Writing sample 1– independent project report

Design and Evaluation of a New Workflow for Customer Money-laundering Risk Ranking

Abstract: Evaluating customers' risk level is an important part of anti-money laundering practice in commercial banks. This project aims to figure out an optimized solution when a small-sized bank, Huaxia Bank, switched its risk ranking workflow to a centralized one as its business expand dramatically. Compared with previous risk ranking process completed within each sub-branch, the bank has to take into consideration the number of experts in a new center, the completion time of whole process of risk ranking, and requirements from its own committee and the external regulatory department. A Discrete Event Simulation (DES) model can precisely describe the decomposition of the centralized risk ranking workflow. Most of the output obtained from the model presented a good fit to real-world situation. However, there are still some improvements needed in order to get a more common conclusion. Finally, we discuss the variation of group size with which the center submitted to the committee each time, and looked into the effect of group size on total time spent in risk ranking in real-life situations.

Keywords: money laundering risk ranking; centralized workflow; Discrete Event Simulation; sensitivity analysis.

1. Background

1.1 Money laundering customer risk ranking

Money laundering is the process of disguising the proceeds of crime and integrating it into the legitimate financial system. It generally involves three steps: placing illicit proceeds into the financial system; separating the criminal proceeds from their origin; and disguising the illicit proceeds through apparently legitimate transactions. In any case, criminals or suspects have to open an account in a bank before their transactions start. And the critical step on the side of the bank is to spot these seemingly abnormal accounts when a customer goes to the counter, demanding to create an account. We say 'seemingly abnormal' because the bank may not have direct evidence to prove that these accounts are involved in money laundering, but does have sufficient reasons to monitor all the transactions associated with these accounts. Then the bank needs to

have a valid working process that enables to give each account a risk level. Usually, money laundering risk ranking scales from low, medium to high. A medium or high level label tells this account is abnormal.

1.2 Two types of workflows for risk ranking

There are two types of workflows that apply to money laundering risk ranking. The first type, the branch-level workflow (figure1), is based on each branch within the bank. That means, any branch just processes risk-ranking with access to information of the customers who open an account in this branch. A manager or several managers are in charge of risk ranking process. They do not need to refer to any information outside the branch. The second type, centralized operation workflow (figure 2), requires a bank to construct a center that specializes in risk-ranking. Under the second pattern, it's no need for branches to evaluate the risk of an account. They help customers open accounts, record customers' information, save all the documents in the system and submitted to the center; Of course, the counter must collect sufficient information needed for analysis. In the center, experts take the responsibility of selecting abnormal accounts that are ready for a committee to determine a final risk level, medium or high. In general, small and medium sized banks adopt the first type of workflow while large-sized ones prefer an operation center even though it probably incurs far more expenses.

Figure1. Branch-level system diagram

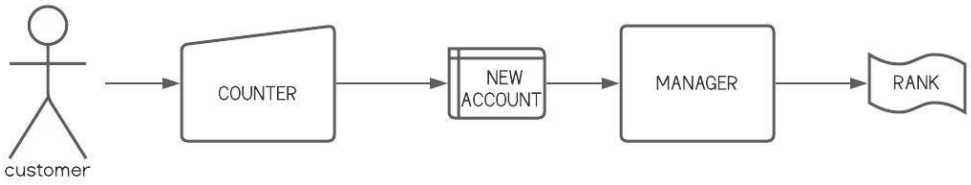
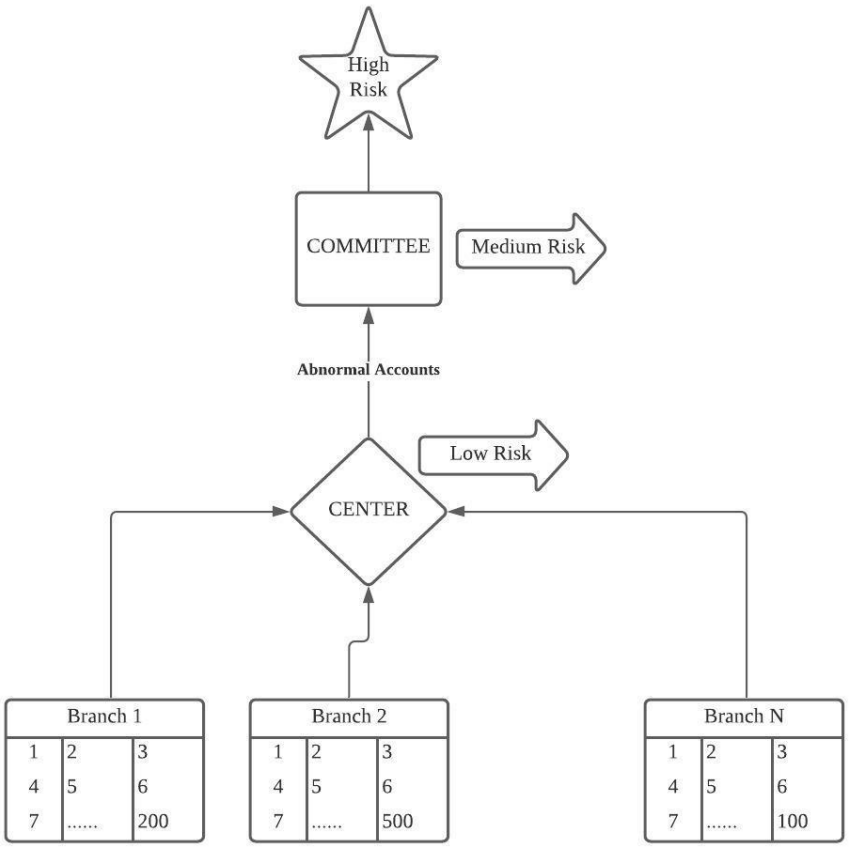


