

上市公司净资产收益率预测分析报告

内容摘要 本报告利用我国上市公司的当年财务指标数据对其下一年度的盈利情况进行分析和预测。从分析结果发现，公司当年的净资产收益率（ROE）、资产周转率（taTurnover）、对数转换的资产总计（logTAssets）、市净率（PB）、营业利润/营业总收入（opTR）、应收账款/营业收入（arR）这 6 个财务指标，对于预测公司下一年的净资产收益率（ROEf）很重要，其中前 5 个指标与下一年的净资产收益率（ROEf）有着正相关关系。这 6 个财务指标主要取决于公司的规模大小、主营业务状况、市场供求状况和资产运营情况。基于本报告的分析结果来看，投资者和管理者可以利用上市公司当年的财务指标了解公司的投资风险、发展状况等信息，从而进行合理的投资规划。

一、研究目的

在金融市场上，如何通过上市公司当年公开可得的财务指标对其来年的盈利进行合理的预测，是一个非常重要的问题。它不仅能帮助投资者了解公司的盈利模式、风险大小以及进行正确的投资，而且能使管理者根据预测结果对企业的发展规划、资源配置等方面对公司进行合理的规划和部署。

目前，市场上通常使用净资产收益率（ROE）来衡量企业的盈利状况，因此，本报告将对我国 A 股上市公司的财务数据进行分析，并尝试对上市公司来年净资产收益率（ROEf）进行简单回归预测，最后根据分析结果提出相关结论和建议。

二、数据来源和相关说明

考虑数据采集的便利性和完整性，在优矿平台上，选取 A 股公司 2015 至 2016 年的数据作为报告样本（样本量分别为 1204、1215）。其中，以 2015 年数据作为训练集，2016 年数据作为测试集。

考虑变量的解释性，以来年的净资产收益率（ROEf）为因变量，选取了下列财务指标作为自变量：

- ❖ 当年的净资产收益率（ROE）：反映当年公司盈利情况。可提出假设，公司当年盈利表现与来年的较为一致。
- ❖ 资产周转率（taTurnover）：反映企业资产管理效率；
- ❖ 营业利润/营业总收入（opTR）：衡量企业经营效率；

- ❖ 债务资本比率（lev）：反映企业长期偿债能力；
- ❖ 营业收入同比增长率（revenueYOY）：评价企业成长状况和发展能力；
- ❖ 市净率（PB）：反应公司净资产与市价的关系；
- ❖ 应收账款/营业收入（arR）：反映企业应收帐款投资效率；
- ❖ 存货/资产总计（inv）：反映企业的存货状况；
- ❖ 对数变换后的资产总计（logTAssets）：反映企业规模。

三、数据预处理

通过对原始数据的初步了解，采用下述方法对缺失值、异常值进行处理：

- ❖ 缺失值剔除：若变量或样本的缺失率大于等于缺失值阈值（0.2），则剔除该特征或者样本；
- ❖ 缺失值填补：使用训练集对应的变量的中位数填补；
- ❖ 极值处理：使用分位数替换的方法进行极值处理（缩尾），即将大于百分位上限阈值（97.5%）和小于下限阈值（2.5%）的值，分别用相应的上下限百分位阈值替换。

