Aalto University
School of Science

MS-E2177 Seminar on Case Studies in Operations Research

Rogue Trading Control

Project Report

Johanna Tavia (Project Manager)
Verneri Lauksio
Ville Kujala

May 31, 2018

Contents

1 Introduction

1.1 Background

Since the collapse in 1995 of Baring Bank, financial organizations have started to develop tools and processes to detect misbehavior in trading. The reason for Baring Bank collapse was due to massive losses of £827 million by one individual trader Nick Leeson (Ross 2018). The 28 year-old trader was responsible to arbitrage profit Nikkei 225 prices between Osaka and Singapore exchanges. He managed to be greatly successful until it was revealed that he had been booking all loss-making transactions to his secret 88888 account. (Customer 2018). Since the financial crisis 2008-2009, banks have gone to great lengths in their bid to stamp out rogue trading, market manipulation and other employee misdoings.(Noonan 2017).

Rogue trading in this project is defined as trading which is operated outside of the established rules of a bank. (PwC 2008). Even if the banking industry is the most regulated industry in the world (Graaf & Kidd 2012), it is important to understand that rogue trading cannot be completely prevented but it can be managed in an acceptable manner by building preventive control systems and risk mitigation strategies (PwC 2008). The solution for the rogue trading problem can be approached from multiple perspectives. According to Graaf & Kidd (2012) common themes around the phenomenon have been culture and governance, trading mandates and limits and reconciliations/confirmations supervision. Moreover, previous incidents have involved also surveilance of off market rates, day 1 P&L, cancelled and amended transactions and IT security. In conclusion, to expose and control rogue trading, financial institutions' risk management need to cover multiple perspectives on both the system and the people level.

1.2 Objectives

In this project, we focus on cancel and amendments (C&A) control, that is one of customer's many trading controls. Often trades are cancelled when there has occurred a common mistake and both parties reach a mutual agreement. However, the C&A trades can also reveal the early recognitions of suspicious trading patterns. The aim of the project is to create and implement a programmatic solution for finding and categorizing suspicious transactions from C&A data. The solution itself cannot reveal rogue traders but can be help to find indicators of suspicious behaviours which could together with other bank's controls, help the customer to reveal possible future frauds.

2 Literature review

2.1 Rogue trading

The purpose of rogue trading is to gain profit by operating outside of the established rules of an institution (PwC 2008). Typically, rogue trading can be divided into the four following cases: Transactions harming the bank by hiding losses or manipulating the profit and loss, transactions harming the client of the institution, transactions used to illegally hide losses to avoid tax payments and transactions used to hide the final beneficial owner of the money for the purpose of money laundering or/and financial crime. (Customer 2018). Steve Allen (2008) explains in his article "Control Lessons from the Societe Generale Fraud" that before the bloom of complex financial computer technologies, the most common way of rogue trading was so called "tickets in the drawer" method where traders simply did not enter some of their trader on the firm's books and records. However, these misbooked trades were usually exposed through an inquiry from a counterparty.

Furthermore, the development of derivatives and growing sophistication of electronic trading strategies have made rogue trading more complex and difficult to detect. (PwC 2008). According to Allen (2008), a common feature for 2000's rogue trading cases is the use of fictitious trades to better hide the large unauthorized positions. These imaginary trades are designed either compensate the risk position of actual trades, make the net risk look small or to create fake profit and loss (P&L) to hide actual earnings. To better understand the use of fictitious trades as a common tool for making the frauds, we presented two following cases of previous rogue trading incidents.

1. In the case of Société Générale, one of Europe's largest banks' trader Jerome Kerviel ended up costing around US$7.2 billion by rogue trading and was uncovered in January 2008. His primary method for creating these unauthorized positions was to make fictitious transactions that hided the risk and P&L of his true trades. By using fictitious trades, he offset the resultant risk to the bank. To do this Kerviel knew when the made trade would be confirmed by bank's control system. Relying on this knowledge, he cancelled the fictitious trade prior to the date that confirmation would be sought. When the incident was found out, there was all together 947 of these cancelled trades because Kerviel has to constantly replace cancelled fictitious transaction with new ones (Moodie 2012).

2. In the case of National Australia Bank, four options traders hide losses by entering into fictitious options deals with an internal counterparty, using off-market prices to gain earnings. Similar to Société Générale, traders knew when to delete the trader after it has compute the profits and was reported to back-office. What is more, at the end there was no checking done for an offsetting internal trades so the traders no longers need to cancel them (Allen 2008).

As summon by their cases, rogue traders have usually exploited various weakness in their company's procedures and systems. A large part of past rogue trading scandals have also been related to unfavourable human characteristics and organisation culture which have driven individuals to make frauds. The cases around rogue trading have usually involved manipulation, lying and misuse of traders' current position (Deloitte 2012). However, within our project is more insightful to focus on system and process perspective.

2.2 Common characteristics and trading control

Examining and recognising some common patterns and characteristics of the most prominent of the rogue traders has been an effective way for financial institutions to flag individuals whose actions may require more control. With this in mind, we summarize six typical patterns which are common for recent rogue trading incidents.

Trade cancellation

Large number of trade cancellations and amendments have related to past rogue trading incidents. In the case of Société Générale, "there was no procedure in place that required control functions to confirm information entered for a trade that was then canceled and nor was there a system in place for red-flagging an unusual level of trade cancellation". However, usually banks monitor cancel-and-correct activity and suspicious actions should be noticed. (Allen 2008). Still, in the age of high-frequency, fully automated and algorithmic "low latency" trading, there has been room of loopholes in the systems (Weber 2011).

Supervision

In many rogue trading incidents, there have been changes in the organisation processes which has caused confusion in people's responsibilities. In the case of Société Générale, Kerviel's immediate supervising manager left the company and the responsibility of validating his position moved to a senior trader, which gave more possibilities of fraudulent activities (Moodie 2008). Often, rogue traders have been involved with middle-office, which has allowed them to gain extensive knowledge of the company's risk controls and how to bypass them (Moodie 2008).

Vocation policy

It is mandatory for traders take two consecutive weeks of holiday in a year and during this time their position is taken by another trader. Rogue trading incidents have revealed that this procedure has not usually been followed. (Allen 2008). What is more, some banks have made exceptions for well performing traders to continue without taking any vacation. In 2002, Allied Irish bank's trader John Rusnak was uncovered for hiding losses of $691 million, was allowed to ignore mandatory vacation policies (Fuerbringer 2002).

Gross positions

Gross position refers to the risk the traders are taking. Usually banks monitor both the gross and the net positions of the traders to see how much risk they are actually taking. If the rogue trader manage to use fictitious trades to hedge the actual positions, the gross risk surveillance would reveal their fraud. (Deloitte 2012). The risk control of Société Générale registered that Kerviel's net risk looked neutral but if the bank control would have surveilled the gross position they would have seen the huge amount of risk he had taken (Allen 2008).

Cash and collateral

There has been usually need of cash and collateral to balance the real trades with fictitious ones (Allen 2008). Such as in JP Morgan rogue trading incident, trader misused the cash and collateral mechanisms to hide his unauthorized, risky positions which gained the bank almost $2.3 billion trading losses (Finma 2012). Monitoring unusual movements at the level of a single trader have revealed misbehaviors in trading (Allen 2008).