Figure2. Centralized system diagram



2. Problem Identification Statement

Huaxia Bank (refer it as Bank X) is a small-sized commercial bank located in Hangzhou, China. It has two braches which conduct risk ranking respectively by two managers. But the managers within a branch assume so many roles that they spent a little time in evaluating money laundering risk of accounts. Some customer accounts were given unreasonable ranks, which may put the bank in the position of a penalty receiver. What's more, the bank would open two new branches soon and they didn't have more candidates. So they wanted to establish an operation center and switched to the centralized workflow with the goals as follows.

a) How many experts should the center hire to handle the same workload as its two branches had completed risk ranking before?
b) If the bank wants to open one or two more branches, do they need to hire more experts in the center?
c) The bank must follow some regulation limitations, among which the most important one is the deadline for risk ranking. When a new account is opened, it has to be given the rank level within the following 14 workdays, i.e 112 hours.
d) The bank has to consider the requirement from committee. The committee

members come from several departments. It's hard for them to aggregate or hold a meeting to discuss and decide the final risk level too often, like every day. Therefore, the center should submit an analysis report serving as a package which includes several abnormal accounts as a whole. In addition, the committee wants the center to reduce the frequency of submission as much as possible.

Actually, in order to prepare a new operation center, Bank X made a survey in its Branch A and finally established an operation center in July, 2016. There was a pilot phase in June, 2016 for conducting survey and collect real data. In this case, building a model and generate some measures can help understand the different conditions which happened in Bank X in the latter part of 2016.

3. Modeling Assumptions & Justifications

In any modeling effort, assumptions need to be made since we cannot expect to capture all the factors influencing the identified problem. The task is simplified by reducing the number of factors under consideration. Then, relationships among the variables may be determined. Again, by assuming relatively simple relationships, we can reduce the complexity of the problem. The assumptions as follows are consistent across the two modeled consecutive processes, the one for opening an account and the other for risk ranking within each type of workflow.

a) Treat private customer and corporate customer follow the same distribution but with different parameters, like the arrival rate and time spent in open an account.
b) The arrival of customers follows one pattern. That means customers come in with the same rate (exponentially distributed) no matter which branch they step in.
c) The process time by a manager or by a committee follow a triangle distribution. This is because in branch-level workflow, there is no exact data of process time by a manager; while in an operation center, no exact time spent by the committee to decide the final risk level. We just have the most likely values, minimum values and maximum values.
d) In branch-level workflow, each branch open accounts and complete risk ranking independently.
e) An expert in the operation center selects a fixed percentage of private and corporate accounts as abnormal ones. These accounts should be submitted to the committee as a group when the number of group members reaches a particular quantity.
f) There is no delay between two processes if the server in next process is idle.
g) As for construction of an operation center, the model only takes the number of experts into consideration. When new accounts are created, it takes 12 hours for a private account or 24 hours for a corporate account to transmit the information recorded by the counters to the anti-money laundering system. Do not consider other factors, like the capacity of anti-money laundering database.

