Project Report

Increase Rate of An Internet Financial Company

Enqun Wang, Yiyan Zhou

May 3, 2016

Project Charter

Enqun Wang (EW), Yiyan Zhou (YZ)

April 1, 2016

Problem Statement

Data from January throughout April, 2016, from First P2P, reflects an *average 1.0370%* five-day incrase rate of trading volume of the platform with around *15% defective values*. Based on the financial professinal knowledge, operation condition of the company, and the financial market environment, it is believed that a five-day increase rate is normaly distributed with constant variance, and a value that *less than 0.83% (assumed to be constant in our project)* is regarded as *a defective value*, which can lead to a reduction in benifit of the company.

Project Objective

Reduce the proportion of defective values from *15%* to *9%* (*67%* improvement) by the end of the second quarter (June 30, 2016), given the normality with constant variance assumption. That is equivalent to increase the five-day increase rate of the trading volume from *1.04%* to the *1.10%*.

Project Team

Name	Role	Comments	Phone
Enqun Wang	Operation Analyst	Organize the project and assign the detailed work to team members	(720) 231 5851
Yiyan Zhou	Operation Analyst	Finish what the leader assigned and give advises	(970) 294 5303
Chee-Yong Tan	Operation executive officer	Adjust the strategies	

Project Definition and Scoping

Metrics (unit of measure):

- The percentage (%) of defective values (five-day incrase rate of trading volume less than 0.83%).

Critical to Satisfaction (linkage to customer):

Individual investors and companies with financing needs are two major types of our clients. An relatively high proportion of defective values (five-day incrase rate of trading volume less than 0.83%) can result financial products of higher prices, therefore individual investors would earn less money and customer companies looking for finance would spend more money. Furthermore, it might lead to a reduction in benifit of the internet finance company.

Defect Definition (include opportunity):

Any five-day increase rate less than 0.83% is regarded as a defective value, which can result financial products of higher prices, therefore individual investors would earn less money and customer companies would spend more money. Furthermore, it might lead to a reduction in benifit of the internet finance company. We assume that the criteria value (0.83%) is a constant in the scope of our project.

Scope of Project:

The project focuses on the first two quarters of 2016, from January 1, 2016 to June 30, 2016. The data is collected at April 20, 2016 from the database of the company, thus we scope this project from January 1 to April 20.

5W2H

Enqun Wang (EW), Yiyan Zhou (YZ)

April 1, 2016

Why?

Why is it important to solve this problem? Why is it important to improve this process? Formulate in terms of expected benefits

A high proportion of defective values (five-day increase rate of trading volumn less than 0.83%) can result to financial products of our company to higher prices; furthermore, investors would earn less and custmoer companies looking for finance would spend more money. Solving this problem can benefit the investors and custmoer companies.

Who?

Who is involved in the process? Who is affected by the problem? Who is interested in solving the problem?

Individual investors and custmor companies are involoved in this process. Low five-day increase rate would cause individual investors earn less and the companies would spend more money.

Where?

Where is the problem located? In which processes do the problems occur?

This problem is located in the process of operation and in the financial market.

When?

When did or does the problem happen? How often did or does the problem happen? When did the problem start?

The opportunity has been exist since November of 2015 and is detected by the Operation Executive Offiecer (OEO) of the company. This kind of issues happed when the strategies are no longer appropriate for the financail market.

What?

What is the defect? What activities, parts and procedures are involved? What happens, when the problems occur?

Any five-day increase rate less than 0.83% is regarded as a defective value, which can result financial products of higher prices, therefore individual investors would eran less money and customer companies looking for finance would spend more money. Furthermore, it might lead to a reduction in benefit of the internet finance company.

How?

How do you know it is a problem? How was the problem identified? How often does it happen?

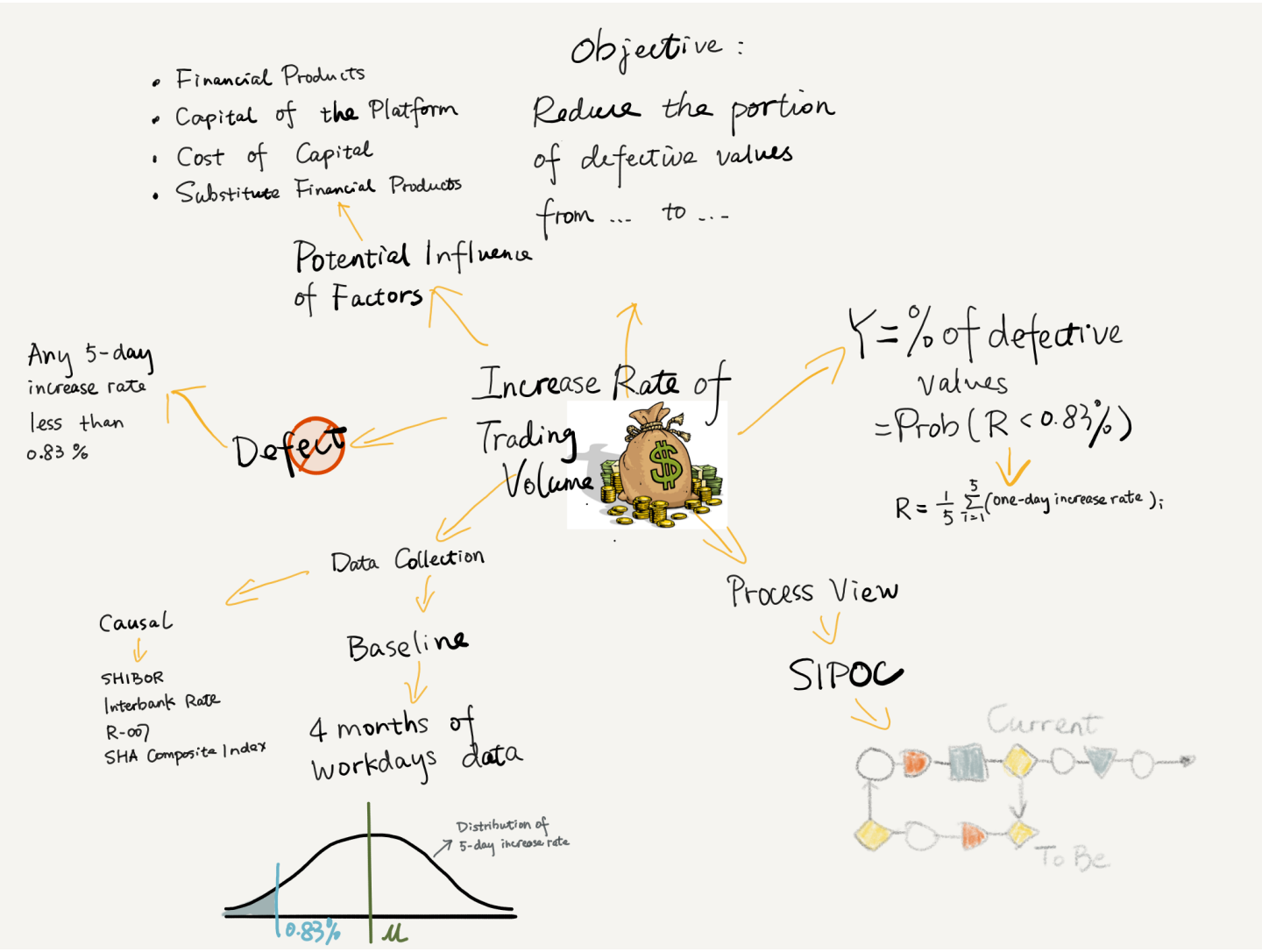
Operation Executive Offiecer (OEO) detected this is a problem and assigned tasks to our team. It is a problem for that it can lead to low increase rate of the trading volume. It is defined as no effective reaction or response to the fluctuation of the finance market. The problem happening follows the periods of the strategies and the fluctuation of the financial market.