P&L

Allen (2008) explains that for hiding loses, rogue traders have generated phantom gains to show profit or fund bonuses. In the case of Societe Generale "Kerviel was reporting trading gains in excess of levels his authorized position taking could have accounted for and this should have given his management and the control functions a warning sign to investigate closely the source of his earnings".

Presented rogue trading characteristics are also summarized in Figure 1.

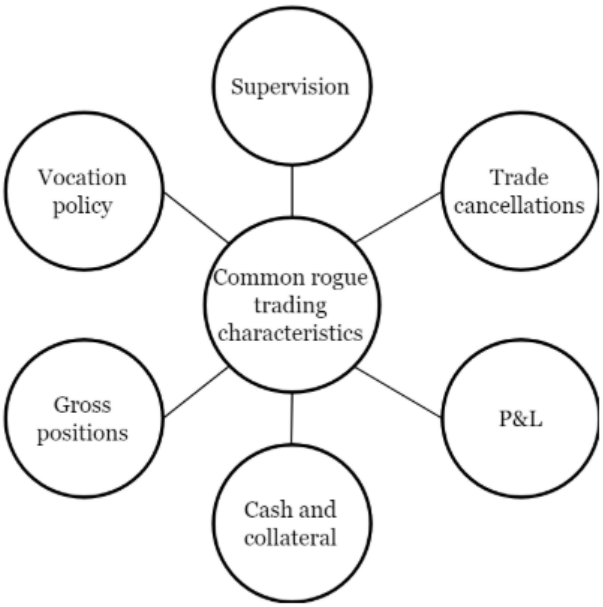

Figure 1: Common rogue trading characteristics

2.3 Our approach

In the light of previous cases of rogue trading such as Societe Generale and National Australian Bank, canceled and amended (C&A) trades play an important part of revealing suspicious trading actions. Especially, when using fictitious trades to cover the real positions, rogue traders have had to cancel and/or amend transactions.

In this project we focus on C&A control which is one of the customers over ten different trading control units. By utilizing C&A data we build a system to find indicators of misbehaving traders and flag them. This could, combined with other control units and more comprehensive investigation, help reveal possible rogue trading actions. Un-

usual level of trading cancellations, suspicious cancellation dates and counterparty activity are some patterns which to investigate. In the next chapter, analysis of the given dataset is presented.

3 Analysis of the dataset

3.1 In general

The dataset given by the customer consisted of data from multiple different sources (trading platforms) with some differing characteristics. Data coming from different data sources had different features, and the combined table given to the authors had therefore a large number of empty cells. This was because the amount of columns varied between data sources; some columns were specific to some data sources. Other clear differences between data from different sources were present too, for instance some parts of the data were comma separated while most of the data was tab separate. Also, date and time formats had variation.

Before any processing of the data, much preprocessing was required to reach a sufficient level of consistency. Preprocessing was primarily done to make the data machine readable for data analysis software, namely R. Tens of hours of manual effort was dedicated for cleaning and standardization. Main issues that where corrected in preprocessing are listed in table 1.

Table 1: Examples of preprocessing done for the dataset.

#	Description
1	All date valued cells had their content transformed to same machine readable date format
2	Cells that where shifted due to non-escaped tabulation characters where corrected to some extent, otherwise a whole row was discarded
3	Rows cut due to a new line character in a comment field on preceding row were connected whenever programmatically feasible
4	Semicolon separated data rows transformed to tab separated
5	All white space sequences longer than one were transformed to length one
6	Leading and trailing white spaces removed
7	Empty cells not reserved for free text where marked as NA

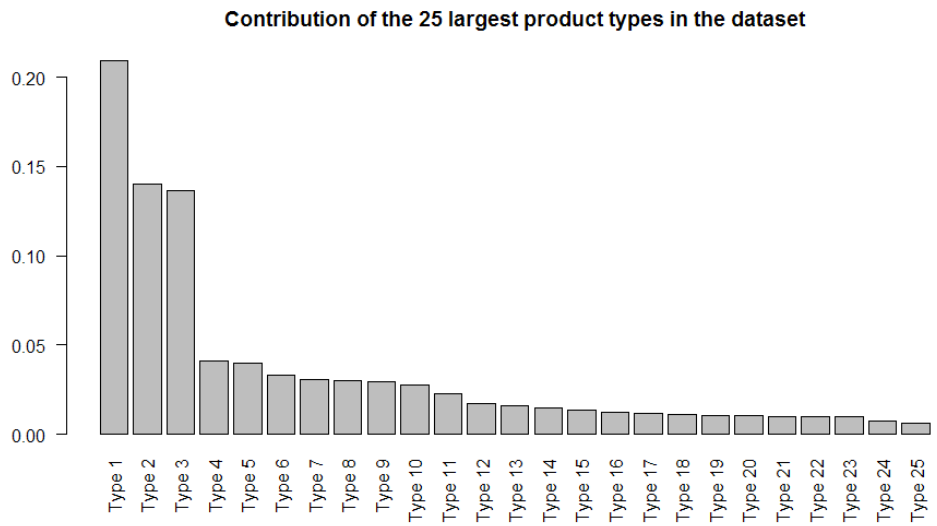

Figure 2: Distribution of dataset rows between the 25 most frequent product types.

At the end, the purpose was to enable the analysis of the data and make the analyzing more efficient. After preprocessing the data to a sufficient level, any further processing was done on a need-to basis just before sending data for a classifier. The processings could be seen as part of each corresponding classifier, but for clarity of the programmatic solution, these two were separated. An example of such processing is the aggregation of specific data features (columns) of a subset of interest of the whole dataset.

3.2 Univariate analysis

With the preprocessing out of the way, some analysis of the dataset was possible. One of the first things to be noticed was the high dimensionality of the categorical features of the data. These non-free text valued columns, which the dataset almost entirely consists of, had an almost unmanageable dimensionality, examples being the 500 product types and over 400 possible types for the amended field. Some columns appeared to have a lot of internal overlap, often because data from different data sources had a different written form for the same term. Nevertheless, the number of factor level combinations in the data is practically infinite, suggesting the need of major aggregation of data before any inference can be carried out. See figures 2, 3 and 4 for the distribution of product types, amended field types and data sources in the dataset.