4. Data Collection and Distribution Analysis

Note: time unit = hours if there is no specific explanation.

Table4.1 Customer arrival times in Branch A

Account Type	Accessible data or measures	Data source	Distribution
	Mean value		
Private accounts	0.5	A survey report done by Branch A; June, 2016 (22 workdays)	exponential
Corporate accounts	3		exponential

Table4.2 Time spent in account opening in Branch A

Account Type	Accessible data or measures			Data source	Distribution
	Most likely	Minimum	Maximum		
Private accounts	0.5	0.25	0.75	A survey report done by Branch A; June, 2016 (22 workdays)	triangle
Corporate accounts	1.5	1	5		triangle

Table4.3 Time spent in risk ranking in Branch A

Account Type	Accessible data description	Data source	Distribution
Private accounts	equally possible from 5 to 15 minutes, no most likely value	Transformed data from timestamps in the Anti-money laundering system; Branch A; June, 2016 (22 workdays)	uniform
Corporate accounts	equally possible from 20 to 45 minutes, no most likely value		uniform

Table4.4 Time spent in risk ranking in the operation center

Account Type	Sample size	Data source	Distribution
Private accounts	200	Transformed data from timestamp in the Anti-money laundering system, the operation center of bank X, Sept 5th-9th, 2016	weibull
Corporate accounts	200		weibull

Table4.5 Time spent by the committee in Bank X

Account Type	Accessible data or measures			Data source	Assumed distribution
	Most likely	Minimum	Maximum		
Personal & corporate accounts	2	1	4	Semi-annual and annual reports of Bank X, in 2016	triangle

Explanation: I have several reports on meeting or conference for the committee to decide the final risk level and they all recorded the starting time and finishing time. But I don't think it's reasonable to fit the time spend on the meeting to a detailed distribution because the bank held meeting every month, sometimes even every quarter so that the data is not enough to generate prediction. Meanwhile, the committee usually held a meeting with several topics for discussion, money-laundering risk ranking just included.

Table4.6 Summary of data distribution

Process type		Open an account	Risk ranking in branch	Risk ranking in center	Final decision by committee
who conducts		counters	managers	expert	committee
Process time	Private accounts	triangle(0.25,0.5,0.75)	UNIF(1/12,1/4)	(5+WEIB(0.439,2.19))/60	TRIA(1,2,4)
	Corporate accounts	triangle(1,1.5,5)	UNIF(1/3,3/4)	(15+WEIB(0.456,8.33))/60	TRIA(1,2,4)

5. Important Analysis of Important Data Streams

5.1 The percentage of the number of abnormal accounts

The data within Bank X doesn't vary a lot. They are around 5% and 3% respectively. But if we have access to many banks other than Bank X, especially the large-sized banks, the percentage could decrease significantly due to their internal management policy and automatic processing of detecting abnormal transactions.

Table 5.1 the percentage of the number of abnormal accounts over all the new created accounts

Account Type		Adopted value	Sample size	Data source
Abnormal accounts percentage	Personal accounts	5%	4	Calculate the mean of observations in Semi-annual and annual reports of Bank X, from 2015 to 2016
	Corporate accounts	3%		

5.2 Distribution analysis of the time spent on risk ranking in the center

The person in charge of risk ranking and selecting abnormal accounts from a large quantity of new created accounts has a great impact on the whole workflow. It's worthwhile to look into the difference if the person is an ordinary manager or an expert on money laundering detection. Obviously, an expert (or a skilled manager who have several years' experience) will show better performance. They spend a quite large amount of time on a small part of accounts which will be probably regarded as abnormal ones. For example, if the identification shows that a person or a company comes from some countries or areas that facilitate money laundering and terrorist financing. The expert in the center may compare their name with the blacklist, search more related information online, and even take a visit to the customers' domestic residence. The series of action takes a longer period of time. In contrast, for many ordinary accounts, they just have a glance of customer information, give them low-level risk label, and set them aside.

