Compressing over-the-counter markets*
Preliminary and incomplete draft

Marco D'Errico[†] Tarik Roukny[‡]

December 30, 2016

Abstract

In this paper, we show both theoretically and empirically that the size of over-the-counter (OTC) markets with outstanding trades can be reduced without affecting individual net positions. First, we find that the networked nature of OTC 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 while preserving individual net positions. We refer to this operation as "compression" and identify feasibility and efficiency criteria, highlighting intermediation as a key factor for excess levels. We show that a trade-off exists between the amount of notional that can be removed from the system and the conservation of trading relationships. Third, we apply our theoretical framework to a unique and comprehensive transaction-level dataset on OTC derivatives. We document large levels of excess across all markets and time. Finally, we show that compression when applied at the global level can reduce a considerable fraction of total notional even under conservative approaches.

*We are grateful to Tuomas Peltonen for extensive comments and conversations. We also thank for comments and suggestions: Andre Sapir, Jorge Abad, Inaki Aldasoro, Christoph Aymanns, Stefano Battiston, Bernard Fortz, Daniel Gros, Renaud Lambiotte, Sergio Masciantonio, Steven Ongena, Ashkan Nikeghbali, Paulina Przewoska, Roberto Stok, Olaf Weeken and participants to the ESRB Joint Expert Group on Interconnectedness meeting, 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 and the BENET 2016 conference. 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 program on derivatives led by the ESRB's interconnectedness expert group, 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.

[†]University of Zurich [marco.derrico@uzh.ch]

[‡]University of Ghent [tarikroukny@gmail.com]

Contents

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 a search and bargaining process. The decentralized feature of OTC markets makes them opaque as market information is often very limited to most agents. As a result of the search friction, dealers play the role of marketmakers and intermediate between buyers and sellers of a given good (Duffie et al., 2005). Several OTC markets have an important role in the economy (Duffie, 2012) and can be very large[1]. The size and lack of transparency of those markets has 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 gross notional of the market and the minimum aggregate amount satisfying 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 such 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 the excess. The main contribution of our paper is to provide a theoretical framework to understand and quantify the redundancy of trades leading to excess, propose methods to remove excess and investigate empirically the efficiency of each approach by applying the framework to a real, unique, transaction-level dataset on over-the-counter derivatives.

From an accounting perspective, a large excess in a market means 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). Let us illustrate the situation with the Figure 1 which maps the network of obligations of an actual OTC market for Credit-Default-Swap (CDS) contracts. Sellers of the CDS are on the left hand-side (green), buyers 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 market share of gross notional per set of market participants. The second line retrieves the average ratio between

[1]For example, 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 reporting to trade repository.

individual net positions and gross 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 only one fifth is explained by the net position. 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. For example, 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 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]. The methods we present in this paper follow the same principle.

Let us illustrate portfolio compression with the stylized example graphically shown in Figure 2(a) of a market consisting of 4 institutions (i , j, k, l) selling and buying the same contract 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 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: $5 - 20 = -15$. We can compress the market by removing the bilateral amount between i and j and reduce the obligations that both firms have with k by 5. The result is illustrated in Figure2(b) where the net position of each firm is equal to the situation while the gross notional of the market is reduced by 15: $x' = 5 + 15 + 10 = 30$.

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 trades outstanding between all counterparties[4]. Naturally, a certain amount of information disclosure is needed in order to run such process. Individual counter-

[3]Formally, the Markets in Financial Instrument Regulation (MiFIR) describes 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. [...] Portfolio compression aims at reducing non-market risks in existing derivatives portfolios without changes in the market risk of the portfolios. ”

[4]In the bilateral case where 2 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.

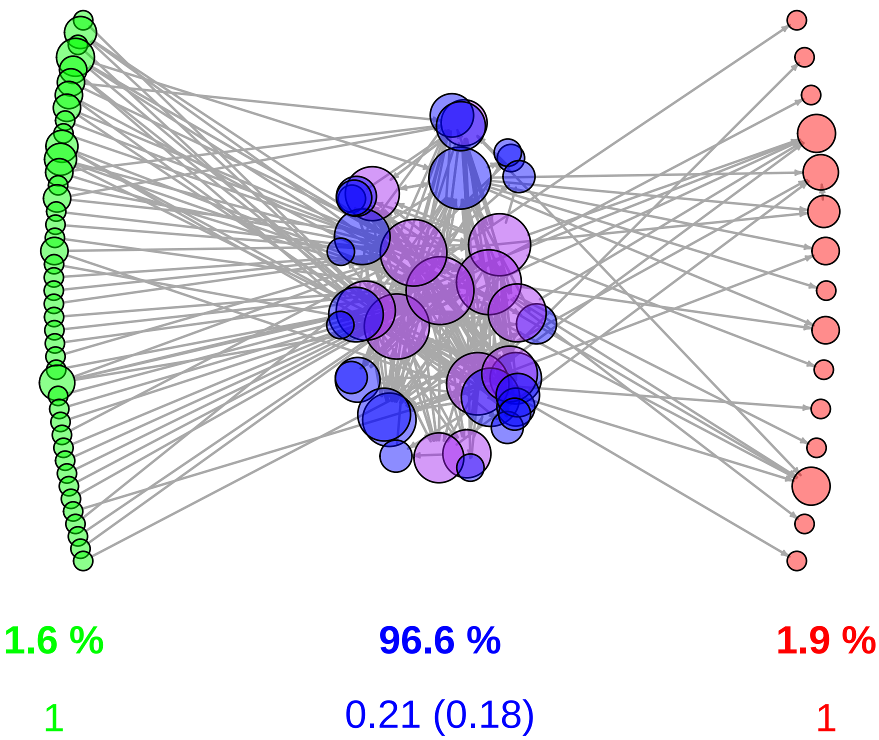
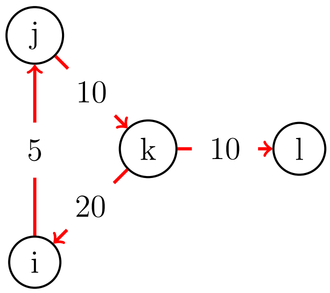
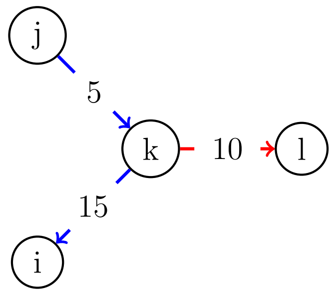


Figure 1: Illustration of an OTC market. All trades are outstanding and relate to Credit-default-swap contracts written on the same government reference entity for the month of April 2016. The data were collected under the EMIR framework and thus contain all trades where at least of counterparty is based in 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.



