

# 生存分析与股指涨跌的概率推断<sup>①</sup>

雷 鸣<sup>1</sup>, 叶五一<sup>2</sup>, 缪柏其<sup>2</sup>, 郭文旌<sup>1</sup>

(1. 南京财经大学金融学院, 南京 210003 2 中国科学技术大学管理学院统计与金融系, 合肥 230026)

**摘要:** 通过研究上证指数连涨连跌的收益率, 发现其服从伽玛分布, 并且可得出股指涨跌的条件概率, 以及极值时的情况, 进一步研究了成交量的影响, 从而说明了技术分析特别是价量分析为何在实际中仍然被广泛运用. 可认为随机游动假说是一种粗糙的近似, 本研究也有助于加深对市场有效性理论的认识.

**关键词:** 生存分析; 伽玛分布; 概率推断; 随机游动

**中图分类号:** F832.5   **文献标识码:** A   **文章编号:** 1007-9807(2010)04-0057-10

## 0 引 言

1900年 Bachelier 提出股市收益率服从正态分布的假设. 然而后来许多学者研究发现许多不符合正态性假设的例子, 也就是说收益率具有尖峰厚尾的特点, 用正态分布进行拟合并不理想, 这也使得基于正态分布的统计推断无效. 进一步的研究表明, 收益率服从稳态分布, 而稳态分布包含的参数过多, 不易于应用. 本文研究的是连涨连跌的收益率.

其次关于有效市场的研究. 在经典金融理论中, 有效市场假说 (EMH) 一直是资本市场理论的基石. Fama<sup>[10-11]</sup> 于 1970 年给出了有效市场的严格定义. 该假说认为, 资本市场的当前价格已充分反映所有的市场公开信息, 今天的收益率与昨天的收益率无关, 且彼此独立, 价格变动是一个随机游动过程. 可见有效市场假说是与随机游动假说密切相关的, 当然正如 Fama 所言, 市场有效并不必需是随机游动. 关于随机游动假说, 已有研究予以了否定. 一种研究认为股市短期存在正相关, 长期存在负相关. 本文的研究也表明, 随机游动只是一种粗糙的近似, 股指收益率之间存在复杂的相关关系. 国内对有效市场的研究很多, 俞乔对

1994 年以前的中国股市有效性研究的结论是中国股市是非有效的; 吴世农<sup>[6]</sup> 指出中国股市当时尚未达到真正意义的弱型效率; 张兵<sup>[7]</sup> 认为 1997 年之前市场无效, 1997 年之后市场呈现弱式有效; 何诚颖从另一个角度也探讨了中国证券市场的有效性. 按照市场弱式有效的理论, 技术分析即是无效的, 但是在实际应用中, 技术分析仍在大量使用, 理论与实践产生了脱节, 这是为什么? 本文的研究恰能从一个角度对此加以解释.

其实有大量研究表明技术分析在一定程度上是有效的. Treynor<sup>[12]</sup> 等的研究表明过去的证券价格结合其它有价值的信息, 对于获得超额收益是有利的; Blume<sup>[13]</sup> 等说明了成交量对于技术分析的适用性; Lo<sup>[9]</sup> 等认为一些技术指标确实提供了一些附加信息从而具有一定的实际价值. 戴洁<sup>[2]</sup> 等论述了技术分析对于股价具有某种程度上的预测力.

生存分析 (Survival Analysis) 是工程、医学和生物学等领域中一个很受关心的内容, 是研究有关生存时间即寿命数据的统计规律的. 生存分析已成为现代数理统计的一个重要分支. 许多统计学家在这一领域做出了大量工作, 尤其是 Cox 的重要贡献. 国外有学者利用生存分析研究一些股

① 收稿日期: 2008-10-07; 修订日期: 2009-02-25.

基金项目: 国家自然科学基金资助项目 (70471071); 国家软科学研究计划资助项目 (2006GXQ3D134); 江苏省高校哲学社会科学基金资助项目 (07SJB790013).

作者简介: 雷 鸣 (1968—), 男, 安徽合肥人, 博士, 副教授. Email: leiming1968@126.com

票在市场上的生存时间与规律<sup>[14]</sup>, 而笔者将生存分析引入对股指的研究<sup>[3]</sup>, 因为统计连涨连跌的收益率, 它可视为股指每次涨跌的“寿命”, 即对生存时间的定义与众不同. 股市指数的连续上涨和下跌可以看作是一种特殊的生存过程, 当股指连续上涨到头转为下跌时, 可以视作上涨的“死亡”; 同样当股指连续下跌到头转为上涨时, 可视为下跌的“死亡”, 股指就是在这两种状态下不停地进行着“生”、“死”相互转化的. 通过生存分析理论中的危险率曲线分析及连涨连跌的相关性分析, 可以看出股指收益率之间并不是独立的, 它们之间存在一种复杂的相关关系, 从而从另一个角度否定了随机游动假说. 进一步地可以得出股指涨跌的概率和条件概率, 从而得出股指涨跌的概率推断, 特别考虑了处于极值时的情形, 同时也研究了成交量的影响, 这为投资者进行投资提供了有价值的参考信息. 因此, 生存分析方法提供了研究股指的一种新视角, 其有助于研究有关股市的一些基础性问题. 本文的研究是在笔者已往研究基础上的延伸.

# 1 历史价格对股指未来涨跌的影响

## 1.1 收益率理论分布的估计与检验

通过研究 1992年 5月 21日至 2007年 4月

11日的上证指数. (舍掉 1992年 5月 21日以前的数据是因为在此前后的上证指数编报方法不同, 无法放到一起作比较.)

采用对数收益率即

$$r = \ln p_t - \ln p_{t-1}$$

连涨的收益率  $x_i$  和连跌的收益率  $y_i$  的数学表达如下:

$$k_0 = \inf\{t: p(t+1) - p(t) < 0\}$$

$$k_{2m+1} = \inf\{t: t > k_{2m}, p(t+1) - p(t) < 0\}$$

$$k_{2m+2} = \inf\{t: t > k_{2m+1}, p(t+1) - p(t) \geq 0\}$$

$$x_{m+1} = \sum_{k_{2m}}^{k_{2m+1}} r(t), m = 0, 1, 2, 3, \dots$$

$$y_{m+1} = \sum_{k_{2m+1}}^{k_{2m+2}} r(t), m = 0, 1, 2, 3, \dots$$

其实股指连涨连跌的天数就对应着统计中所述的游程的个数, 只不过这里研究的是每一个游程中收益率的和, 即连涨几天中或连跌几天中收益率的和, 称之为连涨连跌的收益率. 从生存分析的角度看, 连涨的收益率  $R_1$  和连跌的收益率  $R_0$  可看作是两个不同的生存过程.