How many?

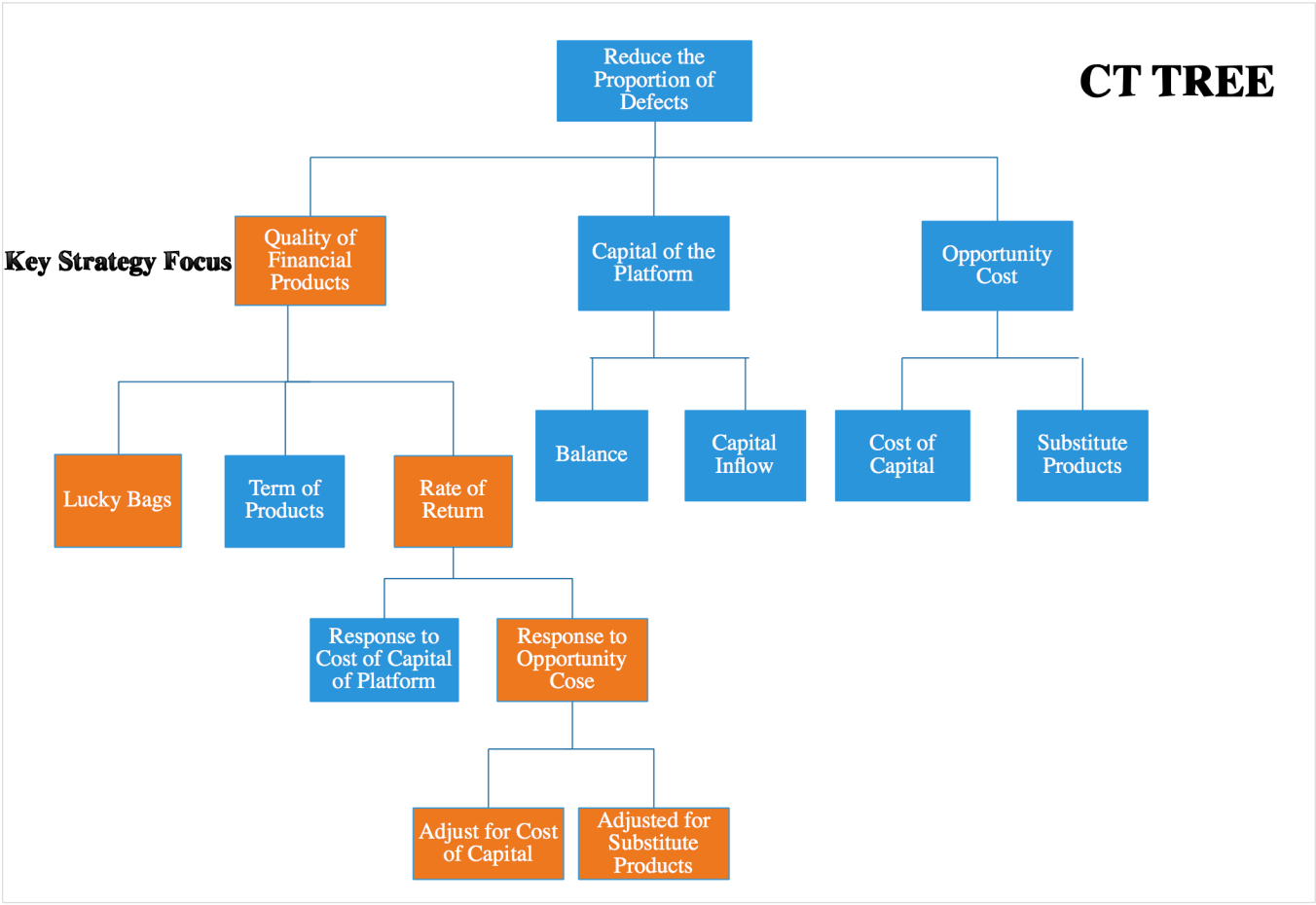
How many defects/units/people? How much money was spent? Do these numbers change or do they stay the same?

Only one defect but related all individual investors and all custmor companies in this project. The number of investors and companies vary based on the significance of the defect. As the compeny is a internet finance compay, the money spent is a kind of strategy and in high frequency.

