

# 基于非参数生存分析模型的上市公司经营困境实证分析

陈鸽林 夏洪胜(博士生导师)

(暨南大学管理学院 广州 510632)

**【摘要】** 本文以1998~2011年沪深两市A股上市公司为研究对象,运用非参数生存分析法即PL法对样本公司随企业年龄变化出现的经营困境风险进行了估计和评价。结果表明,上市公司年龄与企业经营业绩之间存在密切关系,随着年龄的增长,企业经营业绩会发生变化,且不同行业之间企业的生存函数存在明显的差别。

**【关键词】** 生存分析 乘积限估计 ST公司

## 一、引言

在生物学、医学及稳定性等相关领域的研究中,年龄都是一个非常重要的解释变量。从生物学的角度来看,随着年龄的增长,生物的物理功能会发生显著的变化。事实上,在企业和产业动态理论研究的文献中已有不少开始关注企业年龄对企业生存的影响。Cressy(2006)、Ericson and Pakes(1995)、Jovanovic(1982)的研究都认为企业失败率随企业年龄增加而发生变化。组织生态学家认为,年龄与企业生存之间存在着几种不同的关系。新企业在创立之初获得了风险投资基金或银行贷款等方面的优势,企业失败的风险并不大。但是,当初始资源优势枯竭时,企业的失败率将会达到峰值,此后失败率开始降低,这是由于不适宜生存环境的企业已经退出了。

还有不少学者研究发现,企业经营超过一定年限后,其退出的概率可能会随着经营时间的推移而增加。有关研究表明,随着企业年龄的增长,一般盈利能力会下降。这是因为随着时间的推移,企业出现老化,即企业组织机构出现僵化。伴随老化过程,企业成本上升、增长放缓、资产渐渐过时、投资和R&D活动下降。同时,随着年龄的增长,企业内部会慢慢产生寻租行为。老企业会出现治理差、董事会规模过大和较高的CEO薪酬等现象。总而言之,企业会面临着一个真正衰老的问题。超出了一定年龄(开始衰老)的企业,其失败率会增加。

本文运用非参数生存分析方法来描述和估计上市公司生存风险随年龄增长而发生变化的动态变化过程,并比较不同行业上市公司经营困境状态的生存率曲线。

## 二、样本选择与数据来源

在使用传统横截面统计模型进行分析时,往往需要对每一个风险样本选取配对的正常样本。但是,配对样本

如何选择,按照什么比例进行选择,都没有统一的标准,因此,在分析过程中,我们很难准确把握配对样本的选取标准,标准设置过高或过低都会对预测的精度产生非常大的影响。同时,选择不同的样本集,常常会得到不同的结果,导致模型的稳健性不牢靠。而生存分析模型具有独特的构造思路,能够避免由于样本配对所产生的一系列问题。加之生存分析模型具有无须选择配对样本的特点,因而可大大减少选取配对样本的工作量,同时也提升了模型的适用性和稳定性。

本文研究按照中国证监会发布的《上市公司行业分类指引》所规定的分类原则与方法,剔除行业门类编码为J的金融类上市公司,这是因为金融类上市公司的经营具有独特的优势,很少因经营业绩严重不佳而受到“特别处理”(ST)。上市公司戴上ST帽子的原因是其财务状况或其他方面出现了异常,这种异常主要是指以下两种情况:一是上市公司经审计连续两个会计年度的净利润均为负值,即亏损严重;二是上市公司最近一个会计年度经审计的每股净资产低于股票面值。根据这一条件,公司上市后第一年第二年基本上不太可能被处以ST,因此,为了更准确地反映上市公司经营困境中的问题,我们选择样本时剔除了2009~2011年两年新上市的公司。

我们以1998~2011年在沪深两市主板交易的1376家A股上市公司作为研究样本。样本公司数据来源于CCER数据库、深圳证券交易所(<http://www.szse.cn/>)和上海证券交易所(<http://www.sse.com.cn>)网站。

## 三、模型选择与评价

本文使用乘积极限法(Product-Limit Method, PL法)估计样本公司经营困境状态的生存曲线。PL法由Kaplan和Meier于1958年提出,该方法又称为Kaplan-Meier法。