In general, the dataset had 45 columns and approximately 550,000 rows. Most of the columns represented categorical variables. There was roughly eight numerical valued

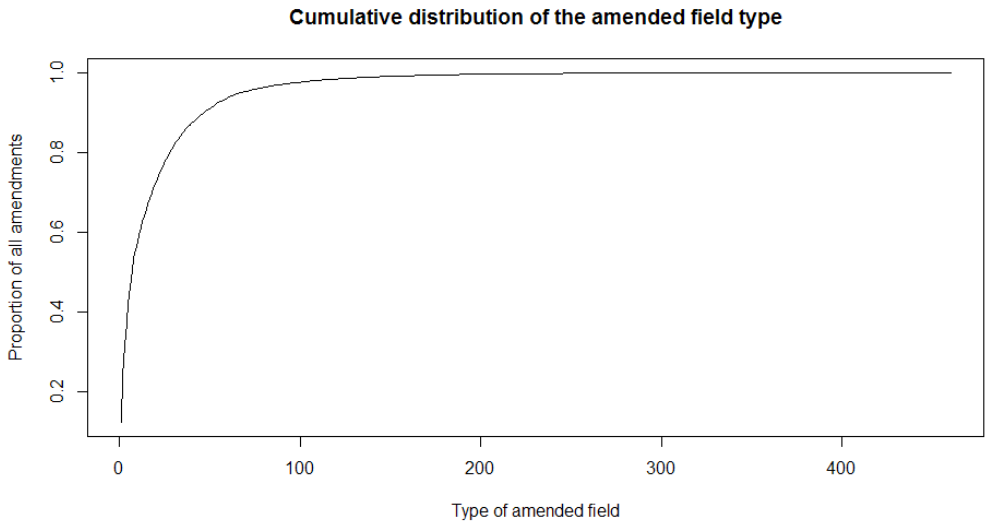

Figure 3: Cumulative distribution of all possible amended field types in the data.

identifier columns, two regular numerical valued columns, and three reserved for date times. Many of the columns for categorical variables seemed to have a inconsistent content, perhaps due to unsuccessful combining of the data from different data sources. Professional terms and abbreviations may contribute to the impression of inconsistency. We recognize that some confusing aspects of the data are the result of our limited knowledge on the domain and the client's systems.

It was noticed, that limiting the inspection of the data to, for example a single product type, or trades where trade dates had been changed, would leave too little data for analysis. This feature of the data was clearly restricting the options for ways of classifying the traders. Using any kind of black box-models would require a lot more preprocessing to get even some results out of them. This lead us to looking into some general approaches to the problem that would take into account as much of the data as possible at the same time.

Another observation on the dataset was that one trade might have multiple rows in the data, as a single row of data is just an amendment, or a cancel, made to a trade. The number of rows for one trade ranged from 1 to about 500. Discussing this with the client, we found out that for some product types a high number of amendments was normal, so analyzing the number of rows per each trade had to be done separately for each product type. As a result, we realized that the product types that most commonly

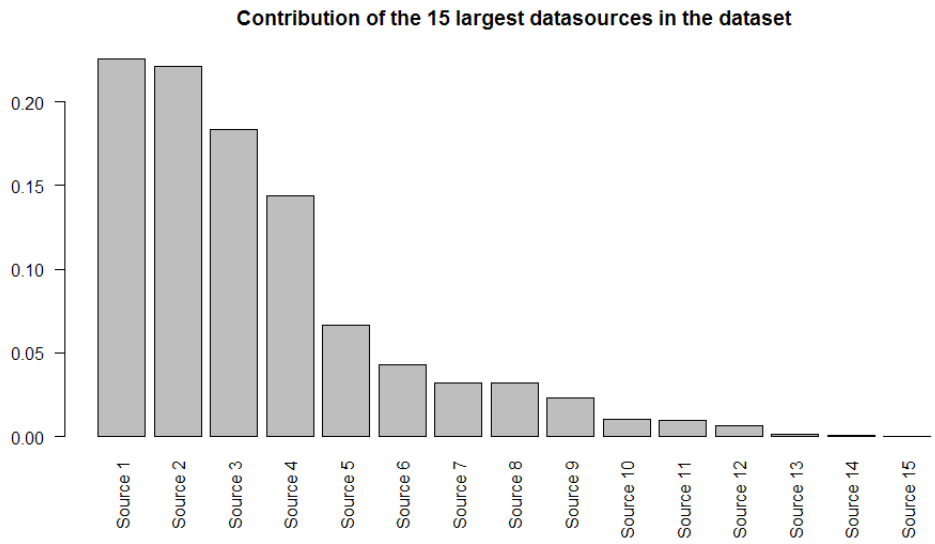

Figure 4: Distribution of dataset rows between the 15 largest data sources.

had a low number of rows did not have any outlying occurrences of trades with high number of rows.

A positive insight that was understood from the data was the fact that it is distributed fairly evenly across the whole time period with some spikes on single days. This makes it viable to further analyze the data for example, by using distributions on shorter time periods. Figure 5 shows how the data is distributed in time.

3.3 Bivariate analysis

As the data was mostly in text, our analysis focused on conditional and unconditional frequencies, and statistics of certain occurrences. To help further analysis, some basic statistics were calculated and inspected for specific variable pairs.

Analyzing variable pairs in the dataset shared, for istance, that on average a trader has C&A actions in two product types (median 1) and with 15 counterparties. On average, one counterparty has C&A actions in one product type and with three traders. These statistics vary somewhat between different product types.

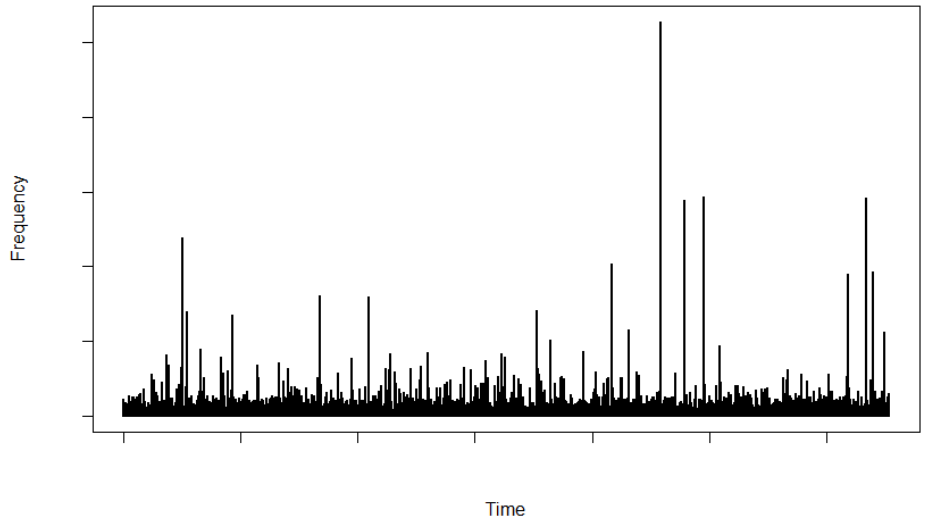

Figure 5: Distribution of the data in the time dimension.

4 Solution proposal

4.1 In general

Without any knowledge on the suspiciousness of each traders behavior, the task of ranking was rather difficult. Statistical learning (machine learning) processes for such problems was categorized as unsupervised learning. The topic was new for all the authors of this study, and it was clear from the beginning that not much of this complex problem domain could be studied by us, given the time constraint. We were largely bound to use any foresight and instinct, in addition to relying in a few assumptions. First assumption being, that the proportion of rogue traders present in the dataset is very small or nonexistent. Our second assumption was, that the proportion of traders behaving suspiciously in any way in a given time frame is rather small. The third assumption was made based on our studies to the literature on rogue trading, that the expected damage of a rogue trader is so high, that a guiding principle for the final solution should be to minimize false negatives.

Examination of the dataset started concurrently with the standardization efforts. Work started from the subset of least ambiguous financial products and their features, i.e. the parameters available in the data. While trying to get insight of the possible ways of going rogue, there was a steady process of finding different classification aspects. Aspects

that could discriminate suspicious behavior from regular one.