(a) Before Compression



(b) After Compression

Figure 2: A graphical example of compression

parties might not know the presence and amount of trades they are not involved in directly. In case institutions refuse 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 take care of the compression analysis. Such entity would recover the portfolio information from each market participant seeking to compress their position, 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 recent regulations have recently supported its adoption[6].

As advertised by compression providers[7], the increasing interest for compres-

[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). 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`).

[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: `\interfootnotelinepenaltyhttp://www.swapclear.com/Images/lchswapcompression.pdf`

sion results from many benefits at the market participant level. Overall, we can distinguish between three major incentives for institutions to engage in compression:

Reduction of counterparty risk As contracts are removed and replaced by new contracts with lower notional amounts, the counterparty risk deriving from the gross exposures to those trades is reduced.

Alleviating regulatory constraints Institutions like banks are subject to capital requirements computed on the basis of gross measurements[8]. Hence, reducing the notional amounts of contracts can help reducing the corresponding capital requirements of market participants.

Improvement of operational management Reducing the number of contracts leads to a reduction of operational risks and easier management: trade count reduction, speed to auction in case of default, less cash-flow needed to settle obligation, 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 in indeed relevant to ensure a proper design and implementation of compression in OTC markets. Furthermore, compression implies a modification of the web of outstanding trades which can affect the risk profiles of market participants and, in turn, the stability of the market as a whole. As monitoring changes in counterparty risk is paramount to both micro and macro-prudential regulation, the effects of compression should not be ignored. The current work seeks to fill that 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 trades, is the *raison d'être* of excess in markets. Dealers are thus at the heart of the generation of redundant trades. However, the level of excess that can be removed (i.e., redundant excess) is a function of potential constraints set by both individuals and regulators, i.e., *compression tolerances*. Hence, compression does not always remove the total amount of excess (i.e., there can be some residual excess after compression). We identify 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 before compression. We show that a trade-off exists

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

between the efficiency of a compression and the level of compression tolerance in the modification of trading relationships.

Finally, we use a unique data set comprising all credit-default-swaps transactions made by EU institutions between 2014 and 2016 in order to quantify the levels of excess exhibited by real OTC markets as well the efficiency of the above mentioned compression methods. We find that the vast majority of markets defined as the set of transactions written on the same reference entity with the same maturity (as illustrated by Figure 1) exhibit levels of excess accounting for more than 50% of the total gross notional. This finding is robust over time. Regarding compression, we show that even the most conservative method achieves more than 50% of excess removal for most markets. Relaxing compression constraints on the intra-dealer segment of markets yields significantly better efficiency (i.e., 70% of excess removal for most markets) with cases of more than 97% of excess removal.

Despite the application of our framework on derivatives markets, our findings and methods can similarly be applied to other over-the-counter markets. As long as a market exhibits fungibility, contingency and intermediation, our framework can help understanding the generation of excess and the possibility of reducing it. Hence markets like credit markets or bond markets are natural 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 and 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 Over-The-Counter (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. Appendices provide proofs of the propositions and lemmas as well as analytical details for the algorithms used in the paper.

2 Literature review

The study of OTC markets structures has gained attention in the last decade prompted by both their role in the 2008 financial crisis and the increase of data availability (Duffie, 2012).

Despite the use of different sets of data 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 intermediary and 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 (Shachar, 2012; Benos et al., 2013; Peltonen et al., 2014; 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 (Craig and Von Peter, 2014; van Lelyveld et al., 2014; Fricke and Lux, 2015).

Institutions acting as dealer are typically large banks (Craig and Von Peter, 2014). (Atkeson et al., 2015) propose a parsimonious theoretical model which generates the above described feature for derivatives markets: they show that banks that enter an OTC derivatives market for intermediation profit incentive 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 credit default swap 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 can be related to the notion of "hot potato" trades, a feature of OTC derivatives observed and modeled by (Burnham, 1991; Flood, 1994; Lyons, 1995, 1997).

Furthermore, those markets are characterized by the concentration of large amounts of notional within the intra-dealer activity. 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) shows that, using worldwide data on CDS, between 70% and 80% of the notional is in the intra-dealer market across reference entities. (Abad et al., 2016) report similar levels for Interest-Rates-Swaps (IRS) markets and Foreign-Exchange (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 OTC derivative markets are "mainly 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 above listed efforts 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).

An explorative study of compression is proposed by (O'Kane, 2014). 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 TriOptima (to the author's claim) which 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. (Benos et al., 2013) use CDS transaction data from the UK to show that monthly reduction breaks in dealers' gross positions are due to compression, indicating the frequency of compression cycles.

In spirit, this paper is also related with works on the "gridlock problem" in payments system and in particular the issues of minimal settlements. Rotemberg (2011) models a system of payment interconnected via due payments and identifies the conditions under which the market can be cleared with minimal endowments of liquid assets. This approach is relevant in the context of this paper 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. (Hayakawa, 2015) expands the model to characterize further the lower and upper bounds of required settlement funds.

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 (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 (Allen and Babus, 2008; 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; Brunnermeier et al., 2013)

However, in this work, we do not build explicit links between compression and systemic risk. Our framework rather helps understanding how to manage counterparty risk reduction, collateral demand and capital requirements in post-trade situations. The focus we take is rather on providing a first comprehensive framework to understand the mechanics underlying such reduction: future work will tackle more systemic risk questions in a more explicit way.

