h}^{dealer}(e_{dh}^{D} - e_{hd}^{D})| + \sum_{d}^{dealer} |\sum_{c^+}^{customer^+} e_{dc^+}^{C} - \sum_{c^-}^{customer^-} e_{c^-d}^{C}|$$

After this decomposition, we can remove the subscripts related to the different networks, and we obtain the general condition for additive excess:

$$\sum_{d}^{dealer} |\sum_{h}^{dealer}(e_{dh} - e_{hd}) + \sum_{c^+}^{customer^+} e_{dc^+} - \sum_{c^-}^{customer^-} e_{c^-d}| =$$

$$\sum_{d}^{dealer} |\sum_{h}^{dealer}(e_{dh} - e_{hd})| + \sum_{d}^{dealer} |\sum_{c^+}^{customer^+} e_{dc^+} - \sum_{c^-}^{customer^-} e_{c^-d}|$$

Hence, the above relationship holds when

1. $\sum_{h}^{dealer}(e_{dh} - e_{hd}) = 0, \quad \forall d \in D$

   or

2. $\sum_{c^+}^{customer^+} e_{dc^+} - \sum_{c^-}^{customer^-} e_{c^-d} = 0, \quad \forall d \in D$

■

A.4 Proposition 3

Proof. Non-conservative compression tolerances allow all possible re-arrangements of edges. Hence, the only condition for non-conservative compression to remove excess (i.e., $\Delta^{c^n}_{\text{red}}(G) > 0$) is merely that excess is non-zero (i.e., $\Delta(G) > 0$). From Lemma 1, we know that positive excess exists in $G = (N, E)$ only when there is intermediation (i.e., $\exists i \in N|\delta(i) = 1$). ■

A.5 Proposition 4

Proof. We proceed by defining a procedure that respects the non-conservative compression constraints and show that this procedure (algorithm) generates a new configuration of edges such that the resulting excess is 0.

Similar to the proof of Proposition 1, consider the three disjoint subsets $N^+ = \{i|v_i^{net} > 0\}, N^- = \{i|v_i^{net} < 0\}$ and $N^0 = \{i|v_i^{net} = 0\}$, such that $N = N^+ \bigcup N^- \bigcup N^0$. Let B be a new set of edges such that:

- $\forall b_{ij} \in B, \quad i \in N^+, j \in N^-$
- $\sum_j b_{ij} = v_i^{net}, \quad \forall i \in N^+$
- $\sum_i b_{ij} = v_j^{net}, \quad \forall j \in N^-$

The market $G' = (N, B)$ is net-equivalent to G while the total gross notional is minimal in virtue of Proposition 1. The nature of the new edges makes G' bipartite (i.e., $\forall b_{ij} \in B, \quad i \in N^+, j \in N^-$), hence, there is no intermediation in G'. The procedure depicted above to obtain B is a meta-algorithm as it does not define all the steps in order to generate B. As a result, several non-conservative compression operation c^n can satisfy this procedure. Nevertheless, by virtue of Proposition 3, each of these non-conservative compression operation lead to $\Delta^{c^n}_{res}(G) = \Delta(G') = 0$ ■

A.6 Proposition 5

Proof. The value of λ will affect the efficiency of the compression. In order to achieve full compression, we show that λ must be above a certain limit. Let us decompose N, the set of nodes, as follows:

$$N^+ = \{i|v_i^{net} > 0\}, \quad N^- = \{i|v_i^{net} < 0\}, \quad N^0 = \{i|v_i^{net} = 0\}$$

As a result, $N = N^+ \cup N^- \cup N^0$. In a case of 0 residual excess, the node with positive net balance can only interact with a node with negative balance (i.e., bi-partite graph):

$$\forall \hat{e}_{ij} \in \hat{E} : i \in N^+, j \in N^-$$

Hence, the maximum possible value of a contract resulting from such compression is:

$$max\{\hat{e}_{ij}\} = max\{|v_i^{net}|\}$$

For the node with $max\{|v_i^{net}|\}$, i^*, the portfolio configuration such that bilateral exposure is minimized is the uniform distribution:

$$max\{\hat{e}_{i^*j}\} = \frac{|v_{i^*}^{net}|}{|N^{-1.sign(v_{i^*}^{net})}|}$$

If the exposure limit λ is set such that this configuration is feasible, we know a solution with 0 residual excess is always feasible.

More generally we see that, a 0 residual solution is possible for any compression tolerance set that satisfies the following conditions:

$$a_{ij} = 0, \quad b_{ij} \geq \frac{|v_{i^*}^{net}|}{|N^{-1.sign(v_{i^*}^{net})}|} \qquad \forall (a_{ij}, b_{ij}) \in \Gamma$$

■

A.7 Proposition 6

Proof. In a conservative compression, we have the constraint:

$$0 \leq e'_{ij} \leq e_{ij} \quad \forall i, j \in N$$

At the individual level, assume i is a customer selling in the market (i.e., $\delta(i) = 0$). Under a conservative approach, it is not possible to compress any of the edges of i. In fact, in order to keep the net position of i constant, any reduction of ε in an edge of i (i.e., $e'_{ij} = e_{ij} - \varepsilon$) requires a change in some other edge (i.e., $e'_{ik} = e_{ik} + \varepsilon$) in order to keep $v_i'^{net} = v_i^{net}$. Such procedure violates the conservative compression tolerance: $e'_{ik} = e_{ik} + \varepsilon > e_{ik}$. The same situation occurs for customers buying. Conservative compression can thus not be applied to node i if $\delta(i) = 0$.

