May 2017

Modeling Over-The-Counter Derivative Trading with and without Central Clearing Parties

Natalie Stohlman Wellen
Worcester Polytechnic Institute

Follow this and additional works at: https://digitalcommons.wpi.edu/mqp-all

Repository Citation

This Unrestricted is brought to you for free and open access by the Major Qualifying Projects at Digital WPI. It has been accepted for inclusion in Major Qualifying Projects (All Years) by an authorized administrator of Digital WPI. For more information, please contact digitalwpi@wpi.edu.
Modeling Over-The-Counter Derivative Trading with and without Central Clearing Parties

Natalie Wellen

May 1, 2017

Abstract

Over-the-counter (OTC) derivatives are believed to have played a significant role in the 2008 financial crisis. In an effort to prevent a similar crisis in the future, the United States Federal Government signed new regulations into law in 2010 that require OTC derivatives to be traded through Central Clearing Parties (CCPs). Theoretically, these regulations should induce a greater amount of stability in the system, though in reality the truth of this statement is unclear. The model we propose in this paper seeks to elucidate the issue by utilizing probabilistic methods, including Gibbs Sampling, to sample different possible OTC derivative trading networks based on the small amount of data available to a regulator. These networks can then be tested with systemic risk analysis to better understand the stability of each type of network: a bilateral trading network without a CCP, a network with one CCP, and one with multiple CCPs. Then, using netted exposures to estimate the amount of risk each institution faces, each type of network is analyzed to better understand if CCPs accomplish the goal of creating a more stable OTC derivative trading network.
## Contents

1 Introduction ........................................ 3
2 Networks ........................................ 8
3 Modeling Overview ................................ 11
4 Bilateral Trading Model ............................ 13
   4.1 Estimating Assets and Liabilities .......... 14
   4.2 Forms of Network Models ..................... 17
   4.3 Initial Feasible Network ....................... 18
   4.4 Gibbs Sampling ................................ 20
   4.5 Modeling Exposures ............................ 22
5 Model with a Single Central Clearing Party .... 24
6 Model with Multiple Central Clearing Parties . 25
7 Netting Efficiency ................................ 26
8 Model Analysis ................................... 28
9 Conclusions ....................................... 43
10 Future Work ..................................... 44
11 Acknowledgements ............................... 45
1 Introduction

In 2007-2008 the world experienced one of the worst financial crashes since the beginning of the Great Depression in 1929 [1]. The burst of the housing bubble caused mortgage backed securities to plummet in value, leading to a chain reaction throughout the global financial sector. First Lehman brothers collapsed in September 2008. This created huge losses for AIG, which would have gone bankrupt if it were not for government intervention that came in the form of substantial bailouts. As of January 30th, 2017, the United States Government has spent a total of $623 billion dollars on the bailouts [2].

Many economists agree that the financial collapse is related to the use of over-the-counter (OTC) derivatives, a type of financial contract. Derivatives allow a financial firm to increase the amount of and/or the likelihood of a profit it may receive. When a derivative is agreed upon between two businesses without being exchanged in a marketplace, it is called an OTC derivative. Derivatives can also contribute to a financial firm owing catastrophic amounts of money. This is because an OTC derivative can be used to either hedge a risk, meaning the chances of a loss are reduced, or to bet on an outcome in a financial market. Before the 2008 crash, the financial sector did not believe that mortgages could default in large groups at a time, which caused many OTC derivatives to be written that bet mortgages would not be defaulted on in large groups. Further, since OTC derivative trades are only limited by the willingness of two firms to trade, not something like how quickly houses can be built, an increasing amount of money was invested in derivative contracts. The financial firms, with the support of risk calculating formulas, believed these bets were certain to work out in their favor [3].

In general, a derivative is a financial instrument that determines its value from something else, such as a loan or stock. It is important to note that derivatives do not create value, such as building a car, but they redistribute how money is spent. There are many different types of derivatives, and OTC derivatives can take infinitely many different forms. For example, one type of derivative is called a forward. A forward requires one party to buy something, called the underlying, at a specific time in the future. Forwards are often sold with commodities as the underlying value, for example a forward may be on soy. The agreement could be that the buying counterparty purchases 1000 bushels of soy for $10 a bushel in three years from now.

The gross notional value of a derivative is the current value of the underlying in the derivative contract. Since the current price of soybeans is $9.78,
in the above example the gross notional amount of the contract is $9780. However this is not the most important value when estimating how much a trader should be prepared to lose in the trade. If we think of the contract as being settled in cash, rather than soybeans, then if the price of soybeans is lower than the contractual amount, the seller of the forward gains the price difference, and the buyer must pay it. If the price is higher, then the seller pays the buyer instead. The amount that one party owes the other if the contract expired today is called exposure. So in the case of the soybean forward, the exposure is $10,000 - $9,780 = $1,220. Most often, exposure amounts are much smaller than the gross notional amounts for a derivative. They are more closely related to the loss a counterparty may experience from a derivative contract. In this way, if a business accrues too much exposure, the likelihood that it goes bankrupt from an unexpected turn in the market increases. This is part of what occurred in 2008.

An important feature of derivatives is their provision of an accessible and more liquid way to spread risk. Liquidity means that an asset can be traded very quickly without loss in value. Cash, for example, is a completely liquid asset.

In many ways, derivatives can behave like an insurance policy, however they do not require a large number of entities to hold similar positions like insurance does to allow for mutualization. A major difference from insurance though, is that one need not own what is being “insured.” This allows flexible hedging and a liquid market, but it also allows derivatives to be used for betting. Credit Default Swaps (CDSs) are a type of derivative that can illustrate this. A CDS is exchanged with another party’s debt as the underlying value, where the seller offers to pay the buyer of the CDS if the owner of the underlying defaults on their debt. In return, the buyer of the CDS pays the seller fixed amounts until the expiration date of the contract or a loan default occurs. When a default occurs, then the seller will usually pay the buyer the amount left to be paid on the underlying loan. In 2007, AIG had sold many CDS contracts with Lehman Brothers’ debt as the underlying value. This meant that when the housing bubble burst and Lehman Brothers collapsed, AIG owed the amount left unpaid on the mortgages that made up much of Lehman Brothers’ debt, and they owed this amount to each party that had traded a CDS with them.