四、描述性分析

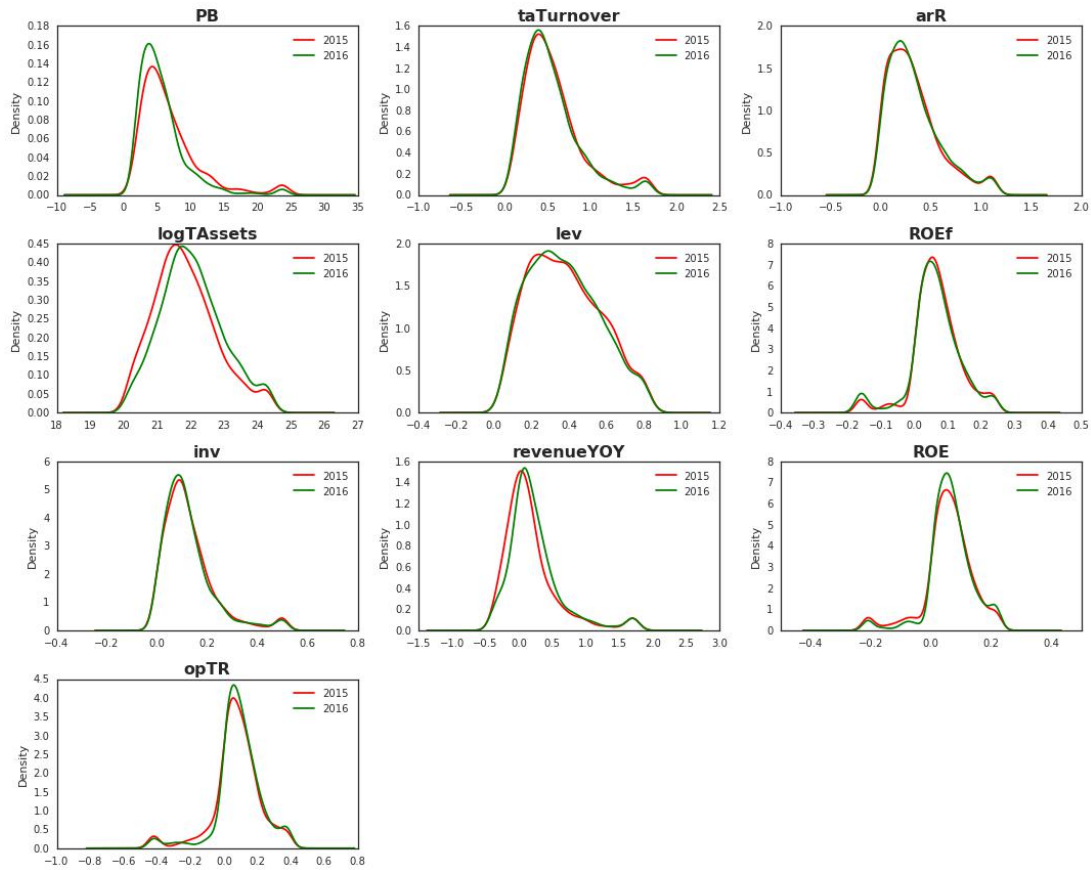
将分别从变量分布情况和变量间的关系来对变量有一个整体的印象。

1. 变量分布情况

表 1、图 1 分别是通过基本统计值、分布趋势图对变量数据的集中度（均值、中位数）和离散度（最小值、最大值、标准差）和分布稳定性进行描述。

表 1 基本描述					
变量名	均值	最小值	中位数	最大值	标准差
PB	6.312	2.004	5.155	24.243	4.235
taTurnover	0.559	0.125	0.483	1.646	0.334
arR	0.335	0.007	0.277	1.105	0.263
logTAssets	21.947	20.205	21.851	24.269	0.940
lev	0.378	0.073	0.363	0.790	0.188
ROEf	0.066	-0.159	0.063	0.237	0.074
inv	0.125	0.002	0.103	0.485	0.100
revenueYOY	0.210	-0.352	0.114	1.705	0.412
ROE	0.061	-0.210	0.061	0.219	0.078
opTR	0.082	-0.421	0.082	0.379	0.142