The only configuration in which a reduction of an edge e_{ij} does not require a violation of the conservative approach and the net-equivalence condition is when i can reduce several edges in order to keep its net balance. In fact, for a node i, the net position is constant after a change $\sum_j e'_{ij} = \sum_j e_{ij} - \varepsilon$ if it is compensated by a change $\sum_j e'_{ji} = \sum_j e_{ji} - \varepsilon$. Only dealers can apply such procedure. Furthermore,

such procedure can only be applied to links with other dealers: a reduction on one link triggers a cascade of balance adjusting that can only occur if other dealers are concerned as customers are not able to re-balance their net position as shown above. Hence, the redundant excess for a conservative approach emerges from intra-dealer links.
Finally, the sequence of rebalancing and link reduction can only finish once it reaches the initiating node back. Hence, conservative compression can only be applied to closed chains of intermediation, that is, a set of links $E^* \subset E$ such that all links have positive values $\prod_{e^* \in E^*} e^* > 0$. ∎

A.8 Lemma 2

Proof. A conservative compression on a closed chain of intermediation $K = (N, E) \to (K, E')$ implies that, in order for the compression to be net equivalent (i.e., $v_i'^{net} = v_i^{net} \;\; \forall i \in N$), a reduction by and arbitrary $\varepsilon \in [0, \max_{ij}\{e_{ij}$ s.t. $(i,j) \in E\}]$ on an edge $e'_{ik} = e_{ik} - \varepsilon$ must be applied on all other edges in the chain: $e' = e - \varepsilon \;\; \forall e' \in E'$.

Overall, reducing by ε one edge, leads to an aggregate reduction of $|E|\varepsilon$ after re-balancing of net positions.

Recall that, in a conservative compression, we have $0 \leq e'_{ij} \leq e_{ij}$. Hence, for each edge, the maximum value that ε can take is e_{ij}. At the chain level, this constraint is satisfied i.f.f. $\varepsilon = min_e\{E\}$. The redundant excess is given by $|E|min_e\{E\}$ and the residual excess is thus

$$\Delta^c_{res}(K) = \Delta(K) - |E|min_e\{E\}$$

∎

A.9 Proposition 7

Proof. From Corollary 4, we know that all the excess is removed from a market when the resulting set of edges E' form a bi-partite structure (i.e., not intermediation). We also know that from Proposition 6, that conservative compression can only be applied to closed chains of intermediation. Hence, given $G = (N, E)$, in order to obtain $\Delta^c_{res}(G) = 0$, we need that (1) all chains are closed chain, to apply conservative compression and (2) all closed chains are balanced, to remove all the excess.

The first condition stems from Proposition 6. The second condition is justified as follows. Consider the special case where $K = (N, E)$ is a closed chain of intermediation such that:

$$e_i = \alpha \quad \forall e_i \in E, \alpha \in R_0^+$$

In this chain, the net position of all nodes $i \in N$ is 0. Hence, removing all the edges satisfies the net-equivalence property and the conservative compression tolerance. As a result, we have $\Delta_{\text{res}}^{c^c}(G) = 0$ simply because $x' = 0$.

Next, consider changing $K = (N, E)$ such that one single edge has a higher value than all the others which remain with the value α:

$$\exists 1! e^* \in E | e^* > \alpha.$$

Following the Lemma 2, we can remove all edges equal to α and modify e^* such that

$$e^{'*} = e^* - \alpha.$$

The market G' has been compressed conservatively and only has one edge left (i.e., $E' = \{e^{'*}\}$). As a result, there is no excess in G' (i.e., no intermediation) and $\Delta_{\text{res}}^{c^c}(G) = 0$.

For a closed chain of any length and heterogeneous edge value distribution, the breaking of intermediation chain can only be done if a node with an edge with values higher than the minimum has the other edge equal to the minimum. Such property is only satisfied when closed chains of intermediation are balanced in the sense of Definition 4.6.2. ■

A.10 Proposition 8

Proof. If there are no entangled chains in $G = (N, E)$, then the following conservative procedure:

1. list all closed chains of intermediation $K_i \in \Pi$ and
2. maximally compress each chain separately,

reaches maximal efficiency. The residual excess is given after aggregating the excess removed on each closed chain separately:

$$\Delta_{res}(G) = \Delta(G) - \sum_{K_i \in \Pi} |E_i| min_e\{E_i\}.$$

If there are entangled chains but the market $G = (N, E)$ is chain ordering proof, compressing chains separately only provides the upper bound as there will be cases where entangled chains will need to be updated (due to the reduction of one or more edges). Hence, we have,

$$\Delta_{res}(G) \leq \Delta(G) - \sum_{K_i \in \Pi} |E_i| \min_{e}\{E_i\}.$$

■

A.11 Proposition 9

Proof. If $\Delta(N, E) = \Delta(N, E^D) + \Delta(N, E^C)$, then we can separate the compression of each market.