Figure3. Time Spent in the Center for Private Accounts

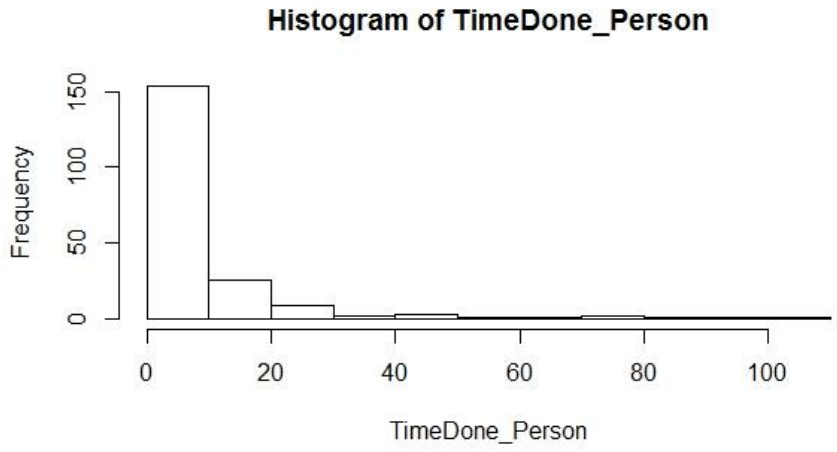


Figure4. Time Spent in the Center for Corporate Accounts

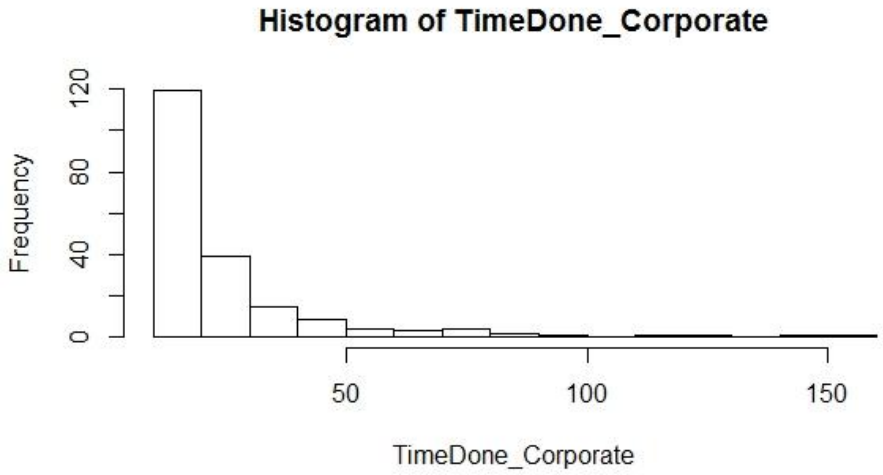


Table 5.2 the distribution of the time spent in risk ranking in the center

Results	TimeDone_Person	TimeDone_Corporate
Distribution Summary	Distribution: Weibull Expression: 5+WEIB(2.19, 0.439) Square Error: 0.001103	Distribution: Weibull Expression: 15+WEIB(8.33, 0.456) Square Error: 0.000510
Chi Square Test	Number of intervals = 3 Degrees of freedom = 0 Test Statistic = 1.27	Number of intervals = 4 Degrees of freedom = 1 Test Statistic = 0.742

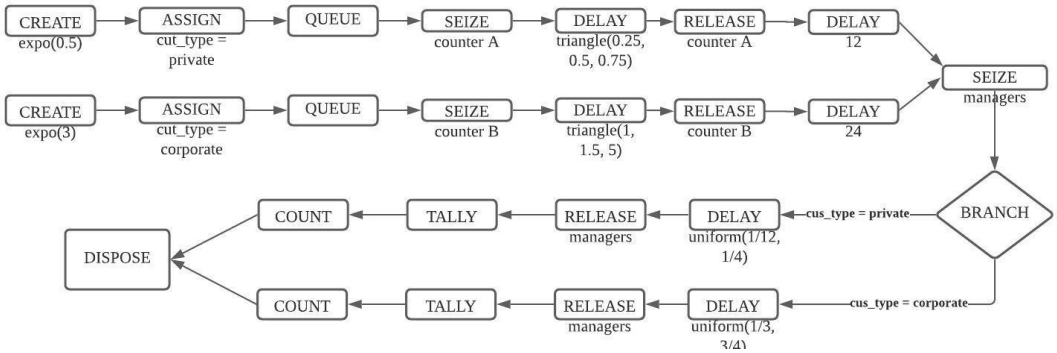
	Corresponding p-value < 0.005	Corresponding p-value = 0.417
Kolmogorov-Smirnov Test	Test Statistic = 0.0266 Corresponding p-value > 0.15	Test Statistic = 0.0499 Corresponding p-value > 0.15
Data Summary	Number of Data Points = 200 Min Data Value = 5 Max Data Value = 106 Sample Mean = 11.2 Sample Std Dev = 14.9	Number of Data Points = 200 Min Data Value = 15 Max Data Value = 158 Sample Mean = 26.6 Sample Std Dev = 22
Histogram Summary	Histogram Range = 5 to 107 Number of Intervals = 14	Histogram Range = 15 to 159 Number of Intervals = 14

6. Flow diagram

In order to simplify the problem, we depicted the flow diagram with two branches as the basic model and extended it to more than two branches.