通过拟合优度检验, 对于一维分布的检验, Kolmogorov-Smirnov 检验较卡方检验为优, 因此采用了前者, 所得结果见文献 [5], 为了本文后面讨论的方便, 将这一结果重又放到这里, 如表 1

表 1 上证指数在不同时期连涨连跌收益率分布的拟合

Table 1 Fitting of successive return of rises and falls of Shanghai Stock Index during the different periods

时间	T + 0时期	T + 1时期	涨停板时期	国有股减持时期	近期
上涨:					
分布	$\Gamma(0.72, 13.4)$	$\Gamma(1.32, 1)$	$\Gamma(1.18, 46.0)$	$\Gamma(0.97, 55.3)$	$\Gamma(1.08, 35.8)$
p value	0.78	0.89	0.97	0.91	0.60
下跌:					
分布	$\Gamma(1.07, 19.9)$	$\Gamma(1.35, 3)$	$\Gamma(0.88, 39.8)$	$\Gamma(1.13, 56.4)$	$\Gamma(0.81, 48.9)$
p value	0.88	0.93	0.95	0.97	0.99

从表 1 可发现用伽玛 ( $\Gamma$ ) 分布可以很好地拟合连涨的收益率  $R_1$  与连跌的收益率  $R_0$  的分布.  $\Gamma$  分布的概率密度函数为

$$f(x) = \frac{\lambda}{\Gamma(\gamma)} (\lambda x)^{\gamma-1} e^{-\lambda x}, x > 0, \gamma > 0, \lambda > 0$$

记为  $\Gamma(\gamma, \lambda)$ , 其中  $\gamma$  为形状参数,  $\lambda$  为尺度参数.

关于  $\Gamma$  分布的矩估计很容易得到, 但众所周知, 极大似然估计 (MLE) 优于矩估计, 对中小样本, 矩估计的效果并不明显比 MLE 差, 但在大样本场合, 矩估计不如 MLE. 在实际应用中, 在上述 MLE 的基础上可以微调参数值, 使与实际数据拟合得更好.

另外, 由表 1 可发现  $X$  和  $Y$  的分布在实行“T+1”前后和实行“涨停板”前后以及国有股减持前后不一样, 即它们在“T+0”时期<sup>②</sup>(1992年 5 月 21 日—1994 年 12 月 30 日), “T+1”时期(1995 年 1 月 1 日—1996 年 12 月 15 日), “涨停板”时期(1996 年 12 月 16 日—2001 年 6 月 30 日)及“国有股减持”时期(2001 年 6 月 30 日—2005 年 11 月 10 日)是不同的, 这反映出股市的政策效应, 同时对近期(2005 年 11 月 11 日—2007 年 4 月 11 日)牛市进行了拟合, 发现上涨和下跌的分布是不一样的, 这从一个侧面表明收益率是非对称的。

### 1.2 股指涨跌的概率与条件概率

根据股指连涨连跌指益率的概率密度函数, 可以得到:

表 2 “涨停板”时期连涨收益率已知时继续上涨的条件概率

Table 2 Conditional probability of the successive return of rises during the rise limit periods

$P(Z = 1   R \geq 0.001)$	$P(Z = 1   R \geq 0.01)$	$P(Z = 1   R \geq 0.02)$	$P(Z = 1   R \geq 0.03)$	$P(Z = 1   R \geq 0.04)$
0.559	0.551	0.551	0.545	0.547
$P(Z = 1   R \geq 0.06)$	$P(Z = 1   R \geq 0.07)$	$P(Z = 1   R \geq 0.08)$	$P(Z = 1   R \geq 0.10)$	
0.537	0.455	0.333	0.143	

可见, 1) 当连涨收益率  $R_1 \geq 0.07$  后, 其继续上涨的条件概率迅速明显地下降, 这是因为此时连涨收益率已经落在极(大)值的范围内, 其继续上涨的条件概率将下降。在后文将对此做进一步的解释; 2) 当连涨收益率落在非极(大)值的范围内, 仔细观察的话, 其继续上涨的条件概率也有微小的变化, 即小涨之后继续上涨的条件概率稍微大一点。这一点并非偶然, 将继续在后文做进一步的解释。

由此看出, 股指连涨收益率的多少, 即历史价格对其之后继续涨跌的概率是有影响的。这一结论同样适用于连跌时的情况。进一步地, 从下面的分析可以看出, 股指连涨(或连跌)收益率的多少, 对其之后上涨(或下跌)落在某一范围内的(条件)概率也是有影响的。

从概率的知识得知

$$P(AB | C) = P(A | BC)P(B | C)$$

事件  $C$  可以表示为  $R_1 \geq r_0$  (或  $R_0 \geq r_0$ ), 事件  $B$  可以表示为  $Z = 1$  (或  $Z = 0$ ), 事件  $A$  可以表示为  $r_0 \leq R_1 \leq r_0 + \Delta r$  (或  $r_0 \leq R_0 \leq r_0 + \Delta r$ )。

$$F(r) = P(X \leq r) = \int_0^r f(x) dx$$

$$\text{或者 } P(X > r) = 1 - F(r)$$

从而知道股指连涨(或连跌)小于若干点的概率或股指连涨(或连跌)大于多少点的概率。同时, 更为有意义的, 还可以得到股指涨跌的条件概率。一共有四种情况, 即:

连涨之后继续涨, 连跌之后继续跌, 连涨之后转为跌, 连跌之后转为涨。先考虑前两种情况。

首先考虑一个简单的问题: 已知连涨收益率为一定的情况下继续上涨的条件概率为多少, 是否为一个常数? 设已知连涨收益率为  $R_1 \geq r_0$ , 上涨表示为  $Z = 1$  下跌表示为  $Z = 0$  即求条件概率  $P(Z = 1 | R_1 \geq r_0)$ 。可以统计这个条件概率, 在“涨停板”时期, 可以得到的结果如表 2

