

选股因子数据的异常值处理和正态转换

— 《金工磨刀石系列之二》



研究结论

- 数据处理是因子选股实证研究的第一步。本报告将结合我们的因子库数据分析比较多种极值处理和正态转换方法。
- 数据极值侦测手段通常有四种：固定比例、均值标准差、MAD 和 Boxplot 方法，实证比较发现经过偏度调整的 Boxplot 方法对异常数值的侦测效率最高，需要人力干预的地方少，建议在实证研究中采用。
- 数据的非正态性会影响 alpah 因子的 zscore 计算和风险因子的因子收益率显著性检验。Box-Cox 变换是最常用的数据正态转换，它的参数 lambda 可以通过极大似然方法估计，但在实际操作中，这个数值只能作为参考，不宜直接采用，转换后因子的经济含义是否便于理解是更值得关注的事情，应优先考虑取对数或取倒数这样简单的转换方式。

风险提示

- 模型失效风险；
- 极端行情对模型的冲击；

报告发布日期

2016 年 06 月 23 日

证券分析师

朱剑涛

021-63325888*6077

zhujiantao@orientsec.com.cn

执业证书编号：S0860515060001

相关报告

常用股票行业分类方法的比较

2016-01-22

东方证券股份有限公司经相关主管机关核准具备证券投资咨询业务资格，据此开展发布证券研究报告业务。

东方证券股份有限公司及其关联机构在法律许可的范围内正在或将要与本研究报告所分析的企业发展业务关系。因此，投资者应当考虑到本公司可能存在对报告的客观性产生影响的利益冲突，不应视本证券研究报告为作出投资决策的唯一因素。

有关分析师的申明，见本报告最后部分，其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

点击进入 <http://www.nibor.com.cn>

目录

一、异常数据的影响与侦测	3
1.1 异常数据的影响	3
1.2 固定比例法	3
1.3 均值标准差法	3
1.4 MAD 法	4
1.5 Boxplot 法	4
1.6 偏度调整 Boxplot 和 MAD 方法对比	5
二、数据正态转换	6
2.1 非正态数据的影响	6
2.2 Box-Cox 变换	7
2.3 正态变换的选择	8
三、总结	9
风险提示	9
参考文献	10

一、异常数据的影响与侦测

1.1 异常数据的影响

数据异常值对线性回归的影响在统计学教科书中多有论述，这里主要讨论其对 α 因子 zscore 得分的影响。

以 2016 年 5 月底全市场股票的 ROE(TTM)数据为例，总共有 2837 个数据，最小值 -10.9，中位数 0.067，均值 0.06，最大值 26.08，标准差 0.5720。可以看到 98% 的数据分布在 $[-0.5, 0.5]$ 这个区间内，如果去掉上下各 1% 的数据，剩余数据的均值为 0.0625，标准差变为 0.1056。异常值使得样本总体均值变化不大，但是标准差却放大了五倍多，因此如果不剔除异常值，大部分股票在 ROE 因子上的 zscore 得分将明显偏小，影响该因子在整个 α 模型中的贡献。

侦测单变量异常值的方法有很多，常用的有以下几种：

1.2 固定比例法

固定比例法的想法是把数据从小到大排列，去掉最小和最大的固定比例数据（例如：1%）来排除异常值。固定比例的设定非常主观，最好是对照数据的分布特点来设定，但这在选股因子数量较多、历史数据较长的情况下可行性太差，而且对于一些数值散落在“合理”区间的数据，强行删除固定比例的数据反而会面临不必要的信息损失。因此这种方法只能作为一种粗略筛法，以获得对异常值数据影响的初步认识，数据建模前需要更精细的异常值侦测工具。

1.3 均值标准差法

这种想法的思路来自于正态分布，假设 $X \sim N(\mu, \sigma^2)$ ，则

$$P(|X - \mu| > k \cdot \sigma) = \begin{cases} 0.317 & k = 1 \\ 0.046 & k = 2 \\ 0.003 & k = 3 \end{cases}$$

通常把离均值 μ 三个标准差以外的数据作为异常值数据。不过要注意的是样本均值和样本标准差都不是稳健统计量，其计算本身受极值的影响就非常大，所以可能会出现一种情况，那就是我们从数据分布图上能非常明显的看到异常点，但按照上面的计算方法，这个异常点可能仍在均值三倍标准差的范围内。因此按照这种方法剔除掉异常值后，需要重新观察数据的分布情况，看是否仍然存在显著异常点，若存在则继续重复上述步骤寻找异常点。

