

关于沪深上市公司资产相关性系数的探讨

联合资信评估有限公司 李英雄

一、绪论

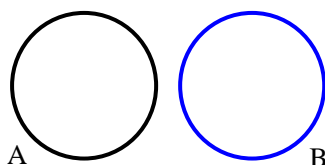
自从马科维茨(1957)提出组合概念以来,相关性系数就成为了学者和金融界人士关注的课题。目前没有公认的经典理论,所以在许多文献中都回避了相关性系数这个问题,将其设为外生变量。相关性系数分为资产相关性系数和违约相关性系数,我们先简单的看一下这两个种相关性系数的概念。

1.1 公司违约相关性

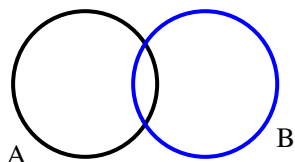
违约相关性是指一个公司对某笔负债的违约情况,与另一个公司对其负债是否违约的似然率。度量的是两个公司一起违约还是单独违约的可能性。不失一般性,我们以两个公司(记为A和B)为例来说明这个概念。违约相关性系数取值有三种可能性:1)系数为正,例如:A公司是B公司的债权人,即B违约加剧了A违约的可能性;2)为零;3)为负,例如:A和B为进行激烈竞争的公司,即A退出市场有利于B的发展。

A和B的违约情况存在四种可能性:1)A违约,B不违约;2)A不违约,B违约;3)A、B都违约;4)A和B都不违约。A和B的联合违约情况可以表示如下:

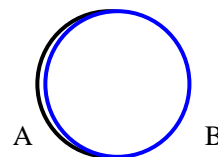
不存在A、B都违约



存在A、B都违约,但是概率较小



存在很大的A、B都违约概率



记A的违约概率为 $P(A)$, B的违约概率为 $P(B)$, A、B都违约的概率为

$P(AB)$, A和B的违约相关性系数为:

$$\rho'_{AB} = \frac{\text{Cov}(A, B)}{\sqrt{\text{Var}(A) \times \text{Var}(B)}}$$

其中, $\text{Var}(A) = \sqrt{P(A) \times (1 - P(A))}$, $\text{Var}(B) = \sqrt{P(B) \times (1 - P(B))}$

$\text{Cov}(A, B) = P(AB) - P(A) \times P(B)$ 。

1.2 公司资产相关性

在结构型违约模型中, 将公司的总资产价值波动视为随机过程, 计算这两个随机过程相关性系数, 该系数就是这两个公司的资产相关性系数。在实际应用中, 采用资产回报率代替资产价值。设A、B公司资产回报率为 $R_A = \{r_A(t)\}_0^T$ 和 $R_B = \{r_B(t)\}_0^T$, 那么A和B的资产相关性为:

$$\rho_{AB} = \frac{\text{Cov}(R_A, R_B)}{\sqrt{\text{Var}(R_A) \times \text{Var}(R_B)}}$$

其中, $\text{Var}(R_A) = \frac{1}{T} \sum_t (r_A(t) - \bar{r}_A)^2$, \bar{r}_A 是 $R_A = \{r_A(t)\}_0^T$ 的期望值; $\text{Var}(R_B)$ 与

$\text{Var}(R_A)$ 的定义相同; $\text{Cov}(A, B) = \frac{1}{T} \sum_t (r_A(t) - \bar{r}_A)(r_B(t) - \bar{r}_B)$ 。

1.3 公司资产相关性与违约相关性的关系

在结构型违约模型中, 有:

$$P(A) = P(r_A(T) < \tilde{r}_A), P(B) = P(r_B(T) < \tilde{r}_B)$$

由资产相关性的定义有

$$P(AB) = P(r_A(T) < \tilde{r}_A, r_B(T) < \tilde{r}_B)$$

$$= \int_0^{\tilde{r}_A} \int_0^{\tilde{r}_B} \phi(r_A(T) < \tilde{r}_A, r_B(T) < \tilde{r}_B, \rho_{AB}) d\tilde{r}_A(t) d\tilde{r}_B(t)。$$

二、相关性系数理论

本文以下部分只讨论资产相关性系数, 如无特别说明, 下文出现的相关性系数均指资产相关性系数。目前, 应用最广泛的相关性理论有历史理论(historic theory)、均值理论(mean theory)和因子模型理论(factor model theory)(KMV 公司称之为全球相关性系数模型 global correlation model)。