PL法是一种比较常用的非参数方法,其对数据的分布没有特殊要求。在计算机的辅助下,该方法可用于小样本、中样本和大样本各种情形的分析。非参数方法还有一个显著特点就是进行生命表分析,适用于按照区间分组或者样本量非常大(例如数以千计)的分析对象。运用PL法,既有PL估计又有生命表估计。PL估计是基于一个个的数据,生命表估计是基于按区间分组的数据。实际上,PL估计可以看成是生命表估计的一种特殊情形,当各个区间只含有一个观测数据时,生命表估计就是PL估计法。

企业失败事件发生的条件概率为:

$$\hat{p}(t_i) = \frac{d_i}{n_i} \quad (1)$$

式中: $d_i$ 为 $t_i$ 期发生经营失败事件的公司数目; $n_i$ 为 $t_i$ 期初仍处于风险状况的公司数目。PL估计下的生存率是指公司在 $t_i$ 期之后仍保持正常经营状态的概率,它是多个时点生存概率的累积,因此其又称为累积生存概率,用 $S(t_i)$ 表示。于是有:

$$\hat{S}(t_i) = (1 - \hat{p}(t_1))(1 - \hat{p}(t_2)) \dots (1 - \hat{p}(t_i)) \quad (2)$$

由此,我们可以估计样本公司失败状态的风险曲线,PL估计法的风险函数估计可以表示如下:

$$\hat{h}(t_i) = \frac{\hat{p}(t_i)}{t_i - t_{i-1}} = \frac{d_i/n_i}{t_i - t_{i-1}} \quad (3)$$

对于公司生存时间,以往文献通常使用两种年龄来测定:一种是以公司成立当年为初始年份来计算年龄;另一种是以公司上市当年为起始年份来计算年龄。由于我们的样本是上市公司,许多在上市之前经过多次合并、公司性质变革等,在这种情况下,使用成立日期作为起始年份来计算年龄并不能完全反映公司生存态势;因此,本文将公司上市当年作为初始年。由于观察期所对应的日历年份是1998~2011年,对于每个样本公司来说,在1998年以前上市的公司,观察的起始期为1998年,即该公司的年龄从1998年算起。1998年以后上市的公司,观察的起始期为1岁,至2011年末仍保持正常经营的公司,我们将其视为截断案例,终止期为公司发生困境(即被ST或\*ST)或截断的年龄。

#### 四、实证分析

1. 全部样本的生存函数与风险函数估计。表1列出了以公司上市当年为初始年份计算方式下各期的生存函数与风险函数,表中期初公司的数目由上期期初的数目加上新进入风险集的公司数目,再减去在上期中发生失败

事件公司的数目及发生截断的公司数目而得到。生存函数由式(2)计算得到,风险函数由式(3)计算得到,生存率标准误差由式(4)计算得到,即:

$$STD(\hat{S}(t_i)) = \hat{S}(t_i) \times \sqrt{\sum \frac{d_i}{(n_i - d_i)}} \quad (4)$$

表1 Kaplan-Meier法风险函数与生存函数估计

公司年龄	期初公司数目	当期发生经营困境的公司数	当期新进入风险集的公司数	当期发生截断的公司数	风险函数h(t <sub>i</sub> )	生存函数S(t <sub>i</sub> )	累积风险函数	生存率标准误差
1	0		831	0		1.000 0		
2	831	1	99	0	0.001 2	0.998 8	0.001 2	0.001 2
3	929	11	136	0	0.011 8	0.997 6	0.013 0	0.003 8
4	1 054	22	77	6	0.020 9	0.985 8	0.033 9	0.005 8
5	1 103	33	70	17	0.029 9	0.965 2	0.063 8	0.007 6
6	1 123	49	65	10	0.043 6	0.936 3	0.107 5	0.009 5
7	1 129	56	62	2	0.049 6	0.895 5	0.157 1	0.010 9
8	1 133	63	3	57	0.055 6	0.851 1	0.212 7	0.012 1
9	1 016	43	10	59	0.042 3	0.803 7	0.255 0	0.012 6
10	924	40	17	61	0.043 3	0.769 7	0.298 3	0.013 2
11	840	45	6	63	0.053 6	0.736 4	0.351 9	0.014 0
12	738	23	0	110	0.031 2	0.696 9	0.383 0	0.014 0
13	605	15	0	61	0.024 8	0.675 2	0.407 8	0.014 3
14	529	15	0	69	0.028 4	0.658 5	0.436 2	0.014 8
15	445	14	0	129	0.031 5	0.639 8	0.467 6	0.015 3
16	302	8	0	108	0.026 5	0.619 7	0.494 1	0.016 0
17	186	5	0	12	0.026 9	0.603 3	0.521 0	0.017 2
18	169	2	0	69	0.011 8	0.587 1	0.532 8	0.017 5
19	98	0	0	75	0.000 0	0.580 1	0.532 8	0.017 5
20	23	0	0	16	0.000 0	0.580 1	0.532 8	0.017 5
21	7	0	0	1	0.000 0	0.580 1	0.532 8	0.017 5
22	6	0	0	6	0.000 0	0.580 1	0.532 8	0.017 5