在样本数据不符合正态分布的情况下，根据 Chebyshev 不等式，我们仍然可以对均值三倍标准差外的样本数据数量占比做出估计

$$P(|X - \mu| \geq k \cdot \sigma) \leq 1/k^2$$

但不等式右边的上界数值过于偏大，在数据分布的偏度和峰度影响下，这种方法可能会把过量的数据认定为异常点。

另外，本报告的第二部分介绍了非正态分布数据的正态转换方法，那是否可以先将数据转换成正态分布，然后再用均值方差法来剔除异常点呢？理论上来说可行，特别是如果后续还要用转换后的数据来做统计分析。但需要注意的是后文介绍的 Box-Cox 变换虽是单调变换，不会改变数据的大小排序关系，但是会改变数据与数据之间的距离大小，所以这些变换可能会使得原始数据的异常点在变换后变得不异常，而原来的正常点却变换成异常值（参考 Seo (2002)里的示例）。另外，Box-Cox 变换并非万能，转换后的数据仍然可能呈现非常明显的非正态特征，而且有时为了保证变换后数据有充分的经济含义，不一定采用最优的 Box-Cox 参数来做变化，转换后数据的正态性更难保证，均值标准差仍然不适用。因此我们在量化选股研究中采取的都是先处理异常值，后做数据转换的作业方式。

1.4 MAD 法

MAD 法是针对均值标准差方法的改进，把均值和标准差替换成稳健统计量，样本均值用样本中位数代替，样本标准差用样本 MAD (Median Absolute Deviation) 代替：

$$md = median(\{x_i, i = 1, 2, \dots, n\})$$

$$MAD = median(\{|x_i - md|, i = 1, 2, \dots, n\})$$

$$MAD_e = 1.483 * MAD$$

通常把偏离中位数三倍 MAD_e (如果样本满足正态分布，且数据量较大，可以证明 $\sigma \approx 1.483 \times MAD$) 以上的数据作为异常值。和均值标准差方法比，中位数和 MAD 的计算不受极端异常值的影响，结果更加稳健。

1.5 Boxplot 法

Boxplot 法由 Turkey(1977)年提出，是一种经验处理方法，假设 Q_1 和 Q_3 分别为数据从小到大排列的 25% 和 75% 分位数，记 $IQR = Q_3 - Q_1$ ，把 $(-\infty, Q_1 - 3 \times IQR) \cup (Q_3 + 3 \times IQR, \infty)$ 区间里的数据标识为异常点。分位数也是稳健统计量，因此 Boxplot 方法对极值不敏感，但如果样本数据正偏严重，且右尾分布明显偏厚时，Boxplot 方法会把过多的数据划分为异常数据，因此 Hubert & Vandervieren (2007) 对原有 Boxplot 方法进行了偏度调整。首先样本偏度定义采用了 Brys(2004) 提出的 MedCouple 方法：

$$md = median(x_i, i = 1, 2, \dots, n)$$

$$mc = median\left(\frac{(x_i - md) - (md - x_j)}{x_i - x_j}, x_i \geq md \text{ \& \& } x_j \leq md\right)$$

然后再给出了经过偏度调整 Boxplot 方法上下限

$$L = \begin{cases} Q_1 - 1.5 * \exp(-3.5 * mc) * IQR & \text{if } mc \geq 0 \\ Q_1 - 1.5 * \exp(-4 * mc) * IQR & \text{if } mc < 0 \end{cases}$$

$$U = \begin{cases} Q_3 + 1.5 * \exp(4 * mc) * IQR & \text{if } mc \geq 0 \\ Q_3 + 1.5 * \exp(3.5 * mc) * IQR & \text{if } mc < 0 \end{cases}$$

区间 $(-\infty, L) \cup (U, \infty)$ 里的数据被定义为异常数据。和原始 Boxplot 方法相比，当样本数据分布右偏时，此法会提升正常数据区间上限的数值；样本数据左偏时，则会降低正常数据区间下限的数值。Hubert & Vandervieren (2007) 仿真生成了不同偏度分布的数据样本，并在其中人为加入一些异常数据，发现调整后 Boxplot 方法对异常数据的识别能力比传统方法要强很多。

