

附录1A 分析报告

对某移动通信校园网中推荐者的推荐价值分析

1.研究目的

通过本公司校园网中推荐者的消费行为数据分析，理解推荐者的推荐价值，为公司后续的营销措施提供参考。

2.背景介绍

当前的移动通信市场竞争日趋激烈，如何稳定现有客户、提高客户忠诚度、降低本公司客户离网率是一个很重要的课题。为此，公司从2007年起，尝试推广校园网计划。希望通过该计划的推广，达到稳定高校在校生客户，培养并提高他们客户价值的目的。

具体来说，校园网计划就是一个客户忠诚度的培养计划。在公司现有客户中，有一大部分是高校在校生，他们本身就是一批非常优良的客户，有稳定的消费（如话费、短信、彩铃等）；此外，他们对各种新鲜的增值业务也乐于尝试。公司的很多新业务都是以高校在校生为切入点的。特别值得关注的是，高校在校生毕业后往往能找到较稳定且收入良好的工作，有潜质成长成为高价值客户。因此，如何深度“套牢”这样一批优质客户是公司很关注的问题。按照校园网运营规则，如果一名高校在校生希望加入校园网，他首先必须是公司的现有客户。此外，还得由已有校园网用户进行短信邀请。接受邀请，则成为校园网的一员。作为回报，所有的校园网内通话资费会变得非常便宜。当然，与网外朋友通话，资费照旧。所以，为了进一步降低自身的资费水平，网内成员有很大的动力邀请朋友加入校园网。而已经加入校园网的成员则发现很难离开，因为大部分朋友及主要的社交网络都还滞留在校园网中，一旦离开，同他们的沟通交流将变得非常昂贵。

经过几年的实践，结果喜忧参半。喜的一面是相关客户的离网率确实有明显下降，忧的是并不是每一个客户都为公司创造了价值。为了推广校园网计划，公司有着巨大的投入，其中包括对推荐者的现金激励。但现有数据表明，有的推荐者确实能够为公司带来优质客户，但有的却带来低价值，甚至负价值客户。为此，我们需要深刻理解：什么样的推荐者能真的带来客户价值，什么样的不能？如果能够有一个良好的判断，公司未来的营销资源的投入将更有目的性。

3.指标设计

我们首先考虑：什么样的指标能够刻画推荐者价值？这是因变量。首先，推荐者本身也是校园网的普通消费者。因此，毋庸置疑，他对企业的直接利润贡献是其价值的一个重要组成部分。为方便起见，简称这部分价值为该客户的“直接价值”。但这不是本研究关注的重点。本研究关注的是推荐者通过推荐其他客户所带来的“间接价值”。举例如下。假如一名推荐者为公司推荐了三名客户，那么，他为企业带来了多少利润呢？这依赖于这三名客户在被推荐前后的行为变化。如果在被推荐加入校园网之前，他们每个月总共贡献利润100元，加入校园网后变成了80元。那么这就是一个失败的推荐者，他的推荐行为为企业带来的利润相对变化为： $(80-100)/100=-20\%$ 。如果在加入校园网后，这三名客户的利润贡献是120元，那么对企业的推荐利润贡献为 $(120-100)/100=20\%$ 。因此，因变量就是某推荐者所有推荐客户在加入校园网前后的相对利润变化。

下面再考虑解释变量。作为一个初步的试探性研究，我们只考虑三个最重要的变量。我们不排除其他有价值变量存在的可能性，那是后续研究的重点。

●通话总量 (X_1)

第一个考虑的解释变量是通话总量，以分钟计。毫无疑问，这是一个很重要的变量，它直接刻画了该推荐者的活跃程度。由于校园网提供非常优惠的通话资费，因此，对那些高通话总量的用户有很强的吸引力。因此，具有高通话总量特征的推荐者也更有可能带来优质的客户。为了分析理解方便，对通话总量做了对数变换。

●大网占比 (X_2)

该变量衡量了在某推荐者的所有通话时长中，有多少发生在该运营商的网内。如果我们将一个人的通话总量看作他的社会关系网络，那么大网占比测算了该推荐者的社会关系网络被现运营商所覆盖的程度。

●小网占比 (X_3)