Intra-dealer market (N, E^D). According to the hybrid compression, the set of constraints in the intra-dealer market is given by a non-conservative compression tolerances set. According to Proposition 4, the residual excess is zero. We thus have:

$$\Delta_{res}^{c^h}(N, E^D) = 0.$$

Intra-dealer market (N, E^D). According to the hybrid compression, the set of constraints in the customer market is given by a conservative compression tolerances set. Since, by construction, the customer market does not have closed chains of intermediation, it is not possible to reduce the excess on the customer market via conservative compression. We thus have:

$$\Delta_{res}^{c^h}(N, E^C) = \Delta(N, E^C).$$

Finally, we obtain

$$\begin{aligned} \Delta_{\text{res}}^{c^h}(N, E) &= \Delta_{\text{res}}^{c^h}(N, E^D) + \Delta_{\text{res}}^{c^h}(N, E^C) \\ &= \Delta(N, E^C) \end{aligned}$$

■

A.12 Proposition 10

Proof. If the market $G = (N, E)$ is such that $\nexists i, j \in N \quad s.t. \quad e_{ij}.e_{ji} > 0$ where $e_{ij}, e_{ji} \in E$ then the compression tolerances will always be:

$$a_{ij} = b_{ij} = \max\{e_{ij} - e_{ji}, 0\} = e_{ij}$$

Hence, $\Delta_{red}^{c^b}(G) = \Delta_{red}(G)$ and thus $\Delta_{res}^{c^b}(G) = 0$. If the market $G = (N, E)$ is such that $\exists i, j \in N \quad s.t. \quad e_{ij}.e_{ji} > 0$ where $e_{ij}, e_{ji} \in E$ then the bilateral compression will yield a market $G' = (N, E')$ where $x^{'} < x$. Hence, $\Delta_{red}^{c^b}(G) < \Delta_{red}(G)$ and thus $\Delta_{res}^{c^b}(G) > 0$ ■

A.13 Proposition 11

Proof. If the market $G = (N, E)$ is such that $\exists i, j \in N \quad s.t. \quad e_{ij}.e_{ji} > 0$ where $e_{ij}, e_{ji} \in E$, then, bilaterally compressing the pair i and j yields the following situation. Before compression, the gross amount on the bilateral pair was $e_{ij} + e_{ji}$. After compression, the gross amount on the same bilateral pair is $|e_{ij} - eji|$. Hence, we have a reduction of gross notional of $2.min\{e_{ij}, e_{ji}\}$. The market gross notional after compression of this bilateral pair is thus given by: $x^{'} = x - 2.min\{e_{ij}, e_{ji}\}$ and the excess in the new market (i.e., residual excess after having bilaterally compressed the pair (i, j)) follows the same change: $\Delta_{res}(G) = \Delta(G) - 2.min\{e_{ij}, e_{ji}\}$. We generalize the result by looping over all pairs and noting that the reduction $min\{e_{ij}, e_{ji}\}$ is doubled counted: pairing by (i, j) and (j, i). Hence, we reach the following expression of the residual excess:

$$\Delta^{c^b}_{res}(G) = \Delta(G) - \sum_{i,j \in N} \min\{e_{ij}, e_{ji}\} \text{ where } e_{ij}, e_{ji} \in E$$

■

A.14 Proposition 12

Proof. We proceed by analyzing sequential pairs of compression operators and show the pairing dominance before generalizing. We start with the bilateral compressor $c^b()$ and the conservative compressor $c^c()$. Let $(a^b_{ij}, b^b_{ij}) \in \Gamma^b$ and $(a^c_{ij}, b^c_{ij}) \in \Gamma^c$ be the set of compression tolerance for the bilateral and conservative compressor, respectively. We have the following relationship:

$$a^c_{ij} \leq a^b_{ij} = b^b_{ij} \leq b^c_{ij} \quad \forall e_{ij} \in E$$

In fact, by definition of each compression tolerance set, we have:

$$0 \leq \max\{e_{ij} - e_{ji}, 0\} \leq e_{ij} \quad \forall e_{ij} \in E$$

Hence, we see that the set of possible values couple for bilateral compression is bounded below and above by the set of conservative compression values. By virtue of linear composition, a solution of the bilateral compression thus satisfies the conservative compression tolerance set. The other way is not true as the lower bound in the bilateral case a^b_{ij} can be equal to $e_{ij} - e_{ji}$ while in the conservative case, we always have that $a^c_{ij} = 0$. Hence, in terms of efficiency, we have that a globally optimal conservative solution is always at least equal, in redundant excess, to the globally optimal bilateral solution: $\Delta^{c^b}_{red}(G) \leq \Delta^{c^c}_{red}(G)$. The case in which the efficiency of $\Delta^{c^c}_{red}(G)$ is higher is a function of the network structure of G. In fact, if the market G only exhibits cycles of length one, we have