Financial firms need some way to protect themselves against the risks associated with trading derivatives. They do this by requiring their counterparty to post collateral. Collateral is cash or another liquid financial asset
that is given to the lending party in case of a default. The problem with most investments is that they are not very liquid, so that selling the investment more slowly could allow the firm to gain much more than trading quickly. Because cash is a completely liquid financial instrument, it performs well as collateral. The amount of collateral necessary for a contract is found with calculations based on probabilities of a default in the contract. The goal is that a large enough fraction of probable losses, which are the exposures, are covered by the collateral a certain percentage of the time. Almost always, the collateral can be rehypothecated, which means it can be reused as collateral by the bank receiving it. The largest problem is that collateral is considered very costly. This is because there is a finite amount of liquid assets available to a financial firm, and they do not want to use all of it as investment collateral. Partially for this reason, financial firms were not prepared with enough collateral in 2008 during the crash.

OTC derivatives that bet on mortgages being repaid were once believed to be a sure money maker, which is partly due to mortgage defaults being viewed as independent events. It was believed that the default on one house loan would not affect the chances of default on another house, especially in different cities [4]. This means that the correlation between the two defaults is assumed to be very low, where correlation describes how well two things can predict one another. In this case, correlation can be thought of as describing underlying economic trends that affect housing prices across the United States. The belief at the time was that housing markets were only locally correlated, so that houses in the same neighborhood are more likely to default than ones in different neighborhoods. The low correlation assumption is part of why large pools of loans from across the United States were considered a sure investment. In 2008, it became clear this thinking was incorrect.

Another reason that derivatives based on mortgages were believed to be a safe investment has to do with the Gaussian Copula formula. This formula can find the risk associated with multiple correlated securities in a relatively simple way. It estimates correlation not through historical data, which is much more difficult and gives a real probability measure, but instead by looking at current market prices for derivative contracts, which gives a risk-neutral probability measure. The formula summarizes this information into a single number meant to elucidate the amount of risk on the asset once a group of securities were bundled together. For instance, this could be a collateralized debt obligation (CDO), a bundle of mortgages that many CDS had
The gross notional and exposure values that OTC derivatives are exchanged at are depicted above. We see that these amounts are growing in the time leading to the financial crash and in the first half of 2008. 

as an underlying leading up to the 2008 financial crisis. This single number from the Gaussian Copula formula is an easily understandable measurement of risk, which helped propagate its widespread use [3]. However, the prices of mortgages in CDOs were calculated with expected future correlation in the market, and correlation is known to change drastically during an economic crisis [5]. Because of how correlation changes, the output of the Gaussian Copula formula caused the CDOs to appear less risky than they actually were. This underestimation of risk in the system was not widely understood until after the financial crash, which is part of why many of the CDOs were rated highly by rating agencies, causing more derivatives to be traded on them. The risk did not disappear from the Gaussian Copula formula output though, instead it was being hidden in the final percent of failure. This meant that over 99% of the time it nothing should go wrong, but when something did, all previous profits could be wiped out and more [3].

The combination of derivatives use for betting, and the misuse of the Gaussian Copula formula partially led to many financial institutions being unprepared for a financial crash; despite warnings from mathematicians including the pioneer behind the use of the formula, David X. Li [3]. In the United States, derivatives are traded in the order of trillions of gross notional value per large financial firm, as can be seen in Figure 1. The exposure of each firm may only be a fraction of the gross notional value, but this still means that each firm would owe billions in dollars. Especially while financial

Figure 1: The gross notional and exposure values that OTC derivatives are exchanged at are depicted above. We see that these amounts are growing in the time leading to the financial crash and in the first half of 2008. [6]
firms were reading the output of the Gaussian Copula formula as a certainty that things were riskless, they assumed less collateral was needed for the contracts. Due to the nature of some of the OTC derivatives, these firms then found themselves owing even more than they believed possible [4].

Despite misuse of OTC derivatives in the past, it would not be wise to simply ban their use. Derivatives have helped develop a booming economy in global trade, and are an important tool for spreading risk when used similarly to insurance. The question then becomes how to regulate derivatives best to avoid the situation we faced in 2008.

One of the major problems regulators face is systemic risk. Systemic risk can be thought of as a domino effect, where if one financial firm goes down, others will follow suit. As we saw with Lehman Brothers and AIG, these two firms were closely linked by the OTC derivatives AIG sold. Thus, when Lehman Brothers went bankrupt, AIG was very likely to, and almost did, follow suit in bankruptcy. However, in a stable economy one financial firm may go bankrupt after making poor trading deals without directly causing any other financial firms to go bankrupt. This means that there is very low systemic risk in the economy.

In response to the global 2008 banking crisis, the Bank for International Settlements (BIS) published a Quarterly Review in September 2009 that recommends the use of Central Clearing Parties (CCPs) to mitigate systemic risk for OTC derivative markets [6]. The recommendation was based on the belief that having all OTC derivative trades going through only a few central parties would make the market more resilient to economic shocks and easier to regulate. CCPs are very similar to a middle man, because they buy a derivative contract from one financial firm and and sell the exact same derivative to another. CCPs also handle and set the amount of collateral for each firm’s OTC derivatives. Each CCP does this by estimating the risk of the contracts it trades, and designing a system to absorb the losses should any of the financial firms it services default. As we can see in Figure 2, CCPs add organization to the market. However, their central location also means that if a CCP fails, the entire market is likely to follow, and so the amount of collateral that a CCPs collect is an important part of their function. If a CCP were to fail though, it is believed that the CCP would be easier to bail out than a multitude of banks, as was done in 2008 and 2009.

In 2010 with the signing of the Dodd-Frank Act, most widely traded OTC derivatives in the United States were required to be traded through a CCP. Research is still ongoing and there is still much to learn before scien-
Figure 2: To the left is a derivative market without a CCP. In this network many financial firms are interconnected, and it is hard to tell which firm is exposed to which other firms. On the right is an OTC market with a CCP. This network is described as a hub and spoke network because everything is only connected to the CCP at the center. [7]

Scientific conclusions can be made about the stability CCPs provide to the OTC derivatives market.

2 Networks

A network is a “group or system of interconnected things” [8]. These things can be communities connected by trading, people connected by social interactions, or computers that can access each other’s files.

In mathematics, a network is a representation of these kinds of relationships. A network consists of two sets, vertices and edges. The set of vertices contains the people or objects in the network. For instance, they could be the cities in a trade network. The set of edges shows a relationship between a pair of vertices. In the trade network, an edge may represent two cities that trade directly with one another. Further, two edges are said to be adjacent if they share a vertex in common, and two vertices are said to be adjacent if they share an edge in common.

In this way many data sets can be represented as a network and described by how the elements are adjacent. Vertices represent the subject of the data collection, and edges connect vertices that have something in common. Further, networks can also be converted into matrices easily. This makes
available certain kinds of mathematical algorithms used to analyze data. Some of these algorithms can find the minimum path between any pair of vertices, or find which vertices are the most connected in the network. The number of edges that are adjacent to a vertex define the connectedness of that vertex.

