



平台交易量模型分析报告

王恩群 营运分析专员

指导: 张勃, 胡亦成, 王泽, 汪楠

营运管理部, 网信理财, 中国先锋金融集团—北京市朝阳区霄云路 28 号院 网信大厦 A 座 7 层

T: (+86) 187 3533 3168 E: wang.enqun@outlook.com

内容摘要

报告模型

高信度模型：

$$\begin{aligned}\text{平台日交易量(百万元)} = & -4.38 - 8.95 \cdot 7 \text{ 天国债回购收盘价}(\%) \\ & + 0.27 \cdot \text{用户总账户余额(百万元)} \\ & + 14.28 \cdot \text{产品加权收益率}(\%) \end{aligned}$$

高预测度模型：

$$\begin{aligned}\text{平台日交易量(百万元)} = & -372.8 + 73.53 \cdot \text{余额宝 7 天年化利率}(\%) \\ & - 13.90 \cdot \text{隔夜SHIBOR}(\%) \\ & + 0.24 \cdot \text{用户总账户余额(百万元)} \\ & + 18.57 \cdot \text{产品加权收益率}(\%) \\ & - 7.05 \cdot 7 \text{ 天国债回购收盘价}(\%) \end{aligned}$$

报告建议

- 一、加强同业产品调研，提高产品创新能力，研发更具竞争力的产品
- 二、研究、采取有效方案，提高用户总体账户余额
- 三、完善市场监测机制，加强宏观政策研判，提高对市场变化的应变能力

目录

内容摘要	2
图标目录	3
引言	4
探索性数据分析	6
模型假设.....	8
模型建立.....	8
模型解释.....	9
模型讨论.....	11
决策建议	12
总结	13
附录	13

图标目录

Table 1 变量说明	5
Table 2 相关系数分析	6
Table 3 BIC 模型解释	9
Table 4 AIC 模型解释.....	10
Table 5 模型特点	11
Figure 1 变量分类	5
Figure 2 红包发放额频数/密度分布图	7
Figure 3 探索性数据分析变量筛选	7

引言

平台交易量从资金角度客观反映了公司的运营现状，作为重要的参考指标为经营层提高进一步决策的参考信息。报告旨在研究 2015 年 3 月初到 6 月底平台交易量变动相关因素，为决策者下一步行动提供决策的参考信息，最终提高公司效益。

报告攫取 2015 年 3 月 4 日至 2015 年 6 月 30 日的历史数据，变量选取考虑四个维度——产品因素、平台资金因素、资金价格因素、以及替代因素，共涉及 12 个参考变量。报告尝试从 12 个参考变量中提取与平台交易量最相关的变量，建立回归模型，进而获取平台交易量变动的相关信息。

