
参赛队号 : (由大赛组委会办公室填写)

2021 年 (第七届) 全国大学生统计建模大赛

参赛学校 : 湖南师范大学

论文题目 : 中国公司于海外上市的退市风险研究
—基于生存分析法

参赛队员 : 冯振声 何婷婷 唐龙

指导老师 : 陈旭

中国公司于海外上市的退市风险研究

—基于生存分析法

目录

一、引言	1
(一) 研究背景	1
(二) 研究意义	1
二、文献综述	2
(一) 退市风险方面的研究	2
(二) 生存分析方面的研究	2
(三) 生存分析方法	3
三、研究设计	4
(一) 指标体系的构建	4
(二) 样本数据的选取	6
(三) 生存时间界定	6
(四) 数据预处理准备工作	6
四、模型的建立与求解	9
(一) 基于 $K-M$ 方法的生存函数的估计	9
(二) 基于 $K-M$ 方法的生存函数点估计的 95%置信区间	13
(三) 基于 COX 回归的比例危险模型	17
五、模型的改进	24
六、研究的优点与不足	25
(一) 研究的优点	25
(二) 研究的不足	26
七、结论与建议	26
(一) 研究的相关结论	26

(二) 研究的相关建议.....	27
八、参考文献.....	28
九、附录.....	29
(一) <i>python</i> 数据统计整理代码	29
(二) <i>python</i> 数据预处理代码	30
(三) <i>R</i> 程序代码	33
(四) 生存函数的 K-M 估计结果汇总表.....	36
(五) 生存函数点估计的置信区间汇总表.....	40
十、致谢.....	44

插图与表格清单

(一) 插图清单

图 1 创业板数据缺失热图.....	7
图 2 主板数据缺失热图	8
图 3 海外主板上市公司生存函数的 $K-M$ 估计.....	10
图 4 海外创业板上市公司生存函数的 $K-M$ 估计	11
图 5 主板组和创业板组生存函数的 $K-M$ 估计.....	12
图 6 主板上市公司生存函数点估计的三种置信区间.....	16
图 7 主板上市公司生存函数点估计的三种置信区间比较	16
图 8 创业板上市公司生存函数点估计的三种置信区间	16
图 9 创业板上市公司生存函数点估计的三种置信区间比较.....	17

(二) 表格清单

表 1 评价指标体系.....	5
-----------------	---

表 2 海外主板上市公司的 $K-M$ 估计及其标准差估计	9
表 3 海外创业板上市公司的 $K-M$ 估计及其标准差估计	10
表 4 $K-M$ 估计生存分析时间的平均值表	13
表 5 主板生存函数的三种 95%置信区间	13
表 6 创业板生存函数的三种 95%置信区间	14
表 7 基本模型回归结果	18
表 8 对可能的调节因素的局部检验	19
表 9 在剔除 X_8 协变量下对可能的调节因素的局部检验	19
表 10 在剔除 X_8 与 X_{11} 下对可能的调节因素的局部检验	20
表 11 在剔除 X_8 、 X_{11} 与 X_2 下对可能的调节因素的局部检验	20
表 12 在剔除 X_8 、 X_{11} 、 X_2 和 X_4 下对可能的调节因素的局部检验	21
表 13 在剔除 X_8 、 X_{11} 、 X_2 、 X_4 和 X_7 下对可能的调节因素的局部检验	21
表 14 在剔除 X_8 、 X_{11} 、 X_2 、 X_4 、 X_7 和 X_6 下对可能的调节因素的局部 检验	22
表 15 在剔除 X_8 、 X_{11} 、 X_2 、 X_4 、 X_7 、 X_6 和 X_1 下对可能的调节因素的 局部检验	22
表 16 各协变量对应的相对风险大小以及其置信区间	23
表 17 时间相依的局部似然比检验	24
表 18 改进的 COX 比例危险模型	25

中国公司于海外上市的退市风险研究

—基于生存分析法

摘要

本文在传统的上市公司退市风险研究中引入生存分析方法，挖掘数据背后的潜在价值，通过定量化手段建立相关退市风险的测度模型，旨在为相关海外上市公司决策营运提供参考，推动我国海外上市公司的发展。本文分别从上市公司营业规模、运营能力与公司资产流动与负债情况三个维度建立评价指标体系。数据整理自 CSMAR 数据库，研究时间范围为 1968 年 6 月 5 日–2021 年 4 月 1 日，并利用 *python* 对研究数据进行统计整理与数据清洗，共得到建模数据 974 例。

首先，我们分别给出了海外主板上市公司与创业板上市公司生存函数的 $K-M$ 估计。基于生存函数的估计结果，可以得出在研究时间范围内，随着上市时间的推移，上市公司面临的退市风险不断增大，主板上市公司在上市后 580 个月左右面临 50% 左右的退市风险。创业板上市公司在上市后 250 个月左右面临 60% 左右的退市风险。通过两样本的假设检验，在研究时间范围内，可以得到主板上市公司的退市风险比创业板上市公司的退市风险小的结论。

其次，我们对以上生存函数的点估计的有效性进行评估，分别给出三种生存函数点估计的置信区间。并且对不同置信区间进行可视化研究分析，得出生存函数的点估计具有很好的估计效果。随后，我们给出了在研究时间范围内，关于海外主板上市公司与创业板上市公司生存时间的平均值估计。

然后，本文探讨了一些协变量因素对海外上市公司退市风险的影响情况，建立了多因素 *COX* 比例危险模型。通过模型可以得出在不考虑其它协变量的作用下，主板上市公司相对于创业板上市公司的相对退市风险为 2.402。在考虑了其它协变量的影响之下，主板上市公司相对于创业板上市公司的退市风险大小

为1.978，相对风险减少了很多。主要原因在于资产总计，管理费用与投资活动产生的现金流量净额是保护因子，流动负债合计是危险因子。进一步本文分别从保护因子与危险因子两个角度给出了相关解释。

最后，我们引入时间相依协变量，对模型进行了改进与客观的评价，并基于我们得出的结论，给出了相关建议。

关键词：退市风险 生存分析 生存函数 COX 比例危险模型

一、 引言

（一） 研究背景

企业的良好发展是国家经济发展的基础，随着全球资本一体化的发展，中国曾几度掀起海外公司上市浪潮^[1]。在一次次浪潮背后，一些企业得到了很好的发展，但同时一些企业也在短时间面临倒闭的风险。中国企业到海外上市，一方面满足了自身的融资需求，同时顺应世界经济一体化发展趋势，使中国企业引进新的管理理念，在制度上更加规范，建立更加现代化的中国企业群。另一方面，中国企业在海外上市也面临着巨大的挑战，而其中之一便是上市后面临着退市的风险。在中国海外上市公司中，有的企业缺乏专业的技术指导，有力的信息支持，上市后往往不稳定，在公司管理经营方面缺乏经验，增大了退市的风险。为此，本文主要对海外主板上市公司与创业板上市公司的退市风险进行定量研究，顺应大数据时代发展需求，充分利用相关统计技术挖掘数据背后的潜在价值。通过定量化手段建立海外公司退市风险的测度模型，旨在为相关海外上市公司决策营运提供参考，推动我国海外上市公司的发展。

（二） 研究意义

本文在传统的上市公司退市风险的研究中引入生存分析方法，既有学术性价值又具有实际应用价值，具体分析如下：

具有学术性价值主要体现在：目前国内关于海外上市公司退市风险的研究主要从定性与定量两个角度展开，而对于定量化研究又多局限于传统的经典统计学方法，主要从静态的角度预测上市公司的退市风险。而本文将生存分析理论应用到海外上市公司退市风险的研究中，提供了一种动态的研究方法，在研究方法上得到创新，提供了新的研究思路。

实际应用的价值主要体现在：本文基于相关学者的研究成果，对建立的生存分析模型不断思考，提出改进。建立的模型可以满足定性研究的需求，同时也很好的展示了海外主板上市公司与企业板上市公司在上市后面临的退市风险大小与风险大小随时间的变化情况。同时，利用 *COX* 比例风险模型进行了多因素风险分析。

二、 文献综述

（一） 退市风险方面的研究

上市公司退市是指其不能够满足上市标准而主动或者被动终止上市的情形。关于影响上市公司退市的原因，目前海内外学者没有给出一致的影响因素评价指标，在相关问题的研究方面主要分为退市风险的定性研究与定量研究^[2]。在定性研究方面，国内学者的研究包括对上市公司退市的制度与政策研究，国外学者主要对上市公司退市的影响因素以及上市公司退市后对自身的影响进行相关研究。上市公司退市风险的定量研究主要方法包括经典的统计学方法和机器学习方法。其中，经典的统计学方法包括 *logistic* 回归模型，*Z-score* 模型，因子分析和多元判别分析等。机器学习方法主要包括支持向量机，决策树与神经网络等。曹强等^[3]以中国 A 股上市公司为研究对象，建立了基于主成分分析的支持向量机的预测模型，并取得了较高的预测精度。虞文美等^[4]基于因子分析建立了弹性反向传播算法的神经网络模型等。

（二） 生存分析方面的研究

生存分析^[5]是将终点事件与该事件发生所经历的时间结合起来研究的一种统计学分析方法。它最初被应用在医学领域，其在处理删失数据方面具有优点，因此得到各领域学者的注意。近年来，生存分析不断得到重视与发展，这门学

科被广泛应用于不同领域，如潘雅琼等^[6]把生存分析技术应用于企业财务危机预警研究中，并取得了较好的实证分析。呈显宏^[7]运用生存分析做贸易便利化研究等等。

生存分析方法具有普适性的特点，并且能很好的处理删失数据问题，提供了一种动态分析事件发生的统计手段。相比于传统的经典统计学而言，其对模型假设与数据要求方面较低，与经典统计方法的静态研究相比，生存分析在分析研究事件发生的动态过程方面也具有明显优势。本文试图将生存分析领域的相关统计技术应用于海外上市公司退市风险的研究上面，同时从多个角度建立上市公司退市风险的评价指标体系，对海外上市公司退市风险随时间的变化情况以及对影响公司退市的协变量因素分析方面展开研究。

（三）生存分析方法

1. 生存分析的数据类型：

右删失数据：一个观测值假定其精确值不知道而只知道其在大于或等于某个时间点之后发生，则称该观测为该时间点处的右删失数据。本文的数据属于广义型右删失数据。

2. 生存分析的基本函数：

生存分析的基本生存函数包括生存函数、累积危险函数与危险率函数，它们相互联系，可以互相推导。

①生存函数

描述生存时间统计特征的基本函数就是生存函数 $S(x)$ ，它反映了观察个体生存至时间 x 的概率，定义为：

$$S(x) = p(X > x) \quad (1)$$

其中 $p(X > x)$ 表示事件发生的概率。

②危险函数

危险函数 $H(x)$ 通常由危险率函数刻画，危险率函数 $h(x)$ 度量了事件在单位时间内的死亡风险，定义为：

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{p(x \leq X < x + \Delta x | X > x)}{\Delta x} \quad (2)$$

式中， X 表示事件发生的时间。

对于连续性随机变量而言，生存函数 $S(x)$ 、危险函数 $H(x)$ 与危险率函数 $h(x)$ 的关系是：

$$S(x) = \exp[-H(x)] = \exp[-\int_0^x h(u) du] \quad (3)$$

在描述某一事件发生的概率随时间的变化时，危险率函数尤为有效。

3. 半参数 COX 比例风险模型：

在假设风险因素不随时间变化的情况下，数据样本量为 n ，依然用 X 表示事件发生的时间。 $Z_j = (z_{j1}, z_{j2}, \dots, z_{jp})'$ 表示协变量向量或者影响因素向量。令 $h(t|Z)$ 表示具有协变量 Z 的个体在时间 t 的危险率。Cox(1972) 的基本模型为：

$$h(t|Z) = h_0(t) \exp(\beta'Z) \quad (4)$$

其中， $h_0(t)$ 表示基准危险率， $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ 是一个参数向量。

假如有两个总体，其协变量值分别记为 Z 与 Z^* ，则两总体危险率之比为：

$$\frac{h(t|Z)}{h(t|Z^*)} \quad (5)$$

其中，可采用偏似然函数估计 β 。

三、 研究设计

(一) 指标体系的构建

本文参考相关文献的指标选取，与我们研究的问题相结合。我们主要从上

上市公司营业规模、运营管理能力与公司资产流动与负债情况三个维度建立评价指标体系。

①上市公司营业规模

在上市公司退市中，财务指标存在统一的标准。我们一般认为，经营规模大的公司更容易达到上市公司标准的营业收入，而经营规模较小的公司更难达到相应的标准。

从这个角度进行考虑，我们分别选取了流动资产合计、公司固定资产净额、公司资产总计与营业总收入共4个指标进行研究。

②营运能力

一个公司的营运能力反映了一个公司获取利润的能力，可以从整体上衡量公司整体的运作效率与内部管理情况。我们一般认为，一个公司的营运能力越强，其更不容易面临退市的风险。

从这个角度考虑，我们分别选取了公司管理费用、公司财务费用、公司利润总额与公司发行股票基本每股收益共4个指标进行研究。

③公司资产流动与负债情况

资不抵债是公司面临退市的常见情况。因此，一个公司的资产与负债情况在很大程度上影响公司面临退市风险的大小。因此，我们选取了公司投资活动产生的现金流量净额、流动负债合计与经营活动产生的现金流量净额3个指标体系进行研究，最终构建如下的指标体系。

