

基于二元极值 Copula 的沪深股市风险相关性分析

陈守东, 孔繁利, 胡铮洋

(吉林大学数量经济研究中心, 吉林大学商学院, 130012)

摘要: 本文介绍了多变量极值分布和极值分布的 Copula 函数, 在此基础上给出了二元 POT 模型, 对联合分布的尾部进行估计。文章将多变量 EVT 和 Copula 相结合对沪、深股市的相关性进行实证研究, 通过运用 Logistic 模型研究得出, 在深市出现最大损失时, 沪市也出现最大损失的概率为 0.715445。

关键词: 多变量 EVT Copula 二元 POT 模型 尾部估计

中图分类号: F224.0

文献标识码: A

1 引言

本文利用极值理论和 Copula (连接函数) 方法对金融市场风险管理做了初步的研究。由于极值理论和 Copula 有严谨的概率统计理论作依托, 因而以其作为核心的度量方法得到了国际金融界的支持和认可。

极值理论是一门用来预测异常现象或者小概率事件风险的模型技术, 能够评估极端情形下金融市场的风险状况。由于仅考虑分布的尾部, 而不是对整个分布进行建模, 从而就避开了分布假设难题, 并且极值理论可以准确地描述分布尾部的分位数, 这有助于处理风险度量中的厚尾问题。在金融、经济、保险等领域, 极端事件、日益频繁的危机及灾难性损失的发生, 给金融市场造成了重大影响, 运用极值理论研究金融市场风险越发重要。

Copula 将联合分布与它们的边际分布连接在一起, 提供了一种描述相依性结构的方法。它的出现和应用为风险分析和多变量时间序列分析提供了一个新的探索方向。首先, 由于不限制边际分布的选择, 可运用 Copula 理论构造灵活的多元分布; 其次, 运用 Copula 理论建立金融模型时, 可将随机变量的边际分布和它们之间的相关结构分离开来研究, 其中, 它们的相关结构可由一个 Copula 函数来描述, 可以捕捉到变量间非线性、非对称的相关关系, 这使建模问题大大简化, 同时也有助于我们对很多金融问题的分析和理解。由于这些性质和特点, 使得 Copula 理论得到广泛应用。

最早的 Copula 是在概率统计理论中出现的, 它的提出可追溯到 Sklar(1959), 此后 Copula 的概念又被 Schweizer & Sklar (1983)中提到。随后 Genest & MacKay (1986)等人又对 Copula 理论进行了发展。Joe (1997)研究了多变量 Copula 理论、模型和相依结构, Nelsen (1999)则比较系统地介绍了 Copula 的定义、构建方法、Archimedean Copulas 及相依性。

Embrechts, P. et al.(1999)探讨了在金融市场中采用线性相关指标度量相依性的局限性, 建议用 Copula 来估计随机变量间的联合分布, 在研究思想上对 Copula 的应用前景做了很好的展望; Bouye, E. et al.(2000), Embrechts, P. et al.(2001)等人用数据拟合 Copula 函数进行了风险分析; Robert De Matteis(2001)对 Copula, 特别是对 Archimedean Copula 及其应用做了比较详尽的总结, 依据 Copula 生成函数的参数, 把 Archimedean Copula 分成不同的类; Patton(2001)构造了马克一美元和日元一美元汇率的对数收益的二元 Copula 模型, 并与相应的 BEKK 模型做了比较, 结果表明 Copula 模型可以更好地描述金融市场间的相关关系; Carpineto et al. (2002)对意大利股市收益率进行了分析, Embrechts et al. (2001), Bouye, E., Durrleman et al. (2001)用 Copula 进行了风险分析,

Bouye, E., Durrleman, V., Nikeghbali, A (2000), Umberto Cherubini, Elisa Luciano, Walter Vecchiato (2004) 详细地介绍了 Copula 理论及其在金融市场中的应用。

Copula 在信用风险领域也有广泛的应用。赖柏志¹参照 Lopez & Marc(1999)²模型和敬永康(2002)³, 针对台湾地区金融机构 2000 年及 2001 年度的资料, 以商业银行作为研究对象, 违约损失率设定为 45% 时, 分别利用 Normal-Copula 及 $t(10)$ -Copula 模拟计算其损失分布, 并分别比较了当置信水平设定为 95% 和 99% 时, 实际损失分布的超限银行数, 发现利用传统的多元正态分布来估计台湾市场的信用风险, 会有明显低估信用风险的情况出现。当改以 $t(10)$ -Copula 来进行模拟时, 可发现在置信水平为 95% 及 99% 时, 超限银行数与理论数值接近, 可以避免信用风险值过度高估或低估的情况发生。除此之外, 在市场风险、资产定价及信用衍生产品的定价上, 近来都有与 Copula 相关的研究出现。