1.6 偏度调整 Boxplot 和 MAD 方法对比

异常数据不是一个完全客观的定义，那些极大和极小的数据可能蕴含了样本分布的重要信息，这些数据的去留决定存在人为主观因素，因此我们只能从这些方法得到的结果哪个更符合研究需求的角度去比较各种方法的优劣。下面比较了偏度调整 Boxplot 和 MAD 两种稳健方法的实际效果。

测试数据是我们数据库里面 44 个因子从 2006.01 至 2016.05 的月度数据，分别用两种方法对每月的因子数据进行检测，考察不同因子平均每个月检测出的异常数据占比，结果如表 1 所示。

表 1：偏度调整 Boxplot 方法和 MAD 方法平均每月侦测到的异常数值比例

Factor	MAD	Boxplot	Factor	MAD	Boxplot
FFMV	13.5%	2.8%	EPS1YGrowth_YOY	13.1%	14.1%
MV	14.5%	3.3%	PEG	13.8%	7.2%
BP_LF	3.6%	2.1%	SalesGrowth_Qr_YOY	7.9%	9.4%
EP_TTM	9.4%	10.1%	ProfitGrowth_Qr_YOY	14.3%	15.1%
EP2_TTM	9.2%	10.5%	EquityGrowth_YOY	18.7%	13.6%
SP_TTM	9.9%	0.9%	OCFGrowth_YOY	13.7%	14.6%
CFP_TTM	10.2%	10.3%	Debt2Asset	1.7%	2.1%
DP_TTM	6.8%	0.6%	Ret1M	3.5%	3.6%
EP_FY1	5.5%	3.8%	Ret3M	3.8%	3.0%
EP_Fwd12M	4.9%	3.3%	PPReversal	2.1%	2.1%
EBIT2EV	6.5%	6.8%	CGO_3M	1.2%	1.8%
DP_FY1	6.6%	1.1%	TO	7.1%	1.9%
ROA	8.8%	8.6%	TO_adj	1.6%	2.0%
ROE	9.0%	8.9%	ILLIQ	5.5%	0.9%
GrossMargin	4.8%	3.5%	Beta	2.2%	1.8%
NetMargin	12.6%	10.9%	IVRCAPM	1.9%	0.9%
AssetTurnover	4.7%	2.2%	IVRFF	1.9%	0.8%
InvTurnover	11.9%	3.3%	IRFF	0.5%	0.0%
GP2Asset	4.2%	2.2%	EarningsRevision_FY1_1M	45.9%	23.9%
ROIC	8.5%	8.1%	EarningsRevision_FY1_3M	14.9%	14.4%
Accrual2NI	14.4%	14.8%	EPSRevision_FY1_1M	43.6%	21.7%
OperatingProfitGrowth_Qr_YOY	13.9%	14.8%	EPSRevision_FY1_3M	14.4%	13.4%

资料来源：东方证券研究所 & Wind 资讯

和 Seo(2002)做的仿真测试结果一致，对于市值和流通市值这样样本偏度较大的数据，MAD 方法识别出的异常数值比例明显高于偏度调整后的 Boxplot 方法；而表 1 右下角的分析师类因子，由于当月没有分析师出报告更新预测，导致因子数值有许多为 0，MAD 方法近乎将一半数据当做

了异常值。整体而言，MAD 方法认定的异常数据数量明显高于 Boxplot 方法，剔除或修改太多数据不利于后续的统计分析。可以尝试人为降低 MAD 方法中的一些阈值设置来降低异常数据占比，但这样有可能使得一些本该剔除的数据未被侦测出来，如果每个因子都设置一个阈值的话，调试的工作量会比较大。相对来说，偏度调整过的 Boxplot 方法对不同因子的适应性更强，建议在实证研究中采用。

本节最后再补充一点，之前提到的这些异常值侦测方法都是针对单变量的，而我们实际做选股模型往往会用到多个因子数据，多个一维的正常数据组合在一起可能变成一个高维的异常数据。例如，10 岁是一个正常年龄，180cm 是一个正常身高，但如果组合在一起 10 岁小孩 180cm 身高则会变成一个异常数据。因此除了单变量的异常值侦测，理论上也需要考察多个因子组合在一起后的高维异常值，不过在做高维数据分析时，需要填充单维度上的缺省数据，而且高维分析很多基于高维联合正态分布假设，有必要先做数据变化再来做异常值分析，具体方法可以参考 Aggarwal (2001)。