3 The market

We consider an Over-The-Counter (OTC) market made of n market participants (institutions) indexed by $N = \{1, 2, ..., n\}$. 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:*

$$v_i^{gross} = \sum_j e_{ij} + \sum_j e_{ji} = \sum_j (e_{ij} + e_{ji}) \tag{1}$$

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:*

$$v_i^{net} = \sum_j e_{ij} - \sum_j e_{ji} = \sum_j (e_{ij} - e_{ji}) \tag{2}$$

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 is given by:*

$$x = \sum_i \sum_j e_{ij} \tag{3}$$

Furthermore, we 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. They are thus 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:

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:*

$$\delta(i) = \begin{cases} 1 & \text{if } \sum e_{ij}.\sum e_{ji} > 0 \quad \textbf{(dealer)} \\ 0 & \text{otherwise} \qquad \textbf{(customer)} \end{cases} \tag{4}$$

In a sense, we generalize the modeling approach of (Atkeson et al., 2015) with regards to market participant types. Note that, as a results, only 3 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 a 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*

$$N = N' \tag{5}$$

and

$$v_i^{net} = v_i'^{net} \qquad \forall i \in N \tag{6}$$

where v_i^{net} and $v_i'^{net}$ 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

$$x' = \frac{1}{2}\sum_{i=1}^{n} |v_i^{net}| = \sum_{i:\ v_i^{net}>0}^{n} v_i^{net} \tag{7}$$

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*

$$\begin{aligned}\Delta(G) &= x - x' && (8)\\ &= \left(\sum_{i=1}^{n}\sum_{j=1}^{n} e_{ij} - \frac{1}{2}\sum_{i=1}^{n} |v_i^{net}|\right) && (9)\\ &= \left(\sum_{i=1}^{n}\sum_{j=1}^{n} e_{ij} - \sum_{i:\ v_i^{net}>0}^{n} v_i^{net}\right) && (10)\end{aligned}$$

Note that Equation (8) and Equation (9) 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 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:*

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

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 centralised echange-traded markets) have different levels of notional excess,

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

Centralised 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 $v_i^{net} = v_i^{gross}, \ \ \forall i \in N$.

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

Even if some over-the-counter 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 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*

$$\delta(i).\delta(j) = 1 \quad \forall e_{ij} \in E^D \tag{11}$$

$$\delta(i).\delta(j) = 0 \quad \forall e_{ij} \in E^C \tag{12}$$

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 states 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*

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

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 results informs us that centralised 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[9].

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).

[9]In the context of clients to a compression service provider, compression tolerances determine how much the clients are willing to stick to their original positions. In derivatives markets, service providers such as TripOptima call those constraints *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.

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

$$c_s(N, E) \succ c_t(N, E) \Leftrightarrow \Delta^s_{red}(G) > \Delta^t_{red}(G) \tag{13}$$

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.:

$$c_s(N, E) \succ c_t(N, E) \Leftrightarrow \rho_s > \rho_t \tag{14}$$

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 heterogenous 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 benchmark 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 benchmark are informative on the role of 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:*

$$a_{ij} = 0 \quad and \quad b_{ij} = +\infty, \quad \forall (a_{ij}, b_{ij}) \in \Gamma, \tag{15}$$

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

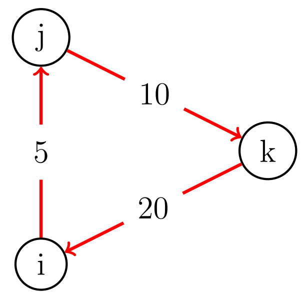


Figure 3: Original configuration the market

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:*

$$a_{ij} = 0 \quad \textit{and} \quad b_{ij} = e_{ij}, \quad \forall (a_{ij}, b_{ij}) \in \Gamma, e_{ij} \in E \tag{16}$$

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).

4.5.3 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 3 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 4(a). Computing the same measurements as before, we obtain:

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

We also obtain the new excess in the market:

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

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.

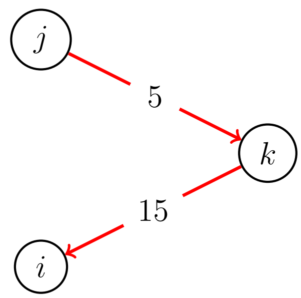
Let us now go back to the initial situation of Figure 3 and adopt a non-conservative approach. We can now create, if needed, new trades. A non-conservative solution is to remove all trades and create 2 new trades: one going from j to i of value 5 and one going from k to i of value 10. We have created a contract that did not exist before between j and i. The resulting market is depicted in Figure 4(b). Computing the same measurements as before, we obtain:

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

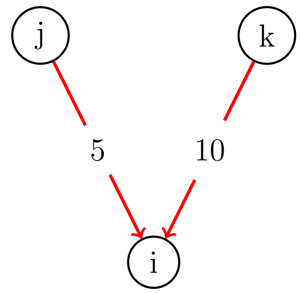
We also obtain the current excess in the market:

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

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.



(a) After conservative compression



(b) After non-conservative compression

Figure 4: Examples of conservative and non-conservative compression approaches.

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

Table 1: Table summarizing the results applying conservative and non-conservative compression on a market with 3 participants.

The results are summarized in Table 1. 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.

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 individuals do. 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:*

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

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 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 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,

$$C = \{c | G' = c(G) : \Delta^c_{res}(G) = 0\} \neq \{\}$$

i.i.f.

$$\lambda \geq \frac{|v^{net}_{i^*}|}{|N^{-1.sign(v^{net}_{i^*})}|}$$

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:

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

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:*

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

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)[10].

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

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.

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

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 preceeding 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:*

$$\Delta_{res}^{c^c}(G) = 0 \tag{17}$$

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:*

$$\Delta_{res}^{c^c}(G) > 0$$

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.