Figure 3 is an example of a trading network. Each vertex is labeled with a city that merchants could trade with. The edges of the network represent trade routes existing between two cities. Further, the edges are labeled with theoretical travel times between the vertices. These are called labels or edge weights, which provide even more information about the relationship between the two vertices. Edge weights are very flexible in what they can represent. Figure 3 could have also used the amount that a merchant could expect to make from trading with the two cities as edge weights. Further, we can represent the information from the network by entering the edge weights into an adjacency matrix. Here entry $ij$ in the matrix is the edge weight for traveling from vertex $i$ to vertex $j$. For example by looking at Mogadishu, Mumbai, and Cilegon, in that order we would obtain the adjacency matrix:

$$
\begin{bmatrix}
0 & 30 & 60 \\
30 & 0 & 20 \\
60 & 20 & 0 \\
\end{bmatrix}
$$

Notice that the adjacency matrix is symmetric since the edges of Figure 3 are undirected.
Figure 4: An example of a directed graph. As can be seen, edges can point in one or both directions between vertices, and not every vertex needs to have adjacent edges.

Edges and their labels can also indicate a specific direction. For example, if there are different expected profits for different directions in the trading network then this can be represented as a directed network. In a directed network each edge has a direction associated with it, allowing two distinct edges between a pair of vertices, one for each direction. This is illustrated in Figure 4 between vertices A and C. For this network, the adjacency matrix with the nodes ordered as A, B, C, D is:

\[
\begin{bmatrix}
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 \\
\end{bmatrix}
\]

Here we see that the matrix is no longer symmetric. Further, since there are no edge weights associated with the network in Figure 4, 1s are used to indicate an edge’s existence and 0s the lack of an edge.

When direction is included and edge weights are distinct in a network, the outcome of algorithms may be different than the undirected or unweighted case. For instance, without edge weights in Figure 3, the minimal path between Mogadishu and Cilegon is the edge adjacent to both. However taking into account the edge weights from travel time, the minimal path is actually from Mogadishu to Mumbai, then from Mumbai to Cilegon, as this will result in less time taken on the trip. Further, if Figure 4 were undirected, we would say that vertices A, C, and D were equally well connected because they are all adjacent to two other vertices. Since the graph is directed though, A is actually the most connected vertex, since it has four adjacent edges, versus
only three at vertices C and D.

In the context of this research, each vertex in the network represents a financial firm that trades in OTC derivatives. The edges represent a trade between two financial firms. The network is directed, where one firm owes the other the edge weight of the edge pointing from the debtor to the creditor. These weights may be either notional or exposure amounts for derivatives traded between the firms.

Figure 2 makes clear the power of visually representing an OTC derivative market, in this case the Credit Default Swap market, where market refers to the accumulation of OTC derivative transactions by a set of banks. On the left, the network trades bilaterally, such as it did before the financial crash, and on the right the same network is represented as trading through a CCP. The blue vertices represent a financial firm that is a net buyer of Credit Default Swaps, and red represent net sellers. The edges represent two firms trading with one another [7]. Representing information in the form of a network is an important tool in summarizing large amounts of data.

Further, a trading network is naturally described by a mathematical network. When organizing the data, it makes sense to organize the information into a table or matrix, which is one of the ways to store a network in a computer. Further, by organizing the information this way it is easy to keep track of the total amount that financial firms are trading by buying or selling, which is given by the column or row sums respectively. This will simplify algorithms used later in the paper such as Gibbs Sampling.

Networks are a powerful mathematical tool. They allow one to organize data visually, effectively summarize data, and use many algorithms designed for networks. This and more makes networks a natural tool to use for OTC derivatives markets.

3 Modeling Overview

Usually when creating a model one relies upon large amounts of data to train and test the model on. However, one of the major challenges of modeling OTC derivatives is that there is very little data available to the public. In fact, the only data consistently available are the gross notional amounts aggregated for each financial institution. In other words, we do not have available the amounts that institutions traded during each transaction, which other institutions were trading partners, nor the total exposure faced by each
institution. The model in this paper takes the gross notional amounts as inputs, and outputs plausible financial trading networks for OTC derivative markets described in the form of exposures.

We begin by imposing structure on the overall market of OTC derivatives. This includes the likelihood that two institutions trade with one another, and the amount of their trades. At this point, plausible bilateral trading networks can be modeled in terms of notional values [9]. This will further be discussed in Section 4 of the paper.

There is however no data to test this or any further steps. To account for this, many different samples are taken of possible bilateral trading networks. In this way, no single market structure is singled out as representative. Instead, a variety of market structures are proposed that can all be tested so that more common market forms will be given more weight since they will show up in the sample more often. This is a result of using Gibbs Sampling to sample the bilateral trading networks, as discussed in Section 4.4 [9].

Once these models are created, the notional amounts need to be converted into exposure amounts. This means to understand how much each financial institution has the potential to lose in the transaction if the other party defaults on the contract. In the past the conversion from notional amount to exposure amount has been modeled with a normal distribution [10], or for the Credit Default asset class of OTC derivatives with a t-distribution with three degrees of freedom [11]. In Section 4.5 this methodology is implemented in a way that reflects the modeled bilateral trading network and ensures a skew symmetric exposure adjacency matrix.

Sections 5 and 6 have a similar goal in that they both model the same market as the bilateral trading network, but impose a different structure through the use of one or more CCPs. Since this paper seeks to compare market structures, the model assumes that the total exposure for each financial institution remains the same regardless of network type. Thus the structure can be compared without confounding from changes in the amounts being traded.

Section 7 discusses a metric that is used to analyze the total amount of exposure that an institution faces across asset classes. This metric is called the netted exposure, which accounts for netting the exposures traded with a single institution, but does not allow netting of exposures between different institutions [10].

This metric is then used in Section 8 to calculate relative netting efficiency and determine which network structure allows banks to minimize the
amount of netted exposure they face. Presumably, lower netted exposures allow the financial institutions in the network to better withstand economic shocks. It is also possible to look at the variance of the netted exposure amounts [12]. Using these two statistics, conclusions are drawn about the most stable network structure.

4 Bilateral Trading Model

The first network structure modeled is the bilateral trading network where each asset class of derivatives has its own network. We start in Section 4.1 by estimating the net notional amounts that each bank trades from the gross notional amounts provided by official statistics. There is very little data available regarding net notional amounts [13]. Even though the data are still used to train the model, the parameters that determine how to sample net notional values from the gross notional values are left flexible, and are one of the model inputs that can be tested. Once the net notionals are sampled, the assets and liabilities of each financial institution can be solved for. Here assets refers to the notional amount for the OTC derivatives that an institution has sold, and liabilities are the notional amount that is bought.