表 1 评价指标体系

维度	变量名称	变量表示
上市公司营业规模	流动资产合计	X_1
	固定资产净额	X_2
	资产总计	X_3
	营业总收入	X_4
营运管理能力	管理费用	X_5

	财务费用	X_6
	利润总额	X_7
	基本每股收益	X_8
公司资产流与负债情况	投资活动产生的现金流量净额	X_9
	流动负债合计	X_{10}
	经营活动产生的现金流量净额	X_{11}

（二）样本数据的选取

本文主要对海外主板上市公司与创业板上市公司退市风险大小情况与风险影响因素进行研究。基于 *CSMAR* 数据库，从中筛选统计了主板上市公司与创业板上市公司的相关指标数据作为此次研究的数据。所有数据在筛选过程中采取控制研究因素以外的其它因素相同，数据指标筛选具有合理性，数据质量具有权威性。

（三）生存时间界定

本文将数据观测时间设定为1968年6月5日–2021年4月1日。生存时间的起点为海外主板上市公司与非主板上市公司上市的最早时间，即1968年6月5日，时间的终点为2021年4月1日，即到该研究数据提取之日截至。若公司在2021年4月1日之前退市，则数据为完全数据；若公司在2021年4月1日之后退市，则数据为右删失数据（广义 *I* 型右删失数据）。我们通过列克西斯调整思想，得到基本右删失数据。

（四）数据预处理准备工作

1. 生存时间转换

将上市公司上市时间与退市时间以月为单位进行整理。引入 *sector* 标识变

量，主板上市公司标识为1，创业板上市公司标识为0。

2. 根据选择的指标删除原始数据中的多余数据，并利用 *python* 完成相关数据统计与预处理工作。具体步骤如下：

①数据整理合并

将各表的数据通过公司 *ID* 进行级联合并，并用均值代表各个公司对应的指标值。整理后共得到创业板上市公司数据 226 例，主板上市公司数据 801 例。

②数据缺失值处理

通过绘制数据缺失值 heatmap，如下图所示，横坐标表示相关指标标识名称，纵坐标表示观察值所在的行数。其中黄色代表缺失值，蓝色代表非缺失值。

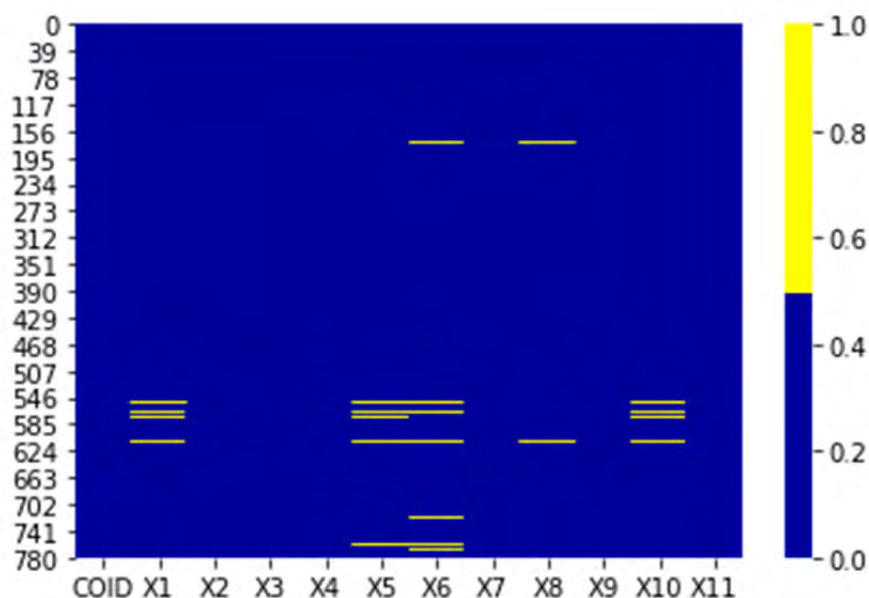


图 1 创业板数据缺失热图

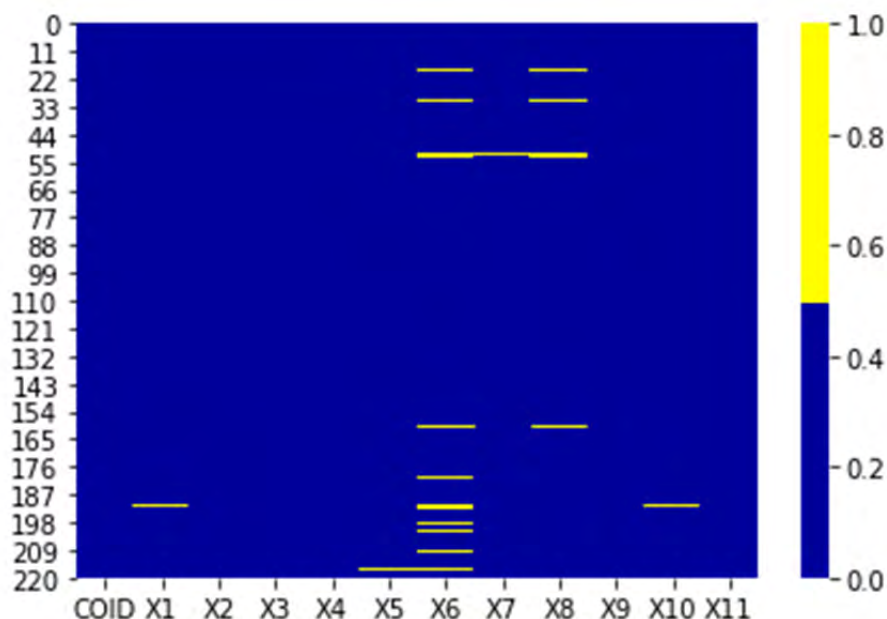


图 2 主板数据缺失热图

根据数据缺失热图可知，在数值型变量中 X_6 与 X_8 缺失值较多，其它数值型变量缺失值很零散。由于数据量充足，因此我们采取的缺失值处理措施为：有缺失就按行删除数据。删除含有缺失值的数据后，最后得到创业板上市公司的数据为 214 例，主板上市公司的数据为 760 例。

③数据异常值处理

统一数据量纲，将风险影响因素变量单位统一为元，对数据进行基本的描述性统计，发现部分数据存在异常值，由于该模型允许数据存在偏斜，因此我们主要处理数据中的较大异常值。即使用数据的 99 分位数替代超过 99 分位数的值，使用数据的 1 分位数替代低于 1 分位数的值。

④数据标准化处理

由于变量之间数值大小相差悬殊，因此对相关指标信息进行 $Z-score$ 标准化处理。

四、模型的建立与求解

(一) 基于 $K-M$ 方法的生存函数的估计

利用统计数据，利用 R 软件对海外上市公司的生存曲线进行估计。首先分别对主板上市公司与创业板上市公司进行研究，探讨这两种类型的上市公司的退市风险情况。之后，我们对比研究了主板上市公司与创业板上市公司的退市风险大小随时间的变化，并对海外主板上市公司与创业板上市公司的退市风险进行了两样本的检验，得出主板上市公司的退市风险比创业板上市公司的退市风险小。进一步，构造生存函数点估计的三种置信区间，给出了生存函数点估计的有效性分析。

1. 海外主板上市公司的生存函数分析

通过收集的数据，基于 $K-M$ 方法，得到海外主板上市公司从上市到退市的生存函数估计，其中包括不同退市时间下 $K-M$ 估计的标准差，部分估计结果见下表。

表 2 海外主板上市公司的 $K-M$ 估计及其标准差估计

研究时间	风险暴露数目	事件数目	乘积限估计	乘积限估计的标准差
1	756	1	0.999	0.00132
3	755	1	0.997	0.00187
16	710	1	0.996	0.00233
20	682	2	0.993	0.00311
30	612	1	0.991	0.00350
32	601	1	0.990	0.00387
33	587	3	0.985	0.00482
34	581	1	0.983	0.00511
43	553	1	0.981	0.00540
45	549	2	0.978	0.00594
46	544	2	0.974	0.00644
47	540	2	0.970	0.00690
49	532	1	0.969	0.00713
50	530	1	0.967	0.00734
53	515	1	0.965	0.00756

进一步，利用 R 软件绘制出主板上市公司的生存曲线，其中横坐标表示公司生存时间，纵坐标表示上市公司随时间变化的生存概率。在本问题的研究背景之下，纵坐标的数值大小表示随着时间的推移，企业不退市的可能性大小。具体结果如下图。

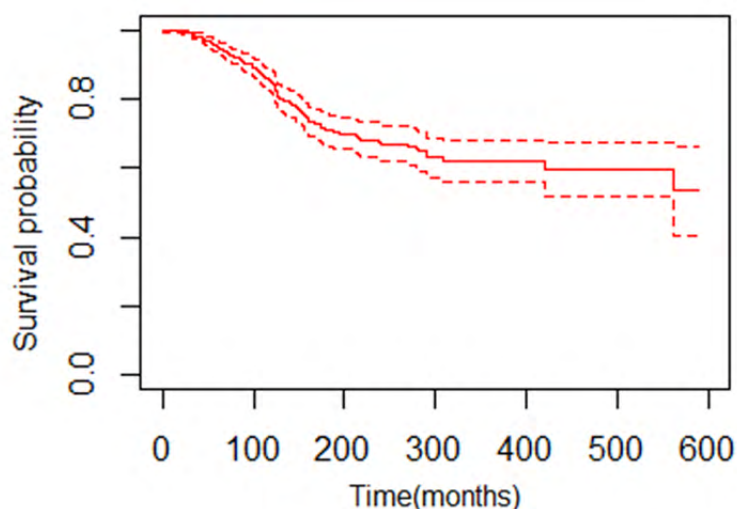


图 3 海外主板上市公司生存函数的 $K-M$ 估计

通过 $K-M$ 生存函数的估计值，我们可以看到随着公司上市时间的推移，上市公司面临的退市风险不断增大，主板上市公司在上市后 580 个月左右面临 50% 左右的退市风险。

2. 海外创业板上市公司的生存函数分析

类似的，得到海外创业板上市公司从上市到退市的生存函数估计，其中包括不同退市时间下 $K-M$ 估计的标准差，具体估计结果见下表。

表 3 海外创业板上市公司的 $K-M$ 估计及其标准差估计

研究 时间	风险暴露 数目	事件 数目	乘积限估计	乘积限估计的 标准差
16	212	1	0.995	0.00471
20	209	1	0.991	0.00667
21	208	1	0.986	0.00816
22	207	2	0.976	0.01050
23	205	1	0.971	0.01148
25	204	2	0.962	0.01320
26	201	1	0.957	0.01397

27	200	1	0.952	0.01470
29	197	1	0.948	0.01540
30	196	2	0.938	0.01669
34	189	3	0.923	0.01851
35	185	1	0.918	0.01907
36	183	1	0.913	0.01961
38	182	3	0.898	0.02113
41	176	2	0.888	0.02208
42	173	1	0.883	0.02254
44	172	2	0.872	0.02342
47	165	1	0.867	0.02387
48	164	1	0.862	0.02430
49	163	1	0.856	0.02472
50	162	2	0.846	0.02552
51	158	2	0.835	0.02630
52	154	1	0.830	0.02668
53	153	1	0.824	0.02705

进一步，利用 R 软件绘制出创业板上市公司的生存曲线，具体如下：

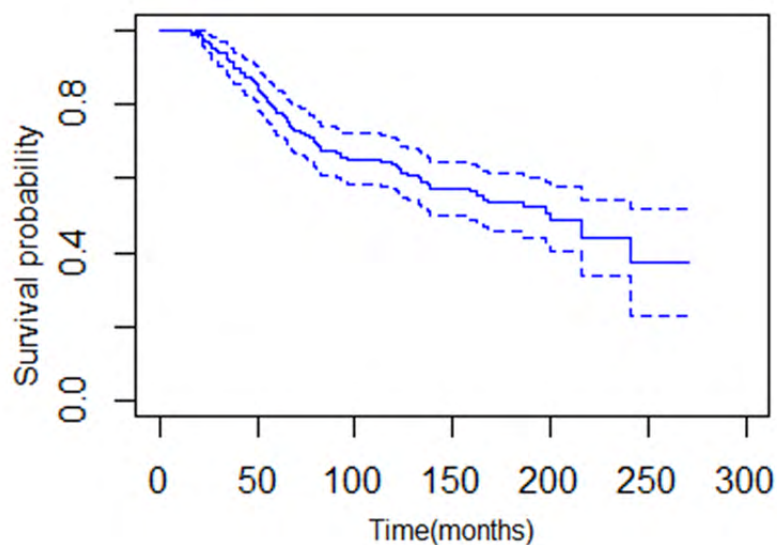


图 4 海外创业板上市公司生存函数的 $K-M$ 估计

通过 $K-M$ 生存函数的估计值，我们可以看到随着主板上市公司上市时间的推移，上市公司面临的退市风险不断增大，创业板上市公司在上市 250 个月左右面临着 60% 左右的退市风险。