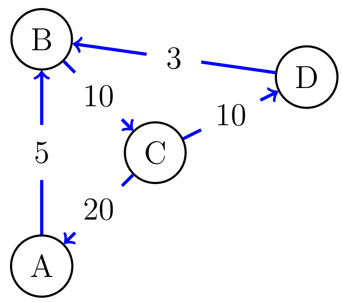


Figure 5: Example of market with entangled chains

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:*

$$E_1 \cap E_2 \neq \{\}$$

An illustration of entangled chains is provided in Figure 5 where the edge BC is share by two chains of intermediation (i.e., ABC and BCD)

Let us now formulate the following assumption on the graph:

Assumption 1. *(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 Assumption, 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

$$\Delta_{res}(G) < \Delta(G) - \sum_{K_i \in \Pi} \Phi(E_{K_i}) \tag{18}$$

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[11].

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 2. *Dealers prefer to keep their intermediation role with customers*

Assumption 3. *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

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

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:*

$$a_{ij} \geq 0 \quad and \quad b_{ij} = e_{ij}, \quad \forall (a_{ij}, b_{ij}) \in \Gamma, e_{ij} \in E^C \tag{19}$$

$$a_{ij} \geq 0 \quad and \quad b_{ij} = +\infty, \quad \forall (a_{ij}, b_{ij}) \in \Gamma, e_{ij} \in E^D \tag{20}$$

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 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

$$\Delta^{c^h}_{res}(N, E) = \Delta(N, E^C) \tag{21}$$

As a result, we see that, in case where 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).

5 Empirical application

5.1 Dataset description

In this Section, we apply the theoretical framework developed in this paper to analyse a unique transaction-level dataset for Credit Default Swaps (CDS) derivatives[12]. The dataset covers all CDS transactions in which at least one counterparty is based in the European Union[13].

There are multiple selection options to aggregate the data. 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 counterparties, the 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 bond 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. 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 case in the following. At the participant level, we select participant 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. Once more, for sake of simplicity and consistency, we do not consider such case in the following. In our framework, a market $G_k = (N_k, E_k)$ is thus defined as the set of participants and outstanding CDS transactions related to a specific reference entity k.

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 references 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 (including large manufacturing groups).

[12]Credit-default-swap contracts are the most used types of credit derivatvies. If 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 Financial Crisis of 2007-2009. For more information, see (Stulz, 2010)

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

5.2 General statistics

An overview of the main statistics is reported in Table 2. In particular, the total notional of the selected 100 entities varies between 210Bn euro and 250Bn euros retaining roughly $30-34\%$ of the total gross notional. The average number of counterparties across the 100 entities is stable and varies between 47 and 59 individual counterparties. The average number of links is also stable and varies between 109 and 129 bilateral outstanting contracts.

Time	Gross notional of 100 top ref. (euros)	Share of gross notional among all ref.	Avg num. of counterparties	Avg num. of contracts
Oct-14	2.1E+11	0.336	55	116.41
Nov-14	2.2E+11	0.326	57	122.10
Dec-14	2.3E+11	0.329	59	128.74
Jan-15	2.4E+11	0.331	59	128.60
Feb-15	2.4E+11	0.329	58	126.91
Mar-15	2.4E+11	0.333	53	115.48
Apr-15	2.1E+11	0.320	47	106.99
May-15	2.2E+11	0.320	47	106.98
Jun-15	2.1E+11	0.323	47	105.99
Jul-15	2.2E+11	0.325	49	107.98
Aug-15	2.1E+11	0.321	52	112.42
Sep-15	2.2E+11	0.328	54	115.46
Oct-15	2.3E+11	0.328	55	118.46
Nov-15	2.4E+11	0.330	57	121.82
Dec-15	2.5E+11	0.321	56	121.60
Jan-16	2.5E+11	0.322	56	121.97
Feb-16	2.5E+11	0.320	56	121.97
Mar-16	2.2E+11	0.312	51	109.04
Apr-16	2.3E+11	0.309	51	110.42

Table 2: General statistics of the dataset.

Table 3 provides further statistics on the market segments (i.e, intra-dealer and customer markets). 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 105 and 130 contracts. Taken as a whole, markets are quite sparse with a density of contracts around 0.15. This means that, on average, only 15% of all possible bilateral contracts are actually present. Interestingly, this measure is almost three times bigger when we only consider the intra-dealer market where the density can reach 0.45. 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 between 73% and 78%. Finally, the last column of Table 3 confirms the very low frequency of customer-customer trades: on average, less then 3% of all contracts are written without a dealer on either side of the trade.

5.3 Quantifying excess and the efficiency of compression

After the general analysis, let us now move to the quantification of the measures introduced in this paper. We start by measuring the level of excess present in the markets at hand as a function of the total gross notional (i.e., $\epsilon(G) = \frac{\Delta(G)}{x}$). Table 4 reports the statistics of excess levels of 6 snapshots equally spread between Oct 2014 and April 16 including minimum, maximum, mean, standard deviation and quartiles. 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 0.806. 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 in the most traded CDS markets (at least by EU institutions) is in excess vis-a-vis market participants' net position.

As non-conservative compression always leads to 0 residual excess, the empirical analysis of the efficiency is trivial: it is equal in every part to the results of Table 4. In addition, this table also provides us with the upper efficiency limit of any compression[14].

In the case of the conservative and hybrid compressions, the results are not trivial and require the implementation of specific algorithms (see the Appendix for a description of each implemented solution). After having implemented each compression algorithm on each market (i.e., for each snapshot, we run the algorithms on each 100 different markets), we compute the efficiency of the compression as a fraction of the total level of excess (reported in Table 4):