报告的具体分析思路如下：

商业目标：平台利率最大化

通过对平台工作重点的调整，有效提高平台交易量，使平台利率最大化。

数据收集：收集潜在相关变量数据

从影响产品投资角度考虑，选出可能影响平台交易量的因素，如产品利率、红包激励、总账户余额、银行利率、替代产品利率，进行数据收集。

数据挖掘：建立回归模型

获得数据后，进行数据可视化，筛选相关变量，建立回归模型，对多个模型进行对比，最后筛选最优模型。

项目计划：应用于制定平台工作重点参考

最终获得平台交易量的相关关系模型，根据模型结论，为决策提供理论依据。

数据收集

报告综合考虑四个分类因素——产品因素、平台资金因素、资金价格因素、以及替代因素，共 12 个变量（Figure1, Table1）。

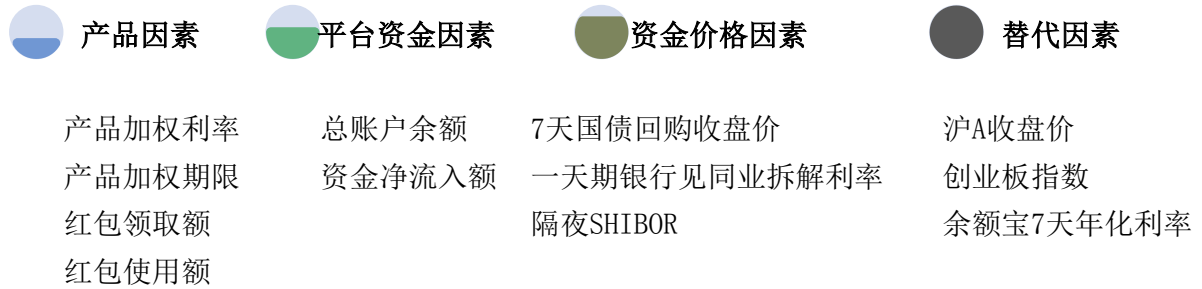


Figure 1 变量分类

变量	变量说明
平台交易量	记录平台一日 24 小时内成交总量，单位为元。
产品加权收益率	以交易金额为权重，记录平台当天产品收益率加权值，单位为%。
产品加权期限	以交易金额为权重，记录平台当天产品期限加权值，单位为月。
红包领取额	记录当天所有用户红包领取总金额，单位为元。
红包使用额	记录当天所有用户红包使用总金额，单位为元。
总账户余额	记录零点时所有用户账户总余额，单位为元。
资金净流入额	记录当天平台（资金流入额－资金流出额），单位为元。
7 天国债回购收盘价	7 天国债回购价收盘价，价格为年利率，单位为%。
一天期银行间同业拆借加权利率	银行间同业拆借市场最终成交的拆借交易利率的加权平均值，单位为%。
隔夜 SHIBOR	银行间同业市场主要做市商成员每日提供交易报价，单位为%。
沪 A 收盘价	上海证券交易所上市以人民币交易股票每日收盘价。
创业板指数	以起始日为一个基准点，按照创业板所有股票的加权平均指数。
余额宝 7 天年化利率	$\text{余额宝年化收益率} = \frac{\text{投资内收益} / \text{本金}}{\text{投资天数} / 365} \times 100\%$

Table 1 变量说明

探索性数据分析

经过相关系数分析发现，一天期银行间同业拆借加权利率、隔夜 SHIBOR、沪 A 收盘价、以及创业板指数相关度较高（Table2）。出于模型精确度的考虑，尝试合并其中部分变量；然而，隔夜 SHIBOR 与沪 A 收盘价的高度负相关缺乏理论支持。因此，报告最终将一天期银行间同业拆借加权利率合并到隔夜 SHIBOR；将创业板指数合并到沪 A 收盘价。

	L ¹	M	N	O
L	1.0000000	0.9998459	-0.9066521	-0.9105291
M		1.0000000	-0.9057128	-0.9102104
N			1.0000000	0.9365823
O				1.0000000

Table 2 相关系数分析

经过变量密度分布分析发现，红包发放额为典型的双峰分布（Figure2）。进一步观察分析红包发放额的金额分步，报告将其以 2 百万为界限，分为低额红包组与高额红包组，设为分类变量。通过分类变量，报告只分析红包的高额与低额对平台交易量的影响，不分析红包发放额具体数额对平台交易量的影响。

1

L: 一天期银行间同业拆借加权利率
M: 隔夜 SHIBOR
N: 沪 A 收盘价
O: 创业板指数

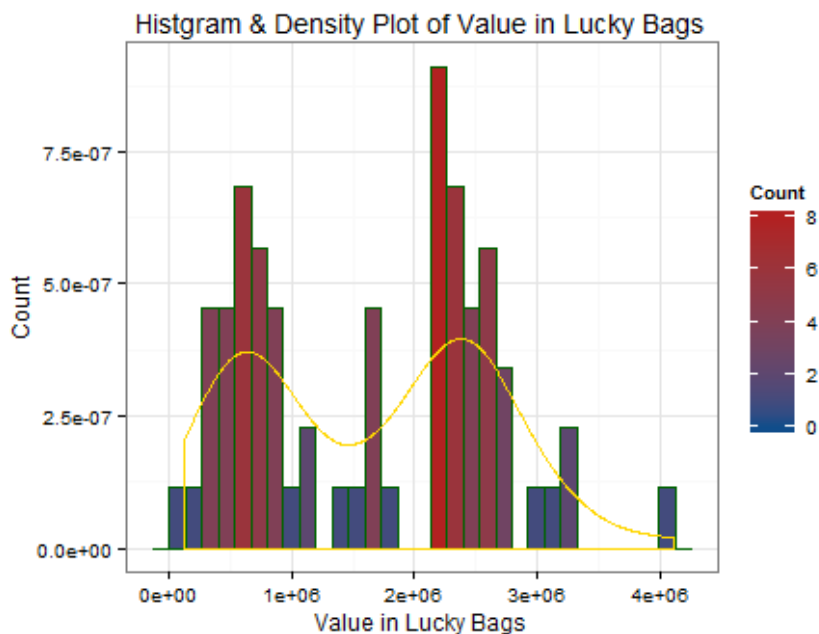


Figure 2 红包发放额频数/密度分布图

根据初步的建模分析发现，沪 A 收盘价的统计学显著程度较高（ $P - \text{value} = 2.91 \times 10^{-5}$ ），然而实际显著程度较低——其他变量恒定下，沪 A 收盘价每上升 1%，平台交易量约增加 402.8 元——因此，报告将沪 A 收盘价从模型中剔除。