Section 4.2 discusses the different assumptions that can be used to create the bilateral trading network. The assets and liabilities serve as the inputs for the trading models developed by Gandy and Veraart [9]. It is assumed that the sum of assets in the network equals the sum of liabilities in the network. In other words, we are assuming that financial institutions are only trading with other financial institutions in the network.
The next step in creating the bilateral trading network is to find an initial feasible network, which is discussed in Section 4.3. To be a feasible trading network, it is required that all edge weights are greater than or equal to zero, and that the sum of assets and liabilities for each institution in the network equals the assets and liabilities used as inputs. One way to solve this problem is by using a max flow algorithm [9].

Afterwards, Gibbs Sampling can be used to build a sequence of networks that converge to match the assumptions made in Section 4.1 and follow the distributions defined in 4.2. Gibbs Sampling uses Markov chains and marginal distributions to achieve this goal. Gibbs Sampling then returns a chain of bilateral trading networks with values in notional amounts of OTC derivative contracts where each network meets the assumptions of the model. All of this is further discussed in Section 4.4.

Finally, to understand the risk associated with these trades, it is important to have the networks in exposure values rather than notional values for the contracts. Section 4.5 discusses how a normal or t-distribution with three degrees of freedom can be used to estimate the exposure values. These estimations are based on the notional values traded between institutions. Also, the adjacency matrix of the bilateral trading network is skew-symmetric when trades are represented in exposure values. If one institution owes another a certain amount, intuitively the other institution is owed the same amount, which is what the skew-symmetry implies. Once in this form, we can analyze the bilateral trading networks for systemic risk.

4.1 Estimating Assets and Liabilities

The only data required to run this model are the gross notional values that each financial institution has traded during a certain time period. Although the aim is to know the assets and liabilities of each financial institution, the data is not available and previous work has been done using only the gross notional values to estimate the net notional values [14].

Gross notional data is often available in quarterly reports for institutions in the United States [15]. Gross notionals can be thought of as the sum of all assets and liabilities of each financial institution. Similarly, net notionals can be thought of as the difference between assets and liabilities. For a trading network with \( n \) institutions, there will be \( n \) gross notional values to keep track of. For this reason, we choose to label each of the financial institutions with an integer between 0 and \( n - 1 \). Then the gross notional data can
be saved as entries in the vector $g$ where $g_i$ is the gross notional value for institution $i$. Likewise the net notional values can be saved in the vector $n$, the assets in $a$, and the liabilities in $l$.

There is no data on the amount of assets and liabilities that each institution trades. However, some data that describe the ratio of net to gross notional amounts are available in European markets [13]. Once a sample of $n$ is taken, we can use equations (1) and (2) to solve for the asset and liability value for each institution

$$a = \frac{g + n}{2}, \quad (1)$$
$$l = \frac{g - n}{2}. \quad (2)$$

To build a method for estimating the net notional values for each institution, we turn to the data that is available to us [13]. By visually estimating the frequencies of each histogram in the data set, we are able to extract numeric values. Then by minimizing the least squares, the parameter $\lambda$ of an exponential distribution is estimated to fit these data. This is done for the positive ratios of net to gross notionals, and separately for the negative ratios. Further, the data are used to estimate how often the ratio is positive, which we call $q$.

It is a requirement that for each financial firm the total amount sold and the total amount bought add to the original gross notional amount given,

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} l_i. \quad (3)$$

Intuitively this makes sense, assuming all the firms that buy and sell OTC derivatives to each other are included in the model, then equation (3) should be true. However, $n$ is randomly sampled from a truncated double exponential distribution, so it is very unlikely that (3) holds. To keep our model consistent with (3), we need to rescale the values of $a$ and $l$ obtained from (1) and (2). This is done by finding the difference in the sums, and then changing every value by an amount proportional to the institution’s current liabilities so that (3) holds.

The last requirement we need to check for $a$ and $l$ is that no entry in either vector is negative. If this occurs, then for that vector we take an equal amount from all other non-zero entries and add it to the negative entry so
Ratios of net to gross notional values make up the histogram. We estimate the data follow a double sided truncated exponential distribution, where the positive truncated exponential has parameter $\lambda = 28$, and the negative one has parameter $\lambda = 15.3$. We further determined that there is a 33% chance of the ratio being positive, such that $q = 0.33$.

that the sum of the vector remains the same, but there are no negative values anywhere.

With a reasonable estimate of $a$ and $l$ we can estimate the trading networks $N$. These networks are saved in the form of matrices and have entries in terms of notional values. The entries of row $i$ represent the liabilities of financial institution $i$, so that the row sum is equal to $l_i$. Similarly, the entries of column $j$ are the assets of financial institution $j$ so that the sum of column $j$ is equal to $a_j$. Thus, each entry $N_{ij}$ is the notional amount in OTC derivatives that institution $i$ has bought from institution $j$. Further, the values along the diagonal must be zero since no financial firm can trade derivatives with itself. Thus we may define a feasible notional trading network for any given $a$ and $l$ as the following:

1. $\forall i, j, N_{ij} \geq 0$,

2. $\forall i \in \{1, 2, \ldots n\}, N_{ii} = 0$,

3. $\forall i \in \{1, 2, \ldots n\}, \sum_{j=1}^{n} N_{ij} = a_i$,
4. \( \forall j \in \{1, 2, \ldots, n\}, \sum_{i=1}^{n} N_{ij} = 1_j. \)

Now that we have defined the constraints on the trading networks, the next step is to define how values are traded between institutions.

### 4.2 Forms of Network Models

Two forms of models developed by Gandy and Veraart are available for use in the bilateral trading model created with methodology from this paper [9]. These models are the basic model and the conjugate distribution model. The basic model imposes structure on the trading network by having the modeler input specific values that describe when and how much institutions will trade. The conjugate distribution model instead uses a prior distribution to sample the values for when and how much institutions trade.

Both of these can also be used to construct tiered models. Tiering allows financial institutions to be sorted into multiple sets with different parameters. This type of model can be used to describe a network with firms of distinctly different sizes. For instance, every OTC market tends to be dominated by only a few institutions, where all others will often make up less than ten percent of the market. In this case, the parameter that determines how much institutions trade can be made much larger for those that trade ninety percent of the market volume than those that trade ten percent of the market volume.

Both the basic and conjugate distribution models are defined by two distributions. One distribution describes the probability that financial institutions decide to trade with one another. The other distribution describes how much financial firms trade after deciding to do so.