With the assumptions mentioned earlier, and the possibilities in the data, a target was set to find a collection of complementing aspects on which the traders could be classified against. Aspects like regularity of intervals of a traders C&A actions, or the distribution of ones C&A actions on the quarters in a year. A diverse set of classifiers can in theory cover multiple points of view to a traders behavior, and by that way, possibly utilize more of the data to more accurately rank the traders. Possibility of irregular, suspicious behavior being captured by the overall solution is then increased. The approach is partly guided by the complex nature of the problem, and the time constraint, due to which it seemed best not to aim to create any complex classification schemes. There are a few reasons for this.

First, novel classifier systems are best used by utilizing standard methods implemented in third party packages dedicated for statistical learning. To make such methods function with the data in hand, it would most likely need exception handling and/or bypassing of gaps and low quality parts in the data. The data was not completely standardized and cleared out of all issues due to the large time resources it had already taken, and would still take, if completeness in this matter was to pursued. Furthermore, both understanding perfectly the inner workings of a third party classifier candidate, and then modifying it or the data to match each other was predicted to be ineffective use of time. In addition, such packages often use lower level languages like C++ to speed the computation. Use of real programming languages like C++ would increase the required time for implementation, compared to using for instance R, and its low level functions mostly written in R, too.

Second, these novel classifiers tend to become less understandable by non-experts, and therefore would make the validation of a classifiers output at the customer side, in this specific case with non-labeled data, even harder. The main emphasis here being inference from the data, rather than forecasting, the use of more human interpretable methods should be justifiable. Especially, because the information from the mechanism built here will be used in conjunction with a multitude of other risk mitigation mechanisms.

Third, all regression and classification schemes need to balance (Hastie et al. 2006,

p. 33) between bias and variance in regard the residuals. That is, with each classifier, one has to avoid fitting random fluctuations, but at the same time try to minimize any systematic error in the results. Novel parametric methods would thereby need careful adjustment of parameters to avoid fitting to the random noise in data. The problem domain is clearly too vast to be understood by the authors in the given time frame in such a level to assess the appropriateness of the level of fit of a parametric method. Proper fitting would, again, likely demand use of method specific 'tuning' heuristics. These are both reasons to avoid complex models, and suggests the use of non-parametric models, when possible.

As a result of the three points presented, a choice was done to build the classification system as a combination of simple parametric methods. The classifiers are easy enough to be interpretable by non-professionals of statistical analysis, and additionally, permit to a sufficient level the exploratory tuning of the small set of parameters they have. Morover, in simplicity, our process can be divided into four parts; exploration of the data, experimenting with the data, implementing the classifiers, and finally tuning the small set of parameters they have.

4.2 Combining the results for the classifiers

Results, that is rankings, from the individual classifiers, have to be combined in order to maximize insight. A natural choice seems to be to consider a fixed percentage of top ranking traders for each classifier. The idea is illustrated in Figure 6.

After ranking the traders on a variety of criteria, the solution gives flags for individuals having a ranking above a given threshold. This can be for instance the top 10 % ranking. Then, all traders with one or more flags form a group of initially suspicious outliers. Individuals with no flags are considered as behaving normally. Traders with one or more flags can be sorted by their flag count, to set the individuals into order of suspiciousness. A graphical explanation of this combining and sorting is shown in Figure 7. By varying the thresholds for the flagging process one can adjust the final result.

4.3 Classifiers

Our proposal for classifying the traders and counterparties consists of six different ways of looking at them. Each one builds on the idea of comparing the data to itself. One of

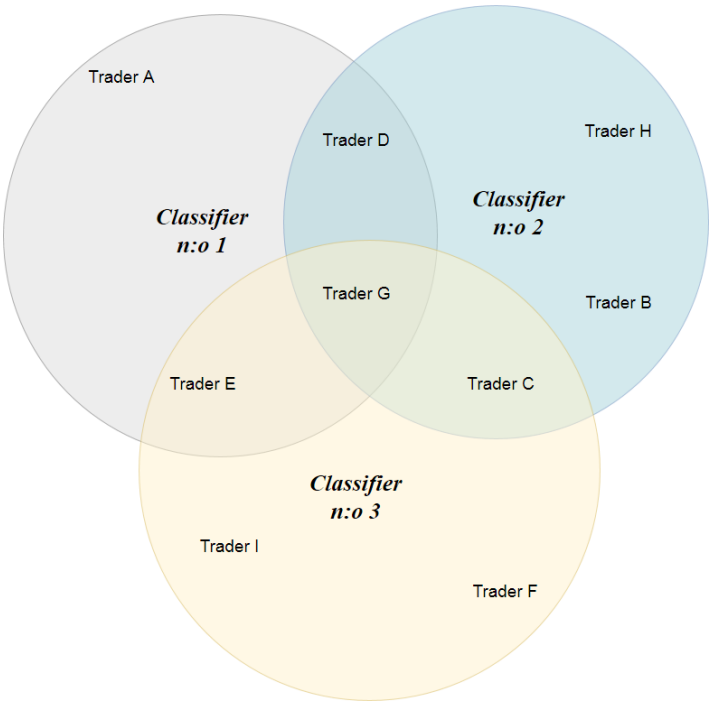

Figure 6: Highest ranking traders of each classifier presented in a Venn diagram. Multiple flags imply a trader occurring in top ranks for multiple classifiers.

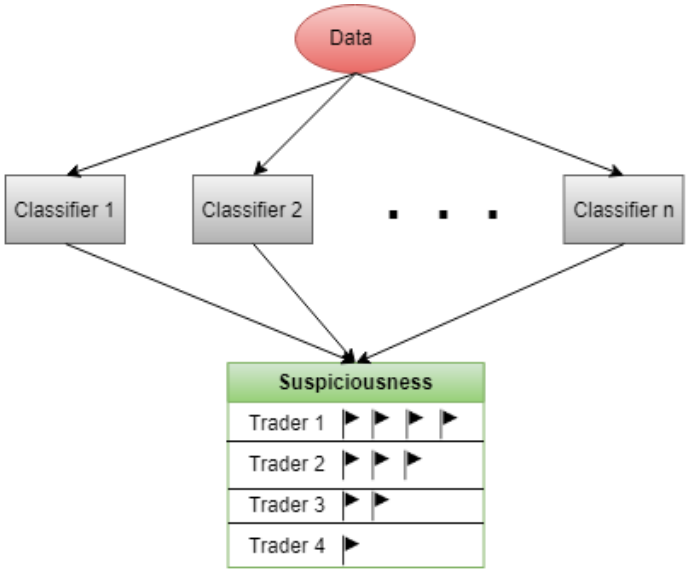

Figure 7: A flowchart of the basic functioning of the software

them, classifier n:o 3, compares the traders against themselves, trying to reveal changes in their behavior. The classifiers rely on statistical assumptions of the behavior of the actors, to allow the use of statistical measures for deviations from a reference. We strove to have the comparisons done with appropriate, representative subsets of the data.