图 1 单变量分布情况



从上述可以看出：

- ❖ **因变量** 公司当年的净资产收益率（ROE）介于-21.0%和 21.9%之间，平均水平为 6.1%，标准差为 7.8%；公司下一年度的净资产收益率（ROEf）介于-15.9%和 23.7%之间，平均水平为 6.6%（平均数）和 6.3%（中位数），标准差为 7.4%。从分布图上来看，大部分公司有正的净资产收益率（ROE）；同当年相比，公司下一年度的净资产收益率（ROEf）稍有增长，且不同公司间的差异略有缩小。
- ❖ **自变量** 从市净率（PB）的均值（6.3）和标准差（4.2）可以看出，市场认为大部分公司具有良好的发展空间，但公司间的差异较大，且易受市场因素影响；资产周转率（taTurnover）的均值（0.56）和标准差（0.33）从侧面反映了大多数公司资产的平均利用水平；应收账款/营业收入（arR）的均值（0.34）和标准差（0.26）可反映出，在相当多的公司的营业收入中应收账款占比较大；对数变化后的资产总计（logTAssets）有较大的取值范围（从 20.21 到 24.27），说明不同的公司在资产规模上的差距较大；值得注意的是，债务资本比率（lev）有较大的取值范围（从 0.07 到 0.79），这表明不同公司间的债务水平差别较大；从存货/资产总计（inv）的均值（0.13）、标准差（0.10）以及取值范围（从 0.00 到 0.49）来看，大部分公司对存货的控制力度较强，以避免出

现高存货率的现象；营业收入同比增长率（revenueYOY）的有较大的取值范围（从-35%到171%），表明了各个公司所处的发展阶段呈现出了多样化现象；营业利润/营业总收入（opTR）的均值（0.08）和标准差（0.14）则反映出了大部分公司的盈利水平。