2.1 历史模型

应用公司过去的资产波动数据, 计算这些公司之间的相关性系数, 并以此作

为这些公司未来的相关性系数。该理论简洁直观，但是计算很繁琐，如果资产池有 n 个公司，我们需要计算 $n \times (n-1)$ 个系数。

2.2 均值模型

将公司按照地区、行业集团划入不同的资产分池，分池的相关性系数为分池内资产的借款公司的两两相关性系数的平均值或者加权平均值，分池之间的相关性系数为两个分池的资产之间的相关性系数的平均值或者加权平均值。由此可以得到地区相关性模型、省份相关性模型、行业集团相关性模型和行业相关性模型等。1984年CHEOL S. EUN和BRUCE G. RESNICK应用股票价格计算的国家之间的相关性系数如下：

	Australia	France	Germany	Japan	Nether-lands	Switzer-land	U.K.	U.S.A.
1. Australia	0.5868							
2. France	0.2867	0.5764						
3. Germany	0.1832	0.3121	0.6534					
4. Japan	0.1525	0.2383	0.3006	0.4162				
5. Netherlands	0.2415	0.3447	0.5091	0.2824	0.6245			
6. Switzerland	0.3588	0.3687	0.4751	0.2819	0.5171	0.6640		
7. U.K.	0.3158	0.3782	0.2996	0.2092	0.3934	0.4315	0.6987	
8. U.S.A.	0.3043	0.2259	0.1707	0.1373	0.2714	0.2728	0.2798	0.4392

不难发现，该系数与2006年公布的新巴塞尔协议的相关性系数有很大的差距，很多国家、国家之间的相关性系数比新巴塞尔协议的相关性系数大。

2.3 因子模型

单因子模型：设公司 i 的收益率 $R_i = \{r_i(t)\}_0^\infty$ 由系统性风险因子 r^f 和公司特殊因子 ε_i 决定，即

$$r_i = \alpha_i + \beta r^f + \varepsilon_i \quad i = 1, 2, \dots, n$$

假设 $Cov(\varepsilon_i, \varepsilon_j) = 0$ ， $i, j = 1, 2, \dots, n, i \neq j$ ； $Cov(\varepsilon_i, r^f) = 0$ ， $i = 1, 2, \dots, n$ 。那么相关性系数为：

$$\rho_{ij} = \frac{\beta_i \beta_j \sigma_f^2}{\sigma_i \times \sigma_j} \quad i, j = 1, 2, \dots, n, i \neq j$$

其中， $\sigma_f^2 = Var(r^f)$ ， $\sigma_i = Var(r_i)$ 。

多因子模型：设公司 i 的收益率 $R_i = \{r_i(t)\}_0^\infty$ 由 k 个风险因子 $r_1^f, r_2^f, \dots, r_k^f$ 和公司特殊因子 ε_i 决定，即

$$r_i = \alpha_i + \beta_i^1 r_1^f + \cdots + \beta_i^k r_k^f + \varepsilon_i$$

假设：1) $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad i = 1, 2, \cdots, n$;

2) $Cov(\varepsilon_i, r_j^f) = 0 \quad i = 1, 2, \cdots, n \quad j = 1, 2, \cdots, k$;

3) $Cov(r_j^f, r_m^f) = 0 \quad m, j = 1, 2, \cdots, k$;

4) $Cov(r_j^f, r_j^f) = 0; \quad j = 1, 2, \cdots, k$;

$$\text{则 } \rho_{ij} = \frac{\sum_{m=1}^k \beta_i^m \beta_j^m (\sigma_f^m)^2}{\sigma_i \times \sigma_j} \quad i, j = 1, 2, \cdots, n, i \neq j$$