该变量是大网占比的有力补充，它衡量的是在推荐者所有发生在大网（即本运营商的网络）内的通话时长中，有多少发生在校园网（即一个更小的网络）内。直观地想，如果一个推荐者的小网占比很高，那么他的主要可被推荐的社会关系网络（请注意，大网以外的客户是不能加入小网的）中的绝大部分已经加入了校园网，因此，该推荐者没有充足的被推荐对象。所以，他能为企业带来的价值应该不大。

以上就是本研究所考虑的三个解释变量，它们显然不全面。由于实际经验的积累，未来一定能够构造出更好的解释变量。

4.描述分析

在正式的模型分析之前，首先对数据作简要描述，以达到了解数据、发现问题的目的。具体分析结果如表1—1所示，从中可以看到的主要结论如下。

表 1—1 各个变量的描述分析

变量名称	样本量	均值	标准差	最大值	中位数	最小值
通话总量	1 123	2.580	0.408	3.601	2.583	0.778
大网占比	1 123	0.846	0.154	1.000	0.898	0.099
小网占比	1 123	0.252	0.211	0.973	0.198	0.000
相对利润变化	1 123	0.193	0.132	0.993	0.187	-0.498

- 样本量是1123个，所有变量都没有缺失。
- 根据第二行统计分析可以看到，客户的平均通话总量为2.580log（分钟），大概对应 $\exp(2.580) = 13.2$ 分钟。如果以两倍标准差计算，绝大多数现有客户的通话总量都保持自在 $\exp(2.580-2\times0.408) = 5.8$ 分钟到 $\exp(2.580+2\times0.408) = 29.8$ 分钟之间。
- 第三行的结果告诉我们，大网占比平均为84.6%，如果以中位数计大约为89.8%，这是一个很高的比例。这说明客户中的绝大部分社会关系网络已经被运营商覆盖。
- 第四行的结果表明，小网覆盖率偏低，以样本平均值计大约为25.2%，而以样本中位数计大概只有19.8%。这说明，样本客户的可推荐社会关系网络，只有很小一部分进入了校园网。因此，未来应该还有可观的待开发空间。
- 最后一行的结果表明，就平均水平而言，推荐者确实为校园网带来了正的相对利润。平均水平以样本均值计为19.3%，以样本中位数计为18.7%。无论如何计算，这都是一个可观的比例。但是同时值得注意的是，其最小值为-49.8%。这说明，确实有推荐者为企业带来了巨大的利润损失。因此，研究并区分哪些推荐者能够为企业带来正价值很有意义。

5.模型分析

在描述分析的基础上，我们通过普通线性回归对各个因素同相对利润变化之间的关系做了模型分析。模型整体的F检验高度显著（p值 < 0.0001），这说明推荐者所带来的推荐价值确实同他的消费行为有关，确实同我们目前考虑的三个因素（即通话总量，大网占比，小网占比）有关。模型拟合优度良好，判决系数为55.7%。相关参数估计以及检验结果如表1—2所示。

表 1—2 各参数估计以及检验结果

变量名称	参数估计	标准误差	<i>t</i> -统计量	<i>p</i> -值
截距项	-0.489	0.026 2	-18.68	<0.000 1
通话总量	0.236	0.006 5	36.13	<0.000 1
大网占比	0.089	0.022 9	3.91	<0.000 1
小网占比	-0.008	0.016 6	-0.47	0.638 0

从表1—2可以看到，除了最后一个解释变量“小网占比”以外，其他的各个因素都高度显著（ p -值 < 0.0001）。具体理解如下：

- 通话总量的系数估计为0.236。这说明，在给定其他特征（即大网占比，小网占比）不变的情况下，高通话总量的推荐者，相比低通话总量的推荐者，能够带来更多的间接价值。

- 大网占比的参数估计为0.089。这说明，在给定其他特征（即通话总量，小网占比）不变的情况下，大网占比高的推荐者，相比大网占比低的推荐者，能够带来更多的利润。

- 最后由于小网占比的 p -值（0.638）高度不显著，因此没有证据证明小网占比同被推荐客户的推荐价值相关。当然，不能排除这种可能性：可能因为样本量不够大，发现不了。所以，目前不对小网占比下任何结论。

6.总结讨论

本研究通过相对利润变化刻画了推荐者价值。通过通话总量、大网占比，以及小网占比描述了推荐者的行为特征。本研究对数据做了描述分析以及回归分析，其中回归分析的判决系数良好。研究发现，通话总量以及大网占比同推荐者的推荐利润正相关，而缺乏足够证据刻画小网占比所起到的作用。