$\Delta^{c^b}_{red}(G) = \Delta^{c^c}_{red}(G)$. Once G exhibits higher length cycles, we have a strict dominance $\Delta^{c^b}_{red}(G) < \Delta^{c^c}_{red}(G)$. Similar reasoning is thus applied to the next pairing: conservative and hybrid compression tolerance sets. Let $(a^h_{ij}, b^h_{ij}) \in \Gamma^h$ be the set of compression tolerance for the hybrid compressor. We have the following nested assembly:

$$a^c_{ij} = a^h_{ij} \quad \text{and} \quad b^c_{ij} = b^h_{ij} \quad \forall e_{ij} \in E^C$$
$$a^c_{ij} = a^h_{ij} \quad \text{and} \quad b^c_{ij} \leq b^h_{ij} \quad \forall e_{ij} \in E^D$$

Where E^C and E^D are the customer market and the intra-dealer market, respectively, with $E^C + E^D = E$. In fact, by definition of the compression tolerance sets in the customer market E^C are the same while for the intra-dealer market we have:

$$a^c_{ij} = a^h_{ij} = 0 \quad \text{and} \quad e_{ij} \leq +\infty \quad \forall e_{ij} \in E^D$$

Similar to the dominance between bilateral and conservative compression, we can thus conclude that: $\Delta^{c^c}_{red}(G) \leq \Delta^{c^h}_{red}(G)$. It is the relaxation of tolerances in the intra-dealer market that allows the hybrid compression to be more efficient than the conservative compression. By virtue of complementarity of this result, the hybrid and non-conservative pairing is straightforward: $\Delta^{c^h}_{red}(G) \leq \Delta^{c^n}_{red}(G)$. As we know from Proposition 4, $\Delta^{c^n}_{red}(G) = \Delta(G)$, we thus obtain the general formulation of weak dominance between the 4 compression operators:

$$\Delta^{c^b}_{red}(G) \leq \Delta^{c^c}_{red}(G) \leq \Delta^{c^h}_{red}(G) \leq \Delta^{c^n}_{red}(G) = \Delta(G)$$

∎

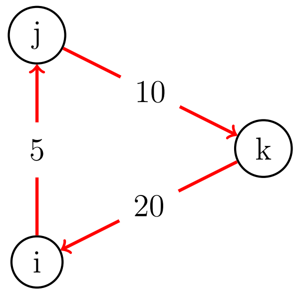


Figure 6: Original configuration the market

B A simple example with 3 market participants

To better articulate the different ways in which portfolio compression can take place according to the conservative and non-conservative approach, let us take the following example of a market made of 3 financial institutions. Figure 6 graphically reports the financial network: the institution i has an outstanding contract sold to j of notional value 5 while buying one from k of notional value 20 and j has an outstanding contract sold to k of notional value 10. For each institution, we compute the gross and net positions:

$$\begin{aligned} v_i^{gross} &= 25 \qquad & v_i^{net} &= -15 \\ v_j^{gross} &= 15 \qquad & v_j^{net} &= +5 \\ v_k^{gross} &= 30 \qquad & v_k^{net} &= +10 \end{aligned}$$

We also obtain the current excess in the market:

$$\Delta(G) = 35 - 15 = 20$$

Let us first adopt a conservative approach. In this case, we can only reduce or remove currently existing trades. A solution is to remove the trade between i and j and adjust the two other contracts accordingly (i.e., subtract the value of the ij contract from the two other contracts). The resulting market is represented in Figure 7(a). Computing the same measurements as before, we obtain:

$$\begin{aligned} v_i^{\prime gross} &= 15 \qquad & v_i^{\prime net} &= -15 \\ v_j^{\prime gross} &= 5 \qquad & v_j^{\prime net} &= +5 \\ v_k^{\prime gross} &= 20 \qquad & v_k^{\prime net} &= +10 \end{aligned}$$

We also obtain the new excess in the market:

$$\Delta_{res}^{cons}(G) = 20 - 15 = 5$$

	Conservative	**Non-conservative**
Total excess	20	20
Redundant excess	15	20
Residual excess	5	0

Table 8: Table summarizing the results applying conservative and non-conservative compression on the market with 3 participants in Figure 7.

We see that, after applying the conservative compression operator that removed the (i, j) contract, we have reduced the excess by 15. It is not possible to reduce the total excess further without violating the conservative compression tolerances. We thus conclude that, for the conservative approach, the residual excess is 5 and the redundant excess is 15.