3. 海外主板上市公司与创业板上市公司的生存函数的比较分析

为对比分析海外主板上市公司与创业板上市公司的退市风险的相对大小情况。我们将二者退市风险随时间变化的情况进行可视化分析，具体如下图。发现海外主板上市公司的退市风险小于海外创业板上市公司的退市风险。进一步，进行两个样本的假设检验问题。检验的原假设是：在研究时间范围内，主板上市公司的生存函数与创业板上市公司的生存函数是相同的。备择假设为：主板上市公司的生存函数与创业板上市公司的生存函数是不相同的。利用 SPSS 进行两样本生存函数的对数秩检验，得到检验统计量 $\chi^2(1) = 42.675$ ，对应的 P 值为 0.000，即在 5% 的显著性水平下，拒绝原假设，可以认为主板上市公司的生存函数与创业板上市公司的生存函数是不相同的。结合二者生存函数的可视化分析，可以得出在研究时间范围内，主板上市公司的退市风险比创业板上市公司的退市风险小的结论。

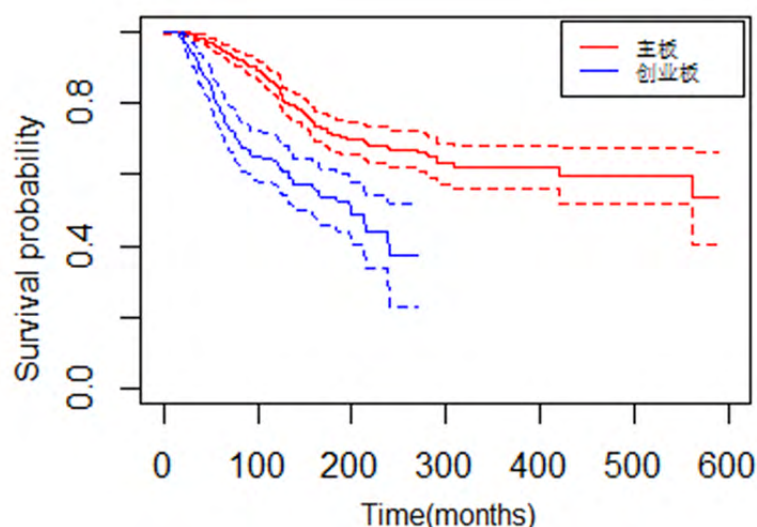


图 5 主board组和创业板组生存函数的 $K-M$ 估计

分析原因在于，主板上市公司在规模上比创业板上市公司要大，其在相关上市标准上也更加严格，因此其上市后面临退市的风险相对较小；而创业板上市公司相比于主板上市公司而言，其门槛较低，比较灵活，但这也增加了上市后的不确定因素，因此面临着较大的退市风险。

进一步，利用生存函数的 $K-M$ 估计，可以得到关于海外主板上市公司与创业板上市公司的退市时间的平均估计，具体结果见下表。在研究范围内，主板上市公司退市时间的平均时间估计为 414.589 个月，95% 的置信区间为 [388.822, 440.355]。创业板上市公司退市时间的平均时间估计为 171.685 个月，95% 的置信区间为 [156.651, 186.719]。

表 4 $K-M$ 估计生存分析时间的平均值表

sector	估算	标准误差	平均值	
			95% 置信区间 下限	95% 置信区间 上限
0	171.685	7.670	156.651	186.719
1	414.589	13.146	388.822	440.355
总体	384.274	11.987	360.779	407.770

(二) 基于 $K-M$ 方法的生存函数点估计的 95%置信区间

对以上生存函数点估计的有效性进行分析，分海外主板上市公司与创业板上市公司给出生存函数点估计的 95% 置信区间。并对不同置信区间进行可视化分析，最后给出模型的置信度与稳健性分析，具体过程与结果如下。

1. 生存函数点估计的三种置信区间

这里分别给出生存函数点估计的线性置信区间、对数变换置信区间与反正弦平方根变换置信区间，部分结果如下表。

表 5 主板生存函数的三种 95%置信区间

研究时间	乘积限估计	乘积限估计的标准差	线性置信区间		对数变换置信区间		反正弦平方根置信区间	
			95%置信下限	95%置信上限	95%置信下限	95%置信上限	95%置信下限	95%置信上限
1	0.999	0.00132	0.996	1.002	0.987	1.000	0.995	1.000

3	0.997	0.00187	0.993	1.001	0.990	0.999	0.992	1.000
16	0.996	0.00233	0.991	1.001	0.988	0.999	0.990	0.999
20	0.993	0.00311	0.987	0.999	0.983	0.997	0.986	0.998
30	0.991	0.0035	0.984	0.998	0.981	0.996	0.983	0.996
32	0.99	0.00387	0.982	0.998	0.979	0.995	0.981	0.996
33	0.985	0.00482	0.976	0.994	0.972	0.992	0.974	0.993
34	0.983	0.00511	0.973	0.993	0.970	0.990	0.972	0.991
43	0.981	0.0054	0.971	0.991	0.967	0.989	0.969	0.990
45	0.978	0.00594	0.967	0.989	0.963	0.987	0.965	0.988
46	0.974	0.00644	0.962	0.986	0.958	0.984	0.960	0.985
47	0.97	0.0069	0.957	0.983	0.954	0.981	0.956	0.982
49	0.969	0.00713	0.955	0.983	0.952	0.980	0.954	0.981
50	0.967	0.00734	0.953	0.981	0.950	0.978	0.952	0.979
53	0.965	0.00756	0.951	0.979	0.947	0.977	0.949	0.978
54	0.963	0.00778	0.948	0.978	0.945	0.975	0.947	0.976
55	0.961	0.008	0.946	0.976	0.943	0.974	0.945	0.975
56	0.957	0.00841	0.941	0.973	0.938	0.970	0.940	0.971
58	0.955	0.0086	0.939	0.971	0.936	0.969	0.938	0.970
59	0.953	0.0088	0.937	0.969	0.933	0.967	0.935	0.968
60	0.952	0.00899	0.935	0.969	0.932	0.966	0.934	0.967
63	0.948	0.00936	0.931	0.965	0.927	0.963	0.929	0.964
64	0.946	0.00955	0.928	0.964	0.925	0.961	0.927	0.962
66	0.944	0.00974	0.926	0.962	0.923	0.959	0.925	0.961
67	0.942	0.00992	0.924	0.960	0.921	0.958	0.922	0.959
68	0.94	0.01011	0.921	0.959	0.918	0.956	0.920	0.957
70	0.938	0.01029	0.919	0.957	0.916	0.954	0.918	0.955

表 6 创业板生存函数的三种 95%置信区间

研究时间	乘积限估计	乘积限估计的标准差	线性置信区间		对数变换置信区间		反正弦平方根置信区间	
			95%置信下限	95%置信上限	95%置信下限	95%置信上限	95%置信下限	95%置信上限
16	0.995	0.00471	0.986	1.004	0.969	0.999	0.982	1.000
20	0.991	0.00667	0.978	1.004	0.962	0.998	0.974	0.999
21	0.986	0.00816	0.970	1.002	0.957	0.995	0.966	0.997

22	0.976	0.0105	0.956	0.996	0.945	0.990	0.952	0.992
23	0.971	0.01148	0.949	0.993	0.939	0.986	0.945	0.989
25	0.962	0.0132	0.937	0.987	0.927	0.980	0.933	0.983
26	0.957	0.01397	0.931	0.983	0.921	0.977	0.927	0.979
27	0.952	0.0147	0.925	0.979	0.915	0.973	0.921	0.976
29	0.948	0.0154	0.919	0.977	0.910	0.970	0.916	0.973
30	0.938	0.01669	0.907	0.969	0.899	0.962	0.904	0.965
34	0.923	0.01851	0.890	0.956	0.882	0.950	0.886	0.953
35	0.918	0.01907	0.884	0.952	0.876	0.946	0.881	0.949
36	0.913	0.01961	0.878	0.948	0.870	0.942	0.875	0.945
38	0.898	0.02113	0.861	0.935	0.854	0.929	0.858	0.932
41	0.888	0.02208	0.850	0.926	0.843	0.921	0.847	0.923
42	0.883	0.02254	0.844	0.922	0.837	0.916	0.841	0.919
44	0.872	0.02342	0.832	0.912	0.826	0.907	0.830	0.909
47	0.867	0.02387	0.826	0.908	0.820	0.902	0.824	0.905
48	0.862	0.0243	0.821	0.903	0.815	0.898	0.819	0.900
49	0.856	0.02472	0.815	0.897	0.809	0.892	0.812	0.895
50	0.846	0.02552	0.804	0.888	0.798	0.883	0.802	0.886
51	0.835	0.0263	0.792	0.878	0.787	0.873	0.790	0.876
52	0.83	0.02668	0.787	0.873	0.781	0.869	0.785	0.871
53	0.824	0.02705	0.780	0.868	0.775	0.863	0.779	0.865

进一步，给出海外主板上市公司三种置信区间的可视化展示，具体如下。

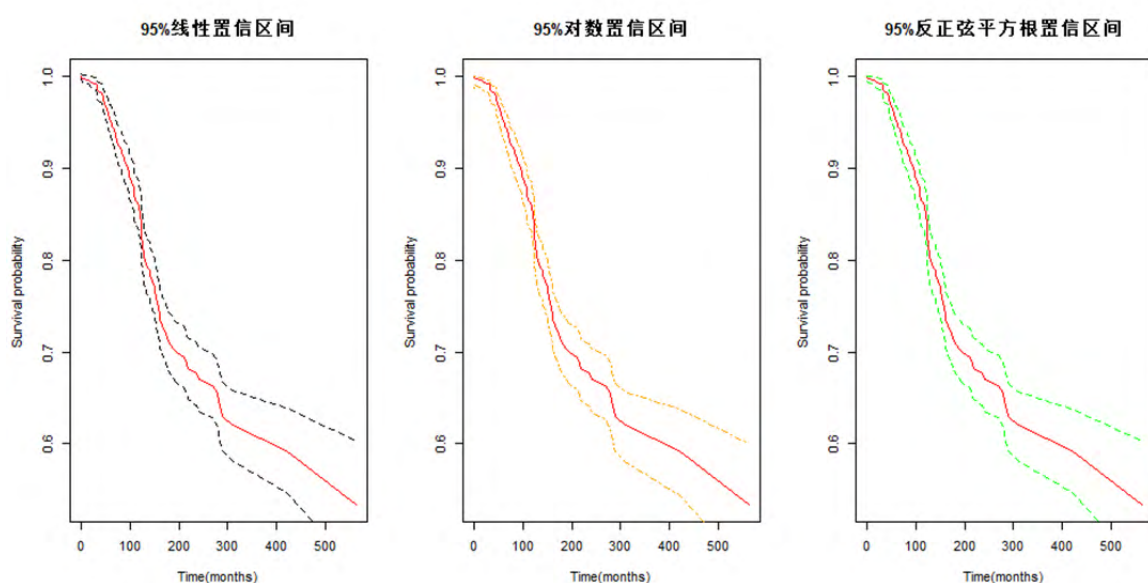


图 6 主板上市公司生存函数点估计的三种置信区间

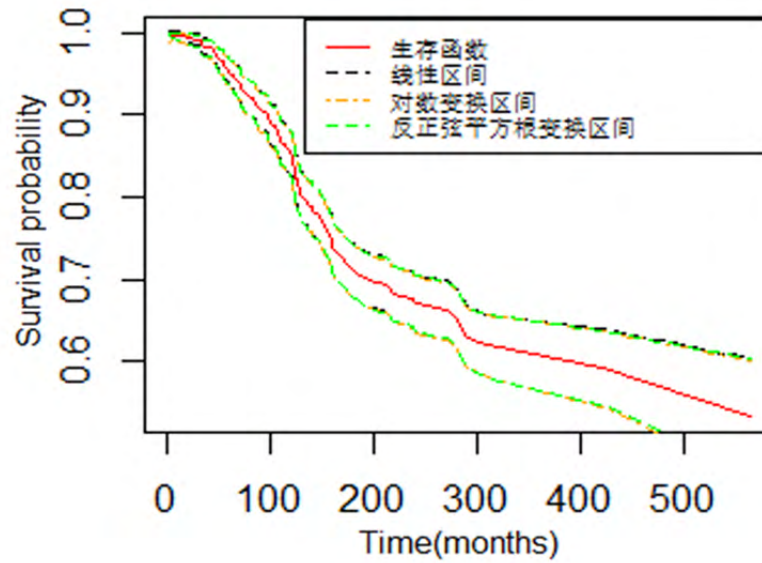


图 7 主板上市公司生存函数点估计的三种置信区间比较

由上图可知，主板上市公司生存函数点估计的三种置信区间十分接近，生存函数前期的点估计的置信区间较窄，后期尾部由于数据的右删失，生存函数的点估计的置信区间较宽。生存函数点估计的置信度在不同的评价指标下具有一致性，从整体上看模型估计结果具有较好的稳健性。