国内对 Copula 的研究还较少, 张尧庭(2002)从理论上探讨了 Copula 在金融上应用的可行性, 指出在不确定线性关系能否正确度量相关关系时, 采用一种更为灵活稳健的相关性分析工具—Copula 技术来分析变量间的相关结构更为可靠。张世英等(2004)研究了 Copula-GARCH 模型对波动性的描述。陈守东等(2006)结合国内证券市场, 基于 Copula 理论、采用 Monte Carlo 模拟方法度量了市场风险。目前, 国内对 Copula 技术及其在相关性和金融风险分析上的应用研究, 特别是结合我国金融市场实际所作的实证研究还不多见。

二、多元极值分布、多元 Copula、多元极值 Copula 函数

1. 多元极值分布

多元极值分布研究的是几个相关总体的极值性质。设 $X_i = (X_{i1}, \dots, X_{id})$, $i = 1, 2, \dots, n$ 是一个 d 维的独立同分布随机向量, 定义 $M_{nj} = \max(X_{n1}, \dots, X_{nd})$ 为第 j 个分量最大值。如果存在正则化的常数序列 $a_n = (a_{n1}, \dots, a_{nd})$, $b_n = (b_{n1}, \dots, b_{nd})$, 其中 $a_{nj} > 0$, 使得当 $n \rightarrow \infty$ 时, 有

$$\lim_{n \rightarrow \infty} \Pr\left(\frac{M_{nj} - b_{nj}}{a_{nj}} \leq x_j, j = 1, \dots, d\right) \rightarrow G(x_1, \dots, x_d) \quad (2.1)$$

上式 G 是一个分布函数, 并且在每个边际上非退化, 则称 G 是一个元极值分布。

De Haan L et al⁴证明了不存在有限参数族来描述多元极值分布, 因此只能在某个模型下进行讨论。De Haan L⁵曾提出过用非参数方法建立模型, 但是这种模型不可微, 而且不能同时考虑边际分布参数与相关结构的估计, 因此一般采用参数模型, 其中较简单而且有广泛应用的是 Logistic 模型⁶。在此模型下, d 元极值分布的分布函数为:

$$G(x_1, \dots, x_d) = \exp\left\{-\left[\sum_{j=1}^d \left(1 + \xi_j \frac{x_j - \mu_j}{\sigma_j}\right)^{-\alpha/\xi_j}\right]^{1/\alpha}\right\} \quad (2.2)$$

其中 μ_j, σ_j, ξ_j 为边际参数, $1 \leq \alpha \leq \infty$ 为相关参数。有关这个模型的性质及参数估计在文献中已经有过大量的研究。

二元极值分布只是在近几年才得到较多的重视, Tawn J A (1988) 总结了前人的工作, 研究了二元极值的各种参数模型, 其中 Gumbel (1960) 提出的 Logistic 模型是最有用的。

¹赖柏志. 关联结构 (Copula) 在信用风险管理之应用, 风险研究小组, Working Paper, 2004.

²Lopez & Marc. Evaluating Credit Risk Models. Journal of Banking and Finance, 1999, 24: 151~165.

³敬永康. 信用风险之模型效果检验—以台湾金融业为例. 货币观测与信用评等, 2002, 35: 20~126.

⁴De Haan L et al. Limit theory for multivariate sample extremes[J]. Zeit.Wahr.verw.cieb. 1977, 40: 317~337.

⁵De Haan L. Extreme in higher dimensions: the model and some statistics[J]. Proc. 45th Sess. I.S.I., 1985.

⁶Gumbel E J. Bivariate exponential distribution[J]. J. Amer. Statist. Assoc., 1960, 55: 698~707.