Let us now go back to the initial situation of Figure 6 and adopt a non-conservative approach. We can now create, if needed, new trades. A non-conservative solution is to remove all trades and create 2 new trades: one going from j to i of value 5 and one going from k to i of value 10. We have created a contract that did not exist before between j and i. The resulting market is depicted in Figure 7(b). Computing the same measurements as before, we obtain:

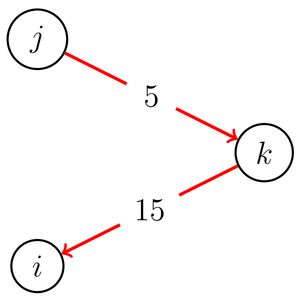
$$\begin{array}{ll} v_i'^g = 15 & v_i'^n = -15 \\ v_j'^g = 5 & v_j'^n = +5 \\ v_k'^g = 10 & v_k'^n = +10 \end{array}$$

We also obtain the current excess in the market:

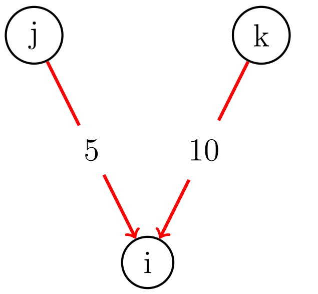
$$\Delta_{res}^{non-cons}(G) = 15 - 15 = 0$$

We observe that we have managed to achieve perfectly efficient compression as there is no more excess of notional in the resulting market while all the net positions have remained untouched. Individual gross positions are now completely in line with the net positions. Nevertheless the solution has generated a new trade (i.e., from j to i). We thus conclude that, for the non-conservative approach, the residual excess is 0 and the redundant excess is 20.

The results are summarized in Table 8. Though simple, the above exercise hints at several intuitive mechanisms and results. In the following sections, we develop further those aspects in a systematic and generalized analysis.



(a) After conservative compression



(b) After non-conservative compression

Figure 7: Examples of conservative and non-conservative compression approaches.

C Compression Algorithms

C.1 Non-Conservative Algorithm

In order to provide a rigorous benchmark, we propose a deterministic non-conservative compression algorithm that achieves perfectly efficient compression. In particular, the solution of the algorithm minimizes the number of trades and maximizes their concentration.

Data: Original Market G=(N,E)
Result: G^* such that $\Delta_v(G^*) = 0$
Let $N^+ = \{s \;\text{ s.t. } v_n^s > 0 \;\text{ and } s \in N\}$ be ordered such that $v_1^{net} > v_2^{net}$;
Let $N^- = \{s \;\text{ s.t. } v_s^{net} < 0 \;\text{ and } s \in N\}$ be ordered such that $v_1^{net} > v_2^{net}$;
Let i = 1 and j = 1;
while $i! = |N^+|$ *and* $j! = |N^-|$ **do**
  Create edge $e^*_{ij} = min(v_i^{net} - \sum_{j'<j} e^*_{ij'}, v_j^{net} - \sum_{i'<i} e^*_{i'j});$
  if $v_i^{net} = \sum_{j'<j} e^*_{ij'}$ **then**
    $i = i + 1;$
  end
  if $v_j^{net} = \sum_{i'<i} e^*_{i'j}$ **then**
    $j = j + 1;$
  end
end

Algorithm 1: A perfectly efficient non-conservative compression algorithm with minimal density

From the initial market, the algorithm constructs two sets of nodes N^+ and N^- which contain nodes with positive and negative net positions, respectively. Note that nodes with 0 net positions (i.e., perfectly balanced position) will become isolated in the intermediation breakdown process. They are thus kept aside from this point on. In addition, those two sets are sorted from the lowest to the highest absolute net position. The goal is then to generate a set of edges such that the resulting network is in line with the net position of each node. Starting from the nodes with the highest absolute net position, the algorithm generates edges in order to satisfy the net position of at least one node in the pair (i.e., the one with the smallest need). For example, if the node with highest net positive position is i with v_i^{net} and the node with lowest net negative position is j with v_j^{net}, an edge will be created such that the node with the lowest absolute net positions does not need more edges to satisfy its net position constraint. Assume that the nodes i and j are isolated nodes at the moment of decision, an edge $e_{ij} = min(v_i^{net}, v_j^{net})$ will thus be generated. In the more general case where i and j might already have some trades, we discount them in the edge generation process: $e^*_{ij} = \min(v_i^{net} - \sum_{j'<j} e^*_{ij'}, v_j^{net} - \sum_{i'<i} e^*_{i'j})$. The algorithm finishes once all the nodes have the net and gross positions equal.

The characteristics of the market resulting from a compression that follows the above algorithm are the following

Given a financial network G and a compression operator c() that is defined by the Algorithm 1, the resulting financial network $G_{min} = c(G)$ is defined as:

$$e_{ij} = \begin{cases} \min(v_n^i - \sum_{j'<j} e_{ij'}, v_n^j - \sum_{i'<i} e_{i'j}), & \text{if } v_n^i \cdot v_n^j < 0 \\ 0, & \text{otherwise} \end{cases}$$

where $i \in V^+ = \{s \;\text{ s.t. } v_n^s > 0\}$ and $j \in V^- = \{s \;\text{ s.t. } v_n^s < 0\}$. Moreover:

- G_{min} is net-equivalent to G
- $\Delta_v(G_{min}) = 0$
- G_{min} has the minimum link density
- G_{min} has maximum trade concentration

C.2 Conservative Algorithm

As we did for the non-conservative case, we now propose and analyze a conservative algorithm with the objective function of minimizing the excess of a given market with two constraints: (1) keep the net positions constant and (2) the new set of trades is a subset of the previous one.