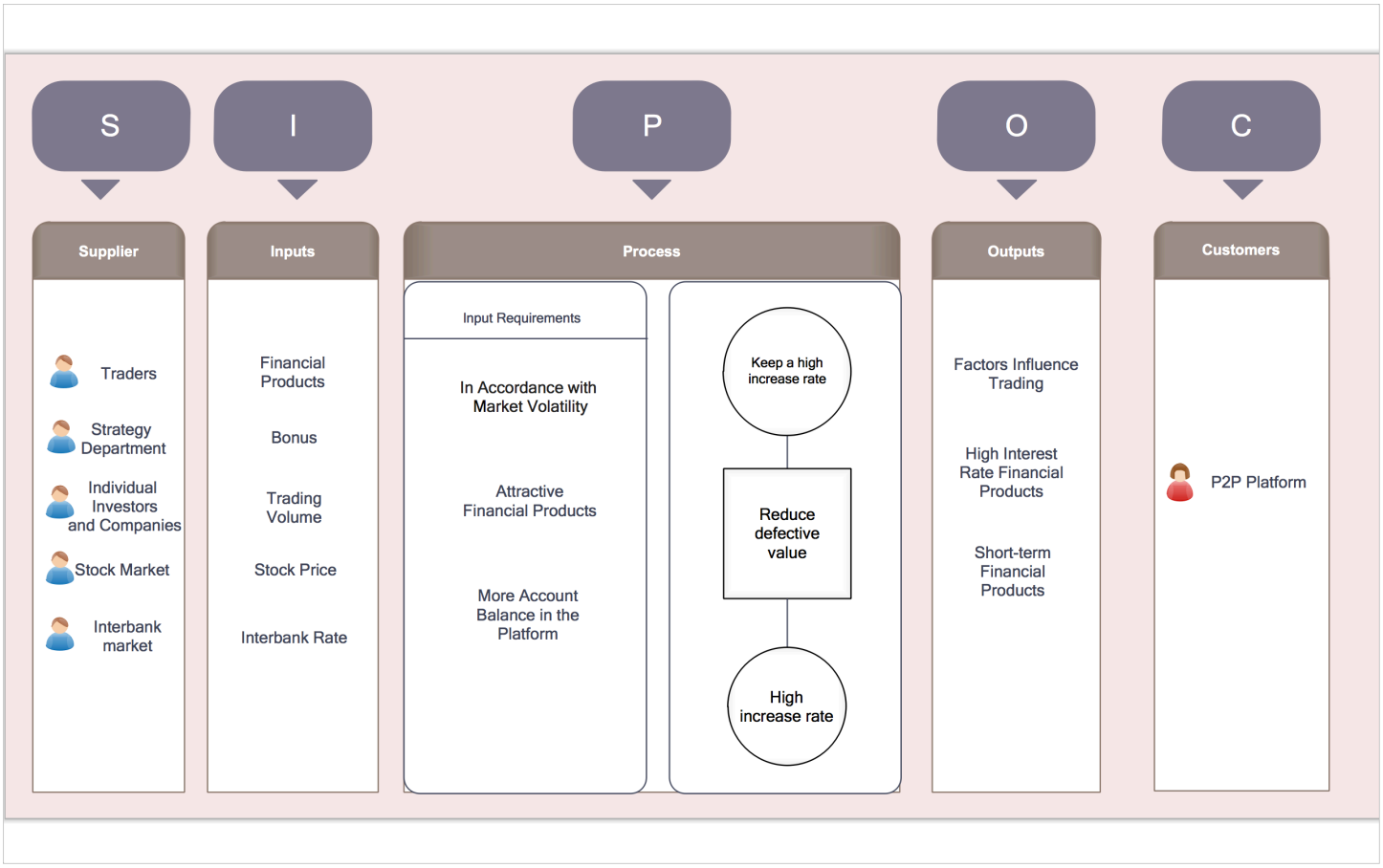
Thought Process Map



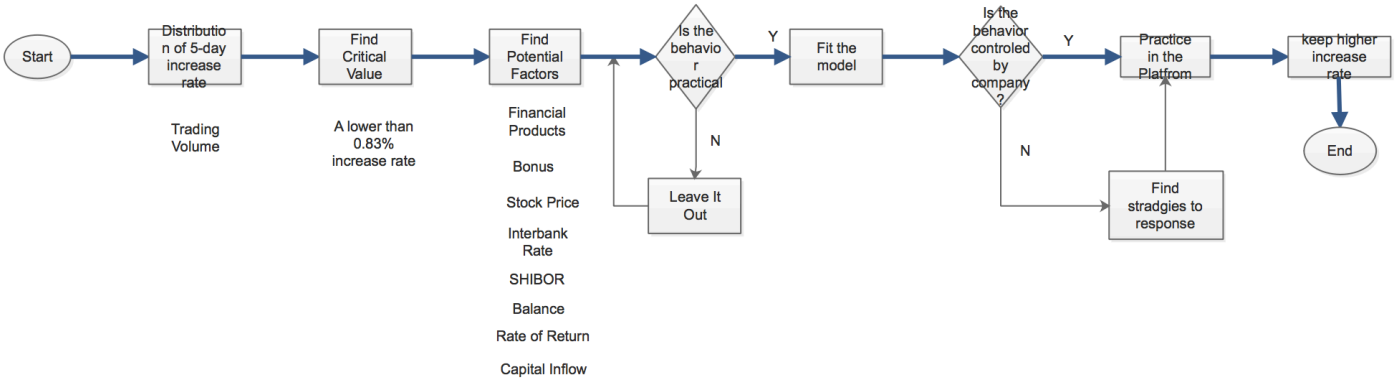
CT TREE


Reduce the Proportion of Defects
Key Strategy Focus
Quality of Financial Products
Capital of the Platform
Opportunity Cost
Lucky Bags
Term of Products
Rate of Return
Balance
Capital Inflow
Cost of Capital
Substitute Products
Response to Cost of Capital of Platform
Response to Opportunity Cose
Adjust for Cost of Capital
Adjusted for Substitute Products

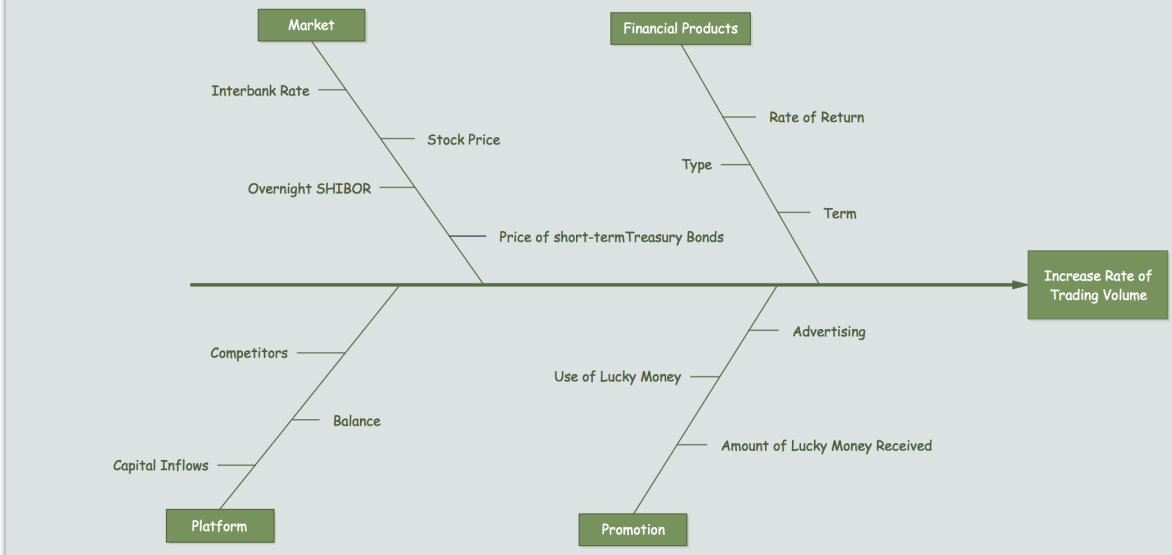
SIPOC


S
I
P
O
C
Supplier
Traders
Strategy Department
Individual Investors and Companies
Stock Market
Interbank market
Inputs
Financial Products
Bonus
Trading Volume
Stock Price
Interbank Rate
Process
Input Requirements
In Accordance with Market Volatility
Attractive Financial Products
More Account Balance in the Platform
Keep a high increase rate
Reduce defective value
High increase rate
Outputs
Factors Influence Trading
High Interest Rate Financial Products
Short-term Financial Products
Customers
P2P Platform