Classifier n:o 1

This classifier uses temporal information in the amendments & cancellations data to assess the abnormality of a traders C&A behavior. The algorithm tests whether the distribution of a traders C&A actions on the weekdays is distributed similarly as the traders product and year specific average. In other words, a goodness-of-fit is assessed with the use of the appropriate chi-squared distribution. Difference from the average C&A action may signify an abnormal trading behavior. The algorithm is given in the pseudo code below:

```
FOR EACH Year, ProductType
    DO  Store the average distribution of C&A actions on the weekdays
        for this Year-ProductType combination
    FOR EACH Trader
        DO  Calculate X2-test statistic for the weekday specific C&A using
            the average obtained above
        DO  Store the log-p-value of the test
    END
END
```

Classifier n:o 2

This classifier also uses the temporal information in the C&A data to assess the abnormality of a traders behavior. This is essentially the same as classifier n:o 1, but works on a quarter level. The algorithm test whether a traders C&A amounts are similarly distributed on the quarters as is the local average. The algorithm is given in the pseudo code below:

```
FOR EACH Year, ProductType
    DO   Store the average distribution of C&A actions on the quarters
         for this Year-ProductType combination
    FOR EACH Trader
         DO  Calculate the X2-test statistic for the quarterly C&A
             amounts using the average obtained above
         DO  Store the log-p-value of the test
    END
END
```

Classifier n:o 3

In this classifier proposal, the number of C&A actions of a trader per year is assumed to be Poisson distributed. The interval between these actions is therefore assumed to be distributed according to an exponential distribution. For each trader, the rate parameter λ of the exponential distribution corresponding to the intervals of ones C&A actions is calculated. A maximum likelihood estimate (Abdulaziz, David 2001) is used. Each pair of consecutive years is compared in regard of the λ-parameter. A year-to-year relative change exceeding the given thresholds (increase or decrease) leads to a flagging of the trader for the latter year. Pseudo-code of the classifier is given below:

```
FOR EACH Trader
   FOR EACH Year Y
      DO   Calculate the lambda-parameter for the rate of the
           traders C&A actions for year Y. At least four
           timepoints required, giving three intervals.
   END
   FOR EACH Year Y
      DO   Calculate the relative change in lambda-parameter
           from year Y-1 to Y
      DO   Store the relative change
   END
END
```

Only traders with at least three intervals (four timepoints) are considered, as less data would not give enough robust maximum likelihood estimates. Quality of the λ-parameter estimate changes as the amount of datapoints for a trader changes.

Classifier n:o 4

This classifier is designed to find abnormality from trader-counterparty pairs. It does it by first counting the number of C&A transactions per each counterparty for every trader. Then it calculates the portion of the transactions that are made with the counterparty that appears most often out of all of a trader's transactions as a percentage. The highest x-percentile is then flagged as suspicious. This classifier is meant to find out if a trader is using/collaborating with a counterparty in a possible rogue trading scheme or if there is something wrong in the process with a particular counterparty. Pseudo-code is given below:

```
FOR EACH Trader
    DO Calculate number of transactions per counterparty
    DO Calculate MAX(counterparty) / all transactions
    DO Sort traders to decreasing order by percentage
END
```

Classifier n:o 5

This is first of two classifiers made to find specific qualities from single rows of data. It counts all the trades of a single trader for which the value date is before the trade date (so called back valued trade). Then it separates those which were canceled from those which were only amended and marks a flag for the trader if he had amended ones and leaves a comment if there are canceled ones. On top of that this classifier also marks the number of amended and canceled back valued trades to separate columns. Pseudo code below:

```
DO  Filter dataset to those rows that have Value date < Trade date
DO  Aggregate filtered data to have only one row per each trade
FOR EACH  Row of filtered and aggregated data
    IF  Trade was canceled
        DO  Add 1 to the traders canceled back valued trades
        IF  Number of trader's canceled bvt:s is 1
            DO  Assign a flag for the trader
        END
    END
    IF  Trade was amended
        DO  Add 1 to the traders amended back valued trades
        IF  Number of trader's amended bvt:s is 1
            DO  Assign a flag for the trader
        END
    END
END
```

Classifier n:o 6

This is the second classifier made to find specific qualities from single rows of data. It finds the trades that were amended after their value date. As is might be possible that the status of the trade (or something else that is not necessarily suspicious) was changed after the value date, this classifier only finds the trades whose value and trade dates were moved after the value date. Flags are given similarly to classifier n:o 5 in that if a trader has any amendments after value date, he gets a flag and the number of these amendments is marked to a separate column. Pseudo-code given below:

```
DO  Filter dataset to those rows that have Value date < Study date
    and Amended field = Trade date or Value date
DO  Aggregate filtered data to have only one row per each trade
FOR EACH  Row of filtered and aggregated data
    DO  Add an amendment after value date to the trader
    IF  Number of trader's amendments after value date = 1
        DO  Assign a flag for the trader
    END
END
```

5 Model evaluation, results

5.1 Testing the classifiers separately

Classifier n:o 1

Taking a logarithm of the p-values of the chi-squared-test performed in this classifier enhances the visual observation of the distance between the traders, and thus makes it easier to differentiate them. By plotting the log-p-values given by this classifier, seen in Figure 8, we can see that most of them are 0 (or close to). This indicates that the classifier is able to find some clear outliers from the data.

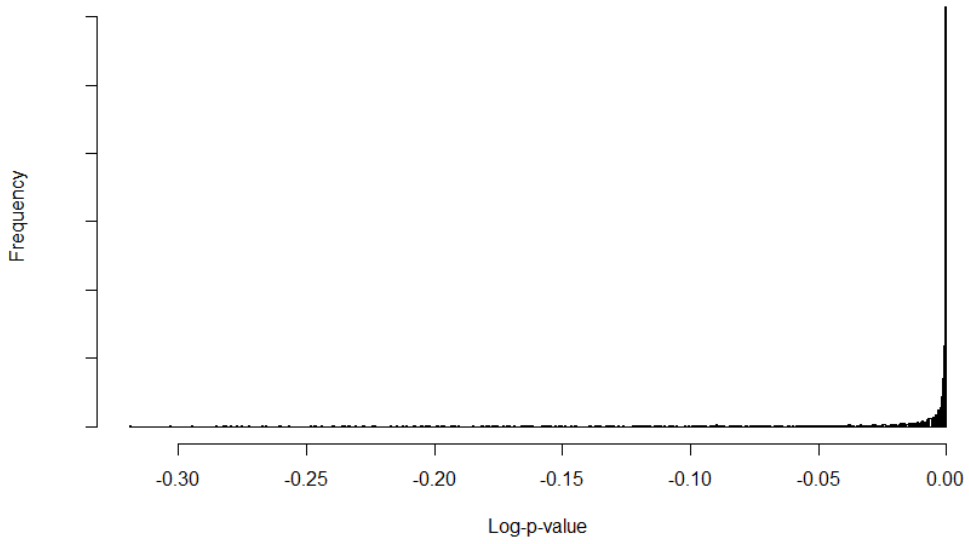

Figure 8: Distribution of the log-p-values from classifier n:o 1

To further validate that the classifier finds the right kinds of outliers, we took some examples and compared them to the average distributions to see if they really are different on the lower end of log-p-values and similar in the higher end. We found out that for all of the years present in our dataset, the average distributions had C&A actions fairly evenly distributed over all weekdays. The traders at the lower end of the log-p-value distribution had high spikes on single days, which is clearly different from the average.

The problem this poses has to do with the number of C&A actions a trader has. As traders who only have one C&A action during the whole year get a distribution with a

high relative spike on one weekday, they automatically get flagged. Additionally, if a trader has C&A actions on only one day in a year, he/she also gets flagged. To decrease the amount of probable false positives, without forgetting the earlier stated guideline of minimizing false negatives, we decided to filter out all traders with C&A actions on less than 5 separate days in a year. This smoothed the curve of the distribution of the log-p values a bit, with a bit more values different than zero relative to values equal to zero, but the form staid the same.