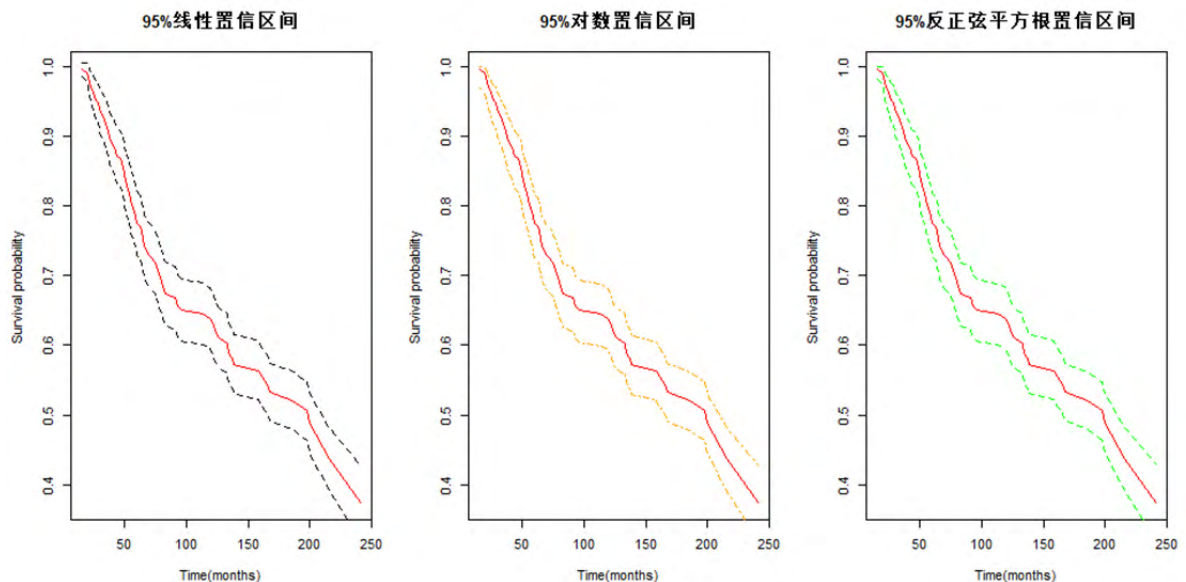


图 8 创业板上市公司生存函数点估计的三种置信区间

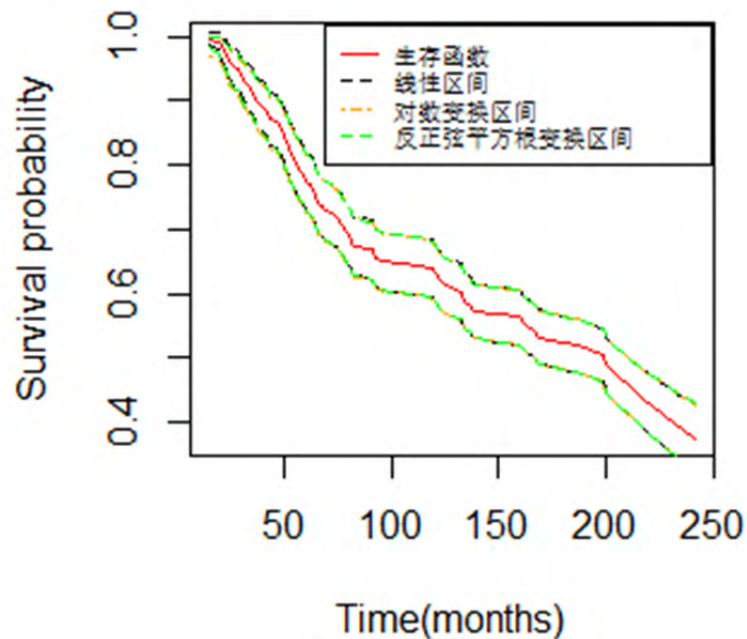


图 9 创业板上市公司生存函数点估计的三种置信区间比较

从创业板生存函数点估计的置信区间来看，三种置信区间十分接近，并且估计的尾部受删失数据影响较小，从整体上模型估计结果效果较好。

（三）基于 COX 回归的比例危险模型

为进一步研究海外上市公司的退市风险与其它协变量之间的关系，本文建立海外上市公司的退市时间与相关评价指标之间的多因素 COX 比例危险模型，并给出相对风险分析，具体结果如下。

1. 风险集划分

基于上文对海外主板上市公司与创业板上市公司生存函数的对比分析，可以确定上市公司类型是海外上市公司退市风险的影响因素之一，故本问题研究的主要因素为海外上市公司类型，这里只研究主板上市与创业板上市两种上市类型。引入的标识变量如下：

$$sector = \begin{cases} 1, \text{主板上市公司} \\ 0, \text{创业板上市公司} \end{cases}$$

潜在调节因素为上文选取的相关评价指标，本文根据上市公司退市风险研

究指标选取的维度，将相关风险因素分为三大风险集。具体如下：

$$\text{第一风险集} \begin{cases} X_1, \text{流动资产合计} \\ X_2, \text{固定资产净额} \\ X_3, \text{资产总计} \\ X_4, \text{营业总收入} \end{cases}$$

$$\text{第二风险集} \begin{cases} X_5, \text{管理费用} \\ X_6, \text{财务费用} \\ X_7, \text{利润总额} \\ X_8, \text{基本每股收益} \end{cases}$$

$$\text{第三风险集} \begin{cases} X_9, \text{投资活动产生的现金流量净额} \\ X_{10}, \text{流动负债合计} \\ X_{11}, \text{经营活动产生的现金流量净额} \end{cases}$$

2. COX 比例模型的建立

Step1: 对基本模型整体检验

以公司退市时间为因变量，公司上市类型为协变量建立比例危险模型，相关回归结果如下：

表 7 基本模型回归结果

因素	方程中的变量					$Exp(B)$
	B	SE	$wald$	df	P 值	
sector	0.876	0.139	39.980	1	0.000	2.402

整体 $wald$ 卡方统计量为 39.980， P 值为 0.000，说明在 5% 的显著性水平下，回归系数显著不为 0，方程具有统计学意义。

Step2: 基于有条件的向后似然比方法或者是向后 $wald$ 方法对风险集变量进行筛选建模，具体建模步骤如下：

表 8 对可能的调节因素的局部检验

步骤 1					
因素	<i>B</i>	<i>SE</i>	<i>wald</i>	<i>df</i>	<i>P</i> 值
<i>sector</i>	0.678	0.143	22.339	1	0.000
流动资产合计 (X_1)	-1.076	0.839	1.643	1	0.200
固定资产净额 (X_2)	0.105	0.213	0.244	1	0.621
资产总计 (X_3)	-4.236	1.642	6.657	1	0.010
营业总收入 (X_4)	0.134	0.267	0.251	1	0.616
管理费用 (X_5)	-1.038	0.472	4.845	1	0.028
财务费用 (X_6)	0.190	0.272	0.488	1	0.485
利润总额 (X_7)	0.249	0.370	0.455	1	0.500
基本每股收益 (X_8)	-0.006	0.058	0.009	1	0.923
投资活动产生的现金流量净额 (X_9)	-1.683	0.624	7.269	1	0.007
流动负债合计 (X_{10})	2.288	0.886	6.659	1	0.010
经营活动产生的现金流量净额 (X_{11})	-0.097	0.393	0.061	1	0.806

由 *wald* 检验结果可知, 风险因素基本每股收益 (X_8) 的 *P* 值为 0.923, 在各协变量局部检验最不显著, 使用向后条件似然比方法, 将该因素剔除。然后在剔除该协变量的条件下, 再做协变量的局部检验, 结果如下:

表 9 在剔除 X_8 协变量下对可能的调节因素的局部检验

步骤 2					
因素	<i>B</i>	<i>SE</i>	<i>wald</i>	<i>df</i>	<i>P</i> 值
<i>sector</i>	0.678	0.143	22.329	1	0.000
流动资产合计 (X_1)	-1.082	0.836	1.675	1	0.196
固定资产净额 (X_2)	0.105	0.213	0.246	1	0.620
资产总计 (X_3)	-4.235	1.639	6.674	1	0.010
营业总收入 (X_4)	0.135	0.267	0.254	1	0.614
管理费用 (X_5)	-1.040	0.472	4.861	1	0.027
财务费用 (X_6)	0.191	0.272	0.490	1	0.484
利润总额 (X_7)	0.248	0.369	0.449	1	0.503
投资活动产生的现金流量净额 (X_9)	-1.682	0.624	7.258	1	0.007

流动负债合计 (X_{10})	2.295	0.884	6.742	1	0.009
经营活动产生的现金流量净额 (X_{11})	-0.098	0.393	0.063	1	0.802

由 *wald* 检验结果可知，经营活动产生的现金流量净额 (X_{11}) 的 *P* 值为 0.802，在各协变量局部检验最不显著，使用向后条件似然比方法，将该因素剔除。在剔除该协变量的条件下，再做协变量的局部检验，结果如下：

表 10 在剔除 X_8 与 X_{11} 下对可能的调节因素的局部检验

步骤 3					
因素	<i>B</i>	<i>SE</i>	<i>wald</i>	<i>df</i>	<i>P</i> 值
<i>sector</i>	0.676	0.143	22.284	1	0.000
流动资产合计 (X_1)	-1.003	0.770	1.696	1	0.193
固定资产净额 (X_2)	0.083	0.193	0.183	1	0.669
资产总计 (X_3)	-4.221	1.660	6.467	1	0.011
营业总收入 (X_4)	0.118	0.260	0.207	1	0.649
管理费用 (X_5)	-1.062	0.464	5.241	1	0.022
财务费用 (X_6)	0.215	0.253	0.723	1	0.395
利润总额 (X_7)	0.210	0.339	0.384	1	0.536
投资活动产生的现金流量净额 (X_9)	-1.641	0.613	7.168	1	0.007
流动负债合计 (X_{10})	2.233	0.846	6.971	1	0.008

由 *wald* 检验结果可知，固定资产净额 (X_2) 的 *P* 值为 0.669，在各协变量局部检验最不显著，使用向后条件似然比方法，将该因素剔除。在剔除该协变量的条件下，再做协变量的局部检验，结果如下：

表 11 在剔除 X_8 、 X_{11} 与 X_2 下对可能的调节因素的局部检验

步骤 4					
因素	<i>B</i>	<i>SE</i>	<i>wald</i>	<i>df</i>	<i>P</i> 值
<i>sector</i>	0.673	0.143	22.171	1	0.000
流动资产合计 (X_1)	-1.055	0.718	2.163	1	0.141
资产总计 (X_3)	-4.064	1.456	7.793	1	0.005

营业总收入(X_4)	0.125	0.259	0.233	1	0.629
管理费用(X_5)	-1.045	0.454	5.290	1	0.021
财务费用(X_6)	0.244	0.231	1.112	1	0.292
利润总额(X_7)	0.196	0.328	0.358	1	0.550
投资活动产生的现金流量净额(X_9)	-1.693	0.583	8.445	1	0.004
流动负债合计(X_{10})	2.213	0.839	6.957	1	0.008

由 wald 检验结果可知,营业总收入(X_4)的 P 值为 0.629 ,在各协变量局部检验最不显著,使用向后条件似然比方法,将该因素剔除。在剔除该协变量的条件下,再做协变量的局部检验,结果如下:

表 12 在剔除 X_8 、 X_{11} 、 X_2 和 X_4 下对可能的调节因素的局部检验

步骤 5					
因素	B	SE	wald	df	P 值
sector	0.669	0.143	21.987	1	0.000
流动资产合计(X_1)	-0.960	0.693	1.918	1	0.166
资产总计(X_3)	-4.155	1.476	7.924	1	0.005
管理费用(X_5)	-0.946	0.400	5.585	1	0.018
财务费用(X_6)	0.228	0.232	0.965	1	0.326
利润总额(X_7)	0.217	0.331	0.430	1	0.512
投资活动产生的现金流量净额(X_9)	-1.682	0.587	8.220	1	0.004
流动负债合计(X_{10})	2.228	0.840	7.028	1	0.008

由 wald 检验结果可知,利润总额(X_7)的 P 值为 0.512 ,在各协变量局部检验最不显著,使用向后条件似然比方法,将该因素剔除。在剔除该协变量的条件下,再做协变量的局部检验,结果如下:

表 13 在剔除 X_8 、 X_{11} 、 X_2 、 X_4 和 X_7 下对可能的调节因素的局部检验

步骤 6					
因素	B	SE	wald	df	P 值
sector	0.676	0.142	22.563	1	0.000
流动资产合计(X_1)	-0.799	0.598	1.785	1	0.182
资产总计(X_3)	-3.763	1.229	9.374	1	0.002

管理费用(X_5)	-0.909	0.395	5.305	1	0.021
财务费用(X_6)	0.207	0.218	0.895	1	0.344
投资活动产生的现金流量净额(X_9)	-1.680	0.584	8.287	1	0.004
流动负债合计(X_{10})	2.033	0.775	6.880	1	0.009

由 *wald* 检验结果可知, 财务费用(X_6)的 *P* 值为 0.344 , 在各协变量局部检验最不显著, 使用向后条件似然比方法, 将该因素剔除。在剔除该协变量的条件下, 再做协变量的局部检验, 结果如下:

表 14 在剔除 X_8 、 X_{11} 、 X_2 、 X_4 、 X_7 和 X_6 下对可能的调节因素的局部检验

步骤 7					
因素	<i>B</i>	<i>SE</i>	<i>wald</i>	<i>df</i>	<i>P</i> 值
<i>sector</i>	0.667	0.142	22.159	1	0.000
流动资产合计(X_1)	-0.807	0.508	2.524	1	0.112
资产总计(X_3)	-3.594	1.073	11.218	1	0.001
管理费用(X_5)	-0.756	0.321	5.536	1	0.019
投资活动产生的现金流量净额(X_9)	-1.717	0.561	9.371	1	0.002
流动负债合计(X_{10})	1.993	0.727	7.522	1	0.006

由 *wald* 检验结果可知,流动资产合计(X_1)的 *P* 值为 0.112 , 在各协变量局部检验最不显著, 使用向后条件似然比方法, 将该因素剔除。在剔除该协变量的条件下, 再做协变量的局部检验, 结果如下:

表 15 在剔除 X_8 、 X_{11} 、 X_2 、 X_4 、 X_7 、 X_6 和 X_1 下对可能的调节因素的局部检验

步骤 8					
因素	<i>B</i>	<i>SE</i>	<i>wald</i>	<i>df</i>	<i>P</i> 值

<i>sector</i>	0.682	0.141	23.282	1	0.000
资产总计 (X_3)	-4.036	1.178	11.735	1	0.001
管理费用 (X_5)	-0.609	0.334	3.321	1	0.068
投资活动产生的现金流量净额 (X_9)	-1.946	0.479	16.540	1	0.000
流动负债合计 (X_{10})	1.282	0.507	6.404	1	0.011

由 *wald* 检验结果可知，在10% 的显著性水平之下，剩下的协变量均通过显著性检验，具体结果如下表，此时停止筛选过程，并用余下的变量建立 *COX* 风险比例模型。根据表15，可以得到最终建立的 *COX* 比例模型为：

$$h(t|X) = h_0(t) \exp\{0.68197 \times \text{sector} - 4.03633 \times X_3 - 0.60867 \times X_5 - 1.94647 \times X_9 + 1.28226 \times X_{10}\} \quad (6)$$

Step3: 相对风险分析

由基本模型的整体检验可知，在不考虑其它协变量的作用下，主板上市公司相对于创业板上市公司的退市风险大小为 2.402，在考虑了其它协变量的影响之下，主板上市公司相对于创业板上市公司的退市风险大小为 1.97778，相对风险减少了很多。这是因为，根据 *COX* 危险比例模型可知，资产总计 (X_3)，管理费用 (X_5) 与投资活动产生的现金流量净额 (X_9) 前面的回归系数均为负值，说明这些协变量都是保护因子，即资产总计，管理费用与投资活动产生的现金流量净额的增加，上市公司退市风险减小。而流动负债合计 (X_{10}) 前面的回归系数是正值，说明流动负债合计是危险因子，该变量的增加，会增大上市公司的退市风险。进一步，我们可以给出主板上市公司与创业板上市公司在控制其它因素不变的情况下，各协变量对应的相对风险大小以及其置信区间，具体结果见下表。

表 16 各协变量对应的相对风险大小以及其置信区间

因素	相对风险大小	相对风险95% 置信下限	相对风险95% 置信上限
<i>sector</i>	1.978	1.499	2.609
资产总计 (X_3)	0.018	0.002	0.178

管理费用 (X_5)	0.544	0.283	1.047
投资活动产生的现金流量净额 (X_9)	0.143	0.056	0.365
流动负债合计 (X_{10})	3.605	1.335	9.732

4. 对保护因子与危险因子的进一步解释

从保护因子的角度分析，资产总计是保护因子，因为一般认为公司的总资产越多，其公司应对风险的能力越强，越容易达到上市公司的各项财务要求，从而保持上市状态。管理费用是保护因子主要是由于管理费用是公司内部运营情况的体现，好的上市公司注重内部人才、资金等各方面的管理，自然管理费用就会更多。投资活动产生的现金流量净额是保护因子，因为一般认为一个公司的资金流动是一个公司资金周转的体现，从某种程度上是企业具有活力的表现。

从危险因子的角度分析，流动负债合计是危险因子，因为公司长时期负债会增大退市风险。

五、模型的改进

考虑建立时间相依的 COX 比例危险模型，在以上 COX 比例危险模型的建立中，我们假定协变量不随时间发生改变。因此，进一步，我们基于表16中选定的协变量，进一步探讨协变量对时间的相依性。通过对每个协变量分别引入时间因素，并定义 $X_i(t) = X_i \times \ln t (i=1,2,3...11)$ ， $sector(t) = sector \times \ln t$ ，然后做局部检验。并得出因子 $sector$ 具有非比例的危险，检验结果具体如下表。

表 17 时间相依的局部似然比检验

因素	局部似然比检验卡方值	df	P 值
$sector$	10.773	1	0.0010
资产合计 (X_3)	0.282	1	0.5954
管理费用 (X_5)	1.857	1	0.1730

投资活动产生的现金流量净额 (X_9)	0.019	1	0.8904
流动负债合计 (X_{10})	0.307	1	0.5795

进一步，在原来的 COX 危险比例模型中引入协变量 $sector(t) = sector \times \ln t$ ，最终得到建模结果如下表。

表 18 改进的 COX 比例危险模型

因素	B	SE	$wald$	df	P 值	$\exp(B)$
$sector$	-3.636	0.970	14.056	1	0.000	0.026
资产合计 (X_3)	-4.277	1.213	12.435	1	0.000	0.014
管理费用 (X_5)	-0.669	0.344	3.773	1	0.052	0.512
投资活动产生的现金流量净额 (X_9)	-2.080	0.500	17.296	1	0.000	0.125
流动负债合计 (X_{10})	1.396	0.530	6.933	1	0.008	4.037
$Sector(t)$	0.680	0.221	9.450	1	0.002	1.974

根据表 18，最终得到改进后的模型为

$$h(t | X) = h_0(t) \exp \{ -3.636 \times sector - 3.636 \times X_3 - 0.669 \times X_5 - 2.080 \times X_9 + 1.396 \times X_{10} + 0.680 \times sector(t) \}$$

模型结果在符号上与前面建模结果具有一致性，进一步验证了模型的正确性。

六、研究的优点与不足

(一) 研究的优点

①本文运用生存分析方法研究海外上市公司退市风险，在研究方法上具有创新性，对理论研究与实际应用都具有重要意义；

②本文研究的数据质量具有权威性，数据指标选取合理，数据筛选时控制了其它因素，同时充分运用 *python* 对数据进行统计整理与数据预处理，清洗后的数据量大，数据质量高，满足建模的相关需求；

③本文运用生存分析动态的分析了海外上市公司上市后面临的退市风险情况，得出的相关结论具有实际意义；

④充分利用 *python*、*R*、*SPSS* 等统计软件建模，提高了建模精度与效率。

（二）研究的不足

①本文在建模的过程中只注重生存分析统计方法的建模，而没有考虑其它经典的统计方法，比如 *logistic* 模型、判别分析模型，在研究方法上略显单一。

②由于时间限制，本文还可以考虑对指标因素进行因子分析，建立基于因子分析的 *COX* 比例风险模型，可能会得到更好的结果。

七、 结论与建议

（一）研究的相关结论

本文主要运用生存分析的方法对海外上市公司面临的退市风险进行研究。研究数据主要来自 *CSMAR* 数据库中关于海外上市公司的统计数据，研究的对象为海外主板上市公司与创业板上市公司。我们主要从上市公司营业规模、运营能力与公司资产流动与负债情况三个维度建立评价指标体系，最终获得11个评价指标。

基于海外主板上市公司与创业板上市公司的生存函数估计，可以得出随着海外上市公司上市时间的推移，上市公司面临的退市风险不断增大。通过两样本的假设检验问题，最终得出在研究时间范围内，主板上市公司的退市风险比创业板上市公司的退市风险小的结论。进一步，在研究范围内，得到主板上市公司退市时间的平均估计为 414.589 个月，95% 的置信区间为 [388.822, 440.355]。创业板上市公司退市时间的平均估计为 171.685 个月，95% 的置信区间为

[156.651,186.719]。

之后，我们对以上生存函数的点估计的有效性进行评估，分别给出了三种生存函数点估计的置信区间。并且对不同置信区间进行的可视化研究分析，得出以上建立生存函数的点估计可信度较高。

接下来，基于以上研究成果，我们引入标识变量 *sector*，建立了多因素 COX 危险比例模型。通过对模型的分析，在不考虑其它协变量的作用下，主板上市公司相对于创业板上市公司的退市风险大小为 2.402。在考虑了其它协变量的影响之下，主板上市公司相对于创业板上市公司的退市风险大小为 1.97778，相对风险减少了很多。因为资产总计，管理费用与投资活动产生的现金流量净额是保护因子，流动负债合计是危险因子。

最后，我们对模型进行了改进与客观的评价。

（二）研究的相关建议

基于此次研究成果，我们给海外上市公司提出相关建议，具体如下：

企业在选择上市类型时应该充分评估自我情况，选择合适的上市类型。

主板上市公司在上市后 580 个月左右面临 50% 左右的退市风险。创业板上市公司在上市 250 个月左右面临着 60% 左右的退市风险。因此，随着企业上市时间的推移，主板上市公司与创业板上市公司都应加大企业内部管理，以降低退市风险。

从实证结果可以得出企业流动负债合计对增加企业退市风险的影响最大，因此，企业应该重视公司负债情况，过多的负债容易导致破产。因此，企业内部应加强对负债情况的自我评估，建立有效的财务预警系统。

八、参考文献

- [1]蔡亚蓉,李璟.中国企业海外上市的回顾与思考[J].中国外汇管理,2001(09):38-39.
- [2]王静楠.基于公司治理结构的中国赴美上市公司退市风险研究[D].东北财经大学,2012.
- [3]曹强,方扶星.上市公司退市风险预警研究——基于主成分分析的支持向量机预测模型[J].铜陵学院学报,2020,19(04):33-37.
- [4]虞文美,方扶星.中小企业板上市公司退市风险预警研究——基于因子分析的 Rprop 神经网络模型分析[J].齐齐哈尔大学学报(自然科学版),2020,36(04):69-73.
- [5]蔡梦.生存分析理论及其应用研究综述[J].价值工程,2016,35(10):19-21.
- [6]潘雅琼,刘艳.科技型企业财务危机预警研究——基于生存分析法[J].会计之友,2021(08):64-70.
- [7]程显宏,李豫新,邹宗森.“一带一盟”对接合作视阈下贸易便利化对出口持续时间的影响研究[J/OL].重庆大学学报(社会科学版):1-16[2021-05-02].<http://kns.cnki.net/kcms/detail/50.1023.C.20210309.1754.004.html>.
- [8]彭非,王伟.生存分析:中国人民大学出版社[M].北京,2004.

九、附录

(一) *python* 数据统计整理代码

```
import pandas as pd
import numpy as np
import os
# 更改文件路劲
os.chdir('D:\数据')
#设置最大显示列数
pd.set_option('display.max_columns', 30)
#设置最大显示行数
pd.set_option('display.max_rows', 500)
import xlrd
workbook = xlrd.open_workbook('data_all.xlsx')
sheet_name = workbook.sheet_names() #返回所有 sheet 的列表
order1 = pd.read_excel('data_all.xlsx',sheet_name='data_one')
order2 = pd.read_excel('data_all.xlsx',sheet_name='data_two')
order3 = pd.read_excel('data_all.xlsx',sheet_name='data_three')
order4 = pd.read_excel('data_all.xlsx',sheet_name='data_four')
order5 = pd.read_excel('data_all.xlsx',sheet_name='data_five')
df1 = pd.merge(left = order1, right=order2, how='inner',
left_on='C0ID', right_on = 'C0ID')
df1.drop(labels = ['CoName', 'StockCode', 'Sector', 'NumExchange',
'ListDate', 'Status', 'DelistDate'],axis = 1,inplace=True)
df11 = df1.groupby('C0ID') #创建分组对象,按照公司 ID 进行分组
df111=df11.mean()
df2 = pd.merge(left = order1, right=order3, how='inner',
left_on='C0ID', right_on = 'C0ID')
df2.drop(labels = ['CoName', 'StockCode', 'Sector', 'NumExchange',
'ListDate', 'Status', 'DelistDate'],axis = 1,inplace=True)
df22 = df2.groupby('C0ID') #创建分组对象,按照公司 ID 进行分组
df222=df22.mean()
df3 = pd.merge(left = order1, right=order4, how='inner',
left_on='C0ID', right_on = 'C0ID')
df3.drop(labels = ['CoName', 'StockCode', 'Sector', 'NumExchange',
'ListDate', 'Status', 'DelistDate'],axis = 1,inplace=True)
df33 = df3.groupby('C0ID') #创建分组对象,按照公司 ID 进行分组
df333=df33.mean()
df01 = pd.merge(left = df111, right=df222, how='inner',
left_on='C0ID', right_on = 'C0ID')
df02 = pd.merge(left = df01, right=df333, how='inner',
```

```

left_on='C0ID', right_on = 'C0ID')
df = pd.merge(left = order1, right=df02, how='inner',
left_on='C0ID', right_on = 'C0ID')
df.to_excel('processingdata 创业
板.xlsx',sheet_name='one',index=False)
dfA = pd.merge(left = order5, right=order2, how='inner',
left_on='C0ID', right_on = 'C0ID')
dfA.drop(labels = ['CoName', 'StockCode', 'Sector', 'NumExchange',
'ListDate', 'Status', 'DelistDate'],axis = 1,inplace=True)
dfAA = dfA.groupby('C0ID') #创建分组对象,按照公司 ID 进行分组
dfAAA=dfAA.mean()
dfB = pd.merge(left = order5, right=order3, how='inner',
left_on='C0ID', right_on = 'C0ID')
dfB.drop(labels = ['CoName', 'StockCode', 'Sector', 'NumExchange',
'ListDate', 'Status', 'DelistDate'],axis = 1,inplace=True)
dfBB = dfB.groupby('C0ID') #创建分组对象,按照公司 ID 进行分组
dfBBB=dfBB.mean()
dfC = pd.merge(left = order5, right=order4, how='inner',
left_on='C0ID', right_on = 'C0ID')
dfC.drop(labels = ['CoName', 'StockCode', 'Sector', 'NumExchange',
'ListDate', 'Status', 'DelistDate'],axis = 1,inplace=True)
dfC
dfCC = dfC.groupby('C0ID') #创建分组对象,按照公司 ID 进行分组
dfCCC=dfCC.mean()
df0A = pd.merge(left = dfAAA, right=dfBBB, how='inner',
left_on='C0ID', right_on = 'C0ID')
df0B = pd.merge(left = df0A, right=dfCCC, how='inner',
left_on='C0ID', right_on = 'C0ID')
dft = pd.merge(left = order5, right=df0B, how='inner',
left_on='C0ID', right_on = 'C0ID')
dft.to_excel('processingdata 主
板.xlsx',sheet_name='one',index=False)