Process Flow Map



C&E Diagram


Market
Interbank Rate
Stock Price
Overnight SHIBOR
Price of short-termTreasury Bonds
Financial Products
Rate of Return
Type
Term
Increase Rate of Trading Volume
Competitors
Balance
Capital Inflows
Platform
Advertising
Use of Lucky Money
Amount of Lucky Money Received
Promotion

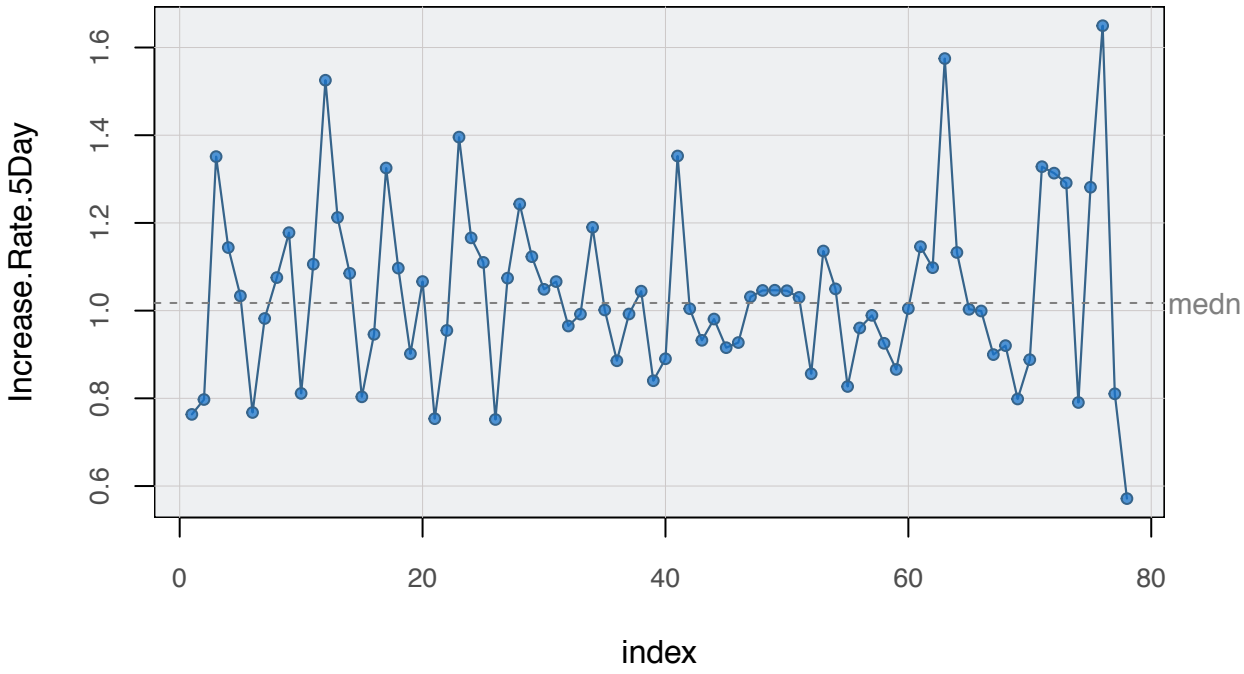
Run Chart

Enqun Wang

April 22, 2016

Run Chart

- The run chart indicates oscillating patterns. Oscillation occurs when the data fluctuates up and down, indicating that the process is not steady. However, this is the common situation in the finance field.




```
## --- Increase.Rate.5Day ---
##
##
## n: 78
## missing: 0
## median: 1.017472
##
## ------------
## Run Analysis
## ------------
##
## size=2  Run  1:    1    2
## size=3  Run  2:    3    4    5
```

```
## size=2  Run  3:    6    7
## size=2  Run  4:    8    9
## size=1  Run  5:   10
## size=4  Run  6:   11   12   13   14
## size=2  Run  7:   15   16
## size=2  Run  8:   17   18
## size=1  Run  9:   19
## size=1  Run 10:   20
## size=2  Run 11:   21   22
## size=3  Run 12:   23   24   25
## size=1  Run 13:   26
## size=5  Run 14:   27   28   29   30   31
## size=2  Run 15:   32   33
## size=1  Run 16:   34
## size=3  Run 17:   35   36   37
## size=1  Run 18:   38
## size=2  Run 19:   39   40
## size=1  Run 20:   41
## size=5  Run 21:   42   43   44   45   46
## size=5  Run 22:   47   48   49   50   51
## size=1  Run 23:   52
## size=2  Run 24:   53   54
## size=6  Run 25:   55   56   57   58   59   60
## size=4  Run 26:   61   62   63   64
## size=6  Run 27:   65   66   67   68   69   70
## size=3  Run 28:   71   72   73
## size=1  Run 29:   74
## size=2  Run 30:   75   76
## size=2  Run 31:   77   78
##
## Total number of runs: 31
## Total number of values that do not equal the median: 78
## Total number of values ignored that equal the median: 0
```