引理：二元极值分布的边际分布是 GEV，其分布函数为：

$$G(x, \mu, \sigma, \xi) = \exp\{-[1 + \xi(x - \mu)/\sigma]^{-1/\xi}\} \quad (2.3)$$

2. 多元 Copula

Sklar 定理：假设一个多维分布函数 H 的边际分布函数为 $F_1(x_1), \dots, F_n(x_n)$ ，则存在一个 Copula 函数 C 满足，

$$H(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (2.4)$$

如果 $F_1(x_1), \dots, F_n(x_n)$ 是连续的，则 Copula 函数是唯一确定的，反之亦然。

由这个定理我们可以推理得出当我们确定了多个金融时间序列的边际分布和选定一个合适的 Copula 函数后，就可以方便的计算出这些金融时间序列的联合分布，这正是 Copula 函数在实际应用研究的优势所在。多维分布的选取和确定其具体形式再也不是我们所要考虑的难题，而只需确定边际分布和选取合适的 Copula 函数。边际分布的确定由于是一维问题，我们已经有许多研究手段可以去实现它，而 Copula 函数的选择则需要我们通过实证结果的分析和一些评判模型的信息准则来确定。

记随机向量 $X = (X_1, \dots, X_d)^T$ 表示 d 维不同的测度在同一时间点上的损失。我们假设这些损失满足联合分布 F ，即：

$$F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d) \quad (2.5)$$

并且单个损失都有连续的边际分布 $F_i(x) = P(X_i \leq x)$ 。由多元 Copula 函数的定义，我们知道对唯一的函数 C ，联合分布能被写成下式：

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (2.6)$$

我们可将多变量 Copula 函数视为单位超立方体值映射到单位区间上的函数或者视为 Copula 是一个具有标准均匀边际分布的多元分布函数。由于在随机变量 X_1, \dots, X_d 是单调递增变换下，Copula 函数 C 是不会改变的。因此，可以将 C 解释为 X 的相关结构。如在 $d = 2$ 情形下，令 $F_1 = F_2 = \Phi$ 为单变量的标准正态分布，我们选择任意 Copula 函数 C （即任何具有均匀边际分布的二元分布），并且运用 C 通过正态边际分布来获得二元分布函数。我们首先选择一个特殊的 Copula 为 Gaussian Copula，表示为 C_ρ^{Ga} ，这个 Gaussian Copula 能够被写成二重积分的形式

$$C_\rho^{Ga}(x, y) = \int_{-\infty}^{\Phi^{-1}(x)} \int_{-\infty}^{\Phi^{-1}(y)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{-\frac{(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right\} ds dt \quad (2.7)$$

其中 ρ 为相关系数，这个 Copula 函数的两个边际分布都为正态分布，积分上限 $\Phi^{-1}(x)$ 为标准正态分布函数的逆。由此我们能得到具有相关系数为 ρ 的二元正态分布。另一个 Copula 是二元 Gumbel Copula，二元 Gumbel Copula 是多元极值分布 Copula 中用途比较大的一个连接函数，并且有简单的形式，

$$C_\alpha^{Gu}(v_1, v_2) = \exp[-\{(-\log v_1)^\alpha + (-\log v_2)^\alpha\}^{1/\alpha}], 1 \leq \alpha < \infty \quad (2.8)$$

参数 α 反映了 X_1, X_2 之间的相关结构：当 $\alpha = 1$ 时，说明 X_1, X_2 之间是相互独立的，有下式 $C_1^{Gu}(v_1, v_2) = v_1 v_2$ 成立； α 越大， X_1, X_2 之间的相关性越强，直到 $\alpha \rightarrow \infty$ 时， X_1, X_2 完全相关。

图 2.1 表明二元分布（边际分布均为标准正态）运用 Gaussian Copula 和 Gumble Copula 这两个不同的相依结构所得到的结果。左图是具有 0.7 的相关系数但却倾向于同时产生 X_1 和 X_2 的极值的二元分布；右图是相关系数为 0.7 的标准二元正态分布。从这样的意义上说，左图对风险管理来说这是更危险的一个分布。基于相关性，这些分布不能被区分开，但他们明显具有不同的相依结构。二元正态是薄尾相关的，Gumbel 分布具有尾部相依⁷。

图 2.1 相关系数为 0.7, Gumble Copula 和 Gaussian Copula 模拟图

3. 多元极值 Copula 函数

用分块样本极大化的多元极值分布来研究 Copula 是理解多元极值理论的一种方式。设 X_1, X_2, \dots, X_n 是随机向量的集合，其中 $X_i = (X_{i1}, \dots, X_{id})^T$ 表示在不同时点上的损失。一个简单的解释是，它表示 d 维工具的负的日收益。至于对大样本单变量的讨论，假设它们在不同时点上的损失是独立的。这样的假设简化了结果，但在一些额外技术条件前提下，如果允许损失的序列相依，则可能再一次简化结果。