图1和图2分别描绘了生存曲线和风险曲线,图中的Kaplan-Meier生存率曲线以生存率为纵轴、公司年龄为横轴。该曲线是公司年龄与公司生存率之间关系的曲线。某一公司任意时刻的生存率都可以根据生存率曲线做出估计,反之也可以由任意生存率来估计其生存时间。生存分析中最常使用的概括性统计量是中位生存时间,一般是指生存率为0.5时所对应的生存时间,它表示有且只有50%的个体可以活过这个时间。中位生存时间可通过将生存率为0.5两侧的左端点连线后进行内插而求得。但从我们的样本来看,由于删失样本占66.66%,超过了总样本的一半,同时最大的样本数据已删失,因此中位生存时间是难以估计的,处理这种情况的实用方法是使用生存时间超过某一个给定时间长度的概率。

从图1的生存曲线来看,有77%以上的公司正常经营年限超过了10年。Kaplan-Meier方法界定删失出现是与

生存时间相互独立的,换句话说,出现删失观察的理由是与死亡的原因无关的。当公司在研究时间结束时仍然活着,那么上述假定当然是满足的。从我们的样本数据来看,出现删失的原因都是观察期结束时,公司依然正常经营,因此满足PL法的假定。上述结果也符合实际情况,上市公司相比较于中小型企业,正常经营的时间一般较长,寿命也比中小型企业相对较长。从生存曲线来看,公司生存率在上市前两年相对较高,比较平稳,上市后4~12年生存率下降较快,14年之后,又相对比较平稳。

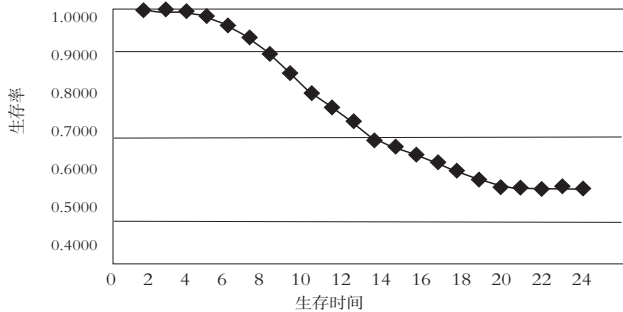


图1 Kaplan-Meier生存函数估计

从风险率来看,样本公司在上市后前8年呈上升趋势,风险率从0升至了0.0556,8-11年间出现了小范围的波动,11年又升到高峰0.0536,此后,风险率快速下滑,13年后开始相对平稳,17年后快速下滑。

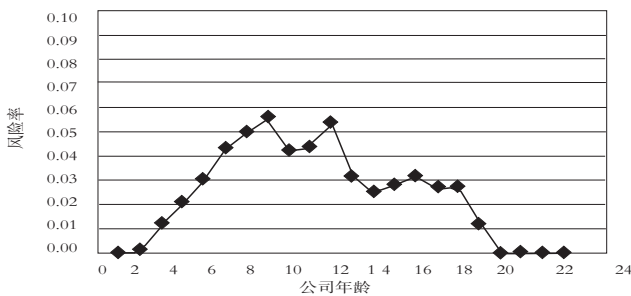


图2 Kaplan-Meier风险函数估计

2.分行业的生存函数估计。表2描述了1998~2011年各行业样本公司数目、生存时间总期数、发生经营困境的期数及经营困境发生率。从表中可以看出,制造业、批发和零售贸易业、信息技术业、综合类、交通和仓储业、电力、煤气及水的生产和供应业、房地产业处于该区间的公司数目较多,而传播与文化产业、建筑业、采掘业处于该区间的公司数目较少。但从经营困境发生率(如图3所示)来看,传播与文化产业和农、林、牧、副、渔业的经营困境发生率相对较高,而电力、煤气及水的生产和供应业、建筑业、房地产业的经营困境发生率较低。这是因为电力、煤气及水的生产和供应业与人们的生活息息相关,是不可或缺的基础产业,一般经营比较稳定;而房地产业和建筑业从中国最近十年房地产行业的繁荣中获得了发展机遇,经营业绩明显改善,企业经营相对比较稳定。

