

连接函数(copula)技术与金融风险分析

张尧庭

ABSTRACT

Copula technique is a kind of comparatively new method of financial risk analysis, whose core is to connect the co-distribution of many random variances with their fringe distributions. This coincides exactly with the method to decompose risks into different components in financial risk analysis.

关键词：金融风险；连接函数；联合分布

目前金融风险是国际、国内都非常关注的课题,风险的度量和分析技术发展得非常快,因为它有强大的市场需求,并且可以较快看到经济效益。不重视它,风险管理失败的后果是惨重的,个人破产,国家动荡不安。传统的统计分析技术在这一方面已有不少贡献,VaR技术就是一个典型的例子,它从出现到现在不到十年的时间,已经广泛流传,成为标准的技术,尽管人们已经发现它有许多不能令人满意的地方,但仍不失为一种分析的工具。近年来,连接函数技术的出现,把金融风险分析推向了一个新的阶段,以本文所附的文献[1],[2],[3]中可以查到,这二三年来,这一类的文章和应用蜂涌而出,有的公司已经开发这一方面的软件,它很快应会变成一种实用的技术,本文就想作一比较通俗的介绍,引入这一技术,让大家来关心、发展它,它正处于开创、发展时期。

copula一词原意是交换、连接的意思。它是把多个随机变量 ξ_1, \dots, ξ_n 的联合分布 $F(x_1, \dots, x_n)$ 与它们各自的边缘分布 $F_{\xi_1}(x_1), \dots, F_{\xi_n}(x_n)$ 相连接,也即是说,连接函数 $C(u_1, \dots, u_n)$ 使

$$F(x_1, \dots, x_n) = C(F_{\xi_1}(x_1), F_{\xi_2}(x_2), \dots, F_{\xi_n}(x_n))$$
等式成立,它是 $F(x_1, \dots, x_n)$ 与 $\{F_{\xi_i}(x_i)\}$ 的连接函数。

可以证明,在很一般的条件下,上式中的多元函数 $C(u_1, \dots, u_n)$ 是存在的,它就是copula,就是连接函数,将多维分布与一维边缘分布联系在一起的函数。这样的函数 $C(u_1, \dots, u_n)$ 在金融风险分析中有什么作用呢?

金融风险分析技术的一个重要的方面,就是将风险进行分解,例如一支股票的风险可以分解为市场风险、板块风险、个股风险与随机扰动风险。市场风险是由在股市上交易的全部股票价格形成的,因此它是由联合分布来决定的,用 $F(x_1, \dots, x_n)$ 表示它。而单个股票价格波动是由边缘分布 $F_i(x_i)$ 描述的, $i=1, 2, \dots, n$ 。有了连

接函数 $C(u_1, \dots, u_n)$,它与边缘分布一维的 $F_i(x_i)$, $i=1, 2, \dots, n$ 无关,它的性质反映了市场的特点。因为用不同的 $F_i(x_i)$ 代 $C(u_1, \dots, u_n)$ 中的 u_i ,就得到不同的联合分布。如果我们考虑的不是整个的市场,而是各个板块,于是每个板块就有各自连接函数 $C(u_1, \dots, u_k)$ 。不同的板块,相应的连接函数 $C(u_1, \dots, u_k)$ 是不同的,它们各自反映了所在板块的特点。这样比较不同板块 $C(u_1, \dots, u_k)$ 的特性,就比较了板块的市场特性。从这里就明显地看出连接函数 $C(u_1, \dots, u_n)$ 在股市风险分解中的作用。

很明显,将上述的每一种股票改成金融资产,市场风险就相当于资产组合中资产结构的风险,因此copula的技术也可用于这一对象。

然而,它的作用不止于此。连接函数有一系列特性,这些特性对我们认识市场内部的相关性有更好的了解,下面我们以二元的为例来逐一说明,更详细的可参看[1],[2],[3]有关内容。有关的一些详尽的证明可以从[1]中去查阅,这里着重给出考虑问题的方法和直观而简略的证明。