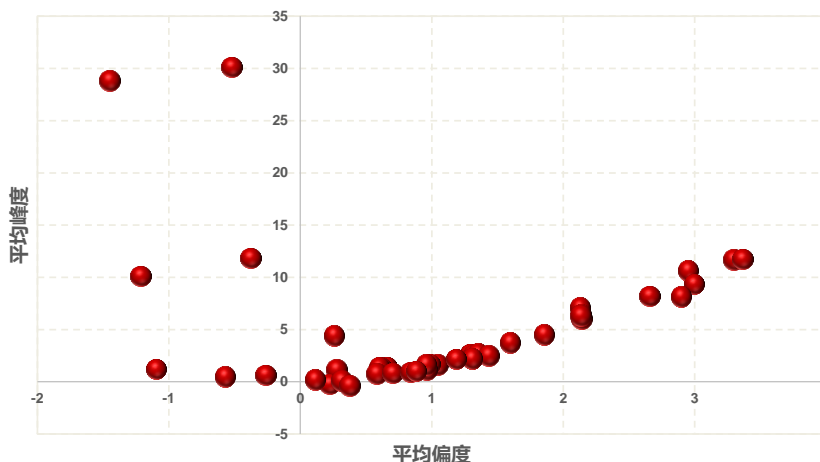
二、数据正态转换

2.1 非正态数据的影响

因子选股过程中至少有两个地方涉及因子数据的正态分布假设：

1. Alpha 因子的 zscore 计算。只有当两个因子的分布都为近似正态分布时，基于均值-标准差标准化得到的 zscore 才有可比性，否则样本偏度和峰度的影响会使得个股在某一个因子上的得分明显偏大或偏小，其在多因子中的效用被显著放大或缩小，也就是 Scinto(2009) 提到的 Percentil Ranking Problem。
2. 风险因子的因子收益率显著性检验。因子收益率由横截面回归得到，其显著性检验有赖于残差项的正态分布假设，虽然线性回归对解释变量与被解释变量的分布没有要求，但当解释变量和被解释变量满足正态分布时，更容易得到符合正态分布的残差。

图 1：44 个股票因子的平均偏度和峰度散点图（2006.01 – 2016.05）



资料来源：东方证券研究所 & Wind 资讯

表 2：44 个股票因子的平均偏度和峰度数值（2006.01 – 2016.05）

Factor	平均偏度	平均峰度	Factor	平均偏度	平均峰度
FFMV	3.30	11.70	EPS1YGrowth_YOY	0.26	4.41
MV	3.36	11.75	PEG	2.89	8.19
BP_LF	1.05	1.64	SalesGrowth_Qr_YOY	0.96	1.18
EP_TTM	0.27	1.20	ProfitGrowth_Qr_YOY	0.60	1.35
EP2_TTM	-0.26	0.65	EquityGrowth_YOY	2.66	8.23
SP_TTM	2.95	10.67	OCFGrowth_YOY	0.58	0.79
CFP_TTM	1.31	2.23	Debt2Asset	0.22	-0.23
DP_TTM	2.14	6.03	Ret1M	0.96	1.66
EP_FY1	1.30	2.66	Ret3M	1.19	2.18
EP_Fwd12M	1.29	2.66	PPReversal	0.70	0.84
EBIT2EV	0.66	1.41	CGO_3M	0.32	0.16
DP_FY1	2.13	7.11	TO	2.14	6.46
ROA	1.32	2.57	TO_adj	0.11	0.22
ROE	0.98	1.68	ILLIQ	2.13	6.37
GrossMargin	1.35	2.69	Beta	-0.57	0.48
NetMargin	1.85	4.54	IVRCAPM	0.84	0.94
AssetTurnover	1.43	2.48	IVRFF	0.88	1.03
InvTurnover	2.99	9.36	IRFF	0.37	-0.36
GP2Asset	1.31	2.23	EarningsRevision_FY1_1M	-0.52	30.14
ROIC	1.59	3.78	EarningsRevision_FY1_3M	-0.38	11.85
Accrual2NI	-1.10	1.22	EPSRevision_FY1_1M	-1.45	28.84
OperatingProfitGrowth_Qr_YOY	0.61	1.40	EPSRevision_FY1_3M	-1.22	10.12