表2 1998~2011分行业公司生存状况统计描述

行业编码	行业名称	公司总数	发生经营困境的公司	生存总期数	发生经营困境的期数	经营困境发生率
A	农、林、牧、副、渔业	38	16	355	107	0.301
B	采掘业	32	5	275	39	0.142
C	制造业	774	263	8 126	1 842	0.227
D	电力、煤气及水的生产和供应业	62	9	724	88	0.122
E	建筑业	30	5	300	34	0.113
F	交通、仓储业	65	12	672	81	0.121
G	信息技术业	84	32	841	211	0.251
H	批发和零售贸易业	100	29	1 123	167	0.149
J	房地产业	56	13	617	68	0.110
K	社会服务业	41	14	417	72	0.173
L	传播与文化产业	11	6	93	36	0.387
M	综合类	83	41	813	235	0.289
总样本		1 376	445	14 356	2 980	0.208

注:①生存时间总期数=行业内公司数×行业内各公司的生存时间期数;②经营困境发生率=发生经营困境的期数/总生存时间期数。

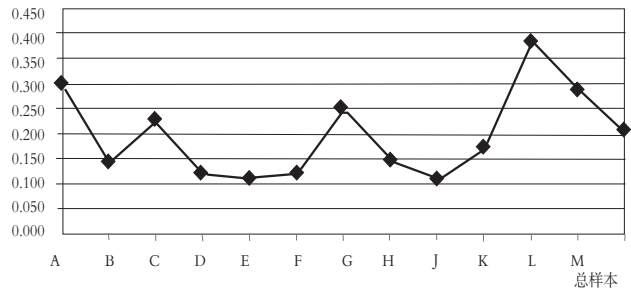


图3 各行业公司经营困境发生率

表2统计结果表明,由于制造业、批发和零售贸易、信息技术业、综合类、交通和仓储业、电力、煤气及水的生产和供应业、房地产业公司数目较多,因而这些行业中发生经营困境公司的数目也相对较多。因此,对这些行业经营困境风险的分析具有一定的典型意义。此外,与其他样本较少的行业相比,上述行业的样本生存时间总期数较多,因此统计结果的可靠性也较强。据此,本文下面仅对制造业、批发和零售贸易、信息技术业、综合类、交通和仓储业、电力、煤气及水的生产和供应业、房地产业进行进一步的分析。在所有样本中,制造业公司数目最多,发生困境的公司数目也最多,因此,我们进行PL分行业比较时,以制造业为基准,将之与相关行业进行比较。

图4描绘出了各行业与制造业生存曲线估计的对比图。从各分图描绘的情况来看,制造业生存概率变化较为平稳,电力、煤气及水的生产和供应业、交通仓储业与制造业相比,生存曲线存在较大差异。从LOG Rank检验来

看,  $P=0.002$ , 通过了显著性检验, 说明制造业的生存风险高于电力、煤气及水的生产和供应业、交通仓储业。从制造业与信息技术业、批发和零售贸易业的比较来看, 它们之间的生存分布差异并不显著。而综合类公司的生存风险大于制造业, 这一结果跟以往的有些研究存在一些区别, 盲目的多元化并没有降低企业陷入经营困境的风险。房地产业的生存概率高于制造业, 这得益于在观察期间我国房地产业的蓬勃发展。