设 (ξ, η) 的联合分布是 $F(x_1, x_2)$,各自的边缘分布为 $\xi \sim G(x), \eta \sim H(y)$,于是由连接函数 $C(u_1, u_2)$ 的性质知道

$$P(\xi < x, \eta < y) = F(x, y) = C(G(x), H(y)) \quad (1)$$
若 $s(x), t(y)$ 分别是 x, y 的严格单调递增函数,因此它们的反函数自然存在而且唯一,我们只看所有涉及的函数均是连续的情形,就可以看出

$$P(s(\xi) < s, t(\eta) < t) = P(\xi < s^{-1}(s), \eta < t^{-1}(t))$$
其中 $s^{-1}(x), t^{-1}(y)$ 是 s, t 的反函数。由于

国家自然科学基金《应用统计》项目资助研究。

$$F_{s(\xi)}(x) = P(s(\xi) \leq x) = P(\xi \leq s^{-1}(x))$$

$$= G(s^{-1}(x))$$

$$F_{t(\eta)}(y) = P(t(\eta) \leq y) = P(\eta \leq t^{-1}(y))$$

$$= H(t^{-1}(y))$$

于是有

$$\begin{aligned} P(s(\xi) \leq x, t(\eta) \leq y) &= P(\xi \leq s^{-1}(x), \eta \leq t^{-1}(y)) \\ &= F(s^{-1}(x), t^{-1}(y)) = C(G(s^{-1}(x)), H(t^{-1}(y))) \\ &= C(F_{s(\xi)}(x), F_{t(\eta)}(y)) \end{aligned}$$

这表示 $(s(\xi), t(\eta))$ 的联合分布相应的连接函数仍然是 (ξ, η) 的连接函数 $C(u_1, u_2)$ 。这证明了连接函数对于随机变量的严格单调增变换是不变的。

利用这个性质, 就可以导出一系列有意义的推论。

1. 连接函数 $C(u_1, \dots, u_n)$ 反映了市场的结构, 它与股票价格的度量单位无关。因为不同的价格度量单位变换只是一个线性变换 $ax+b$, 且 $a>0$ 它是严格单调增的变换。

2. 由连接函数 $C(\cdot, \dots, \cdot)$ 导出的相关性指标, 比常用的相关系数更加合乎人们的要求。因为常用的相关系数实际上是线性变换下不变的一种相关性指标, 涉及到非线性函数的相关性, 它就会导出错误的结论。例如 ξ, η 是两个标准正态随机变量, $E\xi = E\eta = 0, \text{Var}(\xi) = \text{Var}(\eta) = 1$, 它们之间的相关系数是 ρ , 但是作变换, 令

$$\xi_* = e^{\xi}, \eta_* = e^{\eta}$$

显然这是一个严格单调增的变换, 计算一下 ξ_* 与 η_* 的相关系数(这里不给出证明, 有兴趣的可以自己验证)就不是 ρ , 而是

$$\text{Cov}(\xi_*, \eta_*) = \frac{e^{\rho} - 1}{e - 1}$$

另一方面, 若 $\xi \sim N(0, 1), \eta = \xi^2$ 显然与 ξ 是关系密切的, 但是 $\text{Cov}(\xi, \eta) = E\xi\eta - (E\xi)(E\eta) = E\xi^3 - E\xi E\xi^2 = 0$, 因此 ξ, η 是不相关的。所以用相关系数度量相关性, 超出了线性相关的范围就会误导。但是有了 $C(u_1, u_2)$ 这样的连接函数, 用它来引出度量相关性的指标, 就是严格单调增变换下的相关性, 比线性相关的范围要宽。

然而由连接函数导出的相关性度量并不是唯一的。首先可以证明以往用的 Kendall 的 τ 和 Spearman 的 ρ 都是单调变换不变的相关性度量, 并引出了其他一些度量相关性的指标。这些从本文所引的文献中均可以查到。

例 1. 设 $(x_1, y_1), (x_2, y_2)$ 是独立同分布的向量, 令

$$\begin{aligned} \tau &= P((x_1 - x_2)(y_1 - y_2) > 0) \\ &\quad - P((x_1 - x_2)(y_1 - y_2) < 0). \end{aligned}$$