经过探索性数据分析，报告初步将以下变量列入模型建模分析。

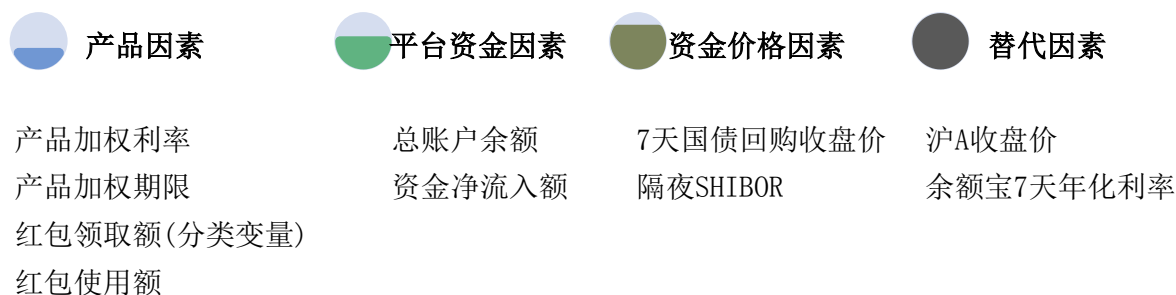


Figure 3 探索性数据分析变量筛选

常规分析

模型假设

报告尝试建立多元线性回归模型分析平台交易量的相关变量，模型建立基于以下假设²：

1. 相关变量与平台交易量成线性关系；
2. 回归模型残差独立分布；
3. 回归模型残差正太分布；
4. 回归模型残差的方差相同。

模型建立

经过初步建模，报告得到多个多元线性模型。报告采用两个筛选标准——Bayesian Information Criterion (BIC) 与 Akaike Information Criterion (AIC) 得到的多元线性回归模型如下：

基于 BIC 的模型：

$$\begin{aligned} \text{平台日交易量(百万元)} = & -4.38 - 8.95 \cdot 7 \text{ 天国债回购收盘价}(\%) \\ & + 0.27 \cdot \text{用户总账户余额(百万元)} \\ & + 14.28 \cdot \text{产品加权收益率}(\%) \end{aligned}$$

基于 AIC 的模型³：

$$\begin{aligned} \text{平台日交易量(百万元)} = & -372.8 + 73.53 \cdot \text{余额宝 7 天年化利率}(\%) \\ & - 13.90 \cdot \text{隔夜SHIBOR}(\%) \\ & + 0.24 \cdot \text{用户总账户余额(百万元)} \\ & + 18.57 \cdot \text{产品加权收益率}(\%) \\ & - 7.05 \cdot 7 \text{ 天国债回购收盘价}(\%) \end{aligned}$$

² 模型假设的合理性检验见附录。

³ 根据 AIC 模型进一步筛选显著程度 (p-value) 强的变量而得，见附录 model2, model3。

模型解释

基于 BIC 的模型：

$$\begin{aligned} \text{平台日交易量(百万元)} = & -4.38 - 8.95 \cdot 7 \text{ 天国债回购收盘价}(\%) \\ & + 0.27 \cdot \text{用户总账户余额(百万元)} \\ & + 14.28 \cdot \text{产品加权收益率}(\%) \end{aligned}$$

变量	相关关系	影响程度 ⁴	自信度	变量增量	交易量变化区间估计	交易量变化点估计
7 天国债回购收盘价	负相关	0.3935	95%	1 个百分点	(-1278) 万元 - (-512) 千万元	-895 万元
用户总账户余额	正相关	0.3866	95%	1 百万元	14 万元 - 40 万元	27 万元
产品加权收益率	正相关	0.2599	95%	1 个百分点	414 万元 - 2441 万元	1428 万元