The basic model assumes that all of the financial institutions trade with each other in exactly the same way. Let \( A \) be the unweighted adjacency matrix for a trading network \( N \). Then the structure we expect see across many trading networks follows the following distributions:

\[
P(A_{ij} = 1) = p_{ij},
\]

\[
N_{ij} | \{A_{ij} = 1\} \sim \text{Exponential}(\lambda_{ij}).
\]

In (5) \( p_{ij} \) is the probability that institution \( i \) will purchase OTC derivatives from institution \( j \). Thus each trade in the network follows a binomial distribution with parameter \( p_{ij} \). This implies that \( \forall i, p_{ii} = 0 \). In general, for the off–diagonal entries of \( p \), they are all equivalent to a single non–zero
value $p$. When there are different values of $p_{ij}$ in $p$, this is considered the tiered model. When this is the case we use the probability that the larger institution trades with another institution as the value of $p_{ij}$.

Further, $\lambda_{ij}$ in (5) is the parameter for an exponential distribution that describes how much a pair of institutions tend to trade with one another. In general, the value of $\lambda_{ij}$ is the same for all $i \neq j$. However, in the case of tiered modeling the $\lambda$ parameter associated with the smaller institution defines how much a pair of institutions trades so that $\lambda$ is symmetric.

The conjugate distribution model makes use of Bayesian statistics to allow the possibility of more complex market structures to be represented in the model while still remaining relatively computationally simple. This is done through the use of prior distributions that are used to sample the parameters $p$ and $\lambda$ of the basic model.

\[
\tilde{p} \sim \text{Beta}(\alpha, \beta), \quad \tilde{\lambda} \sim \text{Gamma}(\gamma, \delta), \quad (6)
\]

\[
p_{ij} = \tilde{p}1(i \neq j), \quad \lambda_{ij} = \tilde{\lambda}, \quad i, j \in N.
\]

Periodically, while the trading networks $N$ are being sampled, new $p$ and $\lambda$ values are chosen based on the distributions in (6). The conjugate distribution model accounts for our ignorance regarding how institutions trade with one another, and it allows trading dynamics to change over time.

These basic and conjugate distribution models describe the expected behavior of the trading networks. This means that after many trading networks have been sampled, which is further discussed in Section 4.4, it is clear that the networks overall follow the distributions in (5) or (6). In other words, these model forms describe the assumptions that are made about the trading networks, and the $N$’s will be chosen so that the model fits these assumptions.

### 4.3 Initial Feasible Network

Equation (3) not only makes sense intuitively, but also acts as the constraint for creating an initial feasible notional trading network. In order to successfully sample instances of $N$ that satisfy the assumptions made in (5) or (6), the algorithm we use first needs an initial network that satisfies the conditions defined in (4). This is true even if the initial network does not fit the model forms discussed in Section 4.2.

To find this initial network we utilize the Edmonds-Karp max flow algorithm [9]. Maximum flow algorithms find the maximum values that can
be sent from one end of a network to another, and during this process they create a network that achieves that maximum flow as a byproduct. It is this byproduct we are interested in.

We do this by setting up a flow network that has an artificial source and sink for the flow. The source connects to a set of vertices where each financial firm has its own vertex to represent its assets, and the sink connects to a separate set of vertices where each financial firm has its own vertex to represent its liabilities. The edges the algorithm creates between these two sets defines a bilateral trading network in notional values, $N$, that satisfies (4). The flow network is pictured in Figure 7.

In this algorithm it is important that (3) holds to ensure that we account for all of the assets and liabilities for the institutions in the initial feasible trading network. Further, the edges connecting the vertices representing assets and liabilities have an infinite capacity so that there is no upper bound on the potential edge weight. This is because the only limitations we want on the system are that all of the liabilities and assets are sent as flow, and that no institutions trade with themselves. We account for the second requirement by not allowing an edge between vertices representing the same institution in the flow network. For further reading on max flow algorithms see [16].

When using Edmonds-Karp, the maximum amount of flow is always sent along each path with positive capacity. This approach often creates a very

---

**Figure 7:** A flow network used during the Edmonds-Karp algorithm. Here the network has $n = 3$ financial institutions in the OTC derivative network. The edges are labeled with their original capacities, if they are unlabeled then they have infinite capacity.
sparse residual network. Due to this sparsity, we will use the Edmonds-Karp algorithm as a second step in generating the initial feasible network. First an Erdös-Rényi graph is sampled with the same parameter as the expected number of edges in the network, which is the arithmetic average for non-diagonal entries of $p$ from Section 4.2. Afterwards, the edges are weighted so that none of the conditions in 4 are broken. This could mean that all of the edges are weighted with the same value as long as the sum of edge weights is not greater than $\sum a$. Then, the difference between the row sum of this matrix and $a$ and the column sum of this matrix and $l$ are used to define the assets and liabilities for the flow network. Afterwards, the two networks are added together so that the edge weight for any non-diagonal entry is the value from the weighted Erdös-Rényi network plus the value from network found with Edmonds-Karp.

This initial feasible network will likely not satisfy the model form that was chosen. However, it will allow us to update entries according to the distributions defined in the model form chosen from Section 4.2, allowing for the convergence of a Markov chain to the desired distribution.

### 4.4 Gibbs Sampling

Once an initial matrix is found that satisfies the conditions of an OTC derivative market network, we utilize Gibbs Sampling to reach the target distribution of either (5) or (6), then return multiple samples of possible trading networks based on the same gross notional data.

Gibbs Sampling is a type of Markov Chain Monte Carlo Method that allows one to sample from distributions for each random variable in the problem while holding the others constant. Further, it can be proven that sampling from the marginal distributions of the random variables during Gibbs Sampling will converge to the joint distribution for all the random variables. Each entry $N_{ij}$ with $i \neq j$ is considered a random variable for this problem. The joint distribution is the model form selected in Section 4.2.

Gibbs Sampling works by creating a Markov chain. A Markov chain is a sequence of samples from a probability distribution where every sample is independent of everything but the immediately preceding sample in the chain. This means that the current trading network you have determines the network that you can get in the next step.

Therefore an initial feasible matrix needs to be given to the algorithm. Gibbs Sampling cannot create a matrix from nothing. Instead, it updates a
Figure 8: A possible cycle update in the four bank network. $\Delta$ is the amount we change the cycle by.

random variable to fit the marginal probability distribution of that random variable. Since this update is based on the current matrix, it is important that the starting matrix is already feasible, meaning it satisfies (4).

To ensure that the trading networks remain feasible after each update, rather than changing the value of individual entries in the adjacency matrix $N$, cycles are randomly chosen to update. For example, if ten is added to the entry in the third row and the first column, then the row and column sums will be ten more than required. So, ten must be subtracted from a different entry in the third row and from a different entry in the first column in order to maintain the correct row and column sums. This can be done for any value $\Delta$ chosen to update the cycle, such as in Figure 8. Every time a cycle in the matrix is chosen and perturbed by $\Delta$, one iteration of Gibbs Sampling is completed and the new $N$ is the next sample. Once enough cycles have been updated, then that $N$ is added to a chain of feasible networks that are returned as the networks that make up the bilateral trading network model.