在未来的后续研究中，可以考虑两方面的改进：第一，在现有三个因素的基础上，采集更多更丰富的影响因素；第二，目前的研究忽略了消费者的生命周期特征，处在不同生命周期的推荐者以及被推荐者可能有不同的规律特征，值得研究。

附录2A 分析报告

地段因素对北京市商品房均价的影响

1.研究目的

通过对来自门户网站公开数据的统计分析，研究地段因素对北京市商品房均价的影响，为未来商品房定价提供理论依据。

2.背景介绍

在我们民族传统古训中，“安居乐业”历来都被放在一个很重要的位置，人们渴望拥有自己房产的愿望无比强烈。随着改革开放的深入、房地产市场的放开，已经很少有人还能够依靠福利分房来解决住房问题，绝大多数人需要在商品房市场上解决自己安居的问题。因此，商品房的价格就成了人们无比关注的一个话题。以北京市为例，在过去的十年中，商品房的均价增幅巨大，步步高升的商品房均价一次又一次地考验着人们心理的底线。人们不禁要问：商品房到底值多少钱？到底应该如何定价？一个合理的范围到底是怎样的？除了政府和百姓关心商品房价格，房地产开发商也会对此非常关注。行业中的俗话说：“没有卖不出去的物业，只有卖不出去的价格。”因此，商品房的定价历来是房地产营销策划的关键。定价过高，虽然会增加单位物业的利润水平，但是会降低销售总量；而定价过低，又无法实现企业利润最大化的目标。因此，房地产开发商也对商品房的合理定价很感兴趣。

那么一套商品房的合理价格到底是多少呢？这个问题实在太复杂，牵扯到的因素太多，但毋庸置疑地段将是决定商品房价格的重要依据。经验告诉我们：越是市中心房子越贵，交通越便利房子越贵，学区越好房子越贵，医疗越便利房子越贵，等等。这些朴素的经验告诉我们，“地段”这个因素的内涵是很丰富的。但这些错综复杂的因素到底如何共同地影响了商品房价格？众多因素中，哪个是最重要的，哪个是次重要的，哪个甚至是不重要的？能否将它们的相对重要性量化出来？如果能够回答这些问题，我们就有可能对新的正在开发的商品房的预期价格有所判断。这就是本研究的主要目的。

3.指标设计

本研究所采用的数据主要来自国内房地产门户网站，主要来源是各个正在销售楼盘的商业广告，这些重要的门户网站包括搜房网、新浪网、网易等。通过手工采集逐个整理，共获得184个有效数据。其中，每一个数据代表一个当时（2010年8月）正在销售的楼盘（请注意：不是某一具体房屋）。因变量是楼盘主打户型的平

均销售价格(万元/平方米)。该指标由楼盘开发商公开宣传。不同开发商对“平均价格”可能有着不同的理解。因此,该数据有一定的合理性,但是不可能绝对精确。为了分析方便,该指标已经被对数变换。除了因变量以外,我们还考虑了下面的解释变量,这些变量主要围绕楼盘的地段信息以及交通便利程度,具体如下:

●环线位置(X_1)

该指标刻画了楼盘所处的环线位置。在本案例中,它总共涉及5个不同的取值2至3环、3至4环、4至5环、5至6环,以及6环以外。如果以2环以内作为北京市传统意义上的市中心,那么环线位置从一个侧面反映了楼盘距离市中心的远近。

●装修状况(X_2)

该指标刻画了楼盘所销售的房屋是毛坯房还是精装修房。因此,该指标有两个可能的取值:精装修以及毛坯。很显然,同一个房屋,精装修的销售价格肯定比毛坯房高。但是,从市场形成的平均水平来看,到底会高多少并不清楚,需要数据来告诉我们。

●所在区县(X_3)

两个类似的楼盘,装修状况相同,所处环线相同,均价就一定一样吗?这还要看它处在哪个行政区县。不同的区县的城市定位是不一样的,相应的商业或者教育资源是不一样的,因此,房价也不会一样。例如,同处4至5环之间,丰台的房价一般来说要比海淀便宜很多。该指标在本案例中有4个不同取值:朝阳、海淀、丰台以及通州。

●有无地铁(X_4)

北京是一个交通极其拥堵的城市。在这个城市中,唯一不会堵车的交通方式就是地铁(包括城铁)。因此,一个楼盘附近是否有地铁站是一个极其重要的指标。该指标有两个取值:有或者无。