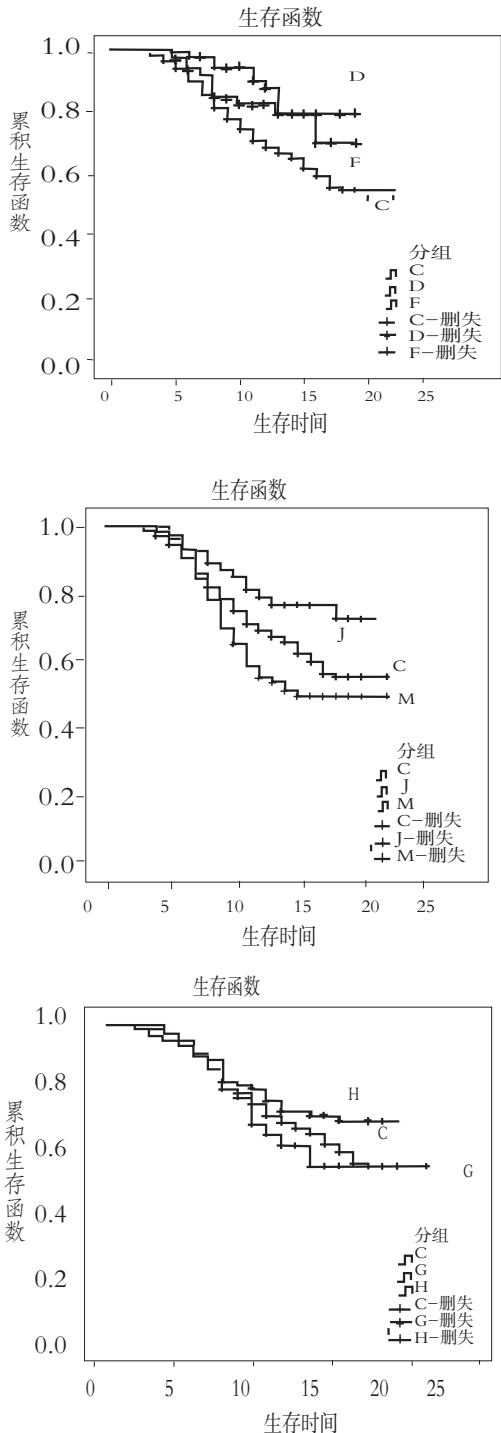


图4 各行业生存曲线估计对比

表3列示了在非参数分析法下, 样本公司分行业各企业的生存曲线的对数秩检验结果。

表3 各行业生存分布的对数秩检验

Log Rank(Mantel-Cox)	卡方	df	Sig.
C、D、F	9.467	1	0.002
C、G、H	3.769	2	0.152
C、J、M	8.010	2	0.018

### 五、结论

根据企业生命周期理论, 为检验企业随着年龄的增长而出现的经营业绩变化, 本文以1998~2011沪深两市(在2009年1月1日之前上市)的1376家非金融类上市公司作为研究样本, 运用生存分析法中比较常用的一种非参数方法(PL法)对中国上市公司随公司年龄增加而出现的的经营困境风险情况进行了估计和评价, 描绘出了样本公司经营困境状态的生存曲线与风险率曲线。

与传统横截面方法相比, 生存分析模型在处理过程中将观察期结束时尚未发生困境的企业作为截断案例进行处理, 它克服了传统方法在失败时间选择上的偏差。从研究结果来看, 企业生存曲线与年龄之间存在较密切的关系。在一般情况下, 在公司上市后的前几年总的状况应当是正常的, 企业的经营在准备上市的惯性推动下不会出现太坏的变化, 但上市的惯性消失后, 企业的风险率逐渐增大, 达到峰值后经过努力扭转风险率会开始下降。

为了进一步了解不同行业企业的生存函数的差别, 本文从样本所涵盖的12个行业门类中选取制造业、批发和零售贸易业、信息技术业、综合类、交通和仓储业、电力、煤气及水的生产和供应业、房地产业做了进一步的分析。结果表明, 不同行业之间企业的生存曲线存在一定的差异, 其中: 综合类公司的生存风险相对较高, 说明过度多元化并没有降低企业经营风险; 制造业生存曲线比较平稳; 房地产业、电力、电力、煤气及水的生产和供应业生存风险相对较低。

### 主要参考文献

1. Arauzo Carod, M. A.. Firm survival: methods and evidence. *Empirica*, 2008; 1
2. 邓晓岚. 非财务视角下中国上市公司经营困境评价模型及实证研究. 华中科技大学博士论文, 2006
3. 彭非, 王伟. 生存分析. 北京: 中国人民大学出版社, 2004
4. Li S., J. Shang et al. Why Do Software Firms Fail? Capabilities, Competitive Actions, and Firm Survival in the Software Industry from 1995 to 2007. *Information Systems Research*, 2010; 3
5. 吴冰. 生存分析及其应用: 以创业研究为例. 上海交通大学学报(哲学社会科学版), 2006; 3