设 $X_i = (X_{i1}, \dots, X_{id})$ ， $i = 1, 2, \dots, n$ 是一个 d 维的独立同分布随机向量，定义 $M_{nj} = \max(X_{j1}, \dots, X_{jd})$ 为第 j 分量的最大值。考虑正则化的最大值向量 $((M_{1n} - b_{1n})/a_{1n}, \dots, (M_{dn} - b_{dn})/a_{dn})^T$ ，其中 $a_{jn} > 0$ ， b_{jn} 是正则化实数序列。如果向量是依分布收敛到一个非退化的分布函数，则多元极值分布一定具有如下的形式：

$$C(H_{\xi_1}(\frac{x_1 - \mu_1}{\sigma_1}), \dots, H_{\xi_d}(\frac{x_d - \mu_d}{\sigma_d})) \tag{2.9}$$

其中，参数 ξ_j ， μ_j 和 σ_j 取某一特定值， C 是一个 Copula， H 为极值分布函数。这类 Copula 称为多元极值分布 Copula。根据 Joe (1997) 的定义，满足以下关系式的 Copula 函数称为极值 Copula 函数：

$$C(u_1^t, \dots, u_d^t) = C^t(u_1, \dots, u_d) \quad , \quad t > 0 \tag{2.10}$$

考察极值 Copula 函数与多元极值理论之间的联系可以发现，他们之间有简单明了的关系：

$$G(x_1^+, \dots, x_n^+, \dots, x_N^+) = C(G_1(x_1^+), \dots, G_n(x_n^+), \dots, G_N(x_N^+)) \tag{2.11}$$

其中， G 为多元极值 $(x_1^+, \dots, x_n^+, \dots, x_N^+)$ 的联合分布， G_i 为非退化的一元分布，而 C 则为极值

⁷Harry Joe. Multivariate Models and Dependence Concepts[M]. Chapman & Hall, 1997.

Copula 函数。没有一个单参数集合包含所有的多变量极值 Copula，但是一些参数 Copula 是满足上面条件的，因此可以作为研究极值相依结构的基础模型。在下面的研究中，我们只是研究二元极值分布和二元极值 Copula 的有关情况。

4. 广义 Pareto 分布

广义 Pareto 分布 (Generalized Pareto Distribution, 简记为 GPD) 是由 Pickands 在 1975 年首次介绍, Davison(1984)、Smith(1984, 1985)、Montfort & Witter(1985)做了进一步研究。它广泛应用于极值分析、拟合保险损失以及可靠性研究领域。广义 Pareto 分布的分布函数含有 ξ , β 两个参数:

$$G_{\xi, \beta}(y) = \begin{cases} 1 - (1 + \frac{\xi y}{\beta})^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp(-\frac{y}{\beta}) & \xi = 0 \end{cases} \quad (2.12)$$

其中, ξ 是重要的形状参数, 而 β 是分布的尺度参数, $\beta > 0$ 。当 $\xi \geq 0$ 时, $y \geq 0$; 当 $\xi \leq 0$ 时, $0 \leq y \leq -\beta/\xi$ 。 $\xi \geq 0$ 时, 广义 Pareto 分布是厚尾的, 这种情形与风险测量是最相关的。

三、二元 POT 模型与参数估计

1. 二元 POT 模型

极值 POT 模型则是对观察值中所有超过某一较大阈值(Threshold)的数据建模。由于 POT 模型有效地使用了有限的极端观察值, 因此通常被认为在实践中是最有用的。设 $F(x)$ 为股价收益率损益分布函数, u 为阈值, $z-u$ 表示超额数, 其超额分布函数记为:

$$F_u(y) = P(X-u \leq y | X > u), \quad 0 \leq y \leq x_0 - u \quad (3.1)$$

其中, $x_0 \leq \infty$ 是 F 的右端点, 超额分布函数表示损失超过阈值的概率, 通过最大的 y 值给出了超过阈值的信息。

显然,

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} \quad (3.2)$$

进而能够得到

$$F(z) = F_u(y)(1 - F(u)) + F(u), \quad z \geq u$$

Pickands 定理 对于一大类分布 F (几乎包括所有的常用分布) 的条件超额分布函数 $F_u(y)$, 存在一个 $G_{\xi, \beta}(y)$, 使得

$$F_u(y) \approx G_{\xi, \beta}(y) = \begin{cases} 1 - (1 + \frac{\xi}{\sigma} y)^{-1/\xi}, \xi \neq 0 \\ 1 - \exp(-y/\sigma), \xi = 0 \end{cases} \quad (3.3)$$