最后值得一提的是,以上的指标设计是不完整的,很多其他重要指标没有考虑。例如,学区情况,很多楼盘保持高价的重要支撑是因为它处在好的学区(中学、小学,甚至幼儿园);对于老人,楼盘附近的医疗卫生状况也很重要,有无好的三甲医院、有无好的社区医院均需考虑。受到数据限制,未能将它们全部考虑在内,后续研究中值得注意。

4.描述分析

在正式的模型分析之前，我们首先对因变量以及解释变量做必要的描述分析。主要结果如表2—1所示。核心发现以及结论如下：

●关于环线分布，首先可以看到的是样本分布。从表2—1中可以看到，3至4环5至6环的样本最多，都是60个。6环以外的样本最少，只有8个，但我们决定还是保留这一组，因为这并不是样本中最重要的一组。除此以外，还可以看到，随着环线位置 X_1 从内往外延伸（从2至3环一直到6环以外），因变量 Y 的均值在逐渐递减。以中位数计，商品房均价从2至3环的 $\exp(1.48) = 4.39$ 万元/平方米，直落到6环以外的 $\exp(0.31) = 1.36$ 万元/平方米，相对差异高达 $4.39/1.36 = 323\%$ 。这说明越靠近市中心，房价越是昂贵。

表 2—1 因变量（对数均价）的描述分析							
因素名称	水平	样本量	均值	标准差	最大值	中位数	最小值
环线位置	2 至 3 环	26	1.441	0.240	1.841	1.482	0.916
	3 至 4 环	60	1.237	0.352	2.152	1.163	0.588
	4 至 5 环	30	1.140	0.212	1.482	1.179	0.788
	5 至 6 环	60	0.860	0.211	1.361	0.854	0.470
	6 环以外	8	0.249	0.341	0.588	0.313	-0.223
装修状况	精装修	100	1.262	0.380	2.152	1.253	0.000
	毛坯	84	0.872	0.262	1.361	0.902	-0.223
所在区县	朝阳	58	1.319	0.377	2.152	1.267	0.693
	丰台	38	1.017	0.233	1.459	1.030	0.501
	海淀	52	1.155	0.266	1.723	1.171	0.560
	通州	36	0.674	0.317	1.253	0.718	-0.223
有无地铁	无	36	1.095	0.323	1.686	1.115	0.000
	有	148	1.081	0.398	2.152	1.030	-0.223

●关于装修状况，首先可以看到样本分布良好，分别为100（精装修）和84（毛坯）。相较而言，精装修房的比例稍高于毛坯房。以中位数计，精装修房的均价为 $\exp(1.25) = 3.49$ 万元/平方米，而毛坯房为 $\exp(0.90) = 2.46$ 万元/平方米。相对差异 $3.49/2.46 = 142\%$ ，比较大。

●关于所在区县，我们发现样本分布大概均匀。相比较而言，朝阳和海淀的样本量稍多，分别为58和52；丰台和通州略少，分别为38和36。各个区县相比，朝阳的房价最贵，以中位数计均价为 $\exp(1.27) = 3.56$ 万元/平方米；海淀次之，均价为 $\exp(1.17) = 3.22$ 万元/平方米。朝阳房价昂贵是因为这里是北京市最重要的商业办公区，而海淀房价昂贵是因为这里的IT产业，以及优厚的教育资源。丰台作为传统意义上的北京市主城区之一，均价也不便宜，其中位数为： $\exp(1.03) = 2.80$ 万元/平方米。最后，通州作为一个正在蓬勃发展的远郊区，目前房价相对偏低，中位数为： $\exp(0.72) = 2.05$ 万元/平方米。

●关于有无地铁，从表2—1中我们发现绝大多数楼盘（148/184）都自称附近

有地铁或者城铁。但是，实际情况并不是这样。仔细研究发现，不同楼盘广告中对“附近”的理解和定义差别巨大。有的楼盘定义的“附近”和我们一般常识所预期的很不一样。因此，我们从房地产门户网站采集到的数据不能准确地反映真实的交通便利程度，相应的统计分析的准确度会大打折扣。这就难怪在表2—1中，对比附近有无地铁的楼盘均价，无论以均值计还是以中位数计，差别都不大。这提醒我们，在后续的分析中，对有无地铁这个变量的理解和判断需要谨慎。