During each iteration of Gibbs Sampling, first the random selection of entries and cycles is necessary. First a value $k$ is randomly chosen where $2 \leq k \leq n$. Next, $k$ rows and columns from $N$ are randomly chosen and systematically arranged to form the cycle that will be updated [9]. Since it is impractical to update every cycle in $N$ before adding a new network to the chain that is returned, cycles are randomly selected in an independent manner that avoids correlation.

Second, entries in $N$ are updated by a change in value, $\Delta$, that is sampled
during each iteration of Gibbs Sampling. A probability mass function (pmf) can be defined to describe the probability that the entries in the notional trading network contain their current values based on the cycle that is chosen to update. Then, the update to the cycle, \( \Delta \), is chosen with probabilities defined by this pmf that describes the target distribution. Thus, every time a cycle is updated, the entries of the cycle are replaced with ones that exactly match the target distribution. These updates create a Markov chain of network matrices, since each update is based only on the current values in \( N \) for the cycle being updated. So with each update the cycle chosen is replaced with another that matches the target distribution. This means that once every entry from the initial feasible matrix has been updated as part of a cycle, we can say the network follows the joint distribution we used from Section 4.2.

Since the initial feasible matrix likely does not match the target distribution, a burn-in period for the method is required to allow the Markov Chain to converge to the target: the first \( b \) samples taken are ignored and not added to the output chain. Burn-ins allow us to take advantage of how Gibbs Sampling will converge to the target distribution as long as the initial network is feasible, without needing the initial network to have the target distribution.

Further, a Markov chain’s consecutive samples will be highly correlated by definition. To remove correlation from the chain of \( N \)s we use to analyze the OTC derivative trading networks, we apply thinning: after a sample is added to the chain, \( t \) samples are ignored before including another in the chain. This allows us to take different samples of \( N \) without the serial correlation of a Markov chain, although it is possible that some small correlation still remains between the samples in the chain.

By utilizing Gibbs Sampling, a chain of any chosen length is returned with different feasible trading networks that each meet specific model forms, which are the target distributions of Gibbs Sampling. Each \( N \) in the returned chain also meets the feasibility constraints for \( N \), and successive samples have little to no correlation. In this way we construct a variety of different OTC trading networks that can eventually be tested for stability.

### 4.5 Modeling Exposures

The final step in modeling bilateral trading of OTC derivative markets is to convert the net notional values into gross notional values. One way of doing this is to use a normal distribution with mean zero and standard deviation
We use a normal and t distribution with three degrees of freedom to describe exposures for OTC derivatives.

equal to the ratio of the gross notional traded between two institutions to the total derivatives traded without the first institution [10]. It has also been suggested that a t-distribution with three degrees of freedom is better when estimating exposures in the credit derivative class [11].

In this model we take advantage of the fact that the chain produced by Gibbs Sampling has estimates for how institutions trade derivatives in notional values. Thus, unlike in previous work, notional trading amounts do not need to be estimated using ratios based on institutional gross notionals[10, 11].

An important assumption in the model is that the exposure from institution \( i \) to \( j \), \((X_{ij})\), is equal to the negative value of exposure from institution \( j \) to \( i \), \((X_{ji})\). Thus

\[
X_{ij} = -X_{ji}. \tag{7}
\]

This means that \( X \) has the property of skew symmetry. Intuitively this makes sense since no institutions should disagree on how much they owe and are owed by one another.

To convert a bilateral trading network in notional values, \( N \), into a bilateral trading network in exposure values, \( X \), we use the following methodology:

\[
h \sim \begin{cases} 
N(0, (\beta t N_{ij})^2) & \text{if } N_{ij} > 0 \\
0 & \text{if } N_{ij} = 0
\end{cases}
\]
\begin{align*}
g &\sim \begin{cases} 
N(0, (\beta_l N_{ji})^2) & \text{if } N_{ji} > 0 \\
0 & \text{if } N_{ji} = 0
\end{cases} \\
X_{ij} &\begin{cases} 
\frac{1}{\sqrt{2}}(h - g) & N_{ij}, N_{ji} > 0 \\
h - g & \text{else}
\end{cases}
\end{align*} 

(8)

In the above, \( \beta_l \) is the illiquidity of asset class \( l \) \([10]\). Also, \( h \) and \( g \) are sampled from the above distribution and then used to calculate the exposure of bank \( i \) to \( j \), for all \( i > j \). This is for the case where the asset class is not credit derivatives. If it is the credit derivative asset class, then we use a \( t(3) \) distribution instead of a normal distribution with the same mean and standard deviations as defined above. In either case, all of the diagonal entries of \( \mathcal{X} \) are zero since institutions cannot trade with themselves, and so they cannot be exposed to themselves. Further we satisfy (7) since we account for both \( N_{ij} \) and \( N_{ji} \) when sampling \( X_{ij} \) and set \( X_{ji} = -X_{ij} \).

We use every \( N \) in the chain as the input to sample \( \mathcal{X} \). At this point we have successfully modeled a chain of possible bilateral trading networks for OTC derivatives. It is important that the networks contain the exposures from institution to institution so that the network can be used for systemic risk analysis. Further, since there is a chain of possible networks, we do not assume that OTC bilateral trading has any single representative form, yet those forms which are more likely to occur will be represented in the chain more often as well.

5 Model with a Single Central Clearing Party

Figure 10: A trading network with a single Central Clearing Party.
With an adjacency matrix for the bilateral trading network, modeling the network with one CCP is just a restructuring of how the trades are transacted. A CCP acts as a middle man in the market, and so now instead of considering the trades to occur directly, all of the transactions are considered to take place with one central institution.

Mathematically this means that for a bilateral trading network the $i^{th}$ row is summed and saved as the $i^{th}$ row 0$^{th}$ column entry of a new trading network adjacency matrix. Likewise the $j^{th}$ column of the bilateral trading network is summed and saved as the $0^{th}$ row $j^{th}$ column entry. This new matrix represents the trading network with a single CCP, where the CCP is the institution represented by the zeroth row and column.

It is important to note that the summed value of exposure faced by each institution remains the same whether they are participating in bilateral trading, or in a market with a single CCP. This means that any risk analysis used will not be confounded with having to take into account how exposures change in different markets as well as how the structure changes.

6 Model with Multiple Central Clearing Parties

![Diagram of a trading network with two Central Clearing Parties](image)

**Figure 11:** A trading network with two Central Clearing Parties.

In reality, there is usually more than one CCP to trade with. For instance, for European markets it is possible to have up to six CCPs for a single derivative asset class and common to have three [17]. Thus, for a more robust comparison of OTC derivative trading with and without CCPs, the
model needs to take into account having multiple CCPs in the same asset class.