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

[14]Note 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

Time	Avg num. dealers	Avg num. customers buying	Avg num. customers selling	Avg num. contracts	Fraction of intra-dealer notional	Total density	Intra-dealer density	Share of intra-customer contracts
Oct-14	18.1	15.57	19.88	116.54	0.763	0.151	0.443	0.003
Nov-14	18.51	16.1	20.7	122.17	0.777	0.155	0.453	0.001
Dec-14	19.22	17.12	21.44	128.76	0.775	0.153	0.443	0.001
Jan-15	19.34	17	21.1	128.63	0.774	0.151	0.438	0.001
Feb-15	19.16	16.8	21.06	126.88	0.779	0.151	0.442	0.001
Mar-15	18.46	15.16	17.44	115.56	0.783	0.157	0.45	0.001
Apr-15	18.03	12.68	14.87	107	0.783	0.158	0.458	0.001
May-15	18.16	12.49	14.56	107.02	0.779	0.156	0.448	0.002
Jun-15	18.36	13.2	14.12	106.05	0.779	0.153	0.433	0.002
Jul-15	18.51	14.32	14.36	107.99	0.763	0.146	0.423	0.002
Aug-15	18.78	14.59	16.79	112.37	0.772	0.147	0.415	0.002
Sep-15	18.95	15.58	17.29	115.72	0.751	0.143	0.409	0.003
Oct-15	19.28	16.06	17.62	118.3	0.762	0.145	0.4	0.002
Nov-15	19.31	16.94	18.76	121.66	0.757	0.145	0.397	0.003
Dec-15	19.5	16.71	18.57	121.62	0.768	0.145	0.396	0.002
Jan-16	19.52	16.88	18.13	121.97	0.770	0.144	0.396	0.002
Feb-16	19.42	17.02	18.08	122	0.758	0.144	0.394	0.002
Mar-16	18.24	14.17	16.64	109.29	0.734	0.142	0.411	0.004
Apr-16	18.62	13.71	16.68	110.62	0.755	0.144	0.409	0.004

Table 3: Dealers/customers statistics.

Total Excess	Oct-14	Jan-15	Apr-15	Jul-15	Oct-15	Jan-16	Apr-16
min	0.422	0.416	0.290	0.262	0.366	0.261	0.292
max	0.811	0.811	0.798	0.820	0.818	0.806	0.781
mean	0.612	0.618	0.612	0.597	0.595	0.568	0.556
stdev	0.087	0.088	0.091	0.096	0.098	0.112	0.098
first quart.	0.561	0.555	0.555	0.527	0.528	0.486	0.499
median	0.617	0.615	0.612	0.610	0.593	0.566	0.565
third quart.	0.665	0.684	0.672	0.661	0.653	0.650	0.635

Table 4: Excess statistics

Table 5 reports the results. On the extremes, both the conservative and the hybrid compression perform with various degrees of efficiency: the minimum amount of excess that 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 compress reaches 90% of excess removal shows the possibility of having very efficient compress despite low 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 the excess from the market, the hybrid compression allowing for greater performances as a result of constraints alleviation (i.e., increase of intra-dealer compression tolerances).

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:

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

Conservative (ρ_c)	Oct-14	Jan-15	Apr-15	Jul-15	Oct-15	Jan-16	Apr-16
min	0.160	0.203	0.141	0.156	0.181	0.119	0.098
max	0.894	0.927	0.924	0.887	0.912	0.911	0.878
mean	0.575	0.631	0.606	0.587	0.566	0.563	0.538
stdev	0.165	0.159	0.161	0.155	0.171	0.178	0.164
first quart.	0.461	0.514	0.525	0.470	0.452	0.444	0.424
median	0.573	0.642	0.604	0.604	0.542	0.556	0.552
third quart.	0.691	0.731	0.737	0.705	0.688	0.690	0.648

Hybrid (ρ_h)	Oct-14	Jan-15	Apr-15	Jul-15	Oct-15	Jan-16	Apr-16
min	0.370	0.460	0.377	0.288	0.266	0.296	0.272
max	0.971	0.986	0.968	0.967	0.977	0.975	0.981
mean	0.731	0.771	0.768	0.754	0.749	0.745	0.764
stdev	0.149	0.129	0.128	0.128	0.142	0.134	0.138
first quart.	0.626	0.692	0.689	0.676	0.647	0.658	0.693
median	0.747	0.794	0.786	0.760	0.780	0.757	0.795
third quart.	0.848	0.869	0.860	0.852	0.856	0.855	0.853

Table 5: Statistics of compression efficiency

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

$$\text{Non-conservative:} \quad \epsilon_n = \sum_{k=1}^{5} \big(w_k \ \epsilon^k(G) \big)$$

$$\text{Hybrid:} \quad \epsilon_h = \sum_{k=1}^{5} \big(w_k \ \rho_h^k \times \epsilon(G) \big)$$

$$\text{Conservative:} \quad \epsilon_c = \sum_{k=1}^{5} \big(w_k \ \rho_c^k \times \epsilon(G) \big)$$

Those ratios can be easily interpreted as the fraction of notional that can be eliminated over all five entities taken individually. Results for these weighted averages are reported in Figure 6. The circled series highligted by the light blue shade represent the weighted non-conservative compression ratio ϵ_n (which coincides with the weighted level of excess); the triangle points represent the weighted hybrid compression ratio ϵ_h; the squared points 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 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.