In the branch-level workflow, the whole process can be decomposed into opening accounts and conducting risk-ranking in a branch. Every branch follows this logic and completes risk ranking independently.

Figur5. Each branch workflow



In the centralized workflow, the whole process can be decomposed into opening accounts in branches and conducting risk-ranking in the center. All the private or corporate accounts created by branches converge to one queue in front of the center respectively.

Figure6(A) Opening accounts in two branches

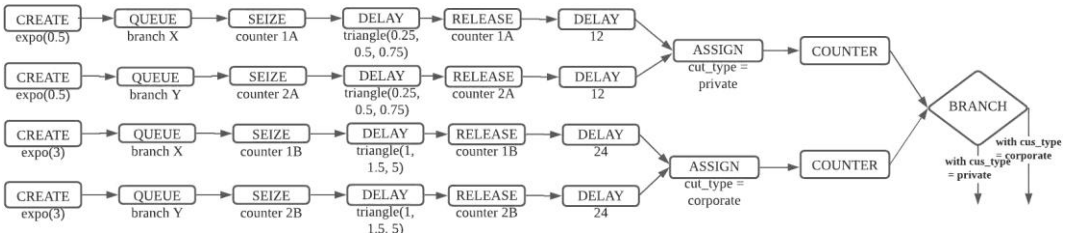
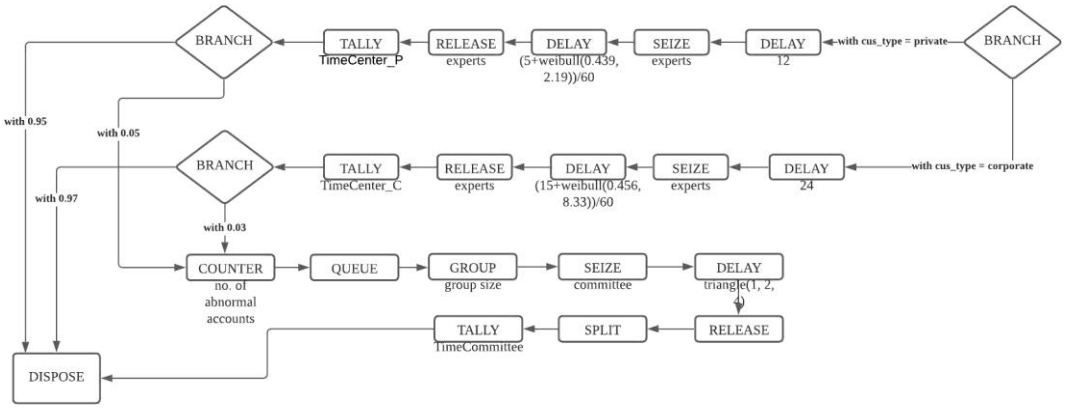


Figure6(B) Risk ranking in the center



7. Run the Model and Define the Notations

Replication setting: 8 hours/per day * 360 days = 2880 hours, no warm-up period.

Notations and statistical measures for centralized operation model

- Group size: the number of abnormal accounts that make up a group for submitting to the committee.
- Submit_times: the times that the center submits a group of abnormal accounts to the committee.
- No .of abnormal accounts = No. of group members * Submit_times
- TimeCommittee: time spent in the committee.
- TimeCenter_C: time spent in the center for corporate accounts
- TimeCenter_P: time spent in the center for private accounts
- total ranking time_C: total time spent on risk ranking for corporate accounts.
- total ranking time_C = TimeCommittee + TimeCenter_C
- total ranking time_P: total time spent on risk ranking for private accounts.

- total ranking time_P = TimeCommittee + TimeCenter_P

7. Simulation Results and Conclusion

The objective of this study was to analyze two types of money-laundering risk ranking workflows in a small-sized commercial bank X. The analysis was intended to provide the management of the bank with a tool to evaluate two operations, i.e., the branch-level and centralized risk ranking workflows, aiming at the reduction of the time spent on risk ranking, and the augment of submission group size while making personnel utilization more efficient. A discrete event simulation (DES) model is the main component of the developed tool. The quantitative input and output gathered while constructing and running the DES model can be utilized by the leadership of the bank.