Table 3 BIC 模型解释

模型具体解释如下：

- 产品加权收益率与平台交易量成正相关关系。在 95% 的自信度下，在 7 天国债回购收盘价及用户总账户余额恒定的情况下，产品加权收益率每提高 1 个百分点，平台交易量增加额在 414 万元到 2441 万元区间，估测值为 1428 万元；
- 用户总账户余额与平台交易量成正相关关系。在 95% 的自信度下，在产品加权收益率及 7 天国债回购收盘价恒定的情况下，用户总账户余额每增加 1 百万元，平台交易量增加额在 14 万元到 40 万元区间，估测值为 27 万元；
- 7 天国债回购收盘价与平台交易量成负相关关系。在 95% 的自信度下，在产品收益率及用户总账户余额恒定的情况下，7 天国债回购收盘价每提高 1 个百分点，平台交易量减少额在 1278 万元到 512 万元区间，估测值为减少 895 万元。

⁴ 根据取标准化回归系数（standardized regression coefficients）绝对值得到；只从数据角度提供影响程度参考值，真实影响程度需综合考虑相关金融专业知识。

基于 AIC 的模型：

$$\begin{aligned} \text{平台日交易量(百万元)} = & -372.8 + 73.53 \cdot \text{余额宝 7 天年化利率}(\%) \\ & - 13.90 \cdot \text{隔夜 SHIBOR}(\%) \\ & + 0.24 \cdot \text{用户总账户余额(百万元)} \\ & + 18.57 \cdot \text{产品加权收益率}(\%) \\ & - 7.05 \cdot \text{7 天国债回购收盘价}(\%) \end{aligned}$$

变量	相关关系	影响程度	自信度	变量增量	交易量变化区间估计	交易量变化点估计
余额宝 7 天年化利率	正相关 ⁵	0.4564	95%	1 个百分点	1567 万元-13199 万元	7353 万元
隔夜 SHIBOR	负相关	0.4243	95%	1 个百分点	(-2666 万元)-(-115 万元)	-1390 万元
用户总账户余额	正相关	0.3447	95%	1 百万元	10 万元 - 38 万元	24 万元
产品加权收益率	正相关	0.3381	95%	1 个百分点	726 万元 - 299 万元	1857 万元
7 天国债回购收盘价	负相关	0.3097	95%	1 个百分点	(-1126 万元)-(-282 万元)	-705 万元

Table 4 AIC 模型解释

⁵ 余额宝 7 天年华利率，不同与预期假设（替代因素），而是与平台交易量成正相关关系的变量；更多讨论见‘模型信息’。

模型讨论

模型选取

报告为决策者提供了两个模型——BIC 筛选标准模型与 AIC 筛选标准模型。两个模型各有侧重，因此各有优缺点。两个模型的特点如下，供决策者根据决策需求选择不同功能的模型。

Table 5 模型特点

	高信度模型（BIC 标准）	高预测度模型（AIC 标准）
特 点	<ol style="list-style-type: none"> 1. 侧重根据已有数据，找到最大限度反应数据信息的模型⁶； 2. 筛选标准严格，信度相对较高，倾向于选择更简单的模型； 3. 存在过分简化的风险。 	<ol style="list-style-type: none"> 1. 侧重根据已有数据，找到预测度最高的模型⁷； 2. 筛选标准相对宽松，预测度相对较高，倾向于选择更复杂的模型； 3. 存在过分拟合、低信度的风险。

简言之，BIC 倾向于找到更接近真实（高信度）的模型，而 AIC 倾向于找到预测性更强（高预测度）的模型。

模型信息

模型反映的产品利率、平台用户总账户余额、以及资金价格与平台交易量的相关关系均符合预期假设；然而余额宝 7 天年华利率与预期假设相反——为正相关关系，从数据角度反映出本平台与余额宝并不存在明显的替代关系。

然而，本平台与余额宝更准确的相关信息需要专业知识以及进一步深入研究，本报告对此不做深入讨论。

⁶ BIC maximizes the posterior likelihood of the data given the model.

⁷ AIC minimizes the Kullback-Leibler divergence between the training data and corresponding predicted values.

决策建议

基于上述相关关系模型，报告提出以下建议供运营层参考：

一、加强同业产品调研，提高产品创新能力，研发更具竞争力的产品

根据公司发展需求及市场环境，合理安排产品利率；

加强同业产品调研，不断提高创新能力，研发更具竞争力的产品。

二、研究、采取有效方案，提高用户总体账户余额

讨论、通过各种途径，合理合法范围内最大限度增加用户数量；

研究、采用有效方案，着重提高用户在平台的资金余额。

三、完善市场监测机制，加强宏观政策研判，提高对市场变化的应变能力

进一步加强宏观政策研判；

密切关注货币市场、资本市场、以及实体经济市场动态，着重把握市场资金需求程度、及资金流向。

总结

报告分析基于 2015 年 3 月 4 日至 2015 年 6 月 30 日历史数据，变量涉及产品因素、平台资金因素、资金价格因素、以及替代因素 4 个维度，共 12 个参考变量。报告尝试从 12 个参考变量中提取与平台交易量最相关的变量，建立回归模型，进而获取平台交易量变动的相关信息。