Data: Original Market G=(N,E)

Result: G^* such that $\Delta_v(G^*) < \Delta_v(G)$ and $E^* \in E$

Let Π be set the of all directed closed chains in G;

Let $G^* = G$;

while $\Pi \neq \emptyset$ **do**

Select $P = (N', E') \in \Pi$ such that $|N'|.min_{e \in E'}(e) = max_{P_i=(N'_i,E'_i) \in \Pi}(|N'_{P_i}|.min_{e \in E'_{P_i}}(e)))$;

$e_{ij} = e_{ij} - min_{e \in E'}(e)$ for all $e_{ij} \in E'$;

$E^* = E^* \setminus \{e : e = min(E')\}$;

$\Pi \setminus \{P\}$

end

Algorithm 2: A deterministic conservative compression algorithm

The algorithm works as follows. First, it stores all the closed chains present in the market. Then, it selects the cycle (i.e., closed chain) that will result in the maximum marginal compression (at the cycle level), that is, the cycle where the combination of the number of nodes and the value of the lowest trades is maximized. From that cycle, the algorithm removes the trade with the lowest notional and subtracts this value from the all the trades in the cycle. It then removes the cycle from the list of cycles and iterates the procedure until the set of cycles in the market is empty.

At each cycle step t of the algorithm, the excess of the market is reduced by:

$$\Delta_t = \Delta_{t-1} - |N'| min_{e \in E'}(e)$$

At the end of the algorithm, the resulting compressed market does not contain directed closed chains anymore: it is a Directed Acyclic Graph (DAG). Hence no further conservative compression can be applied to it.

C.3 Bilateral Algorithm

We now briefly describe the bilateral algorithm. In this case we can only compress bilaterally redundant positions between pairs in the market. This algorithm is straightforward and follows directly from Proposition 11:

Data: Original Market G=(N,E)
Result: G^* such that $\Delta_v(G^*) < \Delta_v(G)$ and $E^* \in E$
Let $G^* = (N, E^* = \{\})$;
for i *and* $j \in N$ **do**
  if $e_{ij} > e_{ji}$ **then**
    $e^*_{ij} = e_{ij} - e_{ji}$
    $E^* \leftarrow e^*_{ij}$
  end
  else
    $e^*_{ji} = e_{ji} - e_{ij}$
    $E^* \leftarrow e^*_{ji}$
  end
end

Algorithm 3: A bilateral compression algorithm

D Programming characterization and optimal algorithm

D.1 Programming formalization

Compression can be seen as the solution of a mathematical program which minimizes a non-decreasing function of gross notional under given net-positions. By introducing constraints on counterparty relationships, we will recover the hybrid and conservative compression.

In particular, let $E\prime$ denote the set of edges after compression and let $f : E\prime \rightarrow \mathbb{R}$ be a non decreasing function, the general compression problem is to find the optimal set $e\prime_{ij}$ in the following program:

Problem 1 (General compression problem)**.**

$$\begin{aligned} \min \quad & f(E') \\ s.t. \quad & \textstyle\sum_j \left(e'_{ij} - e'_{ji}\right) = v_i, \forall i \in V \quad \textit{[net position constraint]} \\ & a_{ij} \leq e'_{ij} \leq b_{ij}, \forall (i,j) \in E \quad \textit{[compression tolerances]} \end{aligned}$$

with $a_{ij} \in [0, \infty)$ and $b_{ij} \in [0, \infty]$. We will refer to E' as the vector of solutions of the problem.

Problem 1 maps all the compression types by translating the compression tolerances (counterparty constraints) and adopting a specific functional form for

f. As we are interested in reducing the total amount of notional, we will set $f(E') = \sum_{ij} e'_{ij}$. The non-conservative compression problem is obtained by setting $e_{ij} \in [0, \infty)$, as follows:

Problem 2 (Non-conservative compression problem)**.**

$$
\begin{aligned}
\min \quad & \textstyle\sum_{ij} e'_{ij} \\
s.t. \quad & \textstyle\sum_j \left(e'_{ij} - e'_{ji}\right) = v_i, \forall i \in N \\
& e'_{ij} \in [0, \infty), \forall (i,j) \in E
\end{aligned}
$$

In problem 2 the tolerances are set to the largest set possible. By further reducing these tolerances for the customer sets, we obtain the hybrid compression problem:

Problem 3 (Hybrid compression problem)**.**

$$
\begin{aligned}
\min \quad & \textstyle\sum_{ij} e'_{ij} \\
s.t. \quad & \textstyle\sum_j \left(e'_{ij} - e'_{ji}\right) = v_i, \forall i \in N \\
& e'_{ij} = e_{ij}, \forall (i,j) \in E^C \\
& e'_{ij} \in [0, \infty), \forall (i,j) \in E^D
\end{aligned}
$$

Last, by further restricting tolerances, we obtain the conservative compression problem:

Problem 4 (Conservative compression problem)**.**

$$
\begin{aligned}
\min \quad & \textstyle\sum_{ij} e'_{ij} \\
s.t. \quad & \textstyle\sum_j \left(e'_{ij} - e'_{ji}\right) = v_i, \forall i \in N \\
& 0 \leq e'_{ij} \leq e_{ij}, \forall (i,j) \in E
\end{aligned}
$$

All problems can be interpreted as standard linear programs, which can be solved in numerous ways. Above, we propose specific closed form solutions for the non-conservative compression problem. For the conservative and hybrid approaches, the general case where the network is not chain ordering proof, a global solution can be obtained via linear programming techniques. We analyze such approach below.