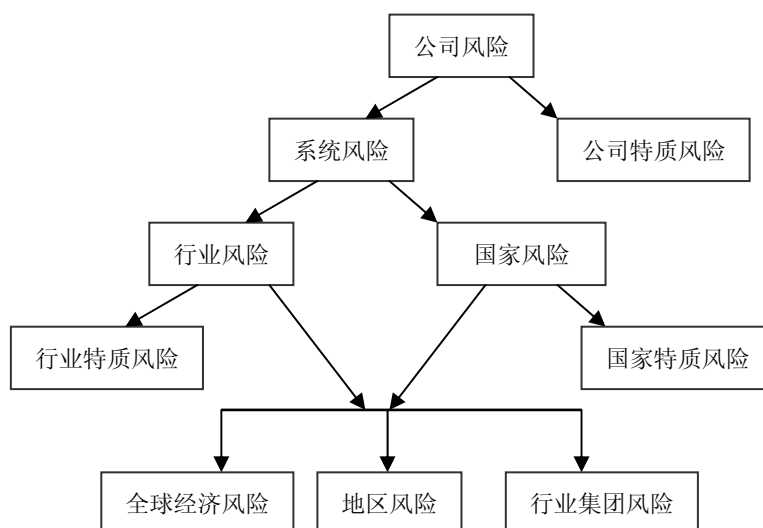
三、风险因子选择

一般的相关性系数模型的因子结构分为三个层次，如下图所示：

第一层次：
组成因子

第二层次：
国家与行业

第三层次：
全球、地区
和行业集团因子



针对中国实际情况，我们选择中国国家因素、地区因素和行业因素。行业按照证监会公布的行业划分标准，将行业的周回报率的相关性较高的行业合并为一个行业集团，详细划分见附录 1，地区划分见附录 2。

采用在深交所和上交所上市的 1711 家公司从 2004 年 4 月 9 日至 2009 年 3 月 17 日的股票周回报率来表征上述各个变量，模拟沪深上市公司的相关性系数。

中国国家因素：选择深交综合指数周回报率和上交综合指数周回报率作为全国经济形势影响沪深所有上市公司的共同因子，分别记为 r^{sz} 和 r^{sh} 。

行业因素：选择 13 个深交行业指数周回报率作为对沪深上市公司回报率的行业风险因子，记为 $r_i^h, i = 1, 2, \cdots, 13$ 。

地区因素：由于没有现成的地区股票指数，所以本人以 2004 年 4 月 9 日的股

价和上市流通股数为基准，计算出 6 个地区的股数周回报率,并以此作为地区风险因子，记为 $r_i^d, i=1,2,\dots,6$ 。

四、计算方法

4.1 地区/行业集团模型

地区/行业集团模型有 13 个待定系数：

1) ρ^B 表示全国基数，即为不同地区、不同行业时的两笔资产的借款人的相关性系数；

2) $\rho_i^R (i=1,2,\dots,6)$ 是地区相关性系数，即两笔资产的借款人处于同一地区行业时，相关性系数在全国基数的基础上再加上该数值；

3) $\rho_j^I (j=1,2,\dots,6)$ 是行业集团相关性系数，即两笔资产的借款人处于同一行业集团时，相关性系数在全国基数的基础上再加上该数值；
则借款人的相关性系数有如下四种情况：

1) 不同行业集团且不同地区的公司之间相关性系数为 ρ^B ；

2) 同处行业集团 j 、不同地区的公司之间相关性系数为 $\rho^B + \rho_j^I$ ；

3) 不同行业集团、同处地区 i 的公司之间相关性系数为 $\rho^B + \rho_i^R$ ；

4) 同处行业集团 j 且同处地区 i 的公司之间相关性系数为 $\rho^B + \rho_j^I + \rho_i^R$ ；

记由该方法计算的相关性系数为 $\hat{\rho}_{ij}$ ，应用均值模型计算的相关性系数为 ρ_{ij} ，

计算 ρ^B 、 $\rho_i^R (i=1,2,\dots,6)$ 和 $\rho_j^I (j=1,2,\dots,6)$ 的标准是最小化 $\sum_{i=1}^{36} \sum_{j=1}^{36} (\hat{\rho}_{ij} - \rho_{ij})^2$ ，

即 $(\rho^B \ \rho_i^R \ \rho_j^I) = \arg \min \sum_{i=1}^{36} \sum_{j=1}^{36} (\hat{\rho}_{ij} - \rho_{ij})^2$ 。