Unlike most of the workflow adopted in large-scale banks, my simulation model in this report accounts for abnormal accounts being treated as a group and going through the final decision phase where, in small scale banks, the committee prefer less frequent receiving the group report.

This consideration allows the model to include the situation that small scale banks have fewer resources to determine the final risk level. On the one hand, the system capacity in small scale banks is not sufficient enough to transform electronic transaction records into visualized data that can be easily recognized by the committee members. On the other hand, the committee members come from different departments or assume more than one role, which makes it difficult for them to often attend meeting at the same time. Therefore, the centralized risk ranking workflow defines the usual practice in small banks: the experts in the operation center are responsible for submitting a combined report written for a group of abnormal accounts to the committee, as included in the DES model in the GROUP, SEIZE, DELAY, RELEAE, and SPLIT steps shown in figure 2.

Under the branch level workflow, unreasonable rank level determination and deficient personnel utilization are salient problems, as evidenced by many real-life cases across the bank X. The results obtained from this study showed that it is possible to follow a centralized risk ranking workflow in order to satisfy the internal and external regulatory requirements using a novel simulation-based approach.

The simulation for each branch under branch-level workflow was used as controls to measure and compare the total number of risk ranking jobs completed under the centralized operation workflow. The basic centralized workflow contains new created account data across 2 branches. The target variables selected were No. of experts, Group size, No. of abnormal accounts, Submit_times, No. of ranked accounts, TimeCommittee, TimeCenter_C, TimeCenter_P, Total ranking time_C, Total ranking

time_P, Expert utilization, Committee utilization. Remember, we want to maximize expert utilization but minimize committee utilization.

Moreover, In order to meet the external regulatory requirement of 14-workday limitation on risk ranking, we have to take the maximum statistics as observations though the average measures are more commonly adopted in many simulation problems. In other words, we observed whether Total ranking time_C or Total ranking time_P is beyond 112 hours by adjusting the group size, and thus obtain the optimized values for target variables when the workflow contain 2 branches. All the results from the basic model are as follows.

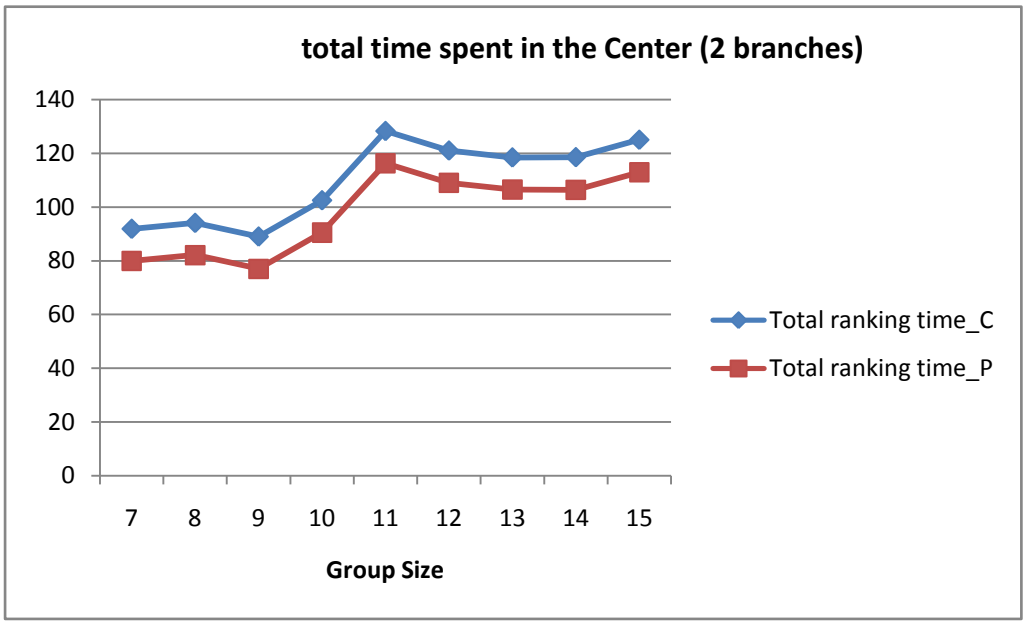
Table7.1 Results from the centralized workflow with two branches' data