5.模型分析

在描述分析的基础上，通过方差分析对各个因素同对数均价之间的关系做了模型分析。由于一些实际考虑，我们对环线位置同装修状况的交互作用特别感兴趣。因此，在模型分析中加入了该交互作用。为了分析简便，没有再考虑其他交互作用。

方差分析表明，模型整体的F \square 检验高度显著（p \square 值<0.0001），这说明北京市商品房的销售均价确实同我们目前考虑的4个因素（即环线位置、装修状况、所在区县，以及有无地铁）中的至少1个因素有关。为了进一步甄别到底哪个因素重要，对各个因素做第2型方差分析见表2—2。从中可以看到，环线位置和装修状况的交互作用不显著，此外有无地铁也不显著（假设5%的显著性水平）。因此，在后面的分析中，将这两个因素剔除。重新拟合的结果表明，剩下的三个因素（即环线位置、装修状况，以及所在区县）在5%的水平下都高度显著。模型拟合优度良好，判决系数为63.4%。相关参数估计以及检验结果如表2—3所示。

表 2—2 第 2 型方差分析检验结果				
变量名称	自由度	残差和	F-统计量	p-值
环线位置	4	4.893	21.85	<0.000 1
装修状况	1	1.188	21.22	<0.000 1
环线位置×装修状况	4	0.315	1.41	0.233 3
所在区县	3	1.372	8.17	<0.000 1
有无地铁	1	0.015	0.27	0.605 4

表 2—3 各参数估计以及检验结果					
因素名称	水平	参数估计	标准误差	t-统计量	p-值
截距项		0.153	0.086	1.77	0.078 8
环线位置	2 至 3 环	1.024	0.117	8.73	<0.000 1
	3 至 4 环	0.962	0.111	8.68	<0.000 1
	4 至 5 环	0.853	0.114	7.51	<0.000 1
	5 至 6 环	0.616	0.096	6.40	<0.000 1
	6 环以外	0.000	NA	NA	NA
装修状况	精装修	0.193	0.042	4.63	<0.000 1
	毛坯	0.000	NA	NA	NA

附录3A 分析报告

财务报表分析与ST预测

1.研究目的

通过分析上市公司的公开财务报表信息，达到预测其未来两年内被ST的可能性，并以此警示投资风险。

2.背景介绍

特别处理（special treatment, ST）是我国股市特有的一项旨在保护投资者利益的政策。具体地说，当上市公司出现财务状况或其他状况异常，导致投资者难以判断公司前景，投资者利益可能受到损害时，交易所要对该公司股票交易实行特别处理。被特别处理的股票每日涨跌幅度是受到限制的。正常情况下，证监会规定一只股票的每日最高涨跌幅为10%，而被特别处理的股票其日涨跌幅被限制在5%以内，这样就通过政策性的限制约束了该股票的日内波动程度。如果把一只股票收益率的波动程度看作其风险的一个重要含义，限制股票的每日涨跌幅度似乎可以在一定程度上控制风险。除了涨跌幅度限制以外，对被特别处理的股票，证监会要求在原股票名称之前加上提醒性注释“ST”。同时，该上市公司的中期报告必须审计。如果一个ST企业仍然持续亏损，那么它将有被退市的风险。

上海证券交易所公布的《上海证券交易所股票上市规则（2008年修订）》第十三章特别处理中对什么样的企业会被ST有详细规定，具体摘录如下：

（一）最近两年连续亏损（以最近两年年度报告披露的当年经审计净利润为依据）；

（二）因财务会计报告存在重大会计差错或者虚假记载，公司主动改正或者被中国证监会责令改正后，对以前年度财务会计报告进行追溯调整，导致最近两年连续亏损；

（三）因财务会计报告中存在重大会计差错或者虚假记载，被中国证监会责令改正但未在规定期限内改正，且公司股票已停牌两个月；

（四）未在法定期限内披露年度报告或者中期报告，且公司股票已停牌两个月；

（五）公司可能被解散；

（六）法院受理关于公司破产的案件，公司可能被依法宣告破产；

(七) 本所认定的其他情形。

由此可见，判定一只股票是否应该被特别处理是一个重大而又复杂的过程。被特别处理可能有很多原因，其中最主要的原因是第一条：“最近两年连续亏损（以最近两年年度报告披露的当年经审计净利润为依据）。”