With this done, we needed to decide the cut point under which traders get flagged. Our evaluation for this is the bottom 5 %. With this value, and our data, about 1 % of the traders in the whole data get a flag every year and in total less than 5% get flags. This goes well together with our assumption that the number of rogue traders in our data is small to nonexistent. Furthermore, we also tested the stability of our classifier by comparing the actual values of the cut point yearly (and from the whole data). This classifier was deemed stable, as the cut points fluctuated only slightly with their midpoint being $\log(p) \approx -0.29$.

Classifier n:o 2

The testing done to this classifier is similar to the testing done to classifier n:o 1, as this classifier functions almost identically to it. For this classifier we also decided to set the cut point at 5 % for the same reasons as classifier n:o 1. This classifier also seems stable with it is cut points fluctuating around $\log(p) \approx -0.46$.

Classifier n:o 3

Our initial implementation of this classifier considered every amendment made by a trader during the same day a separate entry when calculating the intervals. This resulted in some cases to very large year-to-year relative changes in the λ-parameter (from about 10^{-1} to 10^{4}). This observation lead to a decision to consider all traders C&A that had happened during the same day as one action. This brought the changes in λ to a more realistic level, and as a consequence, improved the distribution of all traders' λ's for our purposes. The distribution of the λ parameters can be seen from Figures 9 and 10. Here the positive valued relative change means a rise in the rate of C&A actions compared to the previous year and negative values mean a drop in rate.

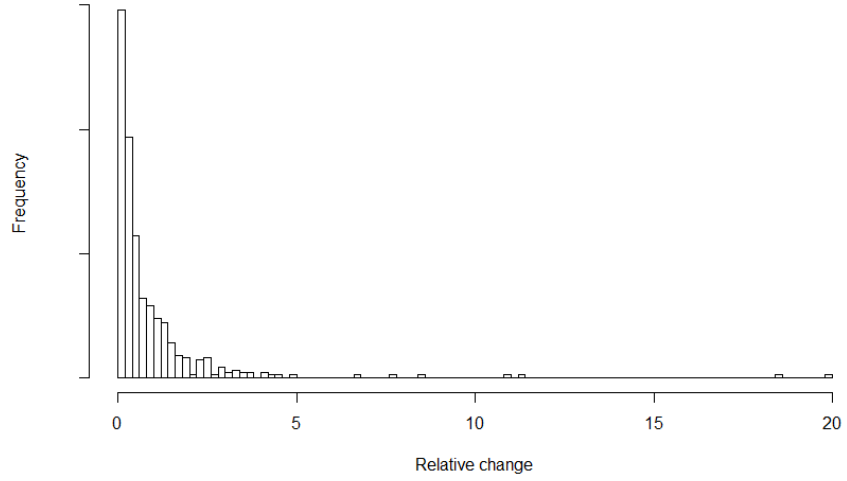

Figure 9: Histogram of the relative changes bigger than 0 from classifier n:o 3

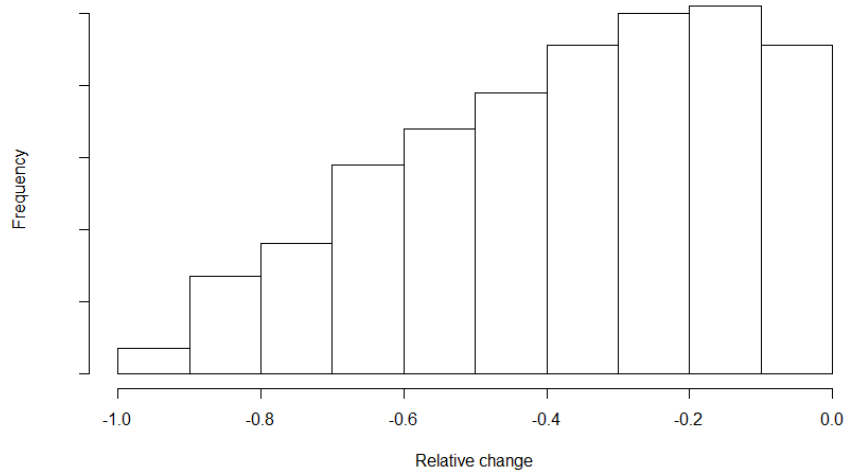

Figure 10: Histogram of the relative changes smaller than 0 from classifier n:o 3

Originally, we had planned to flag every trader (for every year) that had doubled or halved their rate of C&A actions. This would mean flagging traders with year-to-year relative change greater than 1 or less than -0.5. Setting a constant year-to-year limit like this was seen a good option, because with percentiles derived from the resultant population distribution, someone would get flagged regardless of their behavior being suspicious or not. But again, setting a constant boundary that actually separates suspicious behavior from non-suspicious behavior does need in depth theoretical knowledge

of the subject.

With the constant limit presented above, this classifier flagged about 3 % of the traders, yearly, and about 10 % of all traders during the whole time interval in the data. These are quite a bit higher proportions than with our first four classifiers. Being aware of the fact that our theoretical knowledge of the domain is not that good, we decided to test the flagging amount of this classifier with the thresholds derived from the resultant population, as was done with the classifiers 1-4.We ended up using the top and bottom 5 % as the thresholds, lowering the number of flags to the same level as the other classifiers. The benefit of this is that all of our classifiers are balanced, but there is a downside. The bottom limit stays quite stably around -0.7 but the upper limit fluctuates quite heavily between 1 and 2. This means that this classifier might not be as stable as one might want, depending largely on the data that is fed to it.

Classifier n:o 4

Initial testing of this classifier further highlighted the issue of flagging traders with a small number of C&A actions. The fact that this classifier views everyone with high percentage of C&A actions with their most-traded-with counterparty as suspicious, means that every trader with only one row in the data gets automatically flagged, as they have 100 % of their C&A actions with the same counter party. Having a lot of traders at 100 % would have also made it hard to differentiate between traders in the upper end of the distribution. To prevent this, we decided to limit the inspection with this classifier to only those traders that have more than 10 rows of data.

Still, we faced the issue of there being a lot of traders that had all of their C&A actions with the same counter party. This can be seen from Figure 11. As many of the traders at 100 % had quite a lot of rows in the data, we decided to automatically flag all of them and remove them from the list of percentages for the percentile calculation. This made it so that it is more plausible to get meaningful percentiles from the upper end of the distribution and actually being able to differentiate suspicious action from non-suspicious action. The data with these traders removed can be seen in Figure 12.