❖ 其他 大部分变量分布趋势呈现出多波峰现象，进一步说明公司间的差异性。

2. 变量间关系

图 2 和图 3 分别通过相关性分析、散点图进一步描述变量间的相互关系。

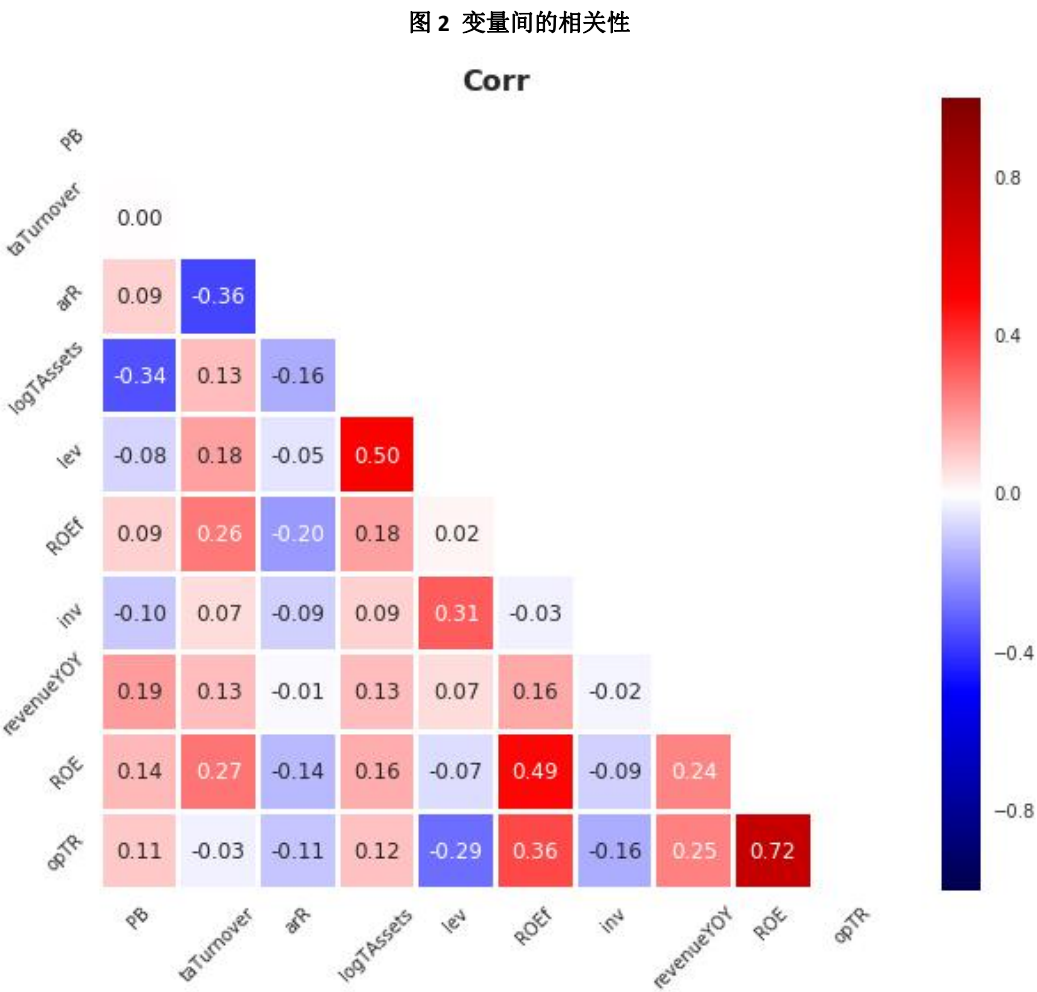
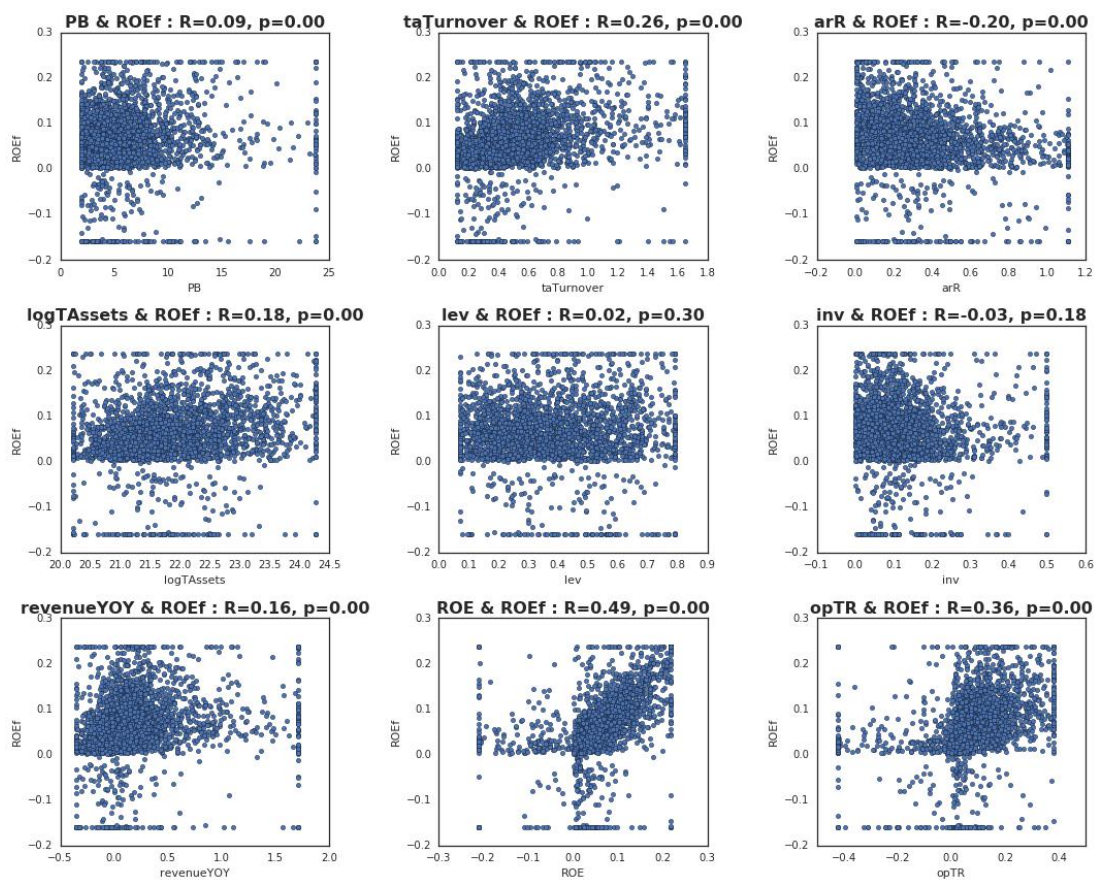