可见, 已知连涨收益率为一定的情况下, 如果继续上涨, 其落在某一范围内的条件概率可分为两部分的乘积, 一部分为连涨收益率为一定的情况下继续上涨的条件概率, 另一部分为在此条件下落在某一范围内的条件概率。从前面的表 2 可以看出, 在非极(大)值的情况下, 已经连续上涨了多少对其后继续上涨的条件概率影响的结果不大, 因此可以近似认为  $P(B | C) \approx P(B)$ , 这样可以得到以下几种情形:

一种情形是连涨之后继续涨

$$\begin{aligned} &P(r_0 \leq R_1 < r_0 + \Delta r, Z = 1 | R_1 \geq r_0) \\ &= P(r_0 \leq R_1 < r_0 + \Delta r | R_1 \geq r_0, Z = 1) \times \\ &P(Z = 1 | R_1 \geq r_0) \\ &\approx P(r_0 \leq R_1 < r_0 + \Delta r | R_1 \geq r_0)P(Z = 1 | \\ &R_1 \geq r_0, Z = 1) \text{ 简记为 } R_1 \geq r_0, \text{ 暗含继续上涨} \\ &\text{这一条件。} \end{aligned}$$

<sup>②</sup> 以股市一项政策的出台作为某一时期起始时间并以之命名。

另一种情形是连跌之后继续跌:

$$\begin{aligned}
& P(r_0 \leq R_0 < r_0 + \Delta r, Z = 0 | R_0 \geq r_0) \\
&= P(r_0 \leq R_0 < r_0 + \Delta r | R_0 \geq r_0, Z = 0) \times \\
& \quad P(Z = 0 | R_0 \geq r_0) \\
&\approx P(r_0 \leq R_0 < r_0 + \Delta r | R_0 \geq r_0)P(Z = 0)
\end{aligned}$$

以上  $R_1$  表示连涨的收益率,  $R_0$  表示连跌的收益率.

第二部分的概率  $P(Z = 1)$  和  $P(Z = 0)$  可以从实际数据中简单地估出来, 即  $P(Z = 1)$  等于这一时期上涨的天数除以总天数,  $P(Z = 0)$  等于这一时期下跌的天数除以总天数. 在下文中重点考虑前一部分的条件概率.

在“涨停板”时期, 得到这两种情况下的条件概率如表 3

表 3 连涨和连跌的两种条件概率

Table 3 Two conditional probability of the successive rises and falls

$P(r_0 \leq R_1 < r_0 + \Delta r   R_1 \geq r_0)$	$\Delta r = 0.01$	$\Delta r = 0.02$	$\Delta r = 0.03$
$r_0 = 0.001$	0.298 8	0.530 5	0.690 6
$r_0 = 0.01$	0.328 8	0.557 2	0.710 6
$r_0 = 0.02$	0.340 3	0.568 8	0.719 8
$r_0 = 0.05$	0.352 9	0.582 6	0.731 4
$P(r_0 \leq R_0 < r_0 + \Delta r   R_0 \geq r_0)$	$\Delta r = 0.01$	$\Delta r = 0.02$	$\Delta r = 0.03$
$r = 0.001$	0.378 3	0.598 7	0.738 1
$r = 0.01$	0.355 7	0.579 7	0.724 1
$r = 0.02$	0.347 7	0.571 8	0.717 8
$r = 0.05$	0.339 2	0.562 4	0.709 8

由上表可知, 在这一时期, 当已知股指连涨收益率大于等于 0.001 时, 如果继续上涨, 那么在区间 0.001 至 0.011 之间转为下跌的条件概率为 0.298 8 而当已知股指连涨收益率大于等于 0.01 时, 如果继续上涨, 那么在区间 0.01 至 0.02 之间转为下跌的条件概率上升为 0.328 8 而当已知股指连涨收益率大于等于 0.05 时, 如果继续上涨, 那么在区间 0.05 至 0.06 之间转为下跌的条件概率上升为 0.352 9 此时比无条件概率即  $P(0 \leq R_1 < 0.01)$  的值 0.288 5 高出约 6 个百分点. 可见已知股指连涨了多少对推断未来继续上涨多少的条件概率是有些影响的, 它与无条件概率有些差别.

很容易得到已知股指连涨收益率大于  $r$ , 之后继续上涨超过  $r + \Delta r$  的概率.

类似地, 可知当股市处于这一时期时, 当已知股指连跌收益率大于等于 0.001 时, 如果继续下跌, 那么在区间 0.001 至 0.011 之间转为上涨的条件概率为 0.378 3 而当已知股指连跌收益率大于等于 0.01 时, 如果继续下跌, 那么在区间 0.01 至 0.02 之间转为下跌的条件概率下降为 0.355 7 而当已知股指连跌收益率大于等于 0.05 时, 如果继

续下跌, 那么在区间 0.05 至 0.06 之间转为上涨的条件概率下降为 0.339 2 比无条件概率  $P(0 \leq R_0 < 0.01)$  的值 0.389 下降了约 5 个百分点. 可见已知股指连跌了多少同样对推断未来继续下跌多少的条件概率也是有些影响的, 也与无条件概率有些差别. 而为何其条件概率出现这样的差别, 在下文中会做出解释.

也可以很容易得到已知股指连跌收益率大于  $r$ , 之后继续下跌超过  $r + \Delta r$  的概率.

### 1.3 极值的情形

根据文献 [4] 中极值的一种定义方法, 即:

假设  $Y_i$  为第  $i$  组观察值的极大值

$$Y_i = \max\{X_{i1}, \dots, X_{im}\}, \quad i = 1, \dots, n$$

从理论上推导出连涨和连跌收益率的极值的危险率曲线 (其定义如下文) 是上翘的.

这里运用极值定义的另一种方法, 因为前一种定义方法在运用中没有此种方便. 记  $N(u)$  为观察值中超过某一个值  $u$  的个数, 极值:

$$\{Y_1, \dots, Y_{N(u)}\} = \{X^{(1)} - u, \dots, X^{(N(u))} - u\}$$

其分布函数定义为

$$F_u(x) = P(X - u \leq x | X > u)$$

由 Pickands, Balkema 和 de Haan 证明了

定理:

$$F_u(X) \approx W_{\xi\beta}(x)$$

$$W_{\xi\beta}(x) = 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}$$

$$\begin{cases} x \geq 0, \xi > 0 \\ 0 \leq x \leq -\beta/\xi, \xi < 0 \end{cases}$$

$W_{\xi\beta}(x)$  称为广义帕累托分布 (GPD).