D.2 Optimal solution of the hybrid and conservative compression problem: the network simplex

If the compression problem aims for a plain constrained minimization of the total notional obligations in the system, the corresponding linear programs 2, 3 and 4 can be resolved in a number of different ways. We will hereby provide a brief explanation of a specific methodology, the *simplex algorithm* and its network implementation (the so-called "network simplex") since it provides ground for interesting interpretations in our context. The compression problems can be thought as finding a minimum cost flow in a network, where the costs of using each link is unitary.[21] In general, we refer the reader to Ahuja et al. (1993) for details on the simplex algorithm, its mathematical properties and the relative proofs.

As shown in the main text, conservative (and hybrid) compression revolves around reducing, or eliminating altogether, closed chains of intermediation (i.e., cycles) in the original market. In the simplex method, this is achieved by finding a *spanning tree* solution. A spanning tree of a market $G = (N, E)$ is a subset of the original market $G^T = (N, E^T)$, with $E^T \subseteq E$, where G^T is a directed acyclic graph that is also connected (i.e. eliminating the orientation, there exists a path between any two nodes). The minimum cost flow problem is then solved by exploring different spanning trees of the original market. Indeed, a key result of network flow theory is that there always exists a cycle-free solution that is also optimal and that it has an optimal spanning tree solution.[22] If the market resembles the customer-dealer structure discussed in the main text, the simplex thus leads to a spanning tree solution which only affects intra-dealer trades (i.e., the subnetwork where cycles lie).

The set of risk tolerances and the levels of individual net positions constitute the constraint space of the compression problem. Consider the optimal solution $e^{'}_{ij}$, and let us partition the set of edges after compression in the following two disjoint subsets: i) the set of free edges (where $0 < e^{'}_{ij} < e_{ij}$), and ii) the set of restricted edges (with $e^{'}_{ij} = \{0, e_{ij}\}$). The notions of "cycle free" and "spanning tree" solutions refer only to these edges. In the case of a spanning tree solution, then every edge not belonging to the tree must be restricted.

Let us now look at the key optimality condition for a given compression solution. First, let us split the edges of the compressed market into the ones in the spanning tree, T; those with zero value L; those that are fully saturated, i.e. $e^{'}_{ij} = e_{ij}$. As a result, let $E^{'} = \{T, L, U\}$ be the *spanning tree partition* of the compressed market. Elaborating on theorem (11.3 in Ahuja et al., 1993), this

[21]There are two main classes of algorithms to solve a minimum cost flow network problem. The first class aims at keeping feasible solutions while striving for optimality, whereas the second keeps optimality while striving for feasibility. The simplex relates to the former.

[22]Provided that the objective function is bounded from below across the constraint space.

structure is optimal if it is feasible (i.e. it satisfies the constraints) and the reduced costs for the arcs in the spanning tree are zero; the reduced costs for edges with zero value are nonnegative; the reduced costs for the set of fully saturated edges are negative. Reduced costs are computed as follows: for each node in the system, define a quantity π_i, which we term the "potential" of i; then the reduced cost of an edge is $c^{\pi}_{ij} = c_{ij} - (\pi_j - \pi_j)$, where c_{ij} is the cost (in the objective function) of one unit of notional between i and j which, since in our case compression aims at reducing total notional, is equal to one ($c_{ij} = 1, \forall(i, j)$).

Reduced costs can be interpreted as the amount of total notional that would increase in case the amount on that edge would increase. Naturally, a strictly negative reduced costs means that we can increase the notional value on that edge and reduce the total notional after compression. The simplex method is based on the above condition and moves along edges in the current tree solution that have negative reduced costs in order to find the optimal solution while maintaining a spanning tree structure.