图 3 自变量与因变量关系



从上述可以看出：

相关性最高的是公司当年的净资产收益率（ROE）与营业利润/营业总收入（opTR），其相关系数高达 0.72，都为反映公司盈利水平的指标。同时，公司当年的净资产收益率（ROE）与下年度的净资产收益率（ROEf）之间，也具有较高的相关系数（0.49），这暗示公司的当年净资产收益率（ROE）对下一年的净资产收益率（ROEf）有一定的预测能力，比较符合常识。最后从散点图看出，数据中存在较多的异常值。

五、数据建模

1. 全模型分析

本报告中用线性回归的分析方法建立模型，以此来寻求自变量（公司当年的 9 项财务指标）与因变量（公司下一年的净资产收益率，ROEf）之间的关系。利用 2015 年的数据和包含 9 个自变量的全模型对因变量进行估计，得到表 2 所示的估计结果。从表 2 可以看到，F 检验的 P 值非常小，表明模型是显著的，即自变量与因变量之间确实存在一定的关系。另外，未调整的判别系数（ R^2 ）为 28.4%，调整后的判别系数（ R^2 ）为 27.9%，这都表明模型对

自变量与因变量之间的关系有一定的解释能力。通过考察各自变量对应的 t 检验的 P 值，在 0.05 的置信区间下，可以判定下一年的净资产收益率（ROEf）与资产周转率（taTurnover）、对数转换的资产总计（logTAssets）、当年的资产周转率（ROE）、市净率（PB）之间有着显著的正相关关系，与应收账款/营业收入（arR）有显著的负相关关系；对于其他变量暂时没有定论。

表 2 全模型

变量名	系数估计值	标准差	P 值
截距	-0.094	0.054	0.084
PB	0.001	0.000	0.010
taTurnover	0.029	0.006	0.000
arR	-0.026	0.007	0.000
logTAssets	0.006	0.003	0.024
lev	0.005	0.013	0.685
inv	-0.024	0.018	0.194
revenueYOY	0.004	0.005	0.407
ROE	0.339	0.034	0.000
opTR	0.024	0.020	0.235
残差项标准差：0.061		模型 F 检验 P 值 < 0.001	
判决系数（R ² ）：0.284		调整的判决系数（R ² ）：0.279	

2. 模型选择与预测

从以上全模型的分析结果容易发现，有 5 个财务指标非常重要，但不能排除其他变量也具有预测能力的可能。因此采用最常用变量选择的方法，即 AIC 和 BIC，来选择具有预测能力的模型。

如果用 AIC 选择模型，可以得到以下模型估计结果，如表 3 所示。

表 3 AIC

变量名	系数估计值	标准差	P 值
截距	-0.111	0.046	0.015
ROE	0.339	0.034	0.000
taTurnover	0.030	0.006	0.000
arR	-0.025	0.007	0.001
logTAssets	0.006	0.002	0.002
PB	0.001	0.000	0.002
opTR	0.026	0.018	0.153
残差项标准差：0.061		模型 F 检验 P 值 < 0.001	
判决系数（R ² ）：0.283		调整的判决系数（R ² ）：0.279	

从表 3 可以看出，AIC 认为当年的净资产收益率（ROE）、资产周转率（taTurnover）、应收账款/营业收入（arR）、对数转换的资产总计（logTAssets）、市净率（PB）、营业利润/营业总收入（opTR）对于预测下一年的资产收益率（ROEf）重要；而且这 6 个变量都在 0.2 的水平下显著，但 AIC 的判别系数与全模型相比无差异。