当“涨停板”时期时, 得到的实证结果为:

连涨的收益率当  $u = 0.06$  时, 极值分布即广义帕累托分布的参数为

$$\xi = -0.4915, \beta = 0.0448$$

连跌的收益率当  $u = 0.08$  时, 极值分布即广义帕累托分布的参数为

$$\xi = -0.1047, \beta = 0.0375$$

计算条件概率:

$$P(r_0 \leq R < r_0 + \Delta r \mid R \geq r_0)$$

其中  $R$  表示连涨或连跌的收益率.

得到的结果如表 4 (表中  $r_0$  的值已减去  $u$ )

表 4 连涨连跌收益率处于极值时的条件概率

Table 4 Conditional probability of the successive return of rises and falls at the extreme value

$P(r_0 \leq R < r_0 + \Delta r \mid R \geq r_0)$	连涨	连跌
$r_0 = 0.001, \Delta r = 0.01$	0.2128	0.2376
$r_0 = 0.01, \Delta r = 0.01$	0.2348	0.2430
$r_0 = 0.02, \Delta r = 0.01$	0.2652	0.2493
$r_0 = 0.03, \Delta r = 0.01$	0.3046	0.2559
$r_0 = 0.05, \Delta r = 0.01$	0.4325	0.2703

可见当股指大涨大跌, 落在极大值范围时, 其在之后某一区间内转为下跌或上涨的概率大大

增加, (当然此时连涨之后继续上涨的条件概率或连跌之后继续下跌的条件概率大大下降, 不能再视为一个常数). 这是因为大涨之后, 获利回吐会增多, 从而抛盘增加使得转为下跌的概率加大; 大跌之后, 逢低吸纳会增多, 从而使得买盘增加使得转为上涨的概率加大. 这与人们的炒股经验相符, 而这里给出了实证的证据, 并从理论上给出了解释

### 1.4 对危险率函数曲线的分析

危险率函数 (Hazard function, 或称失效率函数)  $h(r)$  简单地说就是条件生存率, 它是生存分析中的一个重要的函数. 人的寿命的危险率函数曲线如浴盆状, 即中间低两头高, 表明人处于幼年 and 老年时死亡的概率较大, 而青壮年时死亡的概率较小. 应用到股市中, 粗略地说, 危险率函数  $h(r)$  表示当股指连涨 (或连跌) 的收益率为  $r$  时, 在往后的单位收益率区间内下跌 (或上涨) 的条件概率, 它表明股指在上涨 (或下跌) 到多少点时下跌 (或上涨) 的可能性大小. 这一指标对于分析和预测股指具有重要的应用价值, 也与前面的结论相关.

连涨的收益率  $R_1$  的危险率函数定义为:

$$h(r) = \lim_{\Delta r \rightarrow 0} \frac{P(r \leq R_1 < r + \Delta r \mid R_1 \geq r)}{\Delta r}$$

而前面得到的条件概率即为:

$$P(r \leq R_1 < r + \Delta r \mid R_1 \geq r) \approx h(r) \times \Delta r$$

该式在  $\Delta r$  较小时近似成立.

在“涨停板”时期, 沪市处于一个牛市时期, 上涨和下跌的理论危险率曲线如图 1

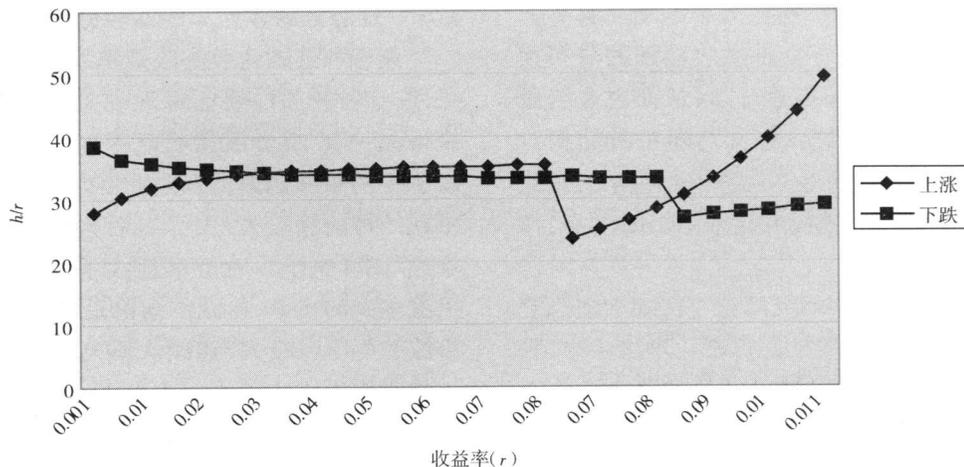


图 1 “涨停板”时期上涨和下跌的危险率曲线

Fig. 1 Hazard curve during the periods of rise limit

从上图可以看出,在“涨停板”时期,即牛市中,当前一天的股指上涨了一点点时,第二天下跌的可能性小一点,因为这一段的危险率曲线是向上弯曲的;当股指上涨较大时比上涨较小时转为下跌的可能性大一点。为什么会这样呢?因为在牛市中,人们对股市看好,当股指上涨一点点时,人们继续看涨,因此股指转为下跌的可能性小一点,而当股指上涨较多时,获利抛盘就会增多,从而使下跌的概率增大。下跌的情况正好相反,牛市时,当股指下跌一点点时,第二天转为上涨的可能性大一点,因为股市看涨;当股指下跌较多时,转为上涨的概率小一点,因为此时人们对股指的预期也发生了变化。当连涨连跌的股指收益率处于极大值时,无论是上涨还是下跌,其危险率曲线都是上翘的<sup>⑤</sup>。表明当股指大涨或大跌时,其转为下跌或上涨的概率要大一点,这与人们的炒股经验相吻合。前文所得条件概率的结果也可以在这里得到解释。通过上述分析,可以根据以前股指的涨跌情况对明天的涨跌作一个概率推断,这种预测比靠单纯抛硬币来决定股指的涨跌要来得好一点,这对于投资者合理把握买卖的时机是有参考价值的。同时也看出这又比单纯在牛市中预测涨,熊市中预测跌来到精细一些。