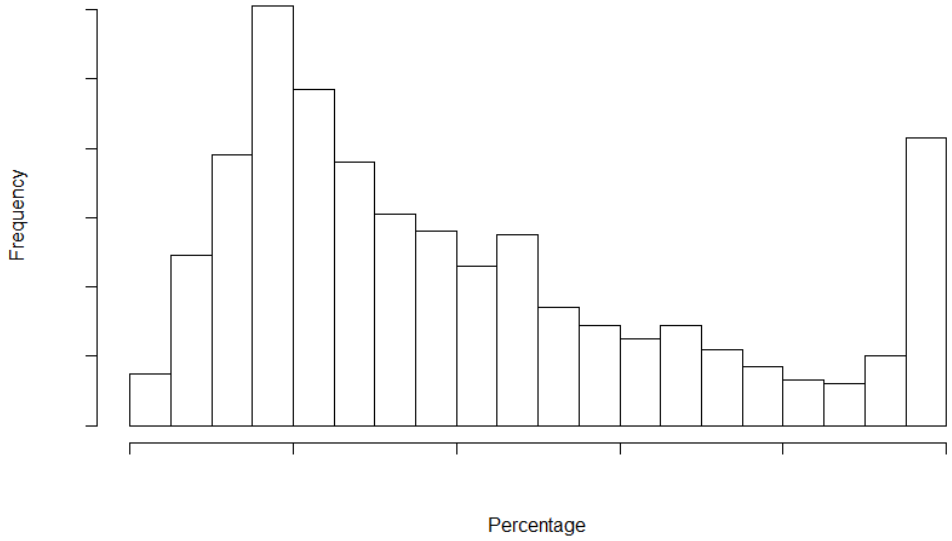

Figure 11: Histogram of the percentages with the most-traded-with counter party

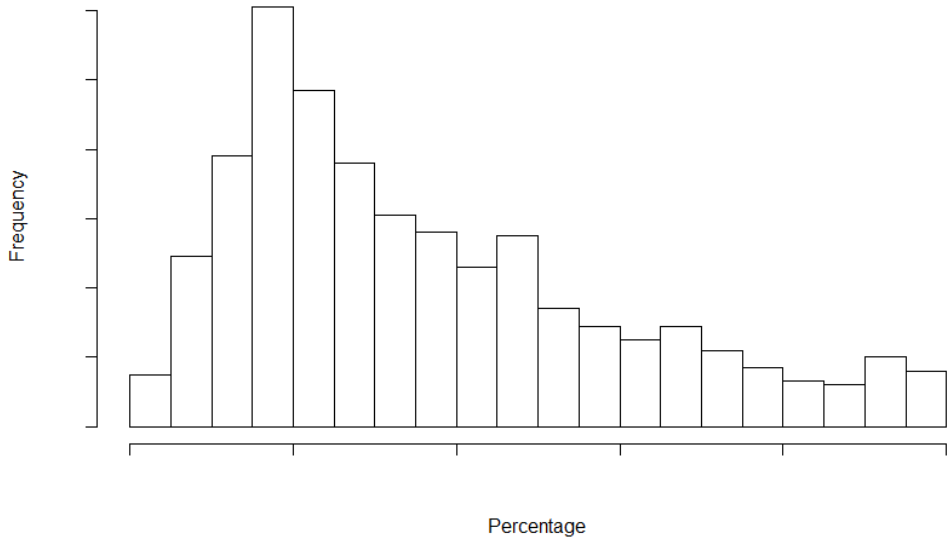

Figure 12: Histogram of the percentages with the most-traded-with counter party with the traders with 100% of their C&A actions with one counter party removed

For this classifier we also decided to flag the top 5-percentile. The percentages of traders flagged at year level, and overall in the time frame of the data are roughly the same as

in the first three classifiers. This classifier also seems quite stable with the cut point percentage fluctuating only slightly around 0.86.

Classifier n:o 5

Initially this classifier was built to give flag every back valued trade a trader has. We quickly noticed this implementation did not give desirable results as some traders have over 500 back valued trades, them then completely cluttering the amount of flags giving too much emphasis on this classifier. We decided to mark the amount of back valued trades to a separate variable and have the flag marking whether the trader has any back valued trades or not.

This seemed to do what it was intended to except that we realized that almost all of the back valued trades that were canceled had a comment saying that they are back valued, meaning that they were already caught in another control system inside the bank. We decided that separating the canceled ones from the ones that were only amended would be a good idea and decided to flag the amended ones and write a comment if the trader has canceled ones as well (or just canceled ones). The decision not to flag canceled back valued trades was mainly due to the fact that with our data, this classifier happened to give the same number of flags as the other classifiers without them.

Classifier n:o 6

Similarly to classifier n:o 5, this classifier originally gave a flag for each amendment after value date. It also ended up cluttering the flag column, so we decided to go with a similar approach as in classifier n:o 5. Marking only one flag if a trader has any amends after value date, with our filtering of the data happened again to give about the same number of flags as the other classifiers. It is good to note that there are other "illegal" amendment types after value date than just the value and trade dates, and adding those to this classifier would increase the amount of flags. As we are not quite sure what they are, we decided to keep this implementation.

5.2 Testing the classifiers jointly

Running all of our classifiers together with the whole data (all years together) resulted in about 20 % of the traders getting at least one flag, with the maximum result being five flags for the top flagged trader. On a year level, about 8 % of the traders were get-

ting flags with the top result varying between three and five flags.

The number of flags a trader can in theory get from the classifiers in a year is six flags. Assuming the trader has trades only in one product type. To narrow the gap between the maximum possible flags one can obtain, and the top results observed, we decided to raise a bit the threshold levels. We ended up raising the threshold percentiles to 7 % for all of the first four classifiers. The average top result is now close to the yearly maximum number of flags.

The results of the analyses were not aligned with the assumption we made in the beginning, which stated that the presence of rogue behavior in the data is close to nonexistent. Nevertheless, most of our classifiers are pretty stable, i.e. not overly dependent on the data used, and thereby making use of the whole range of ranks is well arguable. In the end, the tuning of the threshold levels raised the top result for a trader to eight (all years) and similarly the proportion of traders getting any flags increased to roughly 22 % (all year).

6 Analysis of the results

The main remark to the results is that the implemented solution places a lot emphasis on the distribution-utilizing classifiers, as there are more of them. Although pattern recognition is a big part of finding rogue trading behavior, the implemented classifiers focus on patterns in only one dimension, quantity. The discussions with customer's representatives lead us to believe that there are also other types of patterns, for example distances between trade and amendment dates for a trader-counter party pairs, that could be investigated to find suspicious behavior, but with the theoretical knowledge and data we have, validating the functioning of these kinds of classifiers would not be possible.

Another remark is that a problem of this complexity to be at least somewhat covered, needs a lot more than these six classifiers. The solution is not even near being an exhaustive scan of all possible rogue behavior observable in the given data. Intercepting even some proportion of it requires a broad set of different viewpoints to look from. In its current state, our model can not be used as an all inclusive C&A control. However, we would like to emphasize that our implemented model is a good framework for building

better C&A control. It considers a few important aspects regarding suspicious behavior but does so on a rather general level. With better theoretical knowledge of trading and the markets, implementing improvements becomes highly possible.

7 Recommendations

As mentioned in the literature review, a robust rogue trading surveillance consist of several different trading control systems. The customer has over twenty different trading controls, of which C&A trading control is just one. However, what emerges from previous rogue trading studies, monitoring of C&A is a key element in the detection of suspicious trading activity (Allen 2008; Finma 2012; McConnell 2014). Even if the build model is quite simple, when testing it with C&A data, some suspicious trading activity was discovered. In the future, flagged traders who for instance appear to be using a suspiciously high number of cancellations, should be taken to closer examination. Checking the transactions with the counterparty is one way of ensure the genuineness of the trade. To better investigate the possible abnormalities, the given flags from C&A data should be combined with other trading data.