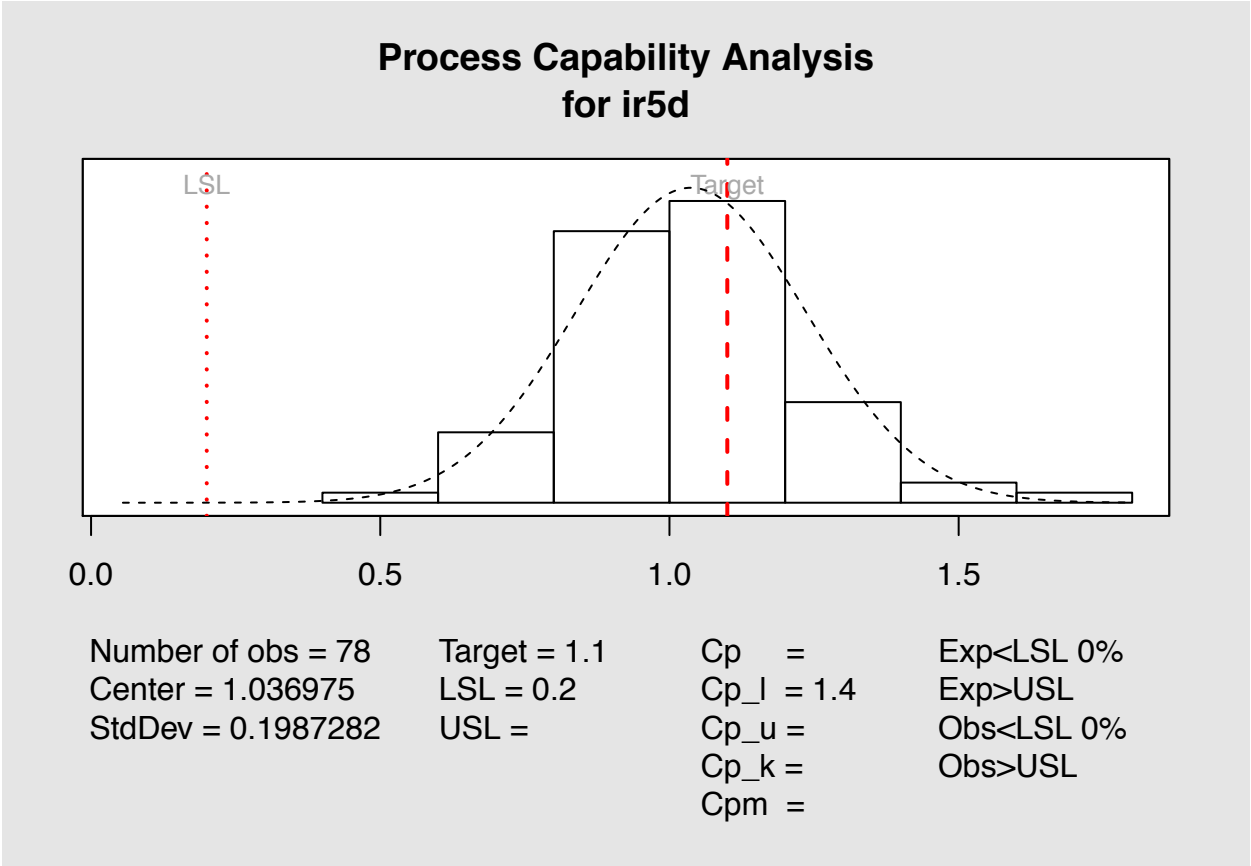
Process Capability Analysis

Enqun Wang

April 20, 2016

Process Capability Analysis

- The Cp_l is 1.404, with a confidence interval (1.208, 1.6), which indicates that the process is capable.




```
##
## Process Capability Analysis
##
## Call:
## process.capability(object = object, spec.limits = c(lsl, usl),     target = target)
##
## Number of obs = 78           Target = 1.1
##         Center = 1.037          LSL = 0.2
##         StdDev = 0.1987         USL =
##
## Capability indices:
##
##       Value   2.5%  97.5%
## Cp
```

```
## Cp_l  1.404  1.208    1.6
## Cp_u
## Cp_k
## Cpm
##
## Exp<LSL 0%    Obs<LSL 0%
## Exp>USL   Obs>USL
```

```
##
## To cite qcc in publications use:
##
##   Scrucca, L. (2004). qcc: an R package for quality control
##   charting and statistical process control. R News 4/1, 11-17.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {qcc: an R package for quality control charting and statistical process control},
##     author = {Luca Scrucca},
##     journal = {R News},
##     year = {2004},
##     pages = {11--17},
##     volume = {4/1},
##     url = {http://CRAN.R-project.org/doc/Rnews/},
##   }
```

Eexploratory Data Analysis

Enqun Wang (EW), Yiyan Zhou (YZ)

April 25, 2016

Preprocess data

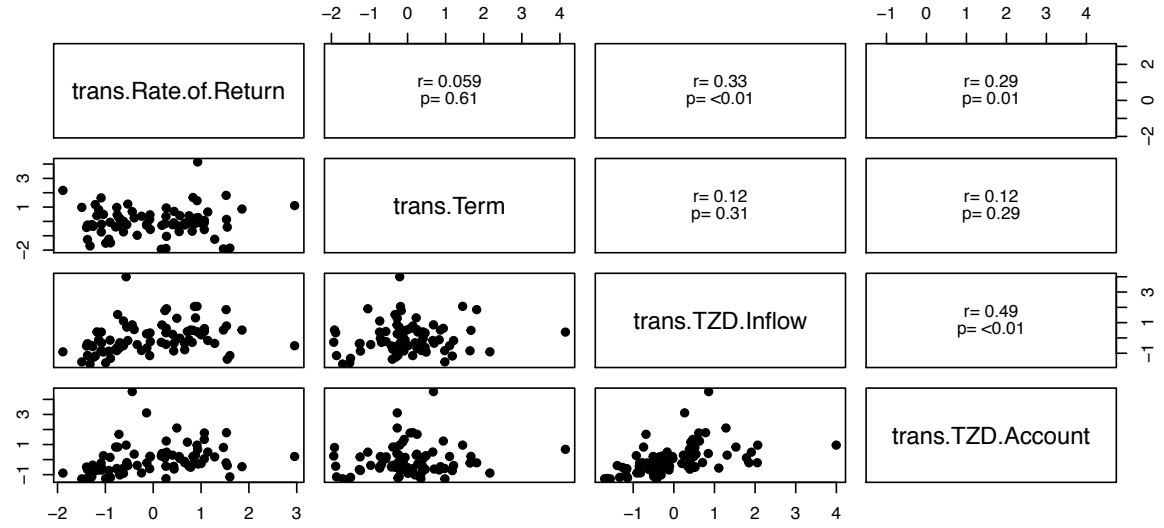
- Based on cause and effect relationship, we divide the variales into four causes: Product Factor, Promotion Factor, Platform Factor, and Market Factor. According to voice of customer (VOC), we would analyze the influence of these factors independently, for that each one represents a different aspect.

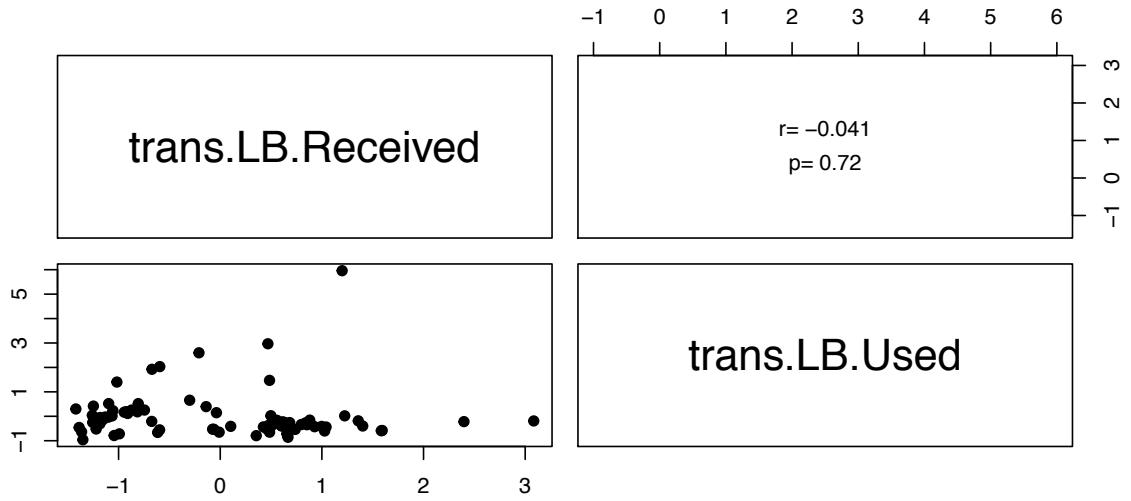
Product Factor	Promotion Factor	Platform Factor	Market Factor
Rate of Return	LB Received	Balance	R.007
Term	LB Used	Capital Inflow	Inerbank Rate
TZD Account			SHIBOR
TZD Inflow			SHA
			GEM