熊市时的危险率曲线正好相反,这里就不作赘述了。

值得一提的是,有时候股指连涨连跌收益率的分布服从指数分布,其危险率曲线是一条水平直线,表明股指下一期的涨跌与前期收益率的大小没有关系(即独立的),指数分布是唯一具有无记忆性的分布,而股指连涨连跌收益率的分布在大多数情况下并不服从指数分布,从而也表明股市是有“记忆”的。指数分布应与随机游动相对应。粗糙地看,可以认为连涨连跌的收益率为指数分布,涨跌之间是独立的,也是无记忆的,但是精确地看,却并不如此。

由此也可以看出本研究结论与传统的随机游动理论不完全一致。传统的随机游动理论认为,股票价格的变化是无法预测的。因为股票价格的变动,是一种偶然发生的任意行为,在统计上是彼此

独立的,前后没有任何联系;股票价格没有记忆,过去的价格不会重复,以后股票价格的变动不会取决于股票价格以前的变动,所以,股票价格的变动趋势是无法预测的。而从本研究可以看出,股指收益率的危险率曲线是弯曲的,也就是说其条件概率与收益率的大小有关,而不是独立的。

从前面的分析也能看出,在不同时期,不同的政策条件下,在不同的市场里,股指涨跌的危险率曲线以及其它生存函数曲线会有所变化,适时分析这种变化,既可以分析政策的影响,也可以发现市场的变化,这正是对股指进行生存分析的意义所在。由此可以看出,对连涨连跌的收益率的分布作变点研究,是极有价值的。由于连涨连跌的收益率服从伽玛分布,故可以通过分布变点的理论对其进行研究,可发现伽玛分布的两个参数可以很好地表示股指涨跌趋势的变化和波动的变化,对其进行实证研究,详细内容可参见文献[5]。通过对变点的研究可见,表示股市趋势变化的参数 $r$ 在股市处于某一趋势时期如牛市(或熊市)时期内其取值相对稳定,另一参数 $\lambda$ 当股市波动发生明显变化时其取值才发生明显的变化,因此,当股市处于某一趋势而波动又未发生明显变化时可以以一个相对稳定的分布对股指的涨跌做出概率推断,而当发生变点的变化时分布的参数需重新估计,这也是本文进行统计推断的一个基础。

在前面已经讨论了连涨之后继续涨和连跌之后继续跌的情况,下面来讨论连涨之后转为跌和连跌之后转为涨的情况。

### 1.5 连涨连跌收益率的联合分布与概率推断

连涨和连跌的股指收益率之间具有一定的相关性,这种相关性要比每日收益率之间的相关性来得大一点。统计股指收益率连涨了多少以后连跌多少的条件概率,可以得出的结论是小涨以后小跌的可能性较大,大涨以后大跌的可能性相对增大。同样也统计收益率连续下跌多少以后连续上涨多少的概率,得到的结论是:连续小跌之后有小涨的可能性较大,连续大跌之后有大涨的可能性相对增大。

用二元Gamma函数来进行拟合,得到在“涨

<sup>⑤</sup>当连涨连跌的收益率处于极值与非极值时,其危险率曲线不连续,这是因为极值的分布与非极值的分布是不一样的。

停板”期间, 连涨连跌收益率的联合分布, 其具体方法见文献 [4], 在这里只引用其结果, 与文献 [4] 中有所不同的是, 经过重新计算可得到这一时期连涨连跌收益率之间的相关系数为 0.3

连涨连跌收益率的理论联合分布  $\tilde{F}(R_b, R_0)$  为

$$\begin{aligned} \tilde{F}(R_b, R_0) &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha)} \times \\ &\int\int\int_{\substack{x+z < \lambda_1 R_1 \\ y+z < \lambda_2 R_0}} x^{\alpha_1-1} e^{-x} y^{\alpha_2-1} e^{-y} z^{\alpha-1} e^{-z} dx dy dz \\ &= \frac{1}{\Gamma(0.88)\Gamma(0.58)\Gamma(0.3)} \times \\ &\int\int\int_{\substack{x+z < 46R_1 \\ y+z < 39.8R_0}} x^{0.88-1} e^{-x} y^{0.58-1} e^{-y} z^{0.3-1} e^{-z} dx dy dz \end{aligned}$$

计算条件概率

$$\begin{aligned} P(R_0 \leq b | R_1 \leq a) &= \frac{P(R_1 \leq a, R_0 \leq b)}{P(R_1 \leq a)} \\ &= \frac{\tilde{F}(R_b, R_0)}{F(R_1)} \end{aligned}$$

可以得出的结果如表 5

表 5 连涨后转为下跌的条件概率

Table 5 Conditional probability after rises to falls

	$P(R_0 \leq b   R_1 \leq a)$
$a = 0.005 \quad b = 0.01$	0.549
$a = 0.01 \quad b = 0.01$	0.523
$a = 0.03 \quad b = 0.01$	0.440
$a = 0.05 \quad b = 0.01$	0.410

可见, 随着连涨收益率幅度的加大, 之后的连跌收益率落在小幅度范围的概率在减小. 同样也可以得到连跌之后转为上涨的条件概率.

这样就完全讨论了股指涨跌的几种情况, 可以看出, 无论是连涨之后继续涨, 连跌之后继续跌, 还是连涨之后转为跌, 连跌之后转为涨, 它们之间互相转换的条件概率都是与历史收益率的大小有些关系的, 而非独立的, 它们之间呈现一种复杂的相关关系. 也就是说, 单就已经连涨连跌了多少这一历史价格信息来看, 也并不是对未来短期的股指涨跌没有影响. 下面还将分析作为历史信息中的另一个重要信息成交量对股指涨跌的影响.