4.2 因子模型

假设借款人 i 的资产回报率 r_i 由 2 个国家因子（深交综合指数、上交综合指数）、6 个地区因子（地区指数）和 13 个行业指数（深交行业指数）决定，其因果关系为如下线性面板数据模型：

$$r_i = \alpha_i + \beta_i^{c1} r^{sz} + \beta_i^{c2} r^{sh} + \sum_{m=1}^6 \beta_m^{Rj} r_j^R + \sum_{k=1}^{13} \beta_i^{Ik} r_k^I + \varepsilon_i$$

假设：模型服从 2.3 小节的多因子模型假设；
 假设：参数是固定的，即该模型为固定影响变系数模型；
 假设：同行业同地区内的公司的系数是相同的。

采用广义差分法估计得到参数系数 α_i 、 β_i^{c1} 、 β_i^{c2} 、 β_m^{Rj} 和 β_i^{lk} 。用历史数据计算各个因子的标准差。

记 σ^{sz} 和 σ^{sh} 分别为深交综合指数标准差和上交综合指数标准差，记 σ_m^R ($m=1,2,\dots,6$) 为地区指数的标准差，记 σ_k^I ($k=1,2,\dots,6$) 为行业指数标准差时，两个借款公司之间的回报率协方差为：

$$Cov(r_i, r_j) = \beta_i^{c1} \beta_j^{c1} \sigma^{sz} + \beta_i^{c2} \beta_j^{c2} \sigma^{sh} + \sum_{m=1}^6 \beta_i^{Rm} \beta_j^{Rm} \sigma_m^R + \sum_{k=1}^{13} \beta_i^{lk} \beta_j^{lk} \sigma_k^I$$

记 σ_i 和 σ_j 分别为公司 i 、 j 的回报率的标准差时，这两个公司之间的相关性系数为：

$$\rho_{ij} = \frac{Cov(r_i, r_j)}{\sigma_i \sigma_j}$$

五、结算结果

4.1 均值模型的结果

地区相关性系数为：

	东北	华北	华东	西北	西南	中南
东北	34.39	35.54	36.12	33.89	33.51	35.93
华北		37.85	37.76	35.25	35.37	37.37
华东			38.89	36.60	36.11	38.60
西北				33.51	32.40	34.84
西南					33.98	35.40
中南						36.83

单位：百分比

行业集团相关性系数为：

	综合制造	交通服务	农业建设	信息传播	金融保险	房产贸易
综合制造	35.68	47.04	44.36	43.57	42.24	43.67
交通服务		46.91	—	41.75	41.37	41.61
农业建设			42.83	36.13	35.05	34.94
信息传播				42.78	34.98	40.47

金融保险					53.91	42.90
房产贸易						41.47

单位：百分比

行业集团/地区相关性系数：

	东北信息传播	华北交通服务	华北信息传播	华北金融保险	华东交通服务	华东金融保险	西北交通服务	西南交通服务
东北信息传播	51.29	56.46	50.12	36.32	49.93	41.40	56.37	55.23
华北交通服务		51.18	41.99	41.31	49.08	43.35	48.12	46.54
华北信息传播			44.93	34.92	46.40	38.78	48.26	48.16
华北金融保险				74.69	43.37	65.08	43.65	37.79
华东交通服务					52.26	48.65	51.46	48.03
华东金融保险						56.51	50.77	45.63
西北交通服务							60.42	59.66
西南交通服务								48.59

由于篇幅限制，行业集团/地区只摘录了一部分

单位：百分比

理论上，对角线上的相关性系数较大上半三角内的系数较小，但是模拟结果不满足完全这个直观现象，原因有以下几条：

- 1) 中国公司的回报率受全国因素影响更大，受地区因素、行业集团因素因子影响较小；
- 2) 中国股市具备特殊性，股价波动不能完全反应沪深上市公司的资产波动情况；
- 3) 计算数据的截取时间段较短，只经历了一个经济周期。