```

(二) python 数据预处理代码

```

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.mlab as mlab
import matplotlib

```

```

import os
# 更改文件路径
os.chdir('D:\数据')
df = pd.read_csv('processingdata 主板.csv')#创业板为 processingdata
创业板.csv
df1=df.drop(labels = ['CoName', 'StockCode', 'Sector',
'NumExchange', 'ListDate', 'Status', 'DelistDate'],axis = 1)
#缺失值热图
cols = df1.columns[:30]
colours = ['#000099', '#ffff00'] # 黄色代表缺失
sns.heatmap(df[cols].isnull(), cmap=sns.color_palette(colours))
df2=df1.dropna(how='any',axis = 0) #只要有缺失, 就删除这一行
df2.columns
df3=df.drop(labels = ['X11', 'X12', 'X13', 'X14', 'X15', 'X16',
'X17', 'X21', 'X22', 'X23', 'X24', 'X25', 'X26', 'X27', 'X28',
'X29', 'X31', 'X32', 'X33', 'X34', 'X35', 'X36'],axis = 1)
df4 = pd.merge(left = df3, right=df2, how='inner',
left_on='C0ID', right_on = 'C0ID')# 内连接
df4.to_excel('endata 主板.xlsx',sheet_name='one',index=False)#处理
创业数据只需要修改相应数据文件名称即可, 即 endata 创业板.xlsx
#数据异常值处理
import numpy as np
import pandas as pd
import os
os.chdir(r'D:\数据')
df = pd.read_csv('endata 创业板.csv',encoding = 'gbk')#endata 主
板.csv
P1 =df.X1.quantile(0.01); P99 = df.X1.quantile(0.99)#计算 99 分位
数与 1 分位数
df['X1_new'] = df['X1']#将处理后的变量存储到新变量中
df.loc[df['X1'] > P99,'X1_new'] = P99
df.loc[df['X1'] < P1,'X1_new'] = P1
P1 =df.X2.quantile(0.01); P99 = df.X2.quantile(0.99)#计算 99 分位
数与 1 分位数
df['X2_new'] = df['X2']#将处理后的变量存储到新变量中
df.loc[df['X2'] > P99,'X2_new'] = P99
df.loc[df['X2'] < P1,'X2_new'] = P1
P1 =df.X3.quantile(0.01); P99 = df.X3.quantile(0.99)#计算 99 分位
数与 1 分位数
df['X3_new'] = df['X3']#将处理后的变量存储到新变量中
df.loc[df['X3'] > P99,'X3_new'] = P99
df.loc[df['X3'] < P1,'X3_new'] = P1
P1 =df.X4.quantile(0.01); P99 = df.X4.quantile(0.99)#计算 99 分位
数与 1 分位数

```

```

df['X4_new'] = df['X4']#将处理后的变量存储到新变量中
df.loc[df['X4'] > P99, 'X4_new'] = P99
df.loc[df['X4'] < P1, 'X4_new'] = P1
P1 = df.X5.quantile(0.01); P99 = df.X5.quantile(0.99)#计算 99 分位数与 1 分位数
df['X5_new'] = df['X5']#将处理后的变量存储到新变量中
df.loc[df['X5'] > P99, 'X5_new'] = P99
df.loc[df['X5'] < P1, 'X5_new'] = P1
P1 = df.X6.quantile(0.01); P99 = df.X6.quantile(0.99)#计算 99 分位数与 1 分位数
df['X6_new'] = df['X6']#将处理后的变量存储到新变量中
df.loc[df['X6'] > P99, 'X6_new'] = P99
df.loc[df['X6'] < P1, 'X6_new'] = P1
P1 = df.X7.quantile(0.01); P99 = df.X7.quantile(0.99)#计算 99 分位数与 1 分位数
df['X7_new'] = df['X7']#将处理后的变量存储到新变量中
df.loc[df['X7'] > P99, 'X7_new'] = P99
df.loc[df['X7'] < P1, 'X7_new'] = P1
P1 = df.X8.quantile(0.01); P99 = df.X8.quantile(0.99)#计算 99 分位数与 1 分位数
df['X8_new'] = df['X8']#将处理后的变量存储到新变量中
df.loc[df['X8'] > P99, 'X8_new'] = P99
df.loc[df['X8'] < P1, 'X8_new'] = P1
P1 = df.X9.quantile(0.01); P99 = df.X9.quantile(0.99)#计算 99 分位数与 1 分位数
df['X9_new'] = df['X9']#将处理后的变量存储到新变量中
df.loc[df['X9'] > P99, 'X9_new'] = P99
df.loc[df['X9'] < P1, 'X9_new'] = P1
P1 = df.X10.quantile(0.01); P99 = df.X10.quantile(0.99)#计算 99 分位数与 1 分位数
df['X10_new'] = df['X10']#将处理后的变量存储到新变量中
df.loc[df['X10'] > P99, 'X10_new'] = P99
df.loc[df['X10'] < P1, 'X10_new'] = P1
P1 = df.X11.quantile(0.01); P99 = df.X11.quantile(0.99)#计算 99 分位数与 1 分位数
df['X11_new'] = df['X11']#将处理后的变量存储到新变量中
df.loc[df['X11'] > P99, 'X11_new'] = P99
df.loc[df['X11'] < P1, 'X11_new'] = P1
df##查看处理后的数据
df.to_excel('data 创业板.xlsx', sheet_name='one', index=False)#导出数据

```

(三) R 程序代码

```
library(survival) # R needs this
library(readxl)
read_xlsx("C:\\Users\\86137\\Desktop\\endata 主板 1.xlsx")
mydata1=read_xlsx("C:\\Users\\86137\\Desktop\\endata 主板 1.xlsx")
read_xlsx("C:\\Users\\86137\\Desktop\\endata 创业板 0.xlsx")
mydata0=read_xlsx("C:\\Users\\86137\\Desktop\\endata 创业板 0.xlsx")
#K-M 估计主板和创业板的生存函数
survtime1=mydata1[,7]
status1=mydata1[,6]
survtime1=apply(survtime1,1,as.numeric)
status1=apply(status1,1,as.numeric)
fit1<-survfit(Surv(survtime1, status1)~1, conf.type="plain")
survtime0=mydata0[,7]
status0=mydata0[,6]
survtime0=apply(survtime0,1,as.numeric)
status0=apply(status0,1,as.numeric)
fit0<-survfit(Surv(survtime0, status0)~1, conf.type="plain")
#输出主板和创业板的生存函数的 K-M 估计及其标准差
print(summary(fit1))
print(summary(fit0))
#作出生存函数图像
plot(fit1, xlab="Time(months)", ylab="Survival probability", xlim =
c(0,600),ylim = c(0,1),col = "red",main="The K-M estimate for 主板
group")
plot(fit0, xlab="Time(months)", ylab="Survival probability",xlim =
c(0,300),ylim = c(0,1),col = "blue",main="The K-M estimate for 创业板
group")
plot(fit1, xlab="Time(months)", ylab="Survival probability",xlim =
c(0,600),ylim = c(0,1),col = "red",main="The K-M estimate for 主板&创
业板")
lines(fit0,type = 'l',col = 'blue', lwd = 1)
legend("topright", inset=.01, c(" 主板 ", " 创业板 "),cex=0.6,lty=c(1,
1),col=c("red", "blue"))

read_xlsx("C:\\Users\\86137\\Desktop\\主板置信区间.xlsx")
data1=read_xlsx("C:\\Users\\86137\\Desktop\\主板置信区间.xlsx")
read_xlsx("C:\\Users\\86137\\Desktop\\创业板置信区间.xlsx")
data0=read_xlsx("C:\\Users\\86137\\Desktop\\创业板置信区间.xlsx")
#作出主板组生存函数的三种 95%置信区间图像
time1=data1[,1]
time1=apply(time1,1,as.numeric)
```



```

St1=data1[,2]
St1=apply(St1,1,as.numeric)
CI11_L=data1[,4]
CI11_L=apply(CI11_L,1,as.numeric)
CI11_U=data1[,5]
CI11_U=apply(CI11_U,1,as.numeric)
CI12_L=data1[,6]
CI12_L=apply(CI12_L,1,as.numeric)
CI12_U=data1[,7]
CI12_U=apply(CI12_U,1,as.numeric)
CI13_L=data1[,8]
CI13_L=apply(CI13_L,1,as.numeric)
CI13_U=data1[,9]
CI13_U=apply(CI13_U,1,as.numeric)

par(mfrow=c(1,3))
plot(time1,St1,xlab="Time(months)", ylab="Survival
probability",type="l",col = "red",lty=1,lwd=1,main="95%线性置信区间")
lines(time1,CI11_L,type = 'l',col = 'black',lty=2,lwd = 1)
lines(time1,CI11_U,type = 'l',col = 'black',lty=2,lwd = 1)
plot(time1,St1,xlab="Time(months)", ylab="Survival
probability",type="l",col = "red",lty=1,lwd=1,main="95%对数置信区间")
lines(time1,CI12_L,type = 'l',col = 'orange',lty=10,lwd = 1)
lines(time1,CI12_U,type = 'l',col = 'orange',lty=10,lwd = 1)
plot(time1,St1,xlab="Time(months)", ylab="Survival
probability",type="l",col = "red",lty=1,lwd=1,main="95%反正弦平方根置
信区间")
lines(time1,CI13_L,type = 'l',col = 'green',lty=20,lwd = 1)
lines(time1,CI13_U,type = 'l',col = 'green',lty=20,lwd = 1)

par(mfrow=c(1,1))
plot(time1,St1,xlab="Time(months)", ylab="Survival
probability",type="l",col = "red",lty=1,lwd=1,main="三种 95%置信区
间的比较(主板)")
lines(time1,CI11_L,type = 'l',col = 'black',lty=2,lwd = 1)
lines(time1,CI11_U,type = 'l',col = 'black',lty=2,lwd = 1)
lines(time1,CI12_L,type = 'l',col = 'orange',lty=10,lwd = 1)
lines(time1,CI12_U,type = 'l',col = 'orange',lty=10,lwd = 1)
lines(time1,CI13_L,type = 'l',col = 'green',lty=20,lwd = 1)
lines(time1,CI13_U,type = 'l',col = 'green',lty=20,lwd = 1)
legend("topright", inset=0.005, c("生存函数","线性区间","对数变换区间",
"反正弦平方根变换区间"),cex=0.5,lty=c(1, 2,10,20),col=c("red",
"black","orange","green"))

```

#作出创业组生存函数的三种 95%置信区间图像

```
time0=data0[,1]
time0=apply(time0,1,as.numeric)
St0=data0[,2]
St0=apply(St0,1,as.numeric)
CI01_L=data0[,4]
CI01_L=apply(CI01_L,1,as.numeric)
CI01_U=data0[,5]
CI01_U=apply(CI01_U,1,as.numeric)
CI02_L=data0[,6]
CI02_L=apply(CI02_L,1,as.numeric)
CI02_U=data0[,7]
CI02_U=apply(CI02_U,1,as.numeric)
CI03_L=data0[,8]
CI03_L=apply(CI03_L,1,as.numeric)
CI03_U=data0[,9]
CI03_U=apply(CI03_U,1,as.numeric)

par(mfrow=c(1,3))
plot(time0,St0,xlab="Time(months)", ylab="Survival
probability",type="l",col = "red",lty=1,lwd=1,main="95%线性置信区间")
lines(time0,CI01_L,type = 'l',col = 'black',lty=2,lwd = 1)
lines(time0,CI01_U,type = 'l',col = 'black',lty=2,lwd = 1)
plot(time0,St0,xlab="Time(months)", ylab="Survival
probability",type="l",col = "red",lty=1,lwd=1,main="95%对数置信区间")
lines(time0,CI02_L,type = 'l',col = 'orange',lty=10,lwd = 1)
lines(time0,CI02_U,type = 'l',col = 'orange',lty=10,lwd = 1)
plot(time0,St0,xlab="Time(months)", ylab="Survival
probability",type="l",col = "red",lty=1,lwd=1,main="95%反正弦平方根置
信区间")
lines(time0,CI03_L,type = 'l',col = 'green',lty=20,lwd = 1)
lines(time0,CI03_U,type = 'l',col = 'green',lty=20,lwd = 1)

par(mfrow=c(1,1))
plot(time0,St0,xlab="Time(months)", ylab="Survival
probability",type="l",col = "red",lty=1,lwd=1,main="三种 95%置信区间
的比较(创业板)")
lines(time0,CI01_L,type = 'l',col = 'black',lty=2,lwd = 1)
lines(time0,CI01_U,type = 'l',col = 'black',lty=2,lwd = 1)
lines(time0,CI02_L,type = 'l',col = 'orange',lty=10,lwd = 1)
lines(time0,CI02_U,type = 'l',col = 'orange',lty=10,lwd = 1)
lines(time0,CI03_L,type = 'l',col = 'green',lty=20,lwd = 1)
lines(time0,CI03_U,type = 'l',col = 'green',lty=20,lwd = 1)
```

```

legend("topright", inset=0.005, c("生存函数", "线性区间", "对数变换区间",
"反正弦平方根变换区间"), cex=0.5, lty=c(1, 2, 10, 20), col=c("red",
"black", "orange", "green"))