根据表3—2中汇总的数据可以看到，在2001—2007年间，随着时间推移，上市企业越来越多，被ST的企业数量也呈现整体上升趋势。但是，相对百分比保持在5%左右。

表 3—2 每年 ST 企业的数量			
被 ST 年度	样本数	ST 样本数	ST 样本数/样本数
2001	624	21	3.37%
2002	738	41	5.56%
2003	819	52	6.35%
2004	882	37	4.20%
2005	922	31	3.36%
2006	1 010	59	5.84%
2007	1 044	46	4.41%
总计	6 039	287	4.75%

由于被特别处理的企业面临退市的风险，因此投资者需要对这类企业多加小心。所以，投资者有必要关心什么样的企业更有可能被ST、它们有什么共同特征、通过正常的财务报表分析能否察觉，这就是本研究要解决的问题。

3.指标设计

为了能够准确预测企业未来的ST可能性，我们考虑了下面的常见财务指标,并对考虑它们的原因做了详细讨论。

●ARA (X₁)

该指标是应收账款与总资产的比例，它反映的是盈利质量。简单地说，两笔不同的生意，同样为企业收获了1个单位的盈利。其中一个带回来的是现金收益，而另外一个带回来的是一个“承诺”，即对方承诺在一定期限内，以一定的方式支付现金。对于企业来讲，这部分收入是应得的，但是还没有立刻兑现，因此称为应收账款。对于绝大多数企业来说，没有应收账款是不可能的。一定的回款期限已经是行业内的惯例。如果对方确实是诚实可信的，能够在约定的时间内完成现金支付，那应收账款不是一个大问题。但是，天有不测风云，应收账款只要还没兑现，就一

那应收账款不是一个大问题。但是，天有不测风云，应收账款只要还没兑现，就一定存在违约的风险。因此，从这个角度看，对于一个企业来说，其资产中应收账款所占的比重应该是越少越好。比重越少说明该企业的盈利质量越好；相反，比重越高说明该企业的盈利质量越差。有学者的研究表明，对我国上市企业而言，应收账款在资产中所占的比率同大股东对小股东的资金侵占挪用也有紧密关系，这也是我们对此变量异常关注的另外一个原因。

●ASSET (X_2)

该指标是对数变换后的资产规模，用于反映公司规模。这不是反映公司规模的唯一指标，甚至不是最好的指标。例如，我们还可以考虑净资产收益率。但是很多经营不善的企业，资不抵债，因此净资产收益率为负数，如何合理解读负的净资产收益率同公司规模之间的抽象关系，没有一个显而易见的答案。此外，可以注意到，对于金融类企业（如基金公司）来说，资产规模其实是一个相当不错的指标。但是对于制造类企业，也许员工数量也是一个不错的选择。因此，依赖于具体问题，什么叫做“公司规模”是一个很复杂的问题，几乎不大可能有一致的答案。我们这里采用了资产规模，为了提高该指标的实际解读能力，做了对数变换。

●ATO (X_3)

该指标是资产周转率。按照定义，它是一个企业在一定时期内（如一年以内）的销售收入净额除以资产平均总额而得。假设两个不同的企业（A和B）都有1个单位的平均资产。在一年以内，A企业总共做了10单生意，而B企业只做了1单。因此，A企业的营业额会高于B企业，A企业的资产周转率也会高于B企业。这说明，A企业对资产的利用率要高于B企业。所以，ATO量化的是一个企业对资产的利用效率。

●ROA (X_4)

该指标是资产收益率。按照定义，它是一个企业在一定时期内（如一年以内）的利润总额除以总资产而得，它反映的是每单位资产能够给企业带来的利润如何。因此，该指标可以看作对企业盈利能力的反映。但它不是反映企业盈利的唯一指标，甚至不是最好的指标。例如，我们还可以考虑净资产收益率，如果净资产不是负数。此外，对于某些特定行业，人们还会关注销售收益率等。总而言之，资产收益率不是最好的盈利指标，但它是一个常用指标。

●GROWTH (X_5)