4.2 行业集团/地区模型的结果

行业集团/地区模型的结果为：

全国基数：37.41%

地区相关性(单位：百分比)：

东北	华北	华东	西北	西南	中南
0.10	4.73	6.63	4.42	0.10	0.10

行业相关性(单位：百分比)：

综合制造	交通服务	农业建设	信息传播	金融保险	房产贸易
8.31	6.69	4.92	9.53	0.31	10.00

行业集团/地区模型的结果有以下几个特点：

1) 该方法计算的相关性系数普遍较高，说明中国上市公司相关性较高，读者在 4.1 节也可以看到这个现象；

2) 全国基数较高，地区相关性系数较小，说明中国上市公司受全国宏观影响程度较大，而受地区影响较小；

3) 行业系数有大有小, 房产贸易相关性系数较高, 说明该行业内的企业受行业因素影响很大, 而金融保险、农业建设相关性系数较小, 说明这些行业内的公司对行业因素变化不太敏感, 更多受到宏观因素的影响。

4.3 因子模型结果

行业/地区的参数:

地区行业 因子	东北信息 传播	华北交通 服务	华北信息 传播	华北金融 保险	华东交通 服务	华东金融 保险	西北交通 服务	西南交通 服务
深交	0.118	0.2869	0.1224	-0.0001	0.1437	0.1801	0.1512	0.1326
上交	0.1649	0.1251	0.1922	-0.0053	0.1728	0.1669	0.1451	0.1519
西南	0.1418	0.1994	0.1638	0.1919	0.1377	0.1448	0.1481	0.1427
中南	0.1655	0.156	0.1058	0.3218	0.1267	0.1413	0.1517	0.1551
采掘业	0.0321	0.1926	0.1293	0.179	0.145	0.1317	0.1425	0.1244
房地产业	0.2906	0.0781	0.2244	0.1563	0.1381	0.1292	0.1554	0.1549
建筑业	0.492	0.1202	0.113	0.0727	0.1491	0.1625	0.1544	0.1402
金融、保险业	0.252	0.2955	0.1448	0.3112	0.1259	0.1416	0.1509	0.1556
制造业	0.2564	0.2255	0.1143	0.1304	0.424	0.504	0.412	0.251

由于篇幅限制, 行业/地区只摘录了一部分; 地区方面: 东北、华北、华东和西北没有包含在表中; 行业方面: 传播与文化产业, 电力、煤气及水的生产和供应业, 交通运输、仓储业, 农、林、牧、渔业, 批发和零售贸易, 社会服务业, 信息技术业和综合类没有包含在表中。

因子方差为:

全国因子:

深交综合指数	上交综合指数
1.878	2.371

单位: 千分之一

地区因子:

东北	华北	华东	西北	西南	中南
2.478	2.445	2.201	2.381	3.149	2.971

单位: 千分之一

行业因子:

采掘业	传播与 文化产 业	电力、 煤气及 水的生 产和供 应业	房地 产业	建筑 业	交通运 输、仓 储业	金融、 保险业	农、林、 牧、渔 业	批发和 零售贸 易	社会服 务业	信息技 术业	制造 业	综合类
4.606	5.327	2.889	3.731	3.667	2.32	5.092	3.157	2.536	2.989	2.522	2.388	2.856

单位: 千分之一

因子模型的计算结果:

	东北信息传播	华北交通服务	华北信息传播	华北金融保险	华东交通服务	华东金融保险	西北交通服务
东北信息传播	14.34	15.61	16.64	15.93	11.25	12.55	8.79
华北交通服务		25.46	23.81	28.01	18.30	21.93	11.93
华北信息传播			26.00	27.83	17.47	20.42	9.77
华北金融保险				33.95	20.18	24.66	11.52
华东交通服务					14.48	17.37	7.43
华东金融保险						22.42	8.54
西北交通服务							14.49

由于篇幅限制，行业/地区只摘录了一部分

单位：百分比

因子模型的结果接近国际评级机构(如惠誉)公布的相关性系数，与新巴塞尔协议公布的相关性系数也较为接近，说明因子模型能够较为准确地测算沪深上市公司的相关性。同时我们也看到，部分同行业同地区分池内的相关性比新巴塞尔协议公布的相关性系数高，说明新巴塞尔协议在国际范围测试的相关性系数在中国局部范围不适合。

行业集团/地区模型测算的相关性系数比因子模型大，也偏离其他机构公布的数据，但是与 2.2 节中 CHEOL S. EUN 和 BRUCE G. RESNICK 测算的系数相近，说明行业集团/地区模型测算的相关性系数普遍较高，这是由于方法的系统性影响。至于哪个模型计算的结果接近实际情况，有待进一步研究和探讨。