即对于充分大的阈值 u , 超额数的分布函数可以用广义 Pareto 分布近似。如果我们能够找到适

宜的阈值 u ，利用最大似然估计方法估计出广义 Pareto 分布 $G_{\xi, \beta}(y)$ 的参数 ξ, β ，并将所有估计量代入上式，当 $z > u$ 时，得到 $F(z)$ 的尾部估计，由 (2.3) 式知：

$$F(z) = (1 - F(u))G_{\xi, \beta}(y) + F(u) = \begin{cases} 1 - \frac{N_u}{n} \left(1 + \xi \frac{z-u}{\hat{\beta}}\right)^{-\frac{1}{\xi}}, \xi \neq 0 \\ 1 - \frac{N_u}{n} \exp\left(-\frac{z-u}{\sigma}\right), \xi = 0 \end{cases} \quad (3.4)$$

2. 阈值 u 的选取和 GPD 参数估计

将所有样本值按从大到小降序排列得顺序统计量 $x_1 > x_2 > \dots > x_n$ ，阈值 u 为介于 x_1 和 x_n 之间的任一统计量。定义平均超额函数为：

$$e(u) = E(x - u | x > u) \quad (3.5)$$

则样本超额函数为：

$$e_n(u) = \frac{\sum_{i=1}^n (x_i - u)^+}{N_u} \quad (3.6)$$

上式中分子表示对任意 u ，在 $x_i > u$ 的条件下， $x_i > u$ 的总和，分母 N_u 表示 $x_i > u$ 的样本个数。

对于 GPD 模型，其平均超额函数为： $e(u) = \frac{\beta + u\xi}{1 - \xi}$ ，显然， $e(u)$ 是关于 u 的线性函数，因而在数轴上两者的关系应是呈线性变化的，根据这一特点可以得到阈值的估计方法。将 u 设为横轴， $e(u)$ 设为纵轴。作散点图，当观测到超过某一临界值 u 后， $e(u)$ 呈明显的线性变化时，则我们可定此 u 值为阈值。也可以利用 Q-Q 图，Hill 图等其他方法来确定阈值。

GPD 参数 ξ, β 的估计方法有极大似然估计方法、矩估计方法等。下面给出符合 GPD 的样本 $\{z_1, z_2, \dots, z_n\}$ 的对数似然函数 $L(\xi, \sigma | z)$ ：

$$L(\xi, \sigma | z) = \begin{cases} -n \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \left(1 + \frac{\xi}{\sigma} y_i\right), \xi \neq 0 \\ -n \ln \sigma - \frac{1}{\sigma} \sum_{i=1}^n y_i, \xi = 0 \end{cases} \quad (3.7)$$

Gumbel Copula 可以用来对二元模型的尾部进行估计。考虑两个风险因子 $(X_1, X_2)^T$ ，它们的联合分布 F 未知，对某个 Copula C ，他们的边际分布为 F_1 和 F_2 ，满足

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

如果有来自这个联合分布的 n 对样本数据, 可以应用前面介绍的一元 POT 模型, 拟合两个边际分布的尾部, 得到尾部估计式:

$$\hat{F}_i(x) = 1 - \frac{N_{u_i}}{n} \left(1 + \hat{\xi}_i \frac{x - u_i}{\hat{\beta}_i}\right)^{-\frac{1}{\hat{\xi}_i}}, x > u_i, i = 1, 2$$

我们用 Gumbel Copula 描述超过阈值的观测值的相依结构, 其中, $\hat{\beta}$ 是相关参数 β 的估计值, 可以通过最大似然方法得到。于是联合分布的尾部估计为:

$$\hat{F}(x_1, x_2) = C_{\hat{\beta}}^{Gu}(\hat{F}_1(x_1), \hat{F}_2(x_2)), x_1 > u_1, x_2 > u_2 \quad (3.8)$$

这是最简单的二元 POT 模型。

四、沪深股市风险的相关性分析

众所周知, 中国股票市场还处于不断发展, 不断完善的过程中, 但其运动规律仍然具有西方成熟股票市场的一些特征, 如市场波动的“聚集性”和“持续性”, 即大的波动后面往往紧跟着大的波动, 小的波动后面往往紧跟着小的波动。由于波动与风险是并存的, 那么大波动必然伴随着高风险, 相应小波动随之就会有低风险。无论是什么样的风险, 对于股票投资者来说都是一个挑战。股票市场也会出现比较大的动荡, 因而研究尾部的分布比整个分布更有必要, 用极值方法分析市场风险, 也就是股市暴涨、暴跌时的收益也就更为有意义。为了便于区分, 把股市上涨称为收益, 股市下跌称为损失。通常我们对损失比较关注, 它关系着企业、个人的经济利益以及企业的生存发展, 因而有必要更深入的分析股市的这种相关性。