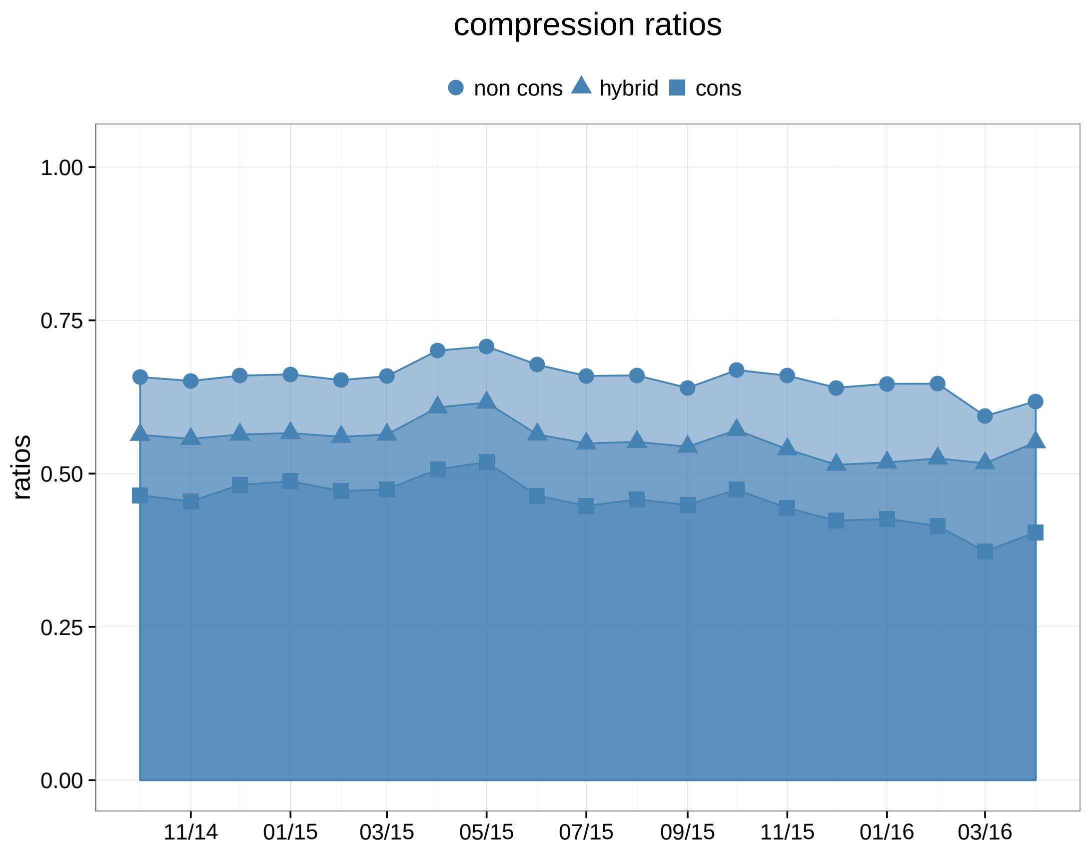


Figure 6: Weighted averages of the non-conservative, hybrid and conservative compression ratios for the top-5 reference entities in our sample.

6 Closing remarks

In this work, we have shown that Over-The-Counter (OTC) markets with contingent trades generate gross volumes that can far exceed the amounts satisfying every market participants' net position. We call this difference *excess*. We show that the activity of dealers acting as intermediary on all transactions is the main factor determining the level of excess in a market. In turn, this excess can be removed while keeping net positions satisfied via a network operation called *compression*. To the best of our knowledge, this work is the first to propose a framework to analyze the *mechanics* of compression (i.e., feasibility and efficiency conditions) highlighting a trade-off between the fraction of excess that can be eliminated from the market and the degree to which counterparties restrain the procedure to pre-existing trading exposures. Furthermore, we report an empirical investigation to quantify the levels of excess and the efficiency of different compression procedures in real OTC markets. We find large levels of excess, important concentration on the intra-dealer activity and great efficiency of compression when relaxing constraints on intra-dealer activities.

The framework is easily adaptable to novelties due to the quick expansion of this technique and the increasing adoption by market participants. An extension of the model would account for multiple networks of exposures, as OTC market participants are often involved in several product classes (Duffie and Zhu, 2011; D'Errico et al., 2016). A natural question, in a similar spirit of (Duffie and Zhu, 2011) is whether, and to what extent, compression can take place on different OTC classes.

Along these lines, as OTC trades are increasingly called for standardization and mandatory central clearing, it is worth investigating the role of Central Clearing Counterparties (CCPs) in determining excess and efficiency of compression. Indeed, CCPs would be a peculiar node in our framework as their net position is always zero. Such analysis is straightforward using our framework.

In our work, we have assumed the perspective of social planner, possessing information on the entire set of trades. In practice gathering information is often imperfect mainly because (i) not all market participants are seeking compression or (ii) compression seekers do not disclose their whole portfolio. Compression providers must therefore operate with a limited set of information, which would naturally reduce the efficiency set from the point of view of a social planner. As such, an analysis on the effects of partial information on the feasibility and efficiency of compression.

Finally, a fundamental item for discussion is the role of compression on systemic risk. As information emerges about the practice and empirical work is done in order to fully characterize its implementation, regulators and policymakers will be endowed with increasingly better tools to understand the implication of this post-trade activity on systemic risk. There are three main stream of this type of research we foresee.

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 (Roukny et al., 2016) or as giving rise to frictions such as payments gridlocks (Rotemberg, 2011). Given these premises, compression seems to have a beneficial effect in mitigating these effects.

However, the elimination of linkage naturally leads to reshaping the underlying network, and potentially concentrates exposures into fewer counterparties. This will lead to changes in counterparty risk, collateral exchange, etc. In particular, as compression cycles become more and more frequent, participants in the OTC market may internalize the arrival of a compression cycle and undercollateralized exposures while expecting their exposures to be reduced. This problem may be further exacerbated if the cycle is unsuccessful and many trades still outstanding: this operational risk can be further increased by volatile market conditions.

Furthermore, the impact of compression on capital is another important aspect. As OTC derivative exposures are computed in the Basel III leverage ratio, banks' capital must include those gross exposures. As compression reduces these 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 requirement frees up unused leverage and increases investments. On the other hand, reducing capital requirement affects the loss absorption capacity of market participants.

Finally, our framework does not pertain uniquely to contingent claims arising from financial institutions. It may also find application also for non-financial firms, thereby allowing to quantify the levels of excess and potential compression in the real economy. Compression may reduce the total amount of outstanding debt due and liquidity needs. On the other hand, understanding the amount of excess in the financial and non-financial systems represents a complementary avenue of research: by identifying potential accumulations of excess, social planners may be able to set countercyclical policies aimed at reducing debts before the underlying network structures becomes critical for the propagation of distress (Kiyotaki et al., 1997).

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. *The American Economic Review*, 105(2):564–608.

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

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. (2008). Networks in finance.

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.

Brunnermeier, M., Clerc, L., El Omari, Y., Gabrieli, S., Kern, S., Memmel, C., Peltonen, T., Podlich, N., Scheicher, M., and Vuillemey, G. (2013). Esrb occasional paper series.

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.

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.

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

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

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

Ford, L. R. and Fulkerson, D. R. (1956). Maximal flow through a network. *Canadian journal of Mathematics*, 8(3):399–404.