资料来源：东方证券研究所 & Wind 资讯

图 1 和表 2 是我们数据库里 44 个因子做完极值处理后每个月因子数据偏度和峰度的平均值，可以看到大部分因子的平均偏度绝对值在 3 以内，平均峰度在 10 以内（正态分布的峰度为 0）。

2.2 Box-Cox 变换

把非正态数据转换为正态数据通常有取对数、开根号、求倒数等方法，这些都可以归为 Box-Cox 变换，形式上可以写作

$$y = \begin{cases} (x^\lambda - 1) / \lambda & \text{if } \lambda \neq 0 \\ \log(x) & \text{if } \lambda = 0 \end{cases}$$

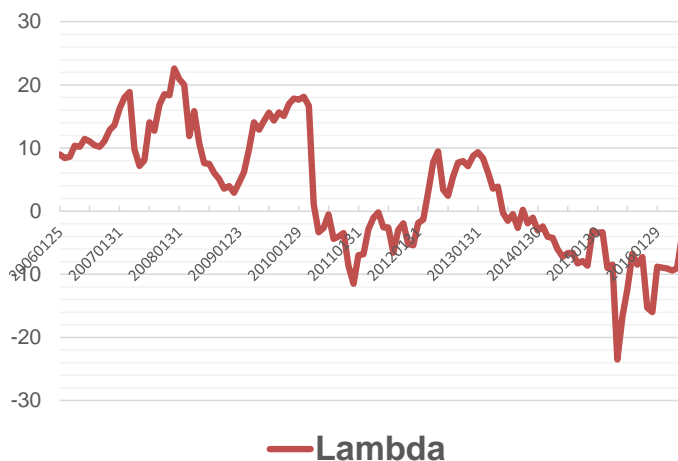
参数 λ 可以通过极大似然法做估计（Box & Cox 1964）。不过 Box-Cox 变换并非万能，变换后的数据仍然可能不满足正态分布，被统计检验拒绝，但是这种变换得到的数据分布往往都有明显的对称性，与正态分布的相似性更高。

另外，Box-Cox 变换针对的是大于零的正数数据，如果 $x < 0$ ，通常有两种做法，一种是给 x 加上一个数值 μ ，使得序列 $x + \mu$ 为正， λ 和 μ 可以都做为参数，通过极大似然方法求解。另一种是 Yeo & Johnson (2000)的做法，对大于零和小于零的数采取不同的幂数变化。

2.3 正态变换的选择

选股因子数据的复杂之处在于它同时具备时间序列和横截面两个维度的变化，由于通常的 alpha 因子 zscore 打分和风险因子中性化处理都在横截面方向进行，因此我们主要对同一个时间点横截面上的因子数据做正态转换处理。但是如果把数据正态转换当成一个纯粹的数学问题，每个横截面上都用极大似然方法估算一个最优的 λ ，会发现有的选股因子的 λ 数值变化非常之大。以 EP 因子的月度数据为例（图 2），我们先把原始的 EP 因子数据加上 1，让数据变正，再来求解每个月的最优 Box-Cox 变换参数 Lambda，可以看到 Lambda 的变化是非常剧烈的，也就是说每个月我们需要对原始的数据都做不同的数据变化。Box-Cox 是一个单调变换，保持了原有数据的相对秩序，因此变换后的因子和原始因子的 IC（秩相关系数）相等；但这种变换会改变 alpha 因子的 zscore，频繁剧烈变换的参数可能会对结果产生负面影响，变换后因子的经济含义也更加难以解释，所以把因子数据正态变换当成一个单纯数学问题的方法不可取。

图 2：EP 因子 Box-Cox 的最优 Lambda (2006.01 – 2016.05)



资料来源：东方证券研究所 & Wind 资讯

我们在做指数增强或者 alpha 对冲组合时往往会做行业和市场风格的风险控制，此时要考察的不再是原始的 alpha 因子是否具备预测能力，IC 是否显著，而应该考察风险调整后的 alpha 因子是否还具备预测能力，风险调整 IC 是否显著。有关风险调整因子的计算方法可以参考 Qian (2005)。