1. 条件分布

应用条件模型 $X|Y=y$ 对数据进行拟合, 分析当给定相关变量 Y 值时相应 X 的条件分布。有两种方法来讨论条件分布: 一是基于联合分布模型, 另一是直接的条件模型。我们只是采用基于联合分布模型的条件分布进行估计。

假定边际分布服从 GEV 分布, 二元极值 Logistic 型描述了 X 与 Y 之间的相关性, 由式(2.3)、(2.8) 容易得到 (X, Y) 的联合分布:

$$F(x, y) = \exp\left(-\left[\left\{1 + \frac{\xi_x(x - \mu_x)}{\sigma_x}\right\}^{-\alpha/\xi_x} + \left\{1 + \frac{\xi_y(y - \mu_y)}{\sigma_y}\right\}^{-\alpha/\xi_y}\right]^{1/\alpha}\right) \quad (4.1)$$

其中, $1 \leq \alpha \leq \infty$

条件分布表达式为:

$$F_{X|Y}(x, y) = \int_{-\infty}^x f(u, y) / f_y(x) du \quad (4.2)$$

$f(x, y)$ 为 $F(x, y)$ 的联合密度函数, 称这种条件分布为 Logistic 条件分布。

2. 数据分析和相关性分析

本节就上海证券交易所与深圳证券交易所极值情况的相关性实例进行分析。其主要目的是针对 $X|Y$ 条件分布模型进行比较, 使我们更好的研究深圳和上海这两个股市波动的相关性。

由于研究的是极值问题, 因而需要采用阈值法对数据进行处理。我们对数据 X, Y 进行变换, 取它的相反数, 这样可以直接应用极值理论对数据进行分析。

3. 参数的确定

对比图 4.1 和图 4.2, 我们确定阈值 $u = 0.0174$, 超过 u 的样本的个数为 186。用极大似然估

计法来估计参数 $\xi = 0.2867022$, $\beta = 0.008266357$, 相应的标准差以及 t 统计量见表 4.1。

对与深市仍采用 Q-Q 图和 Hill 图来确定阈值, 得到阈值 $u = 0.020664$, 超过 u 的样本的个数为 157。用极大似然估计法来估计参数 $\xi = 0.3881023$, $\beta = 0.0088427$, 具体结果见图 4.3 和图 4.4 及表 4.2。

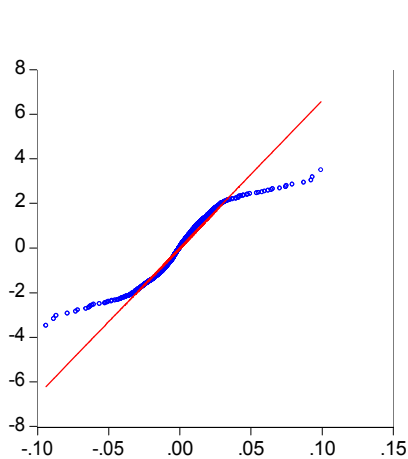


图 4.1 Q-Q 图

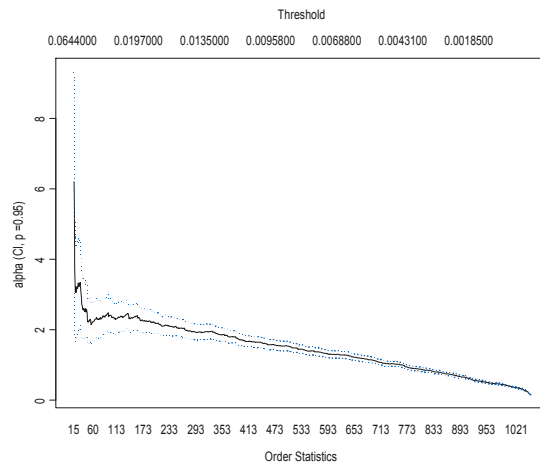


图 4.2 Hill

表 4.1 静态情况下参数估计结果

| 参数 | 估计值 | 标准差 | t 