于是 τ 就度量了 x 与 y 变化的一致性程度。可以证明:

$$\tau = 2P((x_1 - x_2)(y_1 - y_2) > 0) - 1 \quad (2)$$

从(2)式看出 τ 在 $[-1, 1]$ 之间。设 (x_1, y_1) 相应的连接函数是 $C(u, v)$, 则可证明

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1. \quad (3)$$

$$(\text{或 } \tau = 4E(C(U, V)) - 1, (U, V) \sim C(u, v))$$

我们看一下 τ 的定义, 考虑 x, y 两个变量的相关性应该如何来度量, 自然会想到首先应考虑的是它们的变化范围尽管不(如人的身高的变化范围与人的体重的变化范围), 但它们的变化趋势是否应该一致? 若一致, 相关性就比较强, 若正好相反, 相关性也是强的。因此要考虑变化, x 的变化用它的独立的两个观察值来反映是合理的, $x_1 - x_2$ 就反映了 x 的变化。同样地, 独立的 y 观察值 y_1, y_2 的差值 $y_1 - y_2$ 就反映了 y 的变化, 若

$$(x_1 - x_2)(y_1 - y_2) > 0$$

它们的变化是一致的; 若

$$(x_1 - x_2)(y_1 - y_2) < 0$$

它们的变化就相反一致的, 因此

$$\begin{aligned} \tau &= P((x_1 - x_2)(y_1 - y_2) > 0) \\ &\quad - P((x_1 - x_2)(y_1 - y_2) < 0) \end{aligned}$$

就反映了变化一致与否的程度。易见上式的值在 $[-1, 1]$ 之间。 $\tau=1$, 就表示 x 的变化与 y 的变化完全一致, 所以很相关; 当 $\tau=-1$, 表示 x 的变化与 y 的反向变化完全一致, 因此很相关, 方向相反。 $\tau=0$ 一半是一致的, 一半是相反一致的, 所以不能判断是否有相关关系。

很明显, 取严格单调增的函数 s 与 t 之后, 自然有

$$\begin{aligned} (s(x_1) - s(x_2))(t(y_1) - t(y_2)) &> 0 \Leftrightarrow \\ (x_1 - x_2)(y_1 - y_2) &> 0 \end{aligned}$$

所以 τ 的值对于严格单调增的变换是不变的, 这就充分说明了 τ 作为度量 x, y 的相关性指标, 它的优点是什么。

类似地, 我们可以看到以下的几个例子, 它们论述的指标是有什么优良的性质。

例 2. 设 (x, y) 有联合分布 $H(x, y)$, 它们相应的边缘分布是 $F(x)$ 与 $G(y)$, 考虑 x_* 与 y_* , $(x_*, y_*) \sim F(x)G(y)$ (即 x_*, y_* 独立)。假定

$$(x, y) \text{ 与 } (x_*, y_*) \text{ 独立,}$$

此时令

$$\begin{aligned} \rho &= 3[P((x - x_*)(y - y_*) > 0) \\ &\quad - P((x - x_*)(y - y_*) < 0)] \end{aligned}$$

则 ρ 就是 Spearman 的相关性指标, 即通常所说的秩相关系数。

这里 x_* 与 x 同分布, 而且独立, 因此可以将 x_* 看成是 x 的一个重复观察, y 与 y_* 的关系也是如此。而 x_* 与 y_* 是独立的, 因此

$$(x - x_*)(y - y_*) > 0$$

表示 (x, y) 的变换与独立的 x_*, y_* 变化相一致, 这个概率的大小自然也反映了一种相关性。很明显, 它也是对严格单调增的变换是不变的, 因而它可以用连接函数 $C(u, v)$ 来表示, 当 (x, y) 的连接函数 $C(u, v)$ 给定后, 可以证明 ρ ,

$$\rho = 12 \int_0^1 \int_0^1 uv dC(u, v) - 3 \quad (4)$$

用 U, V 分别表示 $[0, 1]$ 上的均匀分布, 此时 $EU = EV = \frac{1}{2}$, $Var(U) = Var(V) = \frac{1}{12}$. 如果 (U, V) 联合分布就是 (4) 中的 $C(u, v)$, 于是

$$E(UV) = \int_0^1 \int_0^1 uv dC(u, v),$$

因此式(4)就是

$$\begin{aligned} \rho &= 12 \left[E(UV) - \frac{1}{4} \right] \\ &= 12 [E(UV) - EUEV] = \frac{1}{12} Cov(U, V) \quad (5) \\ &= Cov(U, V) / \sqrt{Var(U) Var(V)} \\ &= \rho(U, V) \end{aligned}$$