风险调整后的因子和原始因子的偏度、峰度都会有显著差别，例如我们之前的因子选股报告里面通常把以下八个因子作为 alpha 因子，在做了行业中性 and 市值中性的风险控制后，因子每个月的偏度和峰度平均值如下表 3 所示，同样作为估值因子，BP_LF 在做了风险调整后，偏度由 3.36 降到了 2.03，但 SP_TTM 的偏度由 2.95 提升为 3.37，峰度也明显增大。

八个 alpha 因子在风险调整后，仅有 BP_LF、SP_TTM 和 TO 因子还有明显的偏度，峰度也明显偏高，因此只需对这三个因子做正态转换。在转换过程中我们参考了表 3 第六列的 Box-Cox 变换最优 lambda 的平均值，但并不是直接采用。对于 BP_LF 和 SP_TTM 因子，我们先把原始数据加 1 使得所有数据变为正数，然后再取倒数；而换手率因子则直接取对数做变换。这样 BP 和 SP 因子就“近似”的变成了 PB 和 PS 因子，但分母加上了股价，使得分母不会是负值，这种变换的经济含义相对来说更加容易理解。而 TO 指标本身都是正值，做对数转换更符合通常的习惯。

表 3：正态变换前后 alpha 因子及其风险调整因子的偏度和峰度变化

Alpha Factors	原始因子		风险调整因子		原始因子Box-Cox变换	原始因子（正态转换后）		风险调整因子（正态转换后）	
	平均偏度	平均峰度	平均偏度	平均峰度	最优Lamda平均值	平均偏度	平均峰度	平均偏度	平均峰度
BP_LF	3.36	11.75	2.03	7.17	-1.32	-0.09	0.13	-0.98	2.52
EP2_TTM	-0.26	0.65	-0.02	2.87	3.42				
SP_TTM	2.95	10.67	3.37	17.61	-2.11	-0.79	0.15	-1.01	1.95
GP2Asset	1.31	2.23	0.96	3.19	-4.35				
ProfitGrowth_Qr_YOY	0.60	1.35	0.39	2.99	-0.12				
PPReversal	0.70	0.84	0.70	1.82	-3.28				
TO	2.14	6.46	1.70	8.77	-31.58	-0.19	0.60	-0.52	1.88
IRFF	0.37	-0.36	0.12	1.29	-0.34				

资料来源：东方证券研究所 & Wind 资讯

从表 3 后面四列可以看到，原始因子经过上述的正态转换后，样本平均偏度和峰度明显下降，对转换后的因子再做风险调整，得到的风险调整因子的偏度和峰度都在可接受的水平范围内。

三、总结

数据处理是因子选股实证研究的第一步，本报告中我们比较发现偏度调整 Boxplot 方法是一种非常高效的数据极值侦测手段，不需要频繁的人力干预。而因子数据的正态转换则是一个半主动的选择过程，Box-Cox 变换的最优参数适合作为一个参考，但不宜直接使用，转换后因子的经济意义是否容易理解是更值得考虑的事情，因此建议首先采用最简单的取对数或取倒数的变换方法，这样变换后的因子可能不是最接近正态分布，但只要峰度和偏度在可接受的范围内，即可使用。如果这样变换后的因子仍呈现明显的非正态，这时再考虑更高阶的 Box-Cox 变换。

风险提示

1. 量化模型基于历史数据分析得到，随着市场演变，模型存在失效风险。
2. 市场极端行情可能会短期内对模型的使用效果造成巨大冲击。

参考文献

- [1]. Aggarwal, C. C., Yu, P. S., (2001), "Outlier Detection for High Dimensional Data", In: Proceedings of the ACM SIGMOD Conference 2001
- [2]. Box, G.E.P., Cox, D.R.. (1964). An analysis of transformations (with discussion). Journal of the Royal Statistical Society - Statistical Methodology 26:211-252.
- [3]. Brys, G., Hubert, M., Struyf, A., (2004). "A robust measure of skewness". Journal of Computational and Graphical Statistics 13, 996–1017
- [4]. Hubert,M.,Vandervieren,E.,(2008),"An adjusted boxplot for skewed distributions", Computational Statistics and Data Analysis vol(52),pp: 5186–5201
- [5]. Scinto, D., Hardin, J., (2009), "Stock Ranking and Portfolio Selection : Revising and Developing Z-scores", working papers.
- [6]. Qian,E., Hua,R., Sorensen, E., (2007), " Quantitative Equity Portfolio Management: Modern Techniques and Applications", Chapman & Hall/CRC
- [7]. Seo, S., (2002), "A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets", master dissertation, Kyunghee University.
- [8]. Tukey (1977) , "Exploratory data analysis". Addison-Wesely.
- [9]. Yeo, In-Kwon and Johnson, Richard (2000). A new family of power transformations to improve normality or symmetry. Biometrika, 87, 954-959。