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.

Hayakawa, H. (2015). Characterization of lower-bound and upper-bound of required settlement fund under real-time gross settlement. *Available at SSRN 2659975*.

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 (forthcoming)*.

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.

van Lelyveld, I. et al. (2014). Finding the core: Network structure in interbank markets. *Journal of Banking & Finance*, 49:27–40.

Yellen, J. (2013). Interconnectedness and systemic risk: Lessons from the financial crisis and policy implications. *Board of Governors of the Federal Reserve System, Washington, DC.*

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 7.

   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:

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

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

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

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

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

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:

$$\sum_{ij} e_{ij} - 0.5 \sum_i |\sum_j (e_{ij} - e_{ji})| = \sum_{ij} a_{ij} - 0.5 \sum_i |\sum_j (a_{ij} - a_{ji})| + \sum_{ij} b_{ij} - 0.5 \sum_i |\sum_j (b_{ij} - b_{ji}$$

$$-0.5 \sum_i |\sum_j (e_{ij} - e_{ji})| = -0.5 \sum_i |\sum_j (a_{ij} - a_{ji})| - 0.5 \sum_i |\sum_j (b_{ij} - b_{ji})|$$

$$\sum_i |\sum_j (e_{ij} - e_{ji})| = \sum_i |\sum_j (a_{ij} - a_{ji})| + \sum_i |\sum_j (b_{ij} - b_{ji})|$$

$$\sum_i |\sum_j (e_{ij} - e_{ji})| = \sum_i |\sum_j (a_{ij} - a_{ji})| + \sum_i |\sum_j (b_{ij} - b_{ji})|$$

$$\sum_i |\sum_j (a_{ij} + b_{ij} - a_{ji} - b_{ji})| = \sum_i |\sum_j (a_{ij} - a_{ji})| + \sum_i |\sum_j (b_{ij} - b_{ji})|$$

$$\sum_i |\sum_j (a_{ij} - a_{ji}) + \sum_j (b_{ij} - b_{ji})| = \sum_i \Big(|\sum_j (a_{ij} - a_{ji})| + |\sum_j (b_{ij} - b_{ji})| \Big)$$

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:

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

$$\sum_i |\sum_j (e_{ij} - e_{ji})| = \sum_i |\sum_j (e^D_{ij} - e^D_{ji})| + \sum_i |\sum_j (e^C_{ij} - e^C_{ji})|$$

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

1) For the whole network we have

$$\sum_i |\sum_j (e_{ij} - e_{ji})| = \sum_d^{dealer} |\sum_j (e_{dj} - e_{jd})| + \sum_c^{customer} |\sum_j (e_{cj} - e_{jc})| \tag{22}$$

$$= \sum_d^{dealer} |\sum_j (e_{dj} - e_{jd})| + \sum_{c^+}^{customer^+} |\sum_j (e_{c^+j} - e_{jc^+})| + \sum_{c^-}^{customer^-} |\sum_j (e_{c^-j} - e_{jc^-})| \tag{23}$$

$$= \sum_d^{dealer} |\sum_j (e_{dj} - e_{jd})| + \sum_{c^+}^{customer^+} |\sum_j (e_{c^+j})| + \sum_{c^-}^{customer^-} |\sum_j (-e_{jc^-})| \tag{24}$$

$$= \sum_d^{dealer} |\sum_j (e_{dj} - e_{jd})| + \sum_{c^+}^{customer^+} \sum_d^{dealer} e_{c^+d} + \sum_{c^-}^{customer^-} \sum_d^{dealer} e_{dc^-} \tag{25}$$

2) For the dealer network we have

$$\sum_i |\sum_j (e^D_{ij} - e^D_{ji})| = \sum_d^{dealer} |\sum_h^{dealer} (e^D_{dh} - e^D_{hd})| \tag{26}$$

3) For the customer network we have

$$\sum_i |\sum_j (e^C_{ij} - e^C_{ji})| = \sum_d^{dealer} |\sum_j (e^C_{dj} - e^C_{jd})| + \sum_{c^+}^{customer^+} |\sum_j (e^C_{c^+j} - e^C_{jc^+})| \sum_{c^-}^{customer^-} |\sum_j (e^C_{c^-j} - e^C_{jc^-})| \tag{27}$$

$$= \sum_d^{dealer} |\sum_j (e^C_{dj} - e^C_{jd})| + \sum_{c^+}^{customer^+} \sum_d^{dealer} e^C_{c^+d} + \sum_{c^-}^{customer^-} \sum_d^{dealer} e^C_{dc^-} \tag{28}$$

Combining equations, we obtain:

$$\sum_d^{dealer} |\sum_j^n (e_{dj} - e_{jd})| = \sum_d^{dealer} |\sum_h^{dealer} (e^D_{dh} - e^D_{hd})| + \sum_d^{dealer} |\sum_c^{customer} (e^C_{dc} - e^C_{cd})| \tag{29}$$

We continue decomposing the different elements.

1) For the whole network:

$$\sum_{d}^{dealer} |\sum_{j}^{n}(e_{dj} - e_{jd})| = \sum_{d}^{dealer} |\sum_{h}^{dealer}(e_{dh} - e_{hd}) + \sum_{c^+}^{customer^+}(e_{dc^+} - e_{c^+d}) + \sum_{c^-}^{customer^-}(e_{dc^-} - e_{c^-d})| \tag{30}$$

$$= \sum_{d}^{dealer} |\sum_{h}^{dealer}(e_{dh} - e_{hd}) + \sum_{c^+}^{customer^+} e_{dc^+} - \sum_{c^-}^{customer^-} e_{c^-d}| \tag{31}$$

2) for the dealer and customer networks:

$$\sum_{d}^{dealer} |\sum_{h}^{dealer}(e^D_{dh} - e^D_{hd})| + \sum_{d}^{dealer} |\sum_{c}^{customer}(e^C_{dc} - e^C_{cd})| = \tag{32}$$

$$\sum_{d}^{dealer} |\sum_{h}^{dealer}(e^D_{dh} - e^D_{hd})| + \sum_{d}^{dealer} |\sum_{c^+}^{customer^+} e^C_{dc^+} - \sum_{c^-}^{customer^-} e^C_{c^-d}| \tag{33}$$