如果用 BIC 选择模型，可以得到以下模型估计结果，如表 4 所示。从表 4 可以看出，BIC 认为当年的净资产收益率（ROE）、资产周转率（taTurnover）、应收账款/营业收入（arR）对于预测下一年的资产收益率（ROEf）重要，相比 AIC，其认为对数转换的资产总计（logTAssets）、市净率（PB）、营业利润/营业总收入（opTR）不重要；而且这 3 个变量都在 0.001 水平下显著，但判别系数略低于 AIC 和全模型。

表 4 BIC

变量名	系数估计值	标准差	P 值
截距	0.038	0.005	0.000
ROE	0.394	0.022	0.000
taTurnover	0.029	0.006	0.000
arR	-0.027	0.007	0.000
残差项标准差: 0.061		模型 F 检验 P 值 < 0.001	
判决系数 (R^2): 0.272		调整的判决系数 (R^2): 0.270	

为了从三个模型（全模型、基于 AIC 选出的模型、BIC）中选出预测能力较好的模型，使用 2016 年的新数据对模型的预测能力进行检验。同时，也考虑最简单的预测方法，即只用当年的净资产收益率（ROE）预测下一年的净资产收益率（ROEf），称之为“直接预测”。通过计算得到模型的预测比较，如表 5 所示。

表 5 预测结果比较

模型	直接预测	全模型	AIC	BIC
平均预测误差	0.0757	0.0655	0.0655	0.0661

从表 5 可以看出，基于模型的预测都要优于“直接预测”，平均预测误差由 7.57% 提升至 6.55% 左右。三个模型（全模型、基于 AIC 选出的模型、BIC）的平均预测误差相差无几，AIC 会优于 BIC 一些。但同全模型相比，AIC/BIC 的模型相对简单，也为了解各财务指标的预测能力的重要性提供理论依据。进一步而言，从好的预测能力、简单和相对保守的角度来看，AIC 所选择的模型能提供更多的理论思考。

六、结论与建议

从上述分析结果来看，我国上市公司的财务指标数据为预测其下一年度的盈利情况提供

了重要的信息，并且表现出一定的预测能力。具体而言，在选取的财务指标中，从保守和谨慎的角度来看，公司当年的净资产收益率（ROE）、资产周转率（taTurnover）、对数转换的资产总计（logTAssets）、市净率（PB）、营业利润/营业总收入（opTR）、应收账款/营业收入（arR）这 6 个财务指标，对于预测公司下一年的净资产收益率（ROEf）很重要，其中前 5 个指标与下一年的净资产收益率（ROEf）有着正相关关系。

进一步发现，这 6 个财务指标主要取决于公司的规模大小、主营业务状况、市场供求状况和资产运营情况。推测原因如下：规模大的公司历史悠久，经营稳定且拥有充沛的物资资源，同时，较高的资产周转率则表明公司资产得到了合理利用，反映了公司的销售能力强，拥有较多的收入来源，而这类公司也往往受投资者所喜爱，因而容易获得较高的净资产收益率。相反，应收账款占比高的公司，通常行业竞争较大，呆账率相对较高。

因此，我们的结论验证了优秀的经营管理能为公司和投资者创造良好的效益和利润，形成良性的市场效应。

[感想]

总体来说，完成一项数据分析工作，主要包含方案确定、数据分析、相关调研、编写报告这四块：

- ❖ 方案确定：包含围绕解决问题所确认的目标，结果使用方向，数据范围与来源等；
- ❖ 数据分析：包含数据采集、数据清洗、描述分析、数据建模、模型预测、模型评估优化等；
- ❖ 相关调研：包含变量含义，异常分析，模型优化方法等；
- ❖ 编写报告：将上述工作内容和分析结果通过文字和图表的方式，呈现给需求方。注意文字的通俗易懂，思路的清晰，数据和业务相结合。

在编写报告时，发现自己存在很多不足：

- ❖ 语言表达较为欠缺，因此报告中很多表述结构引用原文；
- ❖ 专业知识积累不够，理论知识与实践不能结合，如模型诊断和参数估计分析；
- ❖ 不能快速透过数据看业务。