这告诉我们, Spearman 的 ρ 是 U 与 V 的线性相关系数, 注意到 U 与 V 与原来分布的关系:

当 $\xi \sim F(x)$, $\eta \sim G(y)$, $(\xi, \eta) \sim H(x, y)$ 时, $F(\xi)$ 与 $G(\eta)$ 就分别是均匀分布, 用 U 代表 $F(\xi)$, V 代表 $G(\eta)$, (U, V) 就遵从 $H(x, y)$ 相应的连接函数 $C(u, v)$, 这样就证明了

$\rho = F(\xi)$ 与 $G(\eta)$ 的线性相关系数。

例 3. 基尼(Gini)关联系数。 这是衡量 x 与 y 变化大小次序不一致程度的指标。如果说 τ 与 ρ 只涉及变化的符号, 增大是“+”, 减少是“-”, 只考虑了“+”、“-”号一致的程度, 那么基尼关联数 r 就更细致考虑了它们顺序的一致性和不一致性。

设 (x, y) 的 n 个样本是 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 。将 x_1, \dots, x_n 按从小到大顺序排列后, x_i 的名次 r_i 称为它的秩, 同样 y_i 在 y_1, \dots, y_n 中的名次(秩)记为 s_i , $i = 1, 2, \dots, n$ 。如果 x, y 的变化是一致的, $|r_i - s_i|$ 就应该

很小, 所以 $\sum_{i=1}^n |r_i - s_i|$ 反映了不一致的程度。如果变化方向相反, 那么 r_i 与 s_i 应处于两端, x_i 处于 r_i 位置时, s_i 应处于倒数第 r_i 位置上, 也即 s_i 在第 $n+1-r_i$ 位置上, 因此 $|r_i + s_i - n - 1|$ 就衡量了相反变化的不一致程度。 $\sum_{i=1}^n |r_i + s_i - n - 1|$ 就是总的不一致性。于是

$$\sum_{i=1}^n |r_i + s_i - n - 1| - \sum_{i=1}^n |r_i - s_i|$$

就反映了两种不一致的差距。将它对所有可能的值求平均, 就得所要的关联系数。

很明显, 严格单调增的变换不会改变 x_i, y_i 的秩, 因为大小顺序关系是不变的, 所以它也只与 (x, y) 的连接函数 $C(u, v)$ 有关, 它的表达式是

$$r = 4 \left\{ \int_0^1 C(u, 1-u) du - \int_0^1 (u - C(u, u)) du \right\} \quad (6)$$

例 4. 中位数相关系数。 在前面相关性指标 τ 的定义中。用中位数代期望, 就是中位数相关系数。注意到中位数对于严格单调增变换是不变的, 或更确切地说是同变的, 因此它也是连接函数的特征值, 它实际上是衡量了

$P((x - \bar{x})(y - \bar{y}) > 0) - P((x - \bar{x})(y - \bar{y}) < 0)$ 的大小, 其中 \bar{x}, \bar{y} 分别是 x, y 的中位数。注意到 (x, y) 的分布函数用 $F(x, y)$ 表示时, 容易证明中位数相关系数 ρ 有下面的表达式:

$$\begin{aligned} \rho &= 2P(x - \bar{x})(y - \bar{y}) > 0 - 1 \\ &= 2[P(x < \bar{x}, y < \bar{y}) + P(x > \bar{x}, y > \bar{y})] - 1 \\ &= 2[F(\bar{x}, \bar{y}) + 1 - G(\bar{x}) - H(\bar{y}) + F(\bar{x}, \bar{y})] - 1 \\ &= 4F(\bar{x}, \bar{y}) - 1 \end{aligned}$$