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

$$\sum_{d}^{dealer} |\sum_{h}^{dealer}(e_{dh} - e_{hd}) + \sum_{c^+}^{customer^+} e_{dc^+} - \sum_{c^-}^{customer^-} e_{c^-d}| = \tag{34}$$

$$\sum_{d}^{dealer} |\sum_{h}^{dealer}(e_{dh} - e_{hd})| + \sum_{d}^{dealer} |\sum_{c^+}^{customer^+} e_{dc^+} - \sum_{c^-}^{customer^-} e_{c^-d}| \tag{35}$$

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 Propostion 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 Propostion 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 minimised 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 compression 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 $\varepsilon \in [0, \max_{ij}\{e_{ij} \text{ s.t. } (i,j) \in E\}]$ 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 Proposotion 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 Proposotion 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 $\Delta^{c^c}_{\text{res}}(G) = 0$ 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 $\Delta^{c^c}_{\text{res}}(G) = 0$.

For a closed chain of any length and heterogenous 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:

$$\Delta_{res}^{c^h}(N, E^C) = \Delta(N, E^C) \tag{36}$$

Finally, we obtain

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

■

B Compression Algorithms

B.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:
>
> $$e_{ij} = \begin{cases} min(v_n^i - \sum_{j'<j} e_{ij'}, v_n^j - \sum_{i'<i} e_{i'j}), & \text{if } v_n^i . v_n^j < 0 \\ 0, & \text{otherwise} \end{cases} \tag{37}$$
>
> 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

B.2 Conservative Algorithm

As we did for the 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 Programming characterization and solution

C.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' denote the set of edges after compression and let $f : E^{'} \to \mathbb{R}$ be a non decreasing function, the general compression problem is to find the optimal set $e^{'}_{ij}$ 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 seen as standard linear programs, which can be solved in numerous ways.[15] We have proposed 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 using linear programming. We analyze such approach below.

[15] In particular, the conservative compression can be mapped into a minimum-cost flow problem. Network flow problem are optimization problems that exploit the specific network-nature of the problem. A number of algorithms have been proposed since the 1950s to solve this type of problem, starting from the well-known Ford and Fulkerson (1956) algorithm (see Ahuja et al. (1993) for a comprehensive reference on network flows). Two main classes of algorithms have been developed in order 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.

C.2 A general conservative compression with global solution

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:

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

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

$$\begin{aligned} \min \ & \mathbf{u}^{\top}\mathbf{e}^{'} \\ \text{s.t. } & \mathbf{Q}\mathbf{e}^{'} = \mathbf{v} \\ & \mathbf{0} \leq \mathbf{e}^{'} \leq \mathbf{e} \end{aligned} \tag{38}$$

$\mathbf{Q}$ 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 38.

In addition, a crucial theorem in flow-network theory is that, if the graph is connected (as in our case), then to each basis of $\mathbf{Q}$ corresponds a spanning tree of G. That implies that for each basis, the corresponding spanning tree will also satisfy 38 and therefore be a feasible solution. The space of basic solutions lies in the space of bases generated by the so-called incidence matrix of the original network and such solutions are spanning trees..

In particular, we find that the set of basic solutions $e^{'}_{ij}$ has specific graph-topological properties for each of the set of dealers and customers. Define the set $E^{'}_0 = \{e^{'}_{ij} \text{ such that } e^{'}_{ij} = 0\}$, the set $E^{'}_1 = \{e^{'}_{ij} \text{ such that } e^{'}_{ij} = e_{ij}\}$ and the set $E^{'}_2 = \{e^{'}_{ij} \text{ such that } 0 < e^{'}_{ij} < e_{ij}\}$. The three sets are disjoint by construction and the the customer sets belong to $E^{'}_1$ because of 38. The spanning tree associated with a basic solution will be formed by $E^{'}_1 \cup E^{'}_2$

This is of crucial importance as it constitutes the key ingredient of the application of the standard simplex algorithm to a network problem. By moving along different basis, the simplex method finds an optimal feasible solution. In a network context, this means to find a basis and add one edge, which creates a cycle and eliminating, if possible, the other edges composing the newly created cycle. The simplex algorithm exploits this graph properties to move along the space of feasible solutions and finding the optimal. In general, we use the simplex algorithm to solve our compression problems. It is important to notice that the simplex method is only one of the many ways of solving the linear program. However, we discuss it here as it gives interesting network-theoretical interpretations. First, an important properties relates to the fact that the basis of the incidence matrix of a graph corresponds to a spanning tree. In this sense, any basic solution of the problem is then a spanning tree. The main idea of the simplex method is to move across the space of spanning trees (i.e. basic solutions) by repeatedly adding an edge that is not in the tree (the basic solution). This will modify the basis and therefore lead to a potential new solution.

If the inserted arc increasing 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. The way this is done in practice is based on the following pivoting approach:

Data: Original Market G=(N,E)
Result: $G^{'}$ such that $\Delta(G^{'}) < \Delta_v(G)$ is minimised and
$\mathbf{Qe}^{'} = \mathbf{v}, \mathbf{0} \leq \mathbf{e}^{'} \leq \mathbf{e}$
Let $G^{'} = G$;
Let $\{E_0^{'}, E_1^{'}, E_2^{'}\}$ be a set partition for a feasible solution;
Let c_{ij} be the reduced costs;
Compute node potentials such that $c_{ij} = \pi_i - \pi_j$;
while $\exists(i,j) : c_{ij} < 0$, (i,j) *not in the base* **do**
    Select (i,j) s.t. $c_{ij} < 0$;
    Find the directed cycle generated by (i,j);
    Add to the cycle an amount of notional at max equal to the residual capacity;
    Recompute $\{E_0^{'}, E_1^{'}, E_2^{'}\}$, potentials and reduced costs.
end

Algorithm 3: Illustration of the network simplex algorithm