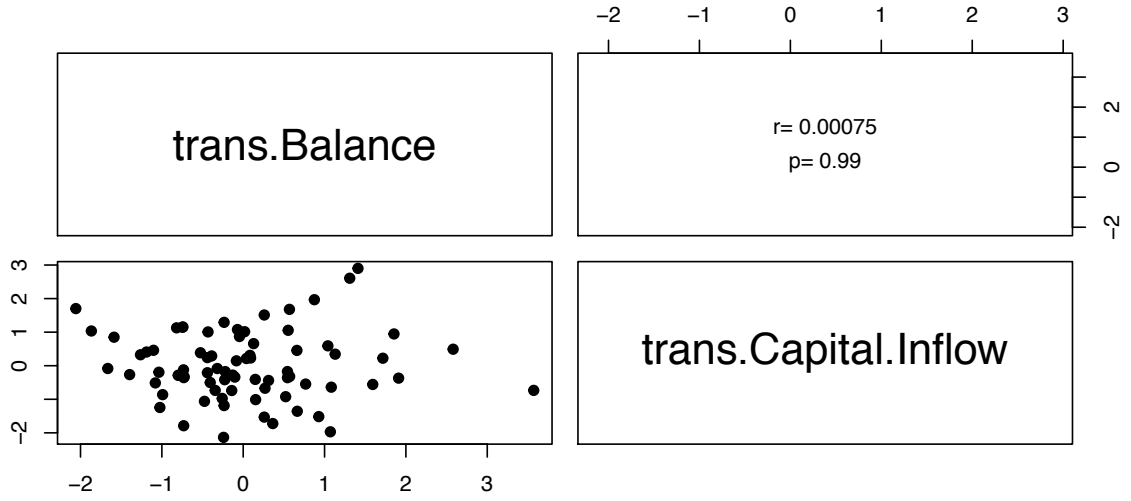
- In order to remove effects of different units, we centered and rescaled the data.

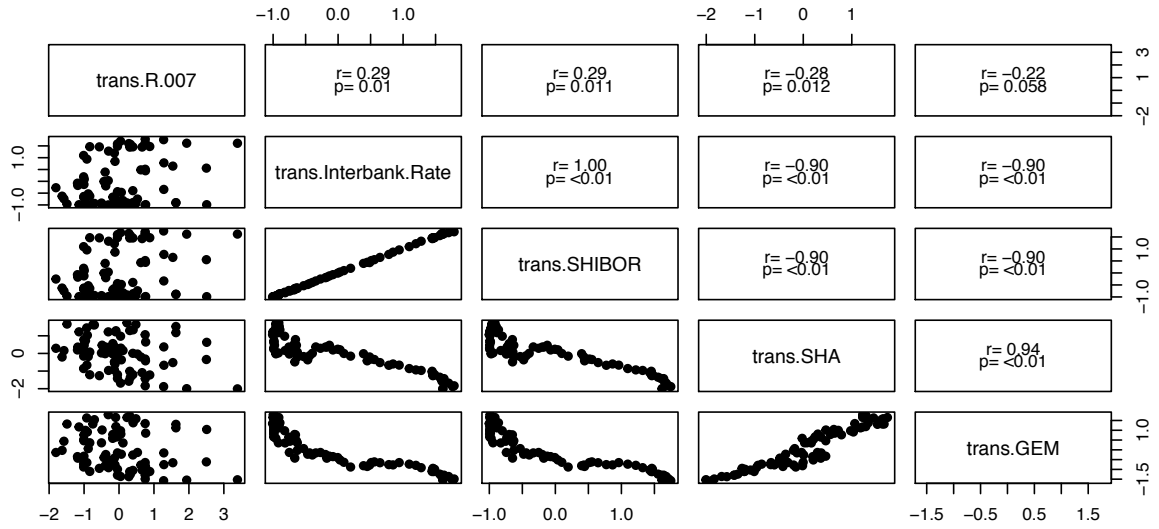
Detect dependent variables

- According to the correlation plot above, we find that `interbank.Rate` , `SHIBOR`, `SHA`, and `GEM` are highly correlated, and that `TZD.Inflow` and `TZD.Account` are highly correlated. So we consider if we could remove some of them.
- Based on the voice of costumers (VOC), we decided to remove `interbank.Rate`, which can be represented by `SHIBOR`; remove `GEM`, which can be represented by `SHA`; and remove `TZD.Inflow`, which can be reflected from `TZD.Account`.