```

(四) 生存函数的 K-M 估计结果汇总表

1. 主板的乘积限估计式及其标准差估计

研究时间	风险暴露数目	事件数目	乘积限估计	乘积限估计的标准差	95%置信区间下限	95%置信区间上限
1	756	1	0.999	0.00132	0.996	1.000
3	755	1	0.997	0.00187	0.994	1.000
16	710	1	0.996	0.00233	0.991	1.000
20	682	2	0.993	0.00311	0.987	0.999
30	612	1	0.991	0.00350	0.985	0.998
32	601	1	0.990	0.00387	0.982	0.997
33	587	3	0.985	0.00482	0.975	0.994
34	581	1	0.983	0.00511	0.973	0.993
43	553	1	0.981	0.00540	0.971	0.992
45	549	2	0.978	0.00594	0.966	0.989
46	544	2	0.974	0.00644	0.961	0.987
47	540	2	0.970	0.00690	0.957	0.984
49	532	1	0.969	0.00713	0.955	0.983
50	530	1	0.967	0.00734	0.952	0.981
53	515	1	0.965	0.00756	0.950	0.980
54	508	1	0.963	0.00778	0.948	0.978
55	507	1	0.961	0.00800	0.945	0.977
56	506	2	0.957	0.00841	0.941	0.974
58	500	1	0.955	0.00860	0.939	0.972
59	497	1	0.953	0.00880	0.936	0.971
60	494	1	0.952	0.00899	0.934	0.969
63	490	2	0.948	0.00936	0.929	0.966
64	480	1	0.946	0.00955	0.927	0.964
66	470	1	0.944	0.00974	0.925	0.963
67	468	1	0.942	0.00992	0.922	0.961
68	467	1	0.940	0.01011	0.920	0.959
70	455	1	0.938	0.01029	0.917	0.958
71	453	2	0.933	0.01066	0.913	0.954
72	451	1	0.931	0.01083	0.910	0.953
73	448	1	0.929	0.01100	0.908	0.951
75	444	2	0.925	0.01135	0.903	0.947

79	435	1	0.923	0.01152	0.900	0.946
82	425	2	0.919	0.01187	0.895	0.942
84	418	1	0.916	0.01204	0.893	0.940
85	414	1	0.914	0.01221	0.890	0.938
87	409	2	0.910	0.01255	0.885	0.934
90	396	2	0.905	0.01290	0.880	0.930
94	390	1	0.903	0.01308	0.877	0.928
96	387	1	0.901	0.01325	0.875	0.926
98	385	2	0.896	0.01359	0.869	0.922
99	382	2	0.891	0.01392	0.864	0.918
100	379	1	0.889	0.01408	0.861	0.916
103	378	1	0.886	0.01424	0.859	0.914
104	377	1	0.884	0.01439	0.856	0.912
106	373	1	0.882	0.01455	0.853	0.910
107	372	1	0.879	0.01470	0.851	0.908
108	370	3	0.872	0.01514	0.843	0.902
109	367	1	0.870	0.01529	0.840	0.900
110	366	1	0.867	0.01543	0.837	0.898
111	364	1	0.865	0.01557	0.835	0.896
113	361	1	0.863	0.01571	0.832	0.893
117	358	1	0.860	0.01585	0.829	0.891
118	357	2	0.855	0.01612	0.824	0.887
120	350	2	0.851	0.01640	0.818	0.883
121	346	2	0.846	0.01667	0.813	0.878
122	344	1	0.843	0.01680	0.810	0.876
123	342	4	0.833	0.01731	0.799	0.867
124	336	4	0.823	0.01780	0.789	0.858
125	325	1	0.821	0.01793	0.786	0.856
126	320	2	0.816	0.01818	0.780	0.851
127	316	3	0.808	0.01855	0.772	0.844
128	313	1	0.805	0.01867	0.769	0.842
129	307	1	0.803	0.01879	0.766	0.840
131	303	1	0.800	0.01891	0.763	0.837
133	300	1	0.797	0.01904	0.760	0.835
134	298	1	0.795	0.01916	0.757	0.832
136	293	1	0.792	0.01929	0.754	0.830
140	278	1	0.789	0.01943	0.751	0.827
141	277	2	0.784	0.01970	0.745	0.822
146	274	2	0.778	0.01997	0.739	0.817
147	272	1	0.775	0.02010	0.736	0.814
150	268	1	0.772	0.02023	0.732	0.812
151	266	2	0.766	0.02049	0.726	0.806
153	259	2	0.760	0.02075	0.720	0.801
156	253	2	0.754	0.02102	0.713	0.796

159	246	2	0.748	0.02129	0.706	0.790
160	242	2	0.742	0.02156	0.700	0.784
161	237	2	0.736	0.02183	0.693	0.779
163	231	1	0.733	0.02197	0.690	0.776
167	221	1	0.729	0.02211	0.686	0.773
168	217	1	0.726	0.02227	0.682	0.770
174	205	2	0.719	0.02261	0.675	0.763
176	202	1	0.715	0.02277	0.671	0.760
179	196	1	0.712	0.02295	0.667	0.757
183	183	1	0.708	0.02315	0.662	0.753
188	172	1	0.704	0.02338	0.658	0.749
195	162	1	0.699	0.02363	0.653	0.746
212	126	1	0.694	0.02409	0.647	0.741
215	121	1	0.688	0.02456	0.640	0.736
218	120	1	0.682	0.02502	0.633	0.731
237	112	1	0.676	0.02552	0.626	0.726
241	108	1	0.670	0.02604	0.619	0.721
270	91	1	0.663	0.02678	0.610	0.715
278	89	1	0.655	0.02749	0.601	0.709
282	85	1	0.647	0.02823	0.592	0.703
290	75	2	0.630	0.03000	0.571	0.689
308	65	1	0.620	0.03106	0.560	0.681
420	23	1	0.593	0.03974	0.516	0.671
563	10	1	0.534	0.06670	0.403	0.665

2. 创业板组的乘积限估计式及其标准差估计

研究时间	风险暴露数目	事件数目	乘积限估计	乘积限估计的标准差	95%置信区间下限	95%置信区间上限
16	212	1	0.995	0.00471	0.986	1.000
20	209	1	0.991	0.00667	0.977	1.000
21	208	1	0.986	0.00816	0.970	1.000
22	207	2	0.976	0.01050	0.956	0.997
23	205	1	0.971	0.01148	0.949	0.994
25	204	2	0.962	0.01320	0.936	0.988
26	201	1	0.957	0.01397	0.930	0.985
27	200	1	0.952	0.01470	0.924	0.981
29	197	1	0.948	0.01540	0.917	0.978
30	196	2	0.938	0.01669	0.905	0.971
34	189	3	0.923	0.01851	0.887	0.959
35	185	1	0.918	0.01907	0.881	0.955
36	183	1	0.913	0.01961	0.875	0.951
38	182	3	0.898	0.02113	0.857	0.939

41	176	2	0.888	0.02208	0.844	0.931
42	173	1	0.883	0.02254	0.838	0.927
44	172	2	0.872	0.02342	0.826	0.918
47	165	1	0.867	0.02387	0.820	0.914
48	164	1	0.862	0.02430	0.814	0.909
49	163	1	0.856	0.02472	0.808	0.905
50	162	2	0.846	0.02552	0.796	0.896
51	158	2	0.835	0.02630	0.784	0.887
52	154	1	0.830	0.02668	0.777	0.882
53	153	1	0.824	0.02705	0.771	0.877
54	151	1	0.819	0.02742	0.765	0.873
55	150	2	0.808	0.02812	0.753	0.863
56	148	1	0.803	0.02845	0.747	0.858
57	146	1	0.797	0.02878	0.741	0.853
58	144	1	0.791	0.02911	0.734	0.849
59	143	1	0.786	0.02943	0.728	0.844
60	142	2	0.775	0.03004	0.716	0.834
63	139	1	0.769	0.03033	0.710	0.829
64	138	1	0.764	0.03062	0.704	0.824
65	137	2	0.753	0.03117	0.691	0.814
66	135	2	0.741	0.03169	0.679	0.804
68	132	1	0.736	0.03195	0.673	0.798
69	131	1	0.730	0.03219	0.667	0.793
72	130	1	0.725	0.03243	0.661	0.788
75	128	1	0.719	0.03267	0.655	0.783
76	127	1	0.713	0.03290	0.649	0.778
77	126	1	0.708	0.03312	0.643	0.773
79	125	2	0.696	0.03354	0.631	0.762
80	123	1	0.691	0.03375	0.624	0.757
81	122	1	0.685	0.03394	0.618	0.751
82	121	1	0.679	0.03413	0.612	0.746
83	120	1	0.674	0.03431	0.606	0.741
91	118	1	0.668	0.03449	0.600	0.736
93	116	2	0.656	0.03484	0.588	0.725
96	114	1	0.651	0.03501	0.582	0.719
113	113	1	0.645	0.03517	0.576	0.714
119	112	1	0.639	0.03533	0.570	0.708
121	111	1	0.633	0.03547	0.564	0.703
122	110	1	0.628	0.03562	0.558	0.697
123	109	1	0.622	0.03575	0.552	0.692
124	105	1	0.616	0.03590	0.546	0.686
127	99	1	0.610	0.03607	0.539	0.680
133	88	1	0.603	0.03632	0.532	0.674
134	87	2	0.589	0.03678	0.517	0.661

137	79	1	0.581	0.03706	0.509	0.654
138	72	1	0.573	0.03742	0.500	0.647
159	59	1	0.564	0.03803	0.489	0.638
162	58	1	0.554	0.03859	0.478	0.630
166	55	1	0.544	0.03918	0.467	0.621
168	52	1	0.533	0.03980	0.455	0.611
186	44	1	0.521	0.04070	0.442	0.601
198	33	1	0.506	0.04242	0.422	0.589
200	30	1	0.489	0.04423	0.402	0.575
216	19	2	0.437	0.05244	0.334	0.540
241	7	1	0.375	0.07324	0.231	0.518

(五) 生存函数点估计的置信区间汇总表

1. 主板组生存函数的三种 95%置信区间

研究时间	乘积限估计	乘积限估计的标准差	线性置信区间		对数变换置信区间		反正弦平方根置信区间	
			95%置信下限	95%置信上限	95%置信下限	95%置信上限	95%置信下限	95%置信上限
1	0.999	0.00132	0.996	1.002	0.987	1.000	0.995	1.000
3	0.997	0.00187	0.993	1.001	0.990	0.999	0.992	1.000
16	0.996	0.00233	0.991	1.001	0.988	0.999	0.990	0.999
20	0.993	0.00311	0.987	0.999	0.983	0.997	0.986	0.998
30	0.991	0.0035	0.984	0.998	0.981	0.996	0.983	0.996
32	0.99	0.00387	0.982	0.998	0.979	0.995	0.981	0.996
33	0.985	0.00482	0.976	0.994	0.972	0.992	0.974	0.993
34	0.983	0.00511	0.973	0.993	0.970	0.990	0.972	0.991
43	0.981	0.0054	0.971	0.991	0.967	0.989	0.969	0.990
45	0.978	0.00594	0.967	0.989	0.963	0.987	0.965	0.988
46	0.974	0.00644	0.962	0.986	0.958	0.984	0.960	0.985
47	0.97	0.0069	0.957	0.983	0.954	0.981	0.956	0.982
49	0.969	0.00713	0.955	0.983	0.952	0.980	0.954	0.981
50	0.967	0.00734	0.953	0.981	0.950	0.978	0.952	0.979
53	0.965	0.00756	0.951	0.979	0.947	0.977	0.949	0.978
54	0.963	0.00778	0.948	0.978	0.945	0.975	0.947	0.976
55	0.961	0.008	0.946	0.976	0.943	0.974	0.945	0.975
56	0.957	0.00841	0.941	0.973	0.938	0.970	0.940	0.971
58	0.955	0.0086	0.939	0.971	0.936	0.969	0.938	0.970
59	0.953	0.0088	0.937	0.969	0.933	0.967	0.935	0.968
60	0.952	0.00899	0.935	0.969	0.932	0.966	0.934	0.967
63	0.948	0.00936	0.931	0.965	0.927	0.963	0.929	0.964
64	0.946	0.00955	0.928	0.964	0.925	0.961	0.927	0.962
66	0.944	0.00974	0.926	0.962	0.923	0.959	0.925	0.961
67	0.942	0.00992	0.924	0.960	0.921	0.958	0.922	0.959
68	0.94	0.01011	0.921	0.959	0.918	0.956	0.920	0.957