Since the model accounts for only large institutions that trade directly through the CCP, derivative contracts are traded in high volume. Further, with a maximum number of CCPs approximately six, there are many more derivatives traded than CCPs. Because of this, it is reasonable to assume that the way distribution that describes how exposure amounts are split between CCPs is a normal distribution.

In the model, a different sample is randomly drawn from the normal distribution for each CCP trading in the asset class. As stated before, part of the usefulness of the model is that the total summed value of exposures per institution remains the same no matter which market structure is being used. So, once the samples are taken we normalize them so that the samples sum to one. These represent the proportion of exposures that are sent through each CCP, and to find the exact amount the sample is multiplied with the entry from the model with one CCP for the institution currently being considered.

Using a normal distribution to sample the proportions, a model of OTC derivative trading with multiple CCPs may be constructed from the model with a single CCP. Thus we have successfully modeled three different market structures for trading OTC derivatives.

7 Netting Efficiency

The goal of modeling OTC derivatives with and without CCPs is to assess the systemic risk of different network structures. One way of doing this is by measuring the netted exposure for an institution in each market structure. Netted exposure are the summation of the positive values of the summation of exposures that are traded with the same institution. It is the sum of positive net exposures because although before we would net positive and negative exposure, an institution cannot guarantee that the timing will sync so that the costs an institution faces from a positive exposure will be offset by a payment resulting from one of the negative exposures.

In the case of bilateral trading networks for each asset class, the exposures for each pair of institutions must first be summed up. Regardless of the asset class the same institutions are trading with one another, so the exposures can be further netted. After this, the netted exposure for each institution is the sum of positive values across the row of the adjacency matrix for that
Figure 12: To calculate the netted exposure for Institution 1, we first sum exposures for the three asset classes, represented by different colored edges, for each pair of institutions that includes Institution 1. Then, the positives of these values are summed together to calculate the netted exposure of Institution 1.

In the case of networks with different CCPs for each asset class, we assume that there is no netting across asset classes. Therefore we assume that each asset class has unique CCPs that only trade in one asset class. This means that in the case of each asset class only trading through one CCP, for each institution netted exposures are calculated by adding the positive exposures from the institution’s corresponding row from each asset class adjacency matrix. When including multiple CCPs in each of the asset classes, the same procedure is taken. The positive entries from the row of each asset class are summed together. This is illustrated in network form in Figure 13. If the CCP trades in multiple asset classes, then the exposures between that CCP and the institution are first netted before the positive part is summed with the exposures from the other CCPs.

Once this is completed, a netted exposure for each institution in each market structure has been computed. To compare the netted exposures of one institution for two different market structures, we take the difference of the two netted exposures, which is called the relative netting efficiency [11]. Thus, for each institution there are \( \binom{s}{2} \) relative netting efficiency values, where \( s \) is the number of different market structures we are comparing. If the netting efficiency is positive, this implies the second market structure is better at reducing exposure, and if the value is negative then the first market structure is better at reducing exposures.

It would be reasonable to expect that if an institution has a smaller
netted exposure, then they are less likely to go bankrupt. Because of this, we recommend market structures that have the highest netting efficiency.

8 Model Analysis

The models in this paper have a lot of flexibility built in. For instance, the parameters for each probability distribution are inputs of the algorithms in the model. This includes the parameters for the double exponential distribution used to estimate net notionals, the \( p \) and \( \lambda \) matrices used in the models designed by Gandy and Veraart, the illiquidity measures as defined by Duffie and Zhu, and the inputs for the normal distribution used to split the exposure values between Central Clearing parties [9, 10].

The data used in the analysis are the gross notional amounts from the fourth quarter of 2016 for the Top 25 commercial banks, savings associations, and trust companies in the United States [18]. When running the model to analyze the OTC derivative market, we use the values estimated in Section 4.1: the probability of a net notional being positive was 33\%, the exponential distribution of positive values had \( \lambda = 28 \), and negative values had \( \lambda = 15.3 \). Then, to describe how the notional values are traded throughout the network, a tiered basic model is used as described in Section 4.2. Here, the \( p \) and \( \lambda \) matrices are estimated using data from the previous year by averaging the total derivatives of the tier and finding its proportion to the total derivatives of all institutions [15]. The tiers are chosen to be the top four banks as the first tier, and the remaining twenty-one as the second tier. Here, we assume

![Figure 13: To calculate the netted exposure for Institution 1 each of the positive edges adjacent to Institution 1 are summed. In this case each asset class is assumed to have different CCPs.](image)

28
that the top four institutions always trade with one another, and trade with
the smaller institutions has a probability of 80%. It is assumed the small
institutions trade with one another 20% of the time. For \( \lambda \), we estimated
that the amount large banks trade with one another follows an exponential
distribution with parameter \( \lambda = \frac{1}{6770000} \) and that the other transactions in
the network have \( \lambda = \frac{1}{19400} \). Here we use the same model for each asset class
and assume independence between asset classes, however this is not necessary
in the model and is another kind of flexibility. For the illiquidity measure,
the values estimated by Cont and Kokholm are used, \( \beta_{\text{credit}} = 0.0098 \) and
\( \beta_{\text{otherwise}} = 0.0039 \) [11]. Finally, for the normal distribution that splits the
exposures between multiple CCPs, the distribution is assumed to have mean
\( = 0 \) and standard deviation \( = 0.33 \).

Using the above values as inputs, we create a chain of 10,000 bilateral
trading networks with the algorithms developed earlier in this paper. The
burn–in is 100,000 and the thinning step is 1,000 during the Gibbs Sam-
pling part of the algorithm. Each of the 10,000 networks in the chain are
converted into a network with one CCP and with three CCPs. This is done
separately for Interest Rate Swap, Foreign Exchange, and Credit Derivative
asset classes. To calculate the gross notionals for each of the asset classes in
OTC derivatives from the available data, we assume that the same propor-
tion of each asset class is exchange traded [18]. The model is then run six
different times to test for consistency in the modeling process.

Once each of the models are created for the three market structures and
three asset classes, the netting efficiencies are calculated for each of the
10,000 markets in the chain. Histograms of these values will be presented on
the next pages.
Single CCP v. Bilateral Trading

**Figure 14:** The histograms show the results of runs one through six for Bank of America. Here a single CCP per asset class always provides the highest netting efficiency.
Figure 15: The histograms show the results of runs one through six for PNC Bank. Here in the market with a single CCP per asset class PNC nearly always has a smaller netted exposure. However, even when it does not have the smallest netted exposure, the market with a single CCP has nearly the same efficiency as bilateral trading.
Figure 16: The histograms show the results of runs one through six for Fifth Third Bank. Similar to PNC Bank, the netted exposure is nearly always less in the network with a single CCP.
Figure 17: The histograms show the results of runs one through six for Huntington National Bank. Even for the smallest of the Top 25 institutions in the United States, a market with a single CCP per asset class almost always has lower netted exposure.