该指标是销售收入增长率。按照定义，它是一个企业在一定时期内（如一年以

以销售总额占营业收入的比率。按照定义,它是企业销售总额与营业收入(XH)之比(内)的销售总额除以前一个时期的销售总额而得,它反映的是企业的增长速度。对于很多新兴的高科技企业,在其成立之初,很难实现盈利,但这并不妨碍企业高速增长。企业的高速增长会反映在什么指标上呢?首先是销售收入,也可能是资产、净资产,还可能是市场占有率等其他非财务指标。企业的高速成长会如何影响其盈利,进而影响其被特别处理的概率呢?这不是一个简单的问题。简而言之,如果企业的销售是盈利的,那么高速成长的销售带来的应该是更好的盈利和更小的特别处理概率。但是,如果企业的销售是亏损的,那么高速成长的销售带来的应该是更大的亏损和更大的特别处理概率。

●LEV (X_6)

该指标是债务资产比率,也叫做杠杆比率。按照定义,它是一个企业债务在其总资产中所占的比率,它反映的是企业的总资产中来自债权人的比率。企业的债务资产比率是如何影响企业盈利,进而影响特别处理的概率呢?这是一个颇有争议的问题。首先,过高的债务资产比显然不好,这会使企业背上沉重的债务负担,企业每年盈利中的一大部分将用于偿还利息,因此损伤盈利。但是,另一方面,几乎没有企业不举债。适当举债能给企业带来很多好处,如很多创业初期的企业,发展势头很好,但是缺乏资金,那么合理举债能够帮助企业迅速成长,占领市场,确立优势。因此,从平均水平上来讲,债务资产比率到底如何影响特别处理概率不是一个显而易见的问题。

●SHARE (X_7)

该指标是企业第一大股东的持股比率,反映的是该企业的股权结构。如果企业的的第一大股东持股比例很高(如大于70%),说明该企业一股独大,其所有者对企业的方方面面具有绝对权威;如果企业的的第一大股东持股比例很低(如小于10%),说明该企业股权分散。企业的股权结构如何影响盈利呢?过度分散的股权结构是不好的,因为这使得所有人都不会真正地关心企业,承担责任;过度集中的股权结构也不好,因为这使得第一大股东有能力侵害小股东的利益。怎样才是一个合理的比例,使得企业的利润最大化、特别处理概率最小化,是一个值得研究的问题。

我们需要特别强调,以上设计的指标体系是有一定的实际意义的,但也可以肯定地说是不完备的。例如,如果我们认为企业的ST状态是一个随着时间变化的动态过程,就应该在解释变量里面加入该企业再前一期的指标。如果我们怀疑企业ST同行业有关系,行业特征也应该作为解释变量考虑进来。所有这些,都可能是我们未来的研究课题。

4.描述分析

在正式的模型分析之前，首先对因变量以及自变量做必要的描述分析。主要结果如表3—3所示。首先讨论表3—3的最后一行ST，从中可以知道，在n=684个样本中ST企业占了5.3%。然后讨论ARA，从表3—3中得知，一般企业平均的应收账款比例为9.5%（以均值计），或者6.8%（以中位数计）。这似乎是一个不大的水平，但是值得注意的是其最大值高达63.5%，这是一个很夸张的比例。从ASSET的结果看，样本中的平均资产规模（以中位数计）为exp（20.70）=9.77亿元。此外还可以看到，企业的平均资产周转率为ATO=52.0%，一般的销售成长速度为GROWTH=11.5%，债务资产比平均保持在LEV=40.6%，平均盈利水平为ROA=5.6%。最后值得注意的是，第一大股东一般持股比例都很高，平均水平为SHARE=46.0%。

表 3—3 因变量的描述分析

变量名称	样本量	均值	标准差	最大值	中位数	最小值
ARA	684	0.095	0.092	0.635	0.068	0.000
ASSET	684	20.780	0.834	24.020	20.700	18.660
ATO	684	0.520	0.363	3.151	0.433	0.003
GROWTH	684	0.115	0.307	0.999	0.102	-0.951
LEV	684	0.406	0.166	0.980	0.407	0.018
ROA	684	0.056	0.039	0.311	0.051	0.000
SHARE	684	46.030	17.680	88.580	44.960	4.160
ST	684	0.053	0.223	1.000	1.000	0.000