注意中位数 \bar{x}, \bar{y} 对应的分布函数值 $G(\bar{x}) = H(\bar{y}) = \frac{1}{2}$, 因此由连接函数的性质, $F(\bar{x}, \bar{y}) = C(G(\bar{x}), H(\bar{y})) = C\left(\frac{1}{2}, \frac{1}{2}\right)$, 于是就有

$$\rho = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1, \quad (7)$$

从上面介绍的这些相关性度量的指标可以看出, 由连接函数可以引出一系列非常有意义的相关性指标。在金融风险的分析中, 更有意义的是随机变量的尾部相关性, 厚尾分布与正态分布的一个有意思的差别, 就是尾部的相关性, 这一特性用连接函数来处理就非常方便。引入条件概率 $P(\eta > u | \xi > u)$, 它反映了股票市场中一种股票价格高涨后, 是否会引起其他股票价格攀升, 从而对股市有较大的影响。用随机变量的语言来描述, 就是 $\xi > u$ 时, $\eta > v$ 的概率是否会发生变化, u, v 相当大时, 就是 ξ, η 尾部的相关性, 用 $\lambda(u)$ 表示 $P(\eta > u | \xi > u)$, 考虑 $u \rightarrow \infty$ 时, $\lambda(u)$ 的极限值, 如果极限存在, 它就反映了尾部相关性的 大小, 因此用

$$\lambda_U = \lim_{u \rightarrow \infty} \lambda(u) \quad (8)$$

表示尾部相关性。很容易看到 $\lambda(u)$ 的极限是严格单调增变换不变的, 所以它的极限是连接函数的性质, 完全可以用连接函数的一种极限值来表示。

在 [2] 中已经证明了

$$\lambda_U = \lim_{u \rightarrow \infty} [P(U > u | U = u) + P(U > u | V = u)] \quad (9)$$

其中 (U, V) 遵从分布 $C(u, v)$, $C(u, v)$ 是连接函数。概率论的知识告诉我们, 当 $\xi \sim F(x)$ 时, 只要 $F(x)$ 是连续函数, 那么 $F(\xi)$ 就是 $[0, 1]$ 上的均匀分布, 而 $F(x)$ 在很多常见的问题中不仅是连续的, 而且还是严格单调变化的增函数, 因此

$$\begin{aligned} P(\xi > x | \eta > x) &= \\ P(F(\xi) > F(x) | F(\eta) > F(x)) \end{aligned}$$

若 ξ 与 η 的边缘分布相同, 那么 $F(\xi)$ 与 $F(\eta)$ 都是均匀分布随机变量, (9) 式的 U 和 V 正是反映了这一事实。可以证明:

对于正态分布, 如果 ξ 与 η 的相关系数用 $\rho_{\xi\eta}$ 来表示, 则它们的尾部相关性 λ_U 是

$$\lambda_U = \begin{cases} 1 & \text{若 } \rho_{\xi\eta} = 1, \\ 0 & \text{若 } \rho_{\xi\eta} < 1. \end{cases} \quad (10)$$

可见正态分布不论相关系数有多大, 只要不是 1, 它的 $\lambda_U = 0$ 。因此用 $\rho_{\xi\eta}$ 描述股票上涨、下跌幅度较大时的相关性, 完全是误导, 过低地估计了有关的风险。然而对非正态的厚尾分布而言, 情况就很不相同, 表 1 是 t 分布相应的尾部相关性 λ_U 取值的一张表(详见[2]), 它明确地显示了自由度对尾部相关性的影响:

表 1 t 分布的 λ_U 值表

		相关系数 ρ				
		-0.5	0	0.5	0.9	1
自由度 ν	2	0.06	0.18	0.39	0.72	1
	0.01	0.08	0.25	0.63	1	
	0	0.01	0.08	0.46	1	
	0	0	0	0	1	

即使 $\rho = 0$, ν 较小时, 尾部相关性却不是 0。注意这个度量指标与相关系数 ρ 有很大的差别。

3. 连接函数可以将研究对象的层次结构给以明确的描述, 并且不同层次的连接函数反映了不同层次的结构。

例如我们可以将股票市场看成的三个层次的一个对象:

个股 — 板块 — 股市整体

板块是由若干个个股组成的, 板块用 α 表示第 α 个板块, $C_\alpha(u_1, u_2, \dots, u_{n_\alpha})$ 表示第 α 个板块中 n_α 个个股的连接函数。而股市整体是由 k 个板块形成的, 因此 $C(v_1, \dots, v_k)$ 可以描述板块与股市联系连接函数, 注意它只是 k 个变量的函数, 每个变量取值在 $[0, 1]$ 之内。很明显, 板块之间的相关性是由连接函数 C 来反映的, α 板块内个股的相关性是由 C_α 的性质来反映的, 所以有了连接函数, 层次结构的分析就非常清楚。