Furthermore, as examined in the literature review, previous rogue trading incident have had common characteristics, which act as a good lessons for control personnel to better prevent the whole phenomenon. In addition to C&A control, enough tight supervision for traders, surveillance of vocation policies and traders gross positions are good to keep in mind. Especially, making sure that traders cannot interfere with the reporting of their own transactions and monitoring traders' system accesses are key principles to remember (Weber 2011). Moreover, monitoring amount cash & collateral flows and suspicious earning patterns and traders performance levels are some general patterns to pay attention. In addition, for future rogue trading detection, the customer should ensure that the data is consistent and accurate so it can be verified easily through reconciliation with external parties. At the beginning of the project, we noticed that the given dataset consisted of several different sources and to be able to analyse it, we had to standardize and clean it with many hours of manual work.

Finally, quite often the rogue trading incidents relate to a system failures. In the case of UBS, front office supervisors monitored C&A trades using an online system, the Super-

visory Control Portal (SCP), which generated daily reports of C&A trades, which were sent to front office supervisors for investigation and sign-off. However, futures trades did not make any SCP alerts due to the lack of a data feed to SCP. System failure was revealed not until many months afterward, when a massive rogue trader incidents was exposed, executed by fictitious future trades (Finma 2012). In the age of high-frequency and algorithmic trading, ensuring that the control systems work by doing stress tests and recruiting computer savvy risk control people, are essential ways of preventing future trading scandals.

References

[1] Abdulaziz, E. and David M. R. (2001) A Bayesian Look at Classical Estimation: The Exponential Distribution. *Journal of Statistics Education*, 9(1).

[2] Allen, S. (2008). *Control Lessons from the Société Générale Fraud*. Bank Accounting & Finance, 21(6), pp. 29-35.

[3] Deloitte. (2012). *Rogue Trading – Risk Management Controls and Culture*. Presentation. Available on: http://docplayer.net/15225833-Rogue-trading-risk-management-controls-and-culture.html.

[4] Authority, S. F. M. S. (2012). *Summary report-FINMA investigation into the events surrounding trading losses of USD 2.3 billion incurred by the investment division of USB AG in London*. 26th November, pp. 13-14

[5] Fuerbringer, J. (2002). *Bank report says trader had bold plot*, The New York Times, 15.3.2002. Available on: https://www.nytimes.com/2002/03/15/business/bank-report-says-trader-had-bold-plot.html.

[6] Graaf, M. & Kidd, J. (2012). *Rogue Trading - Risk Management Controls and Culture*. Presentation, Global Association of Risk Professionals. Available on: http://docplayer.net/15225833-Rogue-trading-risk-management-controls-and-culture.html.

[7] James, G., Hastie, T., Witten, D. & Tibshirani, R. (2006) *Introduction to Statistical Learning with applications in R*. New York. Springer.

[8] McConnell, P. *Dissecting the JPMorgan whale:a post-moderm*, Journal of Operational Risk, 9(2), pp. 59-100.

[9] Noonah, L. (2017). *Rogue traders will always bank on getting away with it*. Financial Times. 19.11.2017 Available at: `https://www.ft.com/content/2103efb6-b4f1-11e7-8007-554f9eaa90ba`.

[10] PwC. (2008). *Rogue Trading*. International presentation for clients.

[11] Ross, J. (2018). *Rogue Trader: How I Brought down Barings Bank and Shook the Financial World by Nick Leeson*, The Academy of Management Review, 22(4), pp. 1006-1010.

[12] Customer. (2018). *Introductory material about rogue trading*. Internally used.

[13] Weber, B.(2011). *High Frequency Trading: The growing threat of rogue trading*, Business Strategy Review, 22(2), pp. 50-53.

Appendix A Example output of our program

Trader	# flags	Description	Back valued trades AMEND	Back valued trades CANCEL	Amendments after value date
Trader 1	5	Amend after value date. Back valued trade AMEND. CP. PREVYEAR YEAR Interval. YEAR SW Weekly distribution	10	0	1
Trader 2	4	Back valued trade AMEND. CP. PREVYEAR YEAR Interval. YEAR Spot Quarterly distribution	2	0	0
Trader 3	3	Back valued trade CANCEL. Back valued trade AMEND. CP. YEAR bond Weekly distribution	11	15	0
Trader 4	3	Amend after value date. PREVYEAR YEAR Interval. YEAR SP Quarterly distribution	0	0	6
Trader 5	3	Amend after value date and YEAR SP Quarterly distribution. YEAR SP Weekly distribution	0	0	2
Trader 6	3	Back valued trade AMEND. YEAR BOND Quarterly distribution. YEAR BOND Weekly distribution	2	0	0
Trader 7	2	CP. YEAR FUTURE Weekly distribution	0	0	0
Trader 8	2	Back valued trade AMEND. PREYEAR YEAR Interval	5	0	0
Trader 9	2	YEAR bond Quarterly distribution. YEAR bond Weekly distribution	0	0	0
Trader 10	1	Back valued trade CANCEL. CP	0	64	0

Appendix B Self-Assessment

Enclosed can be found a short self-assessment addressing following questions:

1. How was the project executed?
2. What was the real amount of effort?
3. In what regard was the project successful?
4. In what regard was it less so?
5. What could have been done better, in hindsight?

Our group was a mixture of mathematical and business skills consisting of industrial engineering and management, computer science and mathematics students. The group had an additional fourth member who by misfortune, had to withdraw from the project. One of the students acted as a project leader and was responsible for keeping contact with the client company and handling the course responsibilities.

1. At the beginning of the project, the group mutually agreed on necessary milestones and made a preliminary schedule for the project. In addition, the group orientated themselves on basic theory of rogue trading. Before starting any modeling, the brainstorming and ideation was done together. In addition, the group met the customer a couple of times to get a better understanding of the project and hear about the bank's ideas and expectations. The project proceeded so that the more computer savvy students analyzed the given data with R and IEM student used her previous knowledge of trading and different financial instruments to come up ideas how to find suspicious trading patterns. Furthermore, the group met the customer several time during the project and asked defined questions by email.

2. We feel that the amount of effort was sufficient in the given time frame and size of the course. The two major parts of the project was finding literature to better understand the theory and at the end, the work for putting the model together. In our opinion, the whole team put a lot of effort on actually understanding the theory and concept of rogue trading. Without the extensive amount of time put on reading previous rogue trading incidents and finding out common patterns, it would have been very challenging to come up any sufficient classifiers.

3. The team was especially satisfied that even if the classifier based model was very sim-

ple and cannot be reveal any actual rogue traders, it still gave reasonable indicators of some traders acting suspiciously. The team also feels that they have succesfully grasped the basics of rogue trading given that the team had very little theoretical knowledge of the subject before the project.

4. As is often the case with such projects, the project plan is not equal with the actual execution. We think that the while the amount of effort was sufficient, the workload could have been more evenly distributed over the spring period. The group members had other courses and work which took resources that could have been used for the project. However, the absence of the fourth group member also affect the amount of work and time we could provide for the course.

5. Team's time management could have been improved because lot of work, particularly the presentations were created quite last minute. Moreover, there could have been executed a better research on mathematical and computer science papers which focus on previous work in C&A control and detecting outliers from the trading data.