More precisely, the simplex exploits the relationship between the spanning tree solutions and the set of bases of the feasible region. Given a market $G = (N; E)$, consider the $|N| \times |E|$ node-edge *incidence matrix* $\mathbf{Q}$, defined as follows. The rows Q are represented by V and the column by E. We index the links by the letter l:

$$q_{il} = \begin{cases} 1 & \text{if the } l\text{-th edge originated from i} \\ -1 & \text{if the } l\text{-th edge terminates in i} \\ 0 & \text{if the } l\text{-th does not include } i. \end{cases}$$

Now, let $\mathbf{e}$ be the vector of all edges, $\mathbf{e}'$ be the optimal solution of the problem, $\mathbf{v}$ the vector of the nodes' net positions, and $\mathbf{u}$ be the vector of all ones. Hence, Problem 4 can be rewritten in the following matrix form

$$\begin{aligned} \min \ & \mathbf{u}^\top \mathbf{e}' \\ \text{s.t. } & \mathbf{Q}\mathbf{e}' = \mathbf{v} \\ & \mathbf{0} \leq \mathbf{e}' \leq \mathbf{e} \end{aligned} \tag{4}$$

$\mathbf{Q}$ is not full rank, but since $\sum_i v_i = \mathbf{u}^\top \mathbf{v} = 0$, then the first set of constraints has one redundant row (that can be eliminated). The set of *bases* of Q are the matrices constituted by $|E| - 1$ linearly independent columns of $\mathbf{Q}$ and therefore each basis represents a subset of E. Each basis is associated to a unique solution of the linear system of equations 4.

In addition, it can be proven that, if the graph is connected (as in our case), then to each basis of $\mathbf{Q}$ indeed corresponds a spanning tree of G. This implies that, for each basis, the corresponding spanning tree will also satisfy constraint 4 and therefore will constitute a feasible solution. The space of basic solutions lies in the space of bases generated by the incidence matrix of the original network and such solutions are spanning trees. This is referred to as the *basis* property for the min-cost flow problem (Ahuja et al., 1993, Theorem 11.10). In the context of a compression problem this can be interpreted as a key correspondence between feasible solutions of the compression problem and spanning trees.

This constitutes the key ingredient of the application of the standard simplex algorithm to a network problem. By moving along the different spanning trees (bases), the network simplex method attempts at find an optimal feasible solution. In a network context, this means to find a basis and add one edge to the current spanning tree, which creates a cycle and eliminating, if possible, the other edges composing the newly created cycle. As observed above, if the inserted arc increases the objective function, then the arc has an associated positive reduced cost; if the arc decreases the objective function, then it has a negative reduced costs; if the arc does not alter the value of the objective function, then its reduced cost is zero.

Below we report the algorithmic expression of the simplex algorithm:

Data: Original Market $G = (N, E)$, set of risk tolerances.
Result: $G^{'}$ such that x' is minimized
begin
  start with an initial tree structure $E^T = (T, L, U)$;
  compute total notional x', reduced costs and node potentials;
  while *there exists some arc* $\notin E^T$ *that violates optimality conditions* **do**
    begin
      choose an edge (i, j) that violates condition ;
      add (i, j) to $E^{'}$ and select the leaving edge (k, l);
      update $E^{'}$, $x^{'}$ and node potentials
    end
  end
end

Algorithm 4: Illustration of the network simplex algorithm

Imprint and acknowlegements

We are grateful to Tuomas Peltonen for extensive comments and discussions at all stages. We also thank for comments and suggestions: Andre Sapir, Per Sjoberg, Mark Yallop, Mathias Dewatripont, Darrell Duffie, Rama Cont, Antoine Mandel, Jorge Abad, Inaki Aldasoro, Christoph Aymanns, Stefano Battiston, Bernard Fortz, Daniel Gros, Renaud Lambiotte, Steven Ongena, Marc Chesney, Ashkan Nikeghbali, Paulina Przewoska, Roberto Stok, Olaf Weeken, Benjamin Vandermarliere and participants to the ESRB Joint Expert Group on Interconnectedness meetings, De Nederlandsche Bank research seminar, the ESRB research seminar, the joint ESRB/BoF/RiskLab SRA conference, the LSE Systemic Risk Centre on Systemic Risk Modeling, the BENET 2016 conference, the 2017 AEA/ASSA Chicago meeting, the Macro Connection MIT Media Lab meeting, the NaXys seminar and the Imperial College Mathematical Finance seminars. The empirical part of this work was produced while the authors were visiting the European Systemic Risk Board. We therefore acknowledge the outstanding support from the ESRB Secretariat and the research programme on derivatives led by the Joint Expert Group on Interconnectedness of the ESRB, chaired by Laurent Clerc and Alberto Giovannini. The authors declare no relevant or material financial interests that relate to the research described in this paper. MD acknowledges support from: FET Project SIMPOL nr. 610704, FET project DOLFINS nr 640772, and the Swiss National Science Foundation grant no. PP00P1-144689.

Marco D'Errico
Department of Banking and Finance, University of Zurich; email: **marco.derrico@uzh.ch**

Tarik Roukny
Media Lab, Massachusetts Institute of Technology and Department of Economics, University of Ghent; email: **roukny@mit.edu**

Postal address	60640 Frankfurt am Main, Germany
Telephone	+49 69 1344 0
Website	**www.esrb.europa.eu**

Note:
The views expressed in ESRB Working Papers are those of the authors and do not necessarily reflect the official stance of the ESRB, its member institutions, or the institutions to which the authors are affiliated.

ISSN	2467-0677 (online)
ISBN	978-92-95081-93-2 (online)
DOI	10.2849/79183 (online)
EU catalogue No	DT-AD-17-011-EN-N (online)