Group Size	7	8	9	10	11	12	13	14	15
TimeCommittee	67.17	69.31	64.32	77.77	103.56	96.17	93.69	94.29	100.27
TimeCenter_C	24.69	24.75	24.66	24.72	24.73	24.84	24.75	24.25	24.75
TimeCenter_P	12.77	12.82	12.70	12.76	12.71	12.86	12.81	12.08	12.72
Total ranking time_C	91.86	94.06	88.98	102.49	128.29	121.01	118.44	118.55	125.02
Total ranking time_P	79.94	82.14	77.02	90.53	116.27	109.03	106.5	106.38	112.99

Such sensitivity analysis guided us to find out the Total Ranking Time, 102.49, which is most close to 112 but not beyond it. Accordingly, we reach the conclusion that the bank gets all the tasks completed within the 14-workday limitation when it has two branches' data and the expert in it submit a combined report encompassing 10 abnormal accounts.

Another important insight from the sensitivity analysis is what the relationship looks like between Total Ranking Time and Group Size. In general, Total Ranking Time increased when the group size increased even though the tendency is not always non-decreasing. In this case, I tested a series of group size, plot two streams of statistics, and guarantee the optimized group size.

Figure 7 Total Ranking Time and Group Size



The last issue is to compare the efficiency of the branch-level and centralized workflows. We call the fifth column as an 'optimized solution' or 'optimized metric' because it includes the maximized number of risk ranking jobs finished under the centralized workflow. This maximized number, 13530 in table 7.2, is almost the same under branch-level workflow with independent two branches (13217 shown under table 7.3). In other words, the center can handle the same workload, or the number of experts is enough. There is no long waiting queue in front of the experts and the committee.

The extended simulation model for centralized workflows respectively dealt with 3 and 4 branches' data. Similarly, running the extended model and conduct sensitivity analysis, we got optimized solutions and list them in table 7.2.

Table 7.2 Optimized Solutions (time unit: hours)

statistics	the center includes 2 branches	the center includes 3 branches	the center includes 4 branches
No. of experts	1	1	2
Group size	10	19	29
No .of abnormal accounts	669	1274	1248
Submit_times	66.90	67.05	43.03
No. of accounts that get a level	13530	20292	26741
TimeCommittee	77.77	84.23	86.58
TimeCenter_C	24.72	24.31	24.66
TimeCenter_P	12.76	12.19	12.56
total ranking time_C	102.49	108.54	111.23

total ranking time_P	90.53	96.41	99.14
Expert utilization	0.5219	0.7865	1.0492 (each:0.5246)
Committee utilization	0.0470	0.0427	0.0322

Table7.3 Results from the branch-level model

Branch	No. of private accounts that are given a risk level	No. of corporate accounts that are given a risk level	Manager utilization
Branch A	5732	933	0.5132
Branch B	5691	861	0.5002
total	11423	1794	--

Note: These two branches (A and B) complete the risk ranking of 13217 accounts (11423+1794=13217).

8. Strengths, weaknesses and possible improvement

The optimized number of experts fits the real life situation very good. Actually, I was the only one person working at the center on July, 2016. Bank X required the center to take on the workload of the two branches during the pilot phase. It sounded like an overwhelming and impossible task, but I finished the risk ranking in time. Then, the center tried to hire more experts when Bank X opened new branches. When the number of branches reached 4, 2 experts can handle all the workload.

From the last column of Table7.2, the center (4 branches) submits a group 43 times in 360 days. The frequency is 8.36 days, which means the committee have to hold a conference almost every week. Unfortunately, it is not the true case according to my working experience. In the latter part of 2016, the committee discussed and examined a group report every two weeks or even every month so that they managed to avoid time conflict between their own work and necessary attending on the conference of final risk ranking decision.

I think there are two main reasons behind this phenomenon. The one is where the abnormal accounts originate from. In reality, any accounts, including new created ones and existing ones, have chances of becoming active and abnormal. A low risk level account can make unusual transactions and be detected again even though it has been inactive for many years. Hence there must have far more abnormal accounts cropping up and ready for re-evaluation, which constitute the base of a larger group size and less submitting frequency. The other point is the judgment from an expert. Different experts may label the same account with different risk levels. Undoubtedly, the percentage of abnormal accounts cannot be a fixed number but will vary a lot instead, which may also increase or decrease the size of abnormal accounts.