## 2 考虑成交量等因素对股指未来涨跌的影响

成交量数据作为一种市场交易中供需情况的具体表现, 可能包含有用的信息能用于对其它感兴趣的金融变量 (例如资产价格、收益率或波动率) 的建模和预测. 许多学者对此进行了研究, 如 Sticckel 与 Verrecchia<sup>[16]</sup> 研究了美国股市成交量对价格的影响; Sivapulle 与 Choi<sup>[15]</sup> 研究了韩国股票市场 KOCSP I 指数收益率与成交量的引导关系; 张维、闫冀楠<sup>[8]</sup> 研究了上证综合指数周收益率与成交量的变动率的引导关系; 陈怡玲、宋逢明<sup>[1]</sup> 对中国股市的价格变动与成交量的关系进行了多层次的实证研究. 他们的研究表明, 股市的日价格变动的绝对量与日交易量呈线性正相关. 本文研究也表明连涨连跌的收益率与成交量之间也具有正相关的关系, 并进一步给出了成交量对股指未来涨跌的概率影响的具体值.

把连涨或连跌时的成交量  $V$ , 即连涨或连跌的每日成交量之和, 作为协变量来研究连涨连跌的股指与成交量之间的关系. 其研究具体见文献 [4], 这里将其推导的过程引用如下, 以便清楚其由来.

假如  $X$  的分布是伽玛分布, 即

$$f(r) = \lambda(\lambda r)^{\gamma-1} e^{-\lambda r} \Gamma(\gamma)$$

那么  $\log(X) = -\log\lambda + \log(\lambda X)$

其中  $\log(\lambda X)$  称为标准对数伽玛分布, 其密度函数为

$$\begin{aligned} e &= \frac{1}{\Gamma(\gamma)} \exp(\gamma W - \exp(W)), \\ &-\infty < W < \infty \end{aligned}$$

一般假设尺度参数  $\lambda$  受协变量即成交量  $v$  的影响, 并且假设是一种线性关系, 而形状参数  $\gamma$  与协变量没有多大影响.

$$\text{即 } Z = \log(X) = -\log\lambda(v) + e$$

$$\text{进一步假设为 } Z = \alpha + \beta v + e$$

可以用极大似然估计和最小二乘法两种方法得到参数值, 但是极大似然法中不容易求解极大似然方程, 而最小二乘法可以很容易得到参数值. 由于这里使用了回归模型, 于是提供了一个分析

框架,即只考虑了成交量对连涨连跌收益率的影响,在Fama的著名三因素模型或其他人的多因素模型中考虑的变量,如市盈率等,甚至包括一些技术指标,其实也可引入这个框架内进行分析.在“涨停板”时期,可以得到连涨情况下:  $\alpha = -4.6$   $\beta = 0.023$  且都是非常显著的,其中成交量的单位为百万手.连跌情况下:  $\alpha = -4.9$   $\beta = 0.028$

且都是非常显著的.

这样在“涨停板”时期,带有协变量即成交量V的连涨收益率  $R_1$  的分布为

$$f_v(r) = \exp(4.6 - 0.023v)^k r^{k-1} \times \exp(-\exp(4.6 - 0.023v)r) / \Gamma(k).$$

可以计算出上证指数在“涨停板”时期不同成交量下连涨收益率的概率如表 6

表 6 不同成交量下连涨收益率在不同区间的概率

Table 6 Probability of the successive return of rises with the different volume at the different interval

成交量 v(百万手)	连涨收益 率 $R_1$	概率 $p$	连涨收益 率 $R_1$	概率 $p$	连涨收益 率 $R_1$	概率 $p$
1	(0 0 01)	0.544	(0 01 0 05)	0.444	(0 05, +∞)	0.012
5	(0 0 01)	0.508	(0 01 0 05)	0.475	(0 05, +∞)	0.0174
10	(0 0 01)	0.464	(0 01 0 05)	0.508	(0 05, +∞)	0.0277
30	(0 0 01)	0.311	(0 01 0 05)	0.577	(0 05, +∞)	0.111
50	(0 0 01)	0.199	(0 01 0 05)	0.538	(0 05, +∞)	0.263
100	(0 0 01)	0.057	(0 01 0 05)	0.254	(0 05, +∞)	0.689

可见,成交量确实对上涨的多少(即寿命)产生影响,即成交量较大时,连涨收益率落在较大值范围内的概率也较大.

更重要的讨论是其条件概率.一种情况是,连涨以后继续涨(或连跌以后继续跌).

即

$$P(r_0 \leq R_1 < r_0 + \Delta r | v = v_0, R_1 \geq r_0) = \frac{P(v = v_0, r_0 \leq R_1 < r_0 + \Delta r)}{P(v = v_0, R_1 \geq r_0)}$$

表 7 不同成交量、不同连涨收益率时的条件概率

Table 7 Conditional probability of the successive return of rises with different volume

成交量 v(百万手)	$r_0 = 0.001, \Delta r = 0.01$	$r_0 = 0.01, \Delta r = 0.01$	$r_0 = 0.02, \Delta r = 0.01$	$r_0 = 0.05, \Delta r = 0.01$
1	$P = 0.558$	$P = 0.590$	$P = 0.601$	$P = 0.611$
5	$P = 0.521$	$P = 0.555$	$P = 0.566$	$P = 0.576$
10	$P = 0.477$	$P = 0.511$	$P = 0.522$	$P = 0.534$
30	$P = 0.322$	$P = 0.353$	$P = 0.365$	$P = 0.377$
50	$P = 0.207$	$P = 0.231$	$P = 0.242$	$P = 0.254$
100	$P = 0.06$	$P = 0.070$	$P = 0.075$	$P = 0.081$

另一种情况是,连涨以后转为下跌(或连跌以后转为上涨).在这种情况下,将连涨连跌收益率的联合分布中的参数  $\lambda$  由固定的变为成交量的变量就可以了,其中  $\lambda$  与成交量的关系已由上述给出.即:

可以计算出“涨停板”时期,不同成交量、不同连涨收益率情况下的条件概率如表 7

可见,成交量的多少对条件概率的大小影响很大,这也表明已知股指连续涨跌了多少与成交量多少这些历史信息对未来股指涨跌的概率是有影响的,对推断其涨跌的概率是有帮助的.

$$\tilde{F}(R_b, R_0) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha)} \times \iiint_{\substack{x+z < \lambda_1^{(v)} R_1 \\ y+z < \lambda_2^{(v)} R_0}} x^{\alpha_1-1} e^{-x} y^{\alpha_2-1} e^{-y} z^{\alpha-1} e^{-z} dx dy dz$$

$$\lambda_1^{(v)} = \exp(4.6 - 0.023v)$$

$$\lambda_2^{(v)} = \exp(4.9 - 0.028v)$$

可得出在“涨停板”时期不同成交量时连涨连跌的条件概率如表 8

由表可见, 连涨之后转为下跌时, 如果连涨时的成交量大, 则其下跌落在小范围内的概率增大. 这种关系一般没有被人注意到.