Since the values are on average negative for all of the runs and the institutions of varying size in Figures 14, 15, 16, and 17, we conclude that one CCP per asset class is relatively more efficient at reducing netted exposures than bilateral trading.
Figure 18: The histograms show the results of runs one through six for Bank of America. Overall Bilateral Trading has a greater relative efficiency when netting exposures than three CCPs per asset class. This is the case about ninety percent of the time.
Figure 19: The histograms show the results of runs one through six for PNC Bank. About fifty percent of the time bilateral trading is better at netting exposures, however, during this fifty percent of the time bilateral trading is much more relatively efficient at it.
Figure 20: The histograms show the results of runs one through six for Fifth Third Bank. On average bilateral trading is more efficient at netting exposures about thirty-eight percent of the time, however it is then much more relatively efficient.
Figure 21: The histograms show the results of runs one through six for Huntington National Bank. On average bilateral trading is only more efficient at netting exposures about thirty percent of the time, however it is much more relatively efficient when this is true.

We find that bilateral trading does a better job at netting exposures for larger institutions, such as in Figure 18. Further, bilateral trading markets have much smaller netted exposures than trading with three CCPs per asset class when bilateral trading has a greater relative netting efficiency. Further, when
trading with three CCPs is relatively more efficient, the netted exposures are very similar. This can especially be seen in Figures 19, 20, and 21. Thus we find that although in most cases it is better for an institution to trade with three CCPs per asset class, in the extreme scenarios bilateral trading networks have much less risk.
Single CCP v. Three CCPs

Figure 22: The histograms show the results of runs one through six for Bank of America. Trading with one CCP per asset class is almost always more relatively efficient than trading with three CCPs per asset class.
Figure 23: The histograms show the results of runs one through six for PNC Bank. Once again it is almost always more efficient to trade with a single CCP per asset class.
Figure 24: The histograms show the results of runs one through six for Fifth Third Bank. Once again it is almost always more efficient to trade with a single CCP per asset class.
Figure 25: The histograms show the results of runs one through six for Huntington National Bank. Once again it is almost always more efficient to trade with a single CCP per asset class.

We conclude that irrelevant of the size of the institution, trading markets with one CCP per asset class are almost always better at reducing the netted exposures for institutions than markets with three CCPs per asset class. This is illustrated in Figures 22, 23, 24, and 25.
9 Conclusions

The model developed in this paper for trading OTC Derivatives with and without CCPs contains a lot of flexibility, yet there are still some assumptions built in. For instance, we assume that within an asset class all of the financial institutions have the same distribution for the ratio of net notionals to gross notionals. Further, in the model it is assumed that in notional values the total assets are equivalent to the total liabilities for each institution. This means that we assume there are no trades with institutions that are not included in the model and do not handle this case. It is also assumed that every financial institution trades with every CCP available. This assumption is justified as long as the model is only used for the institutions that make the largest trades in the market, such as the top 25 in the United States. Finally, we assume that each asset class of derivatives is traded independently of the others, so that no correlation exists between asset classes for a pair of institutions that trade.

Despite these assumptions, the model designed in this paper is flexible enough to handle most market situations. Although we did not do this in Section 8, the different asset classes do not need to have the same market structures. It is possible to have a different number of CCPs in different asset classes, and to assume the notional trading market described by the $p$ and $\lambda$ in Section 4.2 are not the same for different asset classes. Moreover, as new knowledge about the OTC derivative trading markets becomes available, this can be accounted for by changing the parameters of the probability distributions that describe different aspects of the market.

It is also important to note that our algorithm produces a chain of markets for each asset class. This means that a variety of network structures is being tested during the analysis phase, as opposed to one. Since it is impossible to know the exact structure of the OTC derivative market from the publicly available data, and markets evolve over time as different derivatives are traded, having a chain of networks is a more realistic method for modeling trading. Having a chain means that no single structure is emphasized in the analysis. Further, the use of Gibbs Sampling to produce the chain means that all of the markets that are analyzed will be plausible as defined in Equation 4 and follow the distributions that are used from Section 4.2.
The main results of the analysis are:

- The best market structure is trading OTC derivatives with one CCP for each asset class.

- The larger the institution, the more efficient a bilateral trading network is compared to a network with three CCPs per asset class. Here efficient means that the netted exposures are smaller in the bilateral trading network.

- Three CCPs per asset class are more efficient the majority of the time for smaller institutions.

- In the cases where bilateral netting is more efficient than three CCPs per asset class, it is relatively much more efficient than when three CCPs per asset class are more efficient.

These results were obtained using netting efficiency as a metric, which assumes that exposure values are a good measurement of the risk an institution faces. However, this does not take into account the increased costs of trading with a CCP. We also suggest further analysis of the model.

10 Future Work

One of the major assumptions in the model is that each asset class of derivatives trades independently from every other class. However, this is very likely not the case in real OTC trading networks. Because of this, a recommendation for the future would be to include correlations between asset classes in the model. This is best applied when modeling how notional amounts are traded in a bilateral trading network.

A further improvement in the model can be made in the algorithm to split assets and liabilities. One of the assumptions is that total liabilities and assets are the same in the market. However, the way this is currently achieved in the algorithm means that $a_i + l_i \neq g_i$ in most cases anymore. Considering the OTC derivative gross notional values are currently estimated, this is not strictly a problem, but in the future it would be reasonable to require assets and liabilities to sum to the gross notional values used as an input to the model.
Beyond these algorithmic changes, it would also be useful to have further analysis on the current model done. Sensitivity analysis of the current input parameters would be especially informative. For instance, it would be informative to know if changing the parameter for describing the net to gross notional ratio changes the results, and by how much. This is especially interesting since the current parameter choice is estimated with a small amount of data meaning it is more likely the model is over-fitting.

There is still much to learn about the OTC derivatives market and how to make it more stable. Different systemic risk measures could also be applied to the model and smaller institutions that participate in client clearing rather than trading directly with the CCP could also be included. We recommend further study of this topic.

11  Acknowledgements

Thanks would like to be given for summer research funding provided by the Clare Boothe Luce program at WPI. Further thanks are given to Shannon Feeley and Khasan Dymov for contributions made during preliminary work.
References


[17] European Securities and Markets Authority, “List of central counterparties authorised to offer services and activities in the union.”