报告依据 AIC 及 BIC 模型筛选标准，选出两个模型。AIC 模型侧重预测性，但信度较低，存在过拟合风险；BIC 模型侧重真实反映数据，信度较高，但存在过分简化风险。

高信度模型（BIC）：

$$\begin{aligned}\text{平台日交易量(百万元)} = & -4.38 - 8.95 \cdot 7 \text{ 天国债回购收盘价}(\%) \\ & + 0.27 \cdot \text{用户总账户余额(百万元)} \\ & + 14.28 \cdot \text{产品加权收益率}(\%) \end{aligned}$$

高预测度模型（AIC）：

$$\begin{aligned}\text{平台日交易量(百万元)} = & -372.8 + 18.57 \cdot \text{产品加权收益率}(\%) \\ & + 0.24 \cdot \text{用户总账户余额(百万元)} \\ & - 7.05 \cdot 7 \text{ 天国债回购收盘价}(\%) \\ & - 13.90 \cdot \text{隔夜SHIBOR}(\%) \\ & + 73.53 \cdot \text{余额宝 7 天年化利率}(\%) \end{aligned}$$

基于模型分析，报告提出以下建议供运营层参考：

- 一、加强同业产品调研，提高产品创新能力，研发更具竞争力的产品
- 二、研究、采取有效方案，提高用户总体账户余额
- 三、完善市场监测机制，加强宏观政策研判，提高对市场变化的应变能力

附录

Model Building

```
Subset selection object
Call: regsubsets.formula(Y ~ C + D + E + F + H + I + K + N, data = r
data)
8 Variables (and intercept)
1 subsets of each size up to 8
Selection Algorithm: exhaustive
```

		C	D	E	F	H	I	K	N
1	(1)	"	"	"	"	"	"	"	"
2	(1)	"	"	"	"	"	"	"	"
3	(1)	"	"	"	"	"	"	"	"
4	(1)	"	"	"	"	"	"	"	"
5	(1)	"	"	"	"	"	"	"	"
6	(1)	"	"	"	"	"	"	"	"
7	(1)	"	"	"	"	"	"	"	"
8	(1)	"	"	"	"	"	"	"	"

```
summary(model0)$cp #Provides Mallow's Cp for each p
[1] 41.868700 17.549792 10.852234 9.490712 8.262521 7.596501 7.3
10034 9.000000
summary(model0)$bic #Provides BIC for each P
[1] -17.95337 -33.39342 -37.03148 -36.06310 -35.11331 -33.70502 -31.
98281 -28.05960
```

Model 1 BIC

```
Call:
lm(formula = Y ~ C + E + I, data = rdata)

Residuals:
    Min       1Q   Median       3Q      Max
-39834467 -16317834  1303892  16482257  48554931

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.379e+07  4.694e+07  -0.933  0.35420
C             1.428e+07  5.078e+06   2.811  0.00645 **
E             2.660e-01  6.375e-02   4.172  8.76e-05 ***
I            -8.952e+06  1.920e+06  -4.663  1.51e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21560000 on 68 degrees of freedom
(3 observations deleted due to missingness)
Multiple R-squared:  0.5285, Adjusted R-squared:  0.5077
F-statistic: 25.41 on 3 and 68 DF, p-value: 3.864e-11
```

confint(model1)

	2.5 %	97.5 %
(Intercept)	-1.374471e+08	4.987644e+07
C	4.141297e+06	2.440905e+07
E	1.387589e-01	3.931713e-01
I	-1.278380e+07	-5.121191e+06

Normalization:

lm.beta(model1)

	C	E	I
	0.2599486	0.3866093	-0.3934698

Model 2 AIC

Call:

lm(formula = Y ~ C + D + E + F + I + K + N, data = rdata)

Residuals:

	Min	1Q	Median	3Q	Max
	-35466790	-14040755	-3807970	14506821	51028382

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-3.205e+08	1.465e+08	-2.188	0.032357	*
C	1.664e+07	5.901e+06	2.820	0.006383	**
D	-5.525e+06	3.634e+06	-1.520	0.133355	
E	2.571e-01	6.928e-02	3.711	0.000434	***
F	1.846e-01	1.020e-01	1.809	0.075099	.
I	-5.278e+06	2.250e+06	-2.346	0.022097	*
K	-1.448e+07	6.264e+06	-2.311	0.024070	*
N	6.879e+09	2.907e+09	2.366	0.021014	*

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20440000 on 64 degrees of freedom
(3 observations deleted due to missingness)

Multiple R-squared: 0.6012, Adjusted R-squared: 0.5576

F-statistic: 13.79 on 7 and 64 DF, p-value: 9.801e-11

Model 3 AIC

call:

```
lm(formula = Y ~ C + E + I + K + N, data = rdata)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-36317140	-16346449	-2165846	14077477	49448629

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.728e+08	1.457e+08	-2.560	0.01278 *
C	1.857e+07	5.664e+06	3.278	0.00167 **
E	2.371e-01	7.002e-02	3.387	0.00120 **
I	-7.046e+06	2.114e+06	-3.333	0.00141 **
K	-1.390e+07	6.389e+06	-2.176	0.03315 *
N	7.353e+09	2.928e+09	2.511	0.01449 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20900000 on 66 degrees of freedom
(3 observations deleted due to missingness)

Multiple R-squared: 0.57, Adjusted R-squared: 0.5375

F-statistic: 17.5 on 5 and 66 DF, p-value: 5.364e-11

confint(model4)

	2.5 %	97.5 %
(Intercept)	-6.636795e+08	-8.201509e+07
C	7.259939e+06	2.987598e+07
E	9.733974e-02	3.769541e-01
I	-1.126717e+07	-2.824875e+06
K	-2.665568e+07	-1.145414e+06
N	1.506742e+09	1.319872e+10

Normalization

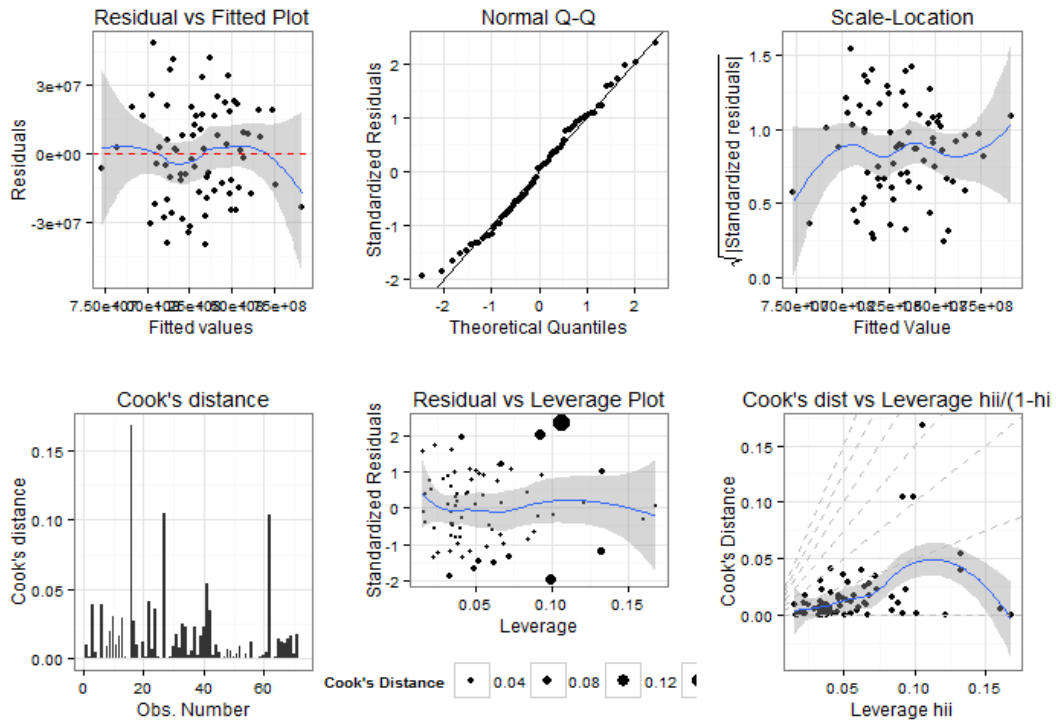
lm.beta(model4)

	C	E	I	K	N
	0.3381194	0.3447189	-0.3096788	-0.4242698	0.4563917

Diagnostics

Model 1

Diagnostic Plots



Model 3

Diagnostic Plots