同样可以得到以一定的置信水平, 已知连涨或连跌收益率以及成交量的情况下, 后市涨跌落在的区间. 如以 90% 的置信水平, 在“涨停板”时期, 已知连涨收益率 (非极值时) 以及成交量的情

表 9 以 90% 的置信水平, 已知连涨收益率及成交量的情况下, 后市涨跌  $\Delta r$  的区间

Table 9 Interval of the future rises and falls with 90% confidence level with the successive returns of rises and volume

连涨收益率 $R_1$	成交量 (百万手)	后市涨跌区间 $\Delta r$	连涨收益率 $R_1$	成交量 (百万手)	后市涨跌区间 $\Delta r$
0.001	1	(-0.039, 0.026)	0.01	1	(-0.041, 0.025)
0.001	10	(-0.037, 0.032)	0.01	10	(-0.040, 0.031)
0.001	30	(-0.039, 0.052)	0.01	30	(-0.040, 0.050)
0.001	50	(-0.038, 0.082)	0.01	50	(-0.039, 0.080)
0.02	1	(-0.045, 0.025)	0.05	1	(-0.052, 0.024)
0.02	10	(-0.043, 0.031)	0.05	10	(-0.050, 0.030)
0.02	30	(-0.042, 0.049)	0.05	30	(-0.048, 0.048)
0.02	50	(-0.041, 0.078)	0.05	50	(-0.045, 0.076)

由上表可知, 历史价格与成交量对未来涨跌的置信区间均是有影响的, 其影响的机理如前所述. 这一结果无疑具有重要的投资参考价值.

### 3 结束语

从上面的分析中可以看出:

(1) 股指涨跌具有一定的趋势, 即牛市和熊市交替出现, 特别是在中国的股市, 这种现象尤其明显. 当已知目前处于何种趋势时, 股指涨跌了多少对后市有一定的影响, 这其实也是技术分析的一个重要假设.

(2) 当出现连续大涨大跌时, 由于人们的心理和投机行为, 也使得其后下跌或上涨的概率增大.

(3) 连涨连跌之间具有一定的相关性, 从而也使得彼此涨跌的概率不独立.

(4) 成交量不仅对当期的股指收益率有影

况下, 后市涨跌  $\Delta r$  的区间如表 9

表 8 不同成交量下连涨后转为下跌的条件概率

Table 8 Conditional probability with different volume after rises to falls

V (百万手)	$P(R_0 \leq b \mid R_1 \leq a)$ $a = 0.01, b = 0.01$
V = 1	0.4652
V = 5	0.4720
V = 10	0.4815
V = 30	0.5164
V = 50	0.5395
V = 100	0.5619

表 9 以 90% 的置信水平, 已知连涨收益率及成交量的情况下, 后市涨跌  $\Delta r$  的区间

Table 9 Interval of the future rises and falls with 90% confidence level with the successive returns of rises and volume

连涨收益率 $R_1$	成交量 (百万手)	后市涨跌区间 $\Delta r$	连涨收益率 $R_1$	成交量 (百万手)	后市涨跌区间 $\Delta r$
0.001	1	(-0.039, 0.026)	0.01	1	(-0.041, 0.025)
0.001	10	(-0.037, 0.032)	0.01	10	(-0.040, 0.031)
0.001	30	(-0.039, 0.052)	0.01	30	(-0.040, 0.050)
0.001	50	(-0.038, 0.082)	0.01	50	(-0.039, 0.080)
0.02	1	(-0.045, 0.025)	0.05	1	(-0.052, 0.024)
0.02	10	(-0.043, 0.031)	0.05	10	(-0.050, 0.030)
0.02	30	(-0.042, 0.049)	0.05	30	(-0.048, 0.048)
0.02	50	(-0.041, 0.078)	0.05	50	(-0.045, 0.076)

响, 而且对之后涨跌幅度的概率也有影响.

在本文中, 不自觉地技术分析找到了某种统计理论的支持. 本研究恰巧说明了为何一方面市场有效性的理论极具影响, 而同时技术分析特别是价量分析在实际中仍被广泛运用, 是有其一定道理的.

由此得到的政策建议如下: 1) 市场有效与否应更加关注有关信息是否能及时披露, 从而利用内幕消息获取超额利润的情形不能频频发生; 2) 市场是否有效要看市场不能被操纵, 从而不能为某些人借以牟取利益, 因此扩大市场规模是有益的; 3) 股市应保持一定比例 (无风险收益加上一定的风险溢价) 的缓慢增长, 不应呈现明显的牛市和熊市的周期, 因为这样离有效市场会偏离得更远. 4) 我国资本市场之所以呈现出明显的趋势性, 即周期性, 与我国资本市场缺乏卖空 (即卖出自己并不拥有的股票) 等机制不无关系, 进行融资融券交易是有必要的, 这样可以使股票价格迅

速地回归到一个合理的价格,避免股价大涨大跌。

本文研究的作用和意义在于:

1) 可以对股指涨跌作出概率推断。

可以知道股指连涨(或连跌)大于多少点的概率;可以知道股指连涨(或连跌)小于若干点的概率;更为重要的,可以得出股指涨跌的条件概率。已经涨跌了多少对之后的涨跌概率是有些影响的,特别是当涨跌幅度处于极大值时;成交量的大小不仅影响当期的涨跌幅度,而且对后期涨跌幅度的概率也是有影响的,从而可以得到在已有历史信息集的情况下股指未来涨跌的一个置信区间。这对投资者进行投资具有重要参考价值。

同样也可以用这种方法对个股进行分析,对高频数据如时数据、分数据,或低频数据如周数据、月数据进行分析。其实本文的研究方法可以引入对很多经济数据的研究,因为很多经济数据都可以从某个角度上看作是“寿命”数据。

2) 生存分析有助于分析股市的政策效应及股市的变化,从而这种分析方法能够对制订宏观政策提供帮助。通过此种方法,还可以探讨股市的一些基本性问题,如随机游动和市场有效性问题。据此对市场有效性进行了一些深入的探讨。