不难看出, 这种层次结构的分析对于可靠性技术也有明显的意义。因为可靠性的系统是层次分明的, 如下图是一个典型的层次

元件 → 部件 → 子系统 → 系统。

实际上串联系统相当于 $C(u_1, \dots, u_n) = \prod_{i=1}^n u_i$, 并联系统

相当于 $C(u_1, \dots, u_n) = 1 - \prod_{i=1}^n (1 - u_i)$ 。所以连接函数的不同性质反映了可靠性系统的不同结构。可以预见, 连接函数的技术会对可靠性的统计分析有一个新的推动。

连接函数的概念使我们对保险业中原有的共同单调 (comonotone) 风险有了更好的理解, 并且开拓了许多可能的应用。

共同单调的随机变量, 是指它的联合分布是一类特殊的分布—fr chet 分布, 设边缘分布是 $F_1(x_1), \dots, F_n(x_n)$, 那么它的联合分布是

$$F(x_1, \dots, x_n) = \min_{1 \leq i \leq n} F_i(x_i) \quad (11)$$

它相应的连接函数 $C(u_1, \dots, u_n)$ 自然就有关系式

$$C(u_1, \dots, u_n) = \min_{1 \leq i \leq n} u_i \quad (12)$$

实际上可以证明, 若 $F(x_1, \dots, x_n)$ 是连续的, 则有单调函数 $f_1, \dots, f_n, f_i(t)$ 均为连续的, 又有随机变量 ξ , 使

$$F(x_1, \dots, x_n) = P(f_1(\xi) < x_1, \dots, f_n(\xi) < x_n)$$

它表示这 n 种风险资产, 等同于一个潜在风险源的若干个风险资产。例如相邻地区的火险就有这种性质。实际上任何一个连接函数 C 都有下述性质:

$$C(u_1, \dots, u_n) \leq \min_{1 \leq i \leq n} u_i$$

因此 fr chet 分布是一个“上界”, 一些风险资产虽然不了解它们真实的 $C(u_1, \dots, u_n)$, 但用 $\min_{1 \leq i \leq n} u_i$ 去代替不会低估有关的情况。它的实际应用与它的下述性质有关, 证明可在 [4] 中找到。

设 (x_1, \dots, x_n) 是随机变量, 且 $E|x_i| < \infty, i = 1, 2, \dots, n$,

$$V = x_1 + \dots + x_n, W = F_1^{-1}(U) + \dots + F_n^{-1}(U),$$

其中 $F_i^{-1}(x)$ 是 x_i 的边缘分布 $F_i(x)$ 的反函数, U 是 $[0, 1]$ 上的均匀分布, 则对任何凸函数 φ , 有

$$E\varphi(V) \leq E\varphi(W) \quad (13)$$

这表示对任何风险的估值, 用 W 来衡量时, 不会比用 V 衡量低。很容易看出, 要求出 V 的分布, 必须从 x_1, \dots, x_n 的联合分布出发, 而 W 只涉及边缘分布, 是可操作的一种估计。因此在保险定价中它是有意义的。

上述结论还可推广在 c_1, c_2, \dots, c_n 均为正数时, 对

$$V^* = c_1 x_1 + \dots + c_n x_n, W^* = c_1 F_1^{-1}(U) + \dots + c_n F_n^{-1}(U)$$

仍然保持上述性质。这就可以看出, 在金融产品的风险管理中, W^* 是可以度量的, 用它来监控管理至少是一种可以参照的内容。

参考文献

- [1] Nelsen, R. B. (1998), An Introduction to Copulas. Lectures Notes in Statistics, 139, Springer Verlag, New York.
- [2] Embrechts, P., Lindskog, F. And McNeil, A. (2001), Modelling Dependence with Copulas and Applications to Risk Management. Dept. of Math. CH-8092, Zürich, Switzerland.
- [3] Bouy, E. (2000), Copulas for Finance, A Reading Guide and Some Applications. City University Business School, London.
- [4] De Vylder, F. E. (2000), Comonotone Risk. 中国人民大学精算数学研讨会演讲稿。

作者简介: 张尧庭, 上海财经大学教授。

(责任编辑: 何平)