70	0.938	0.01029	0.919	0.957	0.916	0.954	0.918	0.955
71	0.933	0.01066	0.914	0.952	0.911	0.950	0.912	0.951
72	0.931	0.01083	0.911	0.951	0.908	0.948	0.910	0.949
73	0.929	0.011	0.909	0.949	0.906	0.947	0.908	0.948
75	0.925	0.01135	0.904	0.946	0.901	0.943	0.903	0.944
79	0.923	0.01152	0.902	0.944	0.899	0.941	0.901	0.942
82	0.919	0.01187	0.898	0.940	0.895	0.938	0.896	0.939
84	0.916	0.01204	0.894	0.938	0.892	0.935	0.893	0.936
85	0.914	0.01221	0.892	0.936	0.889	0.933	0.891	0.934
87	0.91	0.01255	0.888	0.932	0.885	0.930	0.887	0.931
90	0.905	0.0129	0.882	0.928	0.879	0.925	0.881	0.927
94	0.903	0.01308	0.880	0.926	0.877	0.924	0.879	0.925
96	0.901	0.01325	0.878	0.924	0.875	0.922	0.877	0.923
98	0.896	0.01359	0.872	0.920	0.869	0.917	0.871	0.919
99	0.891	0.01392	0.867	0.915	0.864	0.913	0.866	0.914
100	0.889	0.01408	0.864	0.914	0.862	0.911	0.863	0.912
103	0.886	0.01424	0.861	0.911	0.859	0.908	0.860	0.909
104	0.884	0.01439	0.859	0.909	0.856	0.907	0.858	0.908
106	0.882	0.01455	0.857	0.907	0.854	0.905	0.856	0.906
107	0.879	0.0147	0.854	0.904	0.851	0.902	0.853	0.903
108	0.872	0.01514	0.846	0.898	0.844	0.896	0.845	0.897
109	0.87	0.01529	0.844	0.896	0.841	0.894	0.843	0.895
110	0.867	0.01543	0.841	0.893	0.838	0.891	0.840	0.892
111	0.865	0.01557	0.839	0.891	0.836	0.889	0.838	0.890
113	0.863	0.01571	0.836	0.890	0.834	0.887	0.836	0.888
117	0.86	0.01585	0.833	0.887	0.831	0.884	0.832	0.885
118	0.855	0.01612	0.828	0.882	0.826	0.880	0.827	0.881
120	0.851	0.0164	0.824	0.878	0.821	0.876	0.823	0.877
121	0.846	0.01667	0.818	0.874	0.816	0.871	0.818	0.872
122	0.843	0.0168	0.815	0.871	0.813	0.869	0.814	0.870
123	0.833	0.01731	0.805	0.861	0.803	0.859	0.804	0.860
124	0.823	0.0178	0.794	0.852	0.792	0.850	0.794	0.851
125	0.821	0.01793	0.792	0.850	0.790	0.848	0.791	0.849
126	0.816	0.01818	0.787	0.845	0.785	0.843	0.786	0.844
127	0.808	0.01855	0.779	0.837	0.777	0.835	0.778	0.836
128	0.805	0.01867	0.776	0.834	0.774	0.833	0.775	0.833
129	0.803	0.01879	0.773	0.833	0.771	0.831	0.773	0.832
131	0.8	0.01891	0.770	0.830	0.768	0.828	0.770	0.829
133	0.797	0.01904	0.767	0.827	0.765	0.825	0.767	0.826
134	0.795	0.01916	0.765	0.825	0.763	0.823	0.765	0.824
136	0.792	0.01929	0.762	0.822	0.760	0.820	0.761	0.821
140	0.789	0.01943	0.759	0.819	0.757	0.817	0.758	0.818
141	0.784	0.0197	0.754	0.814	0.752	0.812	0.753	0.813
146	0.778	0.01997	0.748	0.808	0.746	0.807	0.747	0.808
147	0.775	0.0201	0.744	0.806	0.743	0.804	0.744	0.805
150	0.772	0.02023	0.741	0.803	0.740	0.801	0.741	0.802
151	0.766	0.02049	0.735	0.797	0.734	0.795	0.735	0.796
153	0.76	0.02075	0.729	0.791	0.727	0.789	0.729	0.790
156	0.754	0.02102	0.723	0.785	0.721	0.783	0.722	0.784

159	0.748	0.02129	0.717	0.779	0.715	0.778	0.716	0.778
160	0.742	0.02156	0.711	0.773	0.709	0.772	0.710	0.773
161	0.736	0.02183	0.705	0.767	0.703	0.766	0.704	0.767
163	0.733	0.02197	0.701	0.765	0.700	0.763	0.701	0.764
167	0.729	0.02211	0.697	0.761	0.696	0.759	0.697	0.760
168	0.726	0.02227	0.694	0.758	0.693	0.756	0.694	0.757
174	0.719	0.02261	0.687	0.751	0.686	0.749	0.687	0.750
176	0.715	0.02277	0.683	0.747	0.682	0.746	0.683	0.746
179	0.712	0.02295	0.680	0.744	0.679	0.743	0.680	0.743
183	0.708	0.02315	0.676	0.740	0.674	0.739	0.676	0.739
188	0.704	0.02338	0.672	0.736	0.670	0.735	0.671	0.736
195	0.699	0.02363	0.667	0.731	0.665	0.730	0.666	0.731
212	0.694	0.02409	0.661	0.727	0.660	0.725	0.661	0.726
215	0.688	0.02456	0.655	0.721	0.654	0.720	0.655	0.720
218	0.682	0.02502	0.649	0.715	0.647	0.714	0.648	0.715
237	0.676	0.02552	0.642	0.710	0.641	0.708	0.642	0.709
241	0.67	0.02604	0.636	0.704	0.635	0.703	0.636	0.704
270	0.663	0.02678	0.628	0.698	0.627	0.696	0.628	0.697
278	0.655	0.02749	0.620	0.690	0.618	0.689	0.619	0.690
282	0.647	0.02823	0.611	0.683	0.610	0.682	0.611	0.682
290	0.63	0.03	0.593	0.667	0.592	0.666	0.593	0.666
308	0.62	0.03106	0.582	0.658	0.581	0.656	0.582	0.657
420	0.593	0.03974	0.547	0.639	0.545	0.638	0.547	0.638
563	0.534	0.0667	0.464	0.604	0.462	0.601	0.464	0.603

2. 创业板组生存函数的三种 95%置信区间

研究时间	乘积限估计	乘积限估计的标准差	线性置信区间		对数变换置信区间		反正弦平方根置信区间	
			95%置信下限	95%置信上限	95%置信下限	95%置信上限	95%置信下限	95%置信上限
16	0.995	0.00471	0.986	1.004	0.969	0.999	0.982	1.000
20	0.991	0.00667	0.978	1.004	0.962	0.998	0.974	0.999
21	0.986	0.00816	0.970	1.002	0.957	0.995	0.966	0.997
22	0.976	0.0105	0.956	0.996	0.945	0.990	0.952	0.992
23	0.971	0.01148	0.949	0.993	0.939	0.986	0.945	0.989
25	0.962	0.0132	0.937	0.987	0.927	0.980	0.933	0.983
26	0.957	0.01397	0.931	0.983	0.921	0.977	0.927	0.979
27	0.952	0.0147	0.925	0.979	0.915	0.973	0.921	0.976
29	0.948	0.0154	0.919	0.977	0.910	0.970	0.916	0.973
30	0.938	0.01669	0.907	0.969	0.899	0.962	0.904	0.965
34	0.923	0.01851	0.890	0.956	0.882	0.950	0.886	0.953
35	0.918	0.01907	0.884	0.952	0.876	0.946	0.881	0.949
36	0.913	0.01961	0.878	0.948	0.870	0.942	0.875	0.945
38	0.898	0.02113	0.861	0.935	0.854	0.929	0.858	0.932

41	0.888	0.02208	0.850	0.926	0.843	0.921	0.847	0.923
42	0.883	0.02254	0.844	0.922	0.837	0.916	0.841	0.919
44	0.872	0.02342	0.832	0.912	0.826	0.907	0.830	0.909
47	0.867	0.02387	0.826	0.908	0.820	0.902	0.824	0.905
48	0.862	0.0243	0.821	0.903	0.815	0.898	0.819	0.900
49	0.856	0.02472	0.815	0.897	0.809	0.892	0.812	0.895
50	0.846	0.02552	0.804	0.888	0.798	0.883	0.802	0.886
51	0.835	0.0263	0.792	0.878	0.787	0.873	0.790	0.876
52	0.83	0.02668	0.787	0.873	0.781	0.869	0.785	0.871
53	0.824	0.02705	0.780	0.868	0.775	0.863	0.779	0.865
54	0.819	0.02742	0.775	0.863	0.770	0.859	0.773	0.861
55	0.808	0.02812	0.763	0.853	0.759	0.848	0.762	0.850
56	0.803	0.02845	0.758	0.848	0.754	0.844	0.757	0.846
57	0.797	0.02878	0.752	0.842	0.748	0.838	0.751	0.840
58	0.791	0.02911	0.746	0.836	0.742	0.832	0.744	0.834
59	0.786	0.02943	0.741	0.831	0.736	0.827	0.739	0.829
60	0.775	0.03004	0.729	0.821	0.725	0.817	0.728	0.819
63	0.769	0.03033	0.723	0.815	0.719	0.811	0.722	0.813
64	0.764	0.03062	0.718	0.810	0.714	0.806	0.717	0.808
65	0.753	0.03117	0.707	0.799	0.703	0.796	0.706	0.797
66	0.741	0.03169	0.695	0.787	0.692	0.784	0.694	0.785
68	0.736	0.03195	0.690	0.782	0.687	0.779	0.689	0.780
69	0.73	0.03219	0.684	0.776	0.681	0.773	0.683	0.775
72	0.725	0.03243	0.679	0.771	0.676	0.768	0.678	0.770
75	0.719	0.03267	0.673	0.765	0.670	0.762	0.672	0.764
76	0.713	0.0329	0.667	0.759	0.664	0.756	0.666	0.758
77	0.708	0.03312	0.662	0.754	0.659	0.751	0.661	0.753
79	0.696	0.03354	0.650	0.742	0.648	0.739	0.650	0.740
80	0.691	0.03375	0.645	0.737	0.643	0.734	0.645	0.735
81	0.685	0.03394	0.639	0.731	0.637	0.728	0.639	0.729
82	0.679	0.03413	0.634	0.724	0.631	0.722	0.633	0.723
83	0.674	0.03431	0.629	0.719	0.626	0.717	0.628	0.718
91	0.668	0.03449	0.623	0.713	0.621	0.711	0.622	0.712
93	0.656	0.03484	0.611	0.701	0.609	0.699	0.611	0.700
96	0.651	0.03501	0.606	0.696	0.604	0.694	0.606	0.695
113	0.645	0.03517	0.601	0.689	0.599	0.687	0.600	0.689
119	0.639	0.03533	0.595	0.683	0.593	0.681	0.594	0.682

121	0.633	0.03547	0.589	0.677	0.587	0.675	0.589	0.676
122	0.628	0.03562	0.584	0.672	0.582	0.670	0.584	0.671
123	0.622	0.03575	0.578	0.666	0.577	0.664	0.578	0.665
124	0.616	0.0359	0.573	0.659	0.571	0.658	0.572	0.659
127	0.61	0.03607	0.567	0.653	0.565	0.652	0.567	0.652
133	0.603	0.03632	0.560	0.646	0.559	0.644	0.560	0.645
134	0.589	0.03678	0.547	0.631	0.545	0.630	0.546	0.631
137	0.581	0.03706	0.539	0.623	0.538	0.622	0.539	0.623
138	0.573	0.03742	0.531	0.615	0.530	0.614	0.531	0.615
159	0.564	0.03803	0.522	0.606	0.521	0.605	0.522	0.606
162	0.554	0.03859	0.512	0.596	0.511	0.595	0.512	0.595
166	0.544	0.03918	0.502	0.586	0.501	0.585	0.502	0.585
168	0.533	0.0398	0.491	0.575	0.491	0.574	0.492	0.574
186	0.521	0.0407	0.479	0.563	0.479	0.562	0.480	0.562
198	0.506	0.04242	0.464	0.548	0.463	0.547	0.464	0.548
200	0.489	0.04423	0.447	0.531	0.446	0.531	0.447	0.531
216	0.437	0.05244	0.392	0.482	0.392	0.481	0.393	0.482
241	0.375	0.07324	0.321	0.429	0.321	0.429	0.322	0.429

十、 致谢

为期两个月的课题研究已经接近尾声，心里充满喜悦的同时，更多的是感动，更想表达的是感谢。

本课题在研究的过程中曾经几次经历了各种困难，导致研究无法正常进行。在这个过程中，我们有过想要放弃的念头。但在老师的支持与同学的鼓励之下，我们还是努力克服各种困难，完成了相关研究，并取得了较为满意的结果。

在整个课题的研究过程中，我们首先要感谢的是陈老师。由于在理论知识上的不足，我们有很多知识需要学习，在这个过程中，陈老师不仅在理论知识上给予我们指导，而且一直鼓励我们，使得我们在建模的时候更加自信。

其次，要感谢为这篇论文提供帮助的小潘同学，在课题开展的前期我们面

临的一个问题就是数据问题，数据的不足导致研究不能正常进行。在这个时候，小潘同学提供了相关数据查找推荐，为我们的研究提供了保障。

最后，要感谢我们自己。在研究过程中，我们不断尝试，相互讨论，在思维的碰撞中不断完善我们的模型，在辛苦工作的同时收获一次次解决问题之后带来的喜悦，在不断的交流磨合之中享受研究的乐趣，收获一拍即合的默契。