注:本文所用软件主要为 S-plus & Eviews  
Matlab Mathematica

## 参考文献:

- [1] 陈怡玲, 宋逢明. 中国股市价格变动与交易量关系的实证研究[J]. 管理科学学报, 2000, 3(2): 62-68  
Chen Yiling Song Fengming An empirical study on the relationship between price changes and trading volume in China stock market[J]. Journal of Management Sciences in China, 2000, 3(2): 62-68 (in Chinese)
- [2] 戴洁, 武康平. 中国股票市场技术分析预测力的实证研究[J]. 数量经济技术经济研究, 2002 (4): 99-102  
Dai Jie Wu Kangping An empirical study on the Chinese stock market on the ability of forecasting with technical analysis[J]. Quantitative Economics and Technical Economics Research, 2002 (4): 99-102 (in Chinese)
- [3] 雷鸣, 缪柏其, 宁静. 运用生存模型对上证指数与成交量的研究[J]. 数理统计与管理, 2003, 22(6): 46-50  
Lei Ming Miao Baiqi Ning Jing Study of Shanghai stock index and the policies effect with survival analysis[J]. Application of Statistics and Management, 2003, 22(6): 46-50 (in Chinese)
- [4] 雷鸣, 缪柏其. 运用生存分析与极值理论对上证指数的研究[J]. 数量经济技术经济研究, 2004 (11): 130-137  
Lei Ming Miao Baiqi Study of the Shanghai stock index with survival analysis and extremum theory[J]. Quantitative Economics and Technical Economics Research, 2004 (11): 130-137 (in Chinese)
- [5] 雷鸣, 谭常春, 缪柏其. 运用生存分析与变点理论对上证指数的研究[J]. 中国管理科学, 2007, 15(5): 1-8  
Lei Ming Tan Changchun Miao Baiqi Study of Shanghai stock index with survival analysis and change point theory[J]. Chinese Journal of Management Science, 2007, 15(5): 1-8 (in Chinese)
- [6] 吴世农. 我国证券市场效率的分析[J]. 经济研究, 1996 (4): 13-19  
Wu Shinong Analysis of the efficiency of Chinese security market[J]. Economic Research, 1996 (4): 13-19 (in Chinese)
- [7] 张兵, 李晓明. 中国股票市场的渐进有效性研究[J]. 经济研究, 2003 (1): 54-61  
Zhang Bing Li Xiaoming Study of the evolving efficiency of Chinese stock market[J]. Economic Research, 2003 (1): 54-61 (in Chinese)
- [8] 张维, 闫冀楠. 关于上海股市量价因果关系的实证探测[J]. 系统工程理论与实践, 1998, 18(6): 111-114  
Zhang Wei Yan Jinnan Empirical detection to the volume-price causality in Shanghai securities exchange[J]. System Engineering Theory & Practice, 1998, 18(6): 111-114 (in Chinese)
- [9] Andrew W. Lo Harry Mamaysky Jiang Wang Foundations of Technical Analysis Computational Algorithms Statistical Inference and Empirical Implementation[J]. Journal of Finance, 2000, 55(4): 1705-1765
- [10] Fama E F Efficient Capital Markets A Review of Theory and Empirical Work[J]. Journal of Finance, 1970, 25(2): 383-417
- [11] Fama E F Efficient Capital Markets II[J]. Journal of Finance, 1991, 46(5): 1575-1617
- [12] Treynor J L Ferguson R. In Defense of Technical Analysis[J]. Journal of Finance, 1985, 40(3): 757-773
- [13] Blume M, Easley D, O'Hara M. Market statistics and technical analysis: The role of volume[J]. Journal of Finance, 1994, 49(1): 153-181

(下转第 75 页)

[27] 李建武, 李敏强. 测试实数编码遗传算法的困难度 [J]. 管理科学学报, 2004, 7(4): 40-44

Li Jianwu, Li Minqiang. Analyzing GA-hardness for real encoding in genetic algorithms [J]. Journal of Management Sciences in China, 2004, 7(4): 40-44. (in Chinese)

## Model calibration of HJM models based on UKF with application

YANG Bao-chen, SU Yun-peng

School of Management, Tianjin University, Tianjin 300072, China

**Abstract** In a multifactor HJM framework, this paper transforms a class of non-Markovian forward rate models with a specific volatility specification into the Markovian representation, which is further cast into a state-space model. Then a maximum likelihood estimator based on the unscented Kalman filter is introduced into the estimation of the term structure models, thus getting the problems of nonlinearity and the existence of latent variables resolved. For the empirical study, a three-factor HJM model is established for the Shanghai Interbank Offered Rate (SHIBOR) market by introducing stochastic market price of risk and a volatility specification appropriate for the market. It is found that the dynamics and volatility structure of SHIBOR are both well captured by the model, and the level and slope factors explain the majority of the variation of the yield curve.

**Key words** Heath-Jarrow-Morton models; unscented Kalman filter; maximum likelihood estimator; Shanghai Interbank Offered Rate

(上接第 66 页)

[14] Pierre Giot and Amin Schowenbacher. IPO, trade sales and liquidations: Modeling venture capital exits using survival analysis [J]. Journal of Banking & Finance, 2007, 31(3): 679-702

[15] Sivapulle P, Choi JS. Testing for linear and nonlinear Ganger causality in the stock price-volume relation: Korean evidence [J]. The Quarterly Review of Economic and Finance, 1999, 39(1): 59-76

[16] Stickle SE, Verrecchia RE. Evidence that trading volume sustains stock price changes [J]. Financial Analysts Journal, 1994, November/December: 57-67

## Survival analysis and the probability inference about the stock index

LEI Ming<sup>1</sup>, YE Wu-yi<sup>2</sup>, MAO Bai-qi<sup>2</sup>, GUO Wen-jing<sup>1</sup>

1. Nanjing University of Finance & Economics, Nanjing 210003, China

2. University of Science & Technology of China, Hefei 230026, China

**Abstract** In this paper, the successive rises and falls of Shanghai stock index returns are analyzed by survival analysis and extremum theory. The distributions of the successive rises and falls of returns are fitted with Gamma distribution. The conditional probability of the rises and falls of the stock index is obtained. Also the influence of the volume is analyzed. We think that the random walk hypothesis is only a rough approximation. And our study helps to better understand the efficient market theory.

**Key words** survival analysis; gamma distribution; probability inference; random walk