分析师申明

每位负责撰写本研究报告全部或部分内容的研究分析师在此作以下声明：

分析师在本报告中对所提及的证券或发行人发表的任何建议和观点均准确地反映了其个人对该证券或发行人的看法和判断；分析师薪酬的任何组成部分无论是在过去、现在及将来，均与其在本研究报告中所表述的具体建议或观点无任何直接或间接的关系。

投资评级和相关定义

报告发布日后的 12 个月内的公司的涨跌幅相对同期的上证指数/深证成指的涨跌幅为基准；

公司投资评级的量化标准

买入：相对强于市场基准指数收益率 15%以上；

增持：相对强于市场基准指数收益率 5%~15%；

中性：相对于市场基准指数收益率在-5%~+5%之间波动；

减持：相对弱于市场基准指数收益率在-5%以下。

未评级 —— 由于在报告发出之时该股票不在本公司研究覆盖范围内，分析师基于当时对该股票的研究状况，未给予投资评级相关信息。

暂停评级 —— 根据监管制度及本公司相关规定，研究报告发布之时该投资对象可能与本公司存在潜在的利益冲突情形；亦或是研究报告发布当时该股票的价值和价格分析存在重大不确定性，缺乏足够的研究依据支持分析师给出明确投资评级；分析师在上述情况下暂停对该股票给予投资评级等信息，投资者需要注意在此报告发布之前曾给予该股票的投资评级、盈利预测及目标价格等信息不再有效。

行业投资评级的量化标准：

看好：相对强于市场基准指数收益率 5%以上；

中性：相对于市场基准指数收益率在-5%~+5%之间波动；

看淡：相对于市场基准指数收益率在-5%以下。

未评级：由于在报告发出之时该行业不在本公司研究覆盖范围内，分析师基于当时对该行业的研究状况，未给予投资评级等相关信息。

暂停评级：由于研究报告发布当时该行业的投资价值分析存在重大不确定性，缺乏足够的研究依据支持分析师给出明确行业投资评级；分析师在上述情况下暂停对该行业给予投资评级信息，投资者需要注意在此报告发布之前曾给予该行业的投资评级信息不再有效。

免责声明

本证券研究报告（以下简称“本报告”）由东方证券股份有限公司（以下简称“本公司”）制作及发布。

本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。本报告的全体接收人应当采取必要措施防止本报告被转发给他人。

本报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的证券研究报告。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的证券研究报告之外，绝大多数证券研究报告是在分析师认为适当的时候不定期地发布。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他投资标的的邀请或向人作出邀请。

本报告中提及的投资价格和价值以及这些投资带来的收入可能会波动。过去的表现并不代表未来的表现，未来的回报也无法保证，投资者可能会损失本金。外汇汇率波动有可能对某些投资的价值或价格或来自这一投资的收入产生不良影响。那些涉及期货、期权及其它衍生工具的交易，因其包括重大的市场风险，因此并不适合所有投资者。

在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任，投资者自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，所有报告版权均归本公司所有。未经本公司事先书面协议授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容。不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

经本公司事先书面协议授权刊载或转发的，被授权机构承担相关刊载或者转发责任。不得对本报告进行任何有悖原意的引用、删节和修改。

提示客户及公众投资者慎重使用未经授权刊载或者转发的本公司证券研究报告，慎重使用公众媒体刊载的证券研究报告。

东方证券研究所

地址：上海市中山南路 318 号东方国际金融广场 26 楼

联系人：王骏飞

电话：021-63325888*1131

传真：021-63326786

网址：www.dfzq.com.cn

Email：wangjunfei@orientsec.com.cn