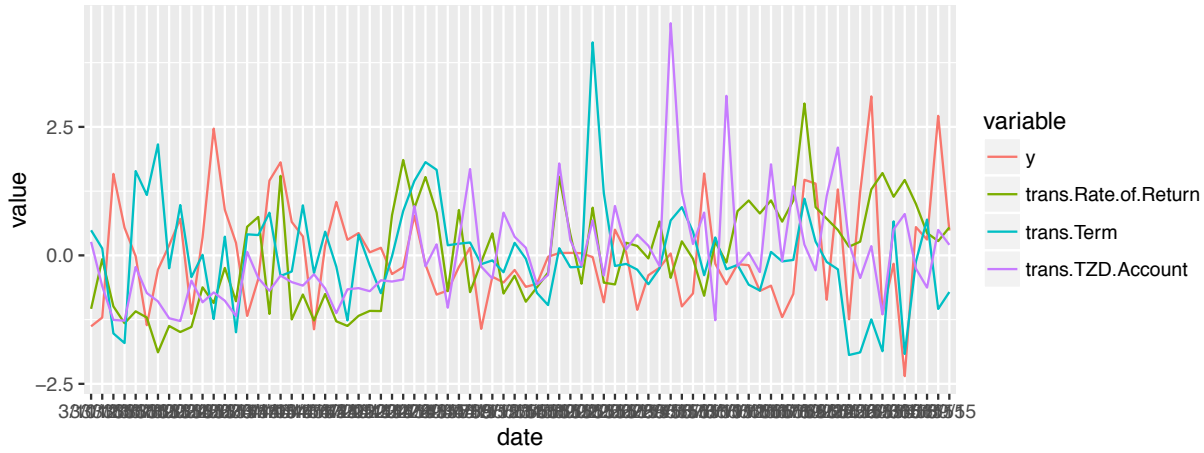

trans.LB.Received
r= −0.041
p= 0.72
trans.LB.Used

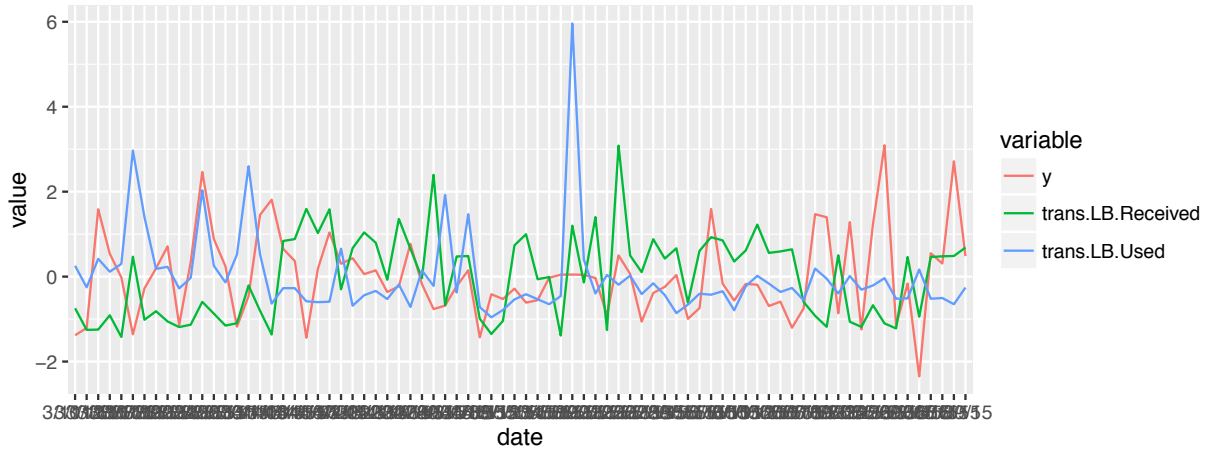

trans.Balance
r= 0.00075
p= 0.99
trans.Capital.Inflow

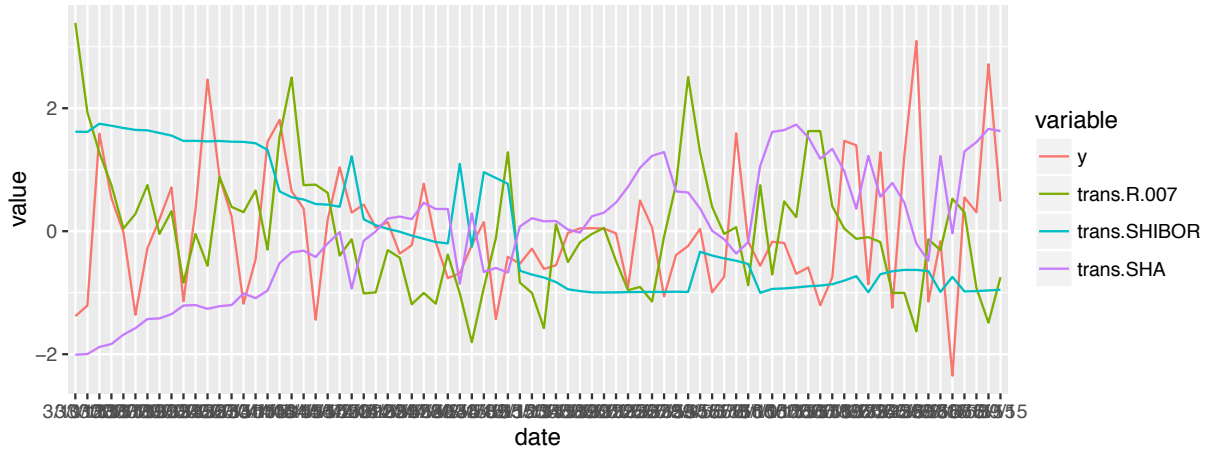


Plot multiple time series

- According to the plots above and VOC, we would remove the variable `SHIBOR` and `SHA`.

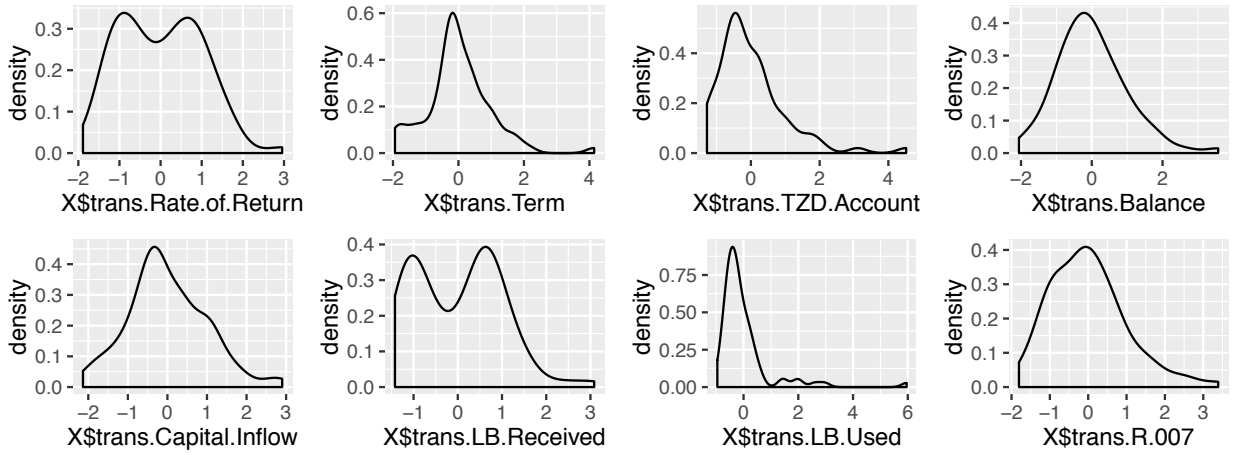






Density plots

- To explore the distributions of the variables, we ploted density plots as follows. It indicates that `Rate of Retrun` and `LB Recevied` are not normally distributed, while others are basicly normal.



First selection of variables

According to the exploratory data analysis, we decide to first elect variables as follows.

Product Factor	Promotion Factor	Platform Factor	Market Factor
1. Rate of Return	1. LB Received	1. Balance	1. R.007
2. Term	2. LB Used	2. Capital Inflow	
3. TZD Account			

Hypothesis Testing

Enqun Wang (EW), Yiyan Zhou (YZ)

April 26, 2016

Exploratory Data Analysis

According to the exploratory data analysis, we decide to first elect variables as follows,

Product Factor	Promotion Factor	Platform Factor	Market Factor
1. Rate of Return	1. LB Received	1. Balance	1. R.007
2. Term	2. LB Used	2. Capital Inflow	
3. TZD Account			

We use the centered data to test the significance of the variables, and find the significant ones.

1. Increase.Rate.5Day ~ Rate.of.Return

```
##
## Call:
## lm(formula = df$trans.Increase.Rate.5Day ~ df$trans.Rate.of.Return)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8454 -0.7421 -0.1314  0.6312  3.4207
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                8.331e-19  1.155e-01   0.000    1.000
## df$trans.Rate.of.Return -5.334e-03  1.162e-01  -0.046    0.964
##
## Residual standard error: 1.007 on 74 degrees of freedom
## Multiple R-squared:  2.845e-05,  Adjusted R-squared:  -0.01348
## F-statistic: 0.002106 on 1 and 74 DF,  p-value: 0.9635
```

The p value of the test is 0.9635, which is greater than 0.05. We should not reject the null hypothesis and conclude that the coefficient of Rate.of.Return is not significant.