我们还将各个解释变量按照ST状态分组做盒状图对比。其中发现变量ARA的组间差异最大，如图3—4所示。

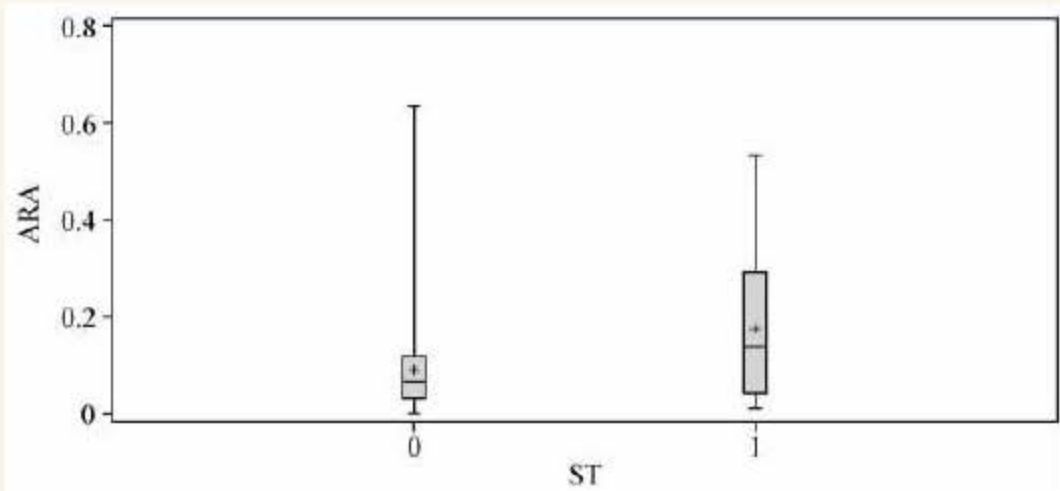


图 3—4 盒状图

从上图可以看到，非ST组（即ST=0）的ARA中位数明显低于ST组（即ST=1）或者说明，ST企业常常伴随着较高的ARA水平（以中位数计，请注意横线的位置），即较差的盈利质量。同样的规律也表现在均值上（请注意“+”的位置）。最后还可以注意到，ST=1组的变异性要比ST=0组的高（请注意盒子的厚度）。所有这一切都表明，ARA这个指标在ST=0组和ST=1组之间的规律是不一样的。因此，

这一切都表明，ARA达1组和非ST=0组和非ST=1组之间的条件是一个正数。因此，我们可以合理地预期该指标对于判断预测企业ST有重要作用。

5.模型分析

在描述分析的基础上，通过方差分析对各个因素同ST状态之间的关系做了逻辑回归模型。参数估计如表3—4所示。从中可以看到只有两个解释变量是显著的。一个是ARA，它的极大似然估计为4.88，是正值。这说明ARA的取值越高（即应收账款比率越高），该企业被特别处理的可能性越大。类似地，LEV的估计量也是显著的（p-值为0.05），其极大似然估计量为2.35，是正值。这说明LEV的取值越高（即债务水平越高），该企业被特别处理的可能性越大。对于其他所有变量，基于现有数据，无法下确定性结论。

表 3—4 各参数估计以及检验结果

因素名称	参数估计	标准误差	卡方统计量	p-值
截距项	-8.869	4.636	3.66	0.055 7
ARA	4.880	1.493	10.69	0.001 1
ASSET	0.247	0.224	1.21	0.271 1
ATO	-0.507	0.658	0.60	0.440 3
ROA	-0.637	6.224	0.01	0.918 5
GROWTH	-0.833	0.567	2.16	0.141 7
LEV	2.354	1.201	3.84	0.050 1
SHARE	-0.011	0.011	0.99	0.318 9

6.预测评估

基于上述模型，可以对每个企业的ST概率予以测算。然后以某阈值为界，将其预测为ST企业（即ST=1）或者非ST企业（即ST=0）。如果以50%为界，我们发现除了一个样本之外，所有样本都会被预测为ST=0。因此总体预测精度优良，错判概率MCR为5.26%；但TPR很差，只有2.8%，对于实际工作没有任何价值。因此，我们重新考虑了加权后的错判概率WMCR，相应地将阈值设为5.26%。因此而产生的总体错判概率为MCR=28.7%，但是，TPR大大提高为TPR=69.4%，同时FPR仍然可以得到一定的控制（FPR=28.5%）。因此，5.26%是一个比较好的阈值，可以推荐。

7.总结讨论

本研究分析上市企业的公开财务报表信息，建立了对企业未来ST状态具有一定预测能力的逻辑回归模型。我们的分析表明企业的盈利质量（以ARA计）是影响企业未来ST可能性的最重要的因素，值得关注。