2. Increase.Rate.5Day ~ Term

```
##
## Call:
## lm(formula = df$trans.Increase.Rate.5Day ~ df$trans.Term)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7531 -0.7831 -0.1170  0.6342  3.1287
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)    2.550e-17  1.141e-01   0.000    1.000
## df$trans.Term -1.564e-01  1.148e-01  -1.362    0.177
##
## Residual standard error: 0.9943 on 74 degrees of freedom
## Multiple R-squared:  0.02445,    Adjusted R-squared:  0.01127
## F-statistic: 1.855 on 1 and 74 DF,  p-value: 0.1773
```

The p value of the test is 0.1773, which is greater than 0.05. We should not reject the null hypothesis and conclude that the coefficient of Term is not significant.

3. Increase.Rate.5Day ~ TZD.Account

```
##
## Call:
## lm(formula = df$trans.Increase.Rate.5Day ~ df$trans.TZD.Account)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7776 -0.7596 -0.0298  0.5272  3.3339
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            6.943e-17  1.102e-01   0.000  1.00000
## df$trans.TZD.Account -2.976e-01  1.110e-01  -2.682  0.00903 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9611 on 74 degrees of freedom
## Multiple R-squared:  0.08857,    Adjusted R-squared:  0.07625
## F-statistic: 7.191 on 1 and 74 DF,  p-value: 0.009031
```

The p value of the test is 0.009031, which is less than 0.05. We should reject the null hypothesis and conclude that the coefficient of TZD.Account is significant.

4. Increase.Rate.5Day ~ LB.Received

```
##
## Call:
## lm(formula = df$trans.Increase.Rate.5Day ~ df$trans.LB.Received)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7086 -0.8061 -0.1410  0.6606  3.2724
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            5.240e-17  1.129e-01   0.000    1.000
## df$trans.LB.Received -2.098e-01  1.137e-01  -1.846    0.069 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9843 on 74 degrees of freedom
## Multiple R-squared:  0.044,  Adjusted R-squared:  0.03108
## F-statistic: 3.406 on 1 and 74 DF,  p-value: 0.06896
```

The p value of the test is 0.04645, which is less than 0.05. We should reject the null hypothesis and conclude that the coefficient of LB.Received is significant.

5. Increase.Rate.5Day ~ LB.Used

```
##
## Call:
## lm(formula = df$trans.Increase.Rate.5Day ~ df$trans.LB.Used)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7526 -0.7464 -0.1423  0.6034  3.4414
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.926e-17  1.148e-01   0.000    1.000
## df$trans.LB.Used  1.066e-01  1.156e-01   0.922    0.359
##
## Residual standard error: 1.001 on 74 degrees of freedom
## Multiple R-squared:  0.01136,    Adjusted R-squared:  -0.001997
## F-statistic: 0.8505 on 1 and 74 DF,  p-value: 0.3594
```

The p value of the test is 0.1695, which is greater than 0.05. We should not reject the null hypothesis and conclude that the coefficient of LB.Used is not significant.

6. Increase.Rate.5Day ~ Balance

```
##
## Call:
## lm(formula = df$trans.Increase.Rate.5Day ~ df$trans.Balance)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8752 -0.7416 -0.1388  0.6520  3.3705
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.082e-17  1.154e-01   0.000    1.000
## df$trans.Balance  3.499e-02  1.162e-01   0.301    0.764
##
## Residual standard error: 1.006 on 74 degrees of freedom
## Multiple R-squared:  0.001224,   Adjusted R-squared:  -0.01227
## F-statistic: 0.0907 on 1 and 74 DF,  p-value: 0.7641
```

The p value of the test is 0.8785, which is greater than 0.05. We should not reject the null hypothesis and conclude that the coefficient of Balanec is not significant.

7. Increase.Rate.5Day ~ Capital.Inflow

```
##
## Call:
## lm(formula = df$trans.Increase.Rate.5Day ~ df$trans.Capital.Inflow)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6154 -0.7388 -0.0793  0.6976  2.1425
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -2.907e-17  1.034e-01   0.000        1
## df$trans.Capital.Inflow  4.457e-01  1.041e-01   4.283 5.47e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9012 on 74 degrees of freedom
## Multiple R-squared:  0.1986, Adjusted R-squared:  0.1878
## F-statistic: 18.34 on 1 and 74 DF,  p-value: 5.467e-05
```

The p value of the test is 5.467e-05, which is less than 0.05. We should reject the null hypothesis and conclude that the coefficient of Capital.Inflow is significant.

8. Increase.Rate.5Day ~ R.007

```
##
## Call:
## lm(formula = df$trans.Increase.Rate.5Day ~ df$trans.R.007)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.62852 -0.78853 -0.06729  0.70667  3.14317
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -6.368e-18  1.113e-01   0.000   1.0000
## df$trans.R.007 -2.678e-01  1.120e-01  -2.391   0.0194 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.97 on 74 degrees of freedom
## Multiple R-squared:  0.07169,    Adjusted R-squared:  0.05915
## F-statistic: 5.715 on 1 and 74 DF,  p-value: 0.01936
```

The p value of the test is 0.03061, which is less than 0.05. We should reject the null hypothesis and conclude that the coefficient of R.007 is significant.

To sum up, LB.Received, Capital.Inflow, TZD.Account and R.007 are four significant variables.

Considering the above four significant the variables, we do model selection by forward stepwise.

```
## Subset selection object
## Call: regsubsets.formula(df$trans.Increase.Rate.5Day ~ df$trans.LB.Received +
##     df$trans.Capital.Inflow + df$trans.TZD.Account + df$trans.R.007,
##     data = df, method = "forward")
## 4 Variables  (and intercept)
##                           Forced in Forced out
## df$trans.LB.Received        FALSE      FALSE
```

```
## df$trans.Capital.Inflow     FALSE      FALSE
## df$trans.TZD.Account        FALSE      FALSE
## df$trans.R.007              FALSE      FALSE
## 1 subsets of each size up to 4
## Selection Algorithm: forward
##          df$trans.LB.Received df$trans.Capital.Inflow df$trans.TZD.Account
## 1  ( 1 ) " "                  "*"                     " "
## 2  ( 1 ) " "                  "*"                     "*"
## 3  ( 1 ) "*"                  "*"                     "*"
## 4  ( 1 ) "*"                  "*"                     "*"
##          df$trans.R.007
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) "*"
```

It seems that the capital inflow is the most significent. We can now select the best model based on the voice of customer.

The importance of the four variables are as followed:

Importance	Variable
1	Capital Inflow
2	Lucky Bag Received
3	R 007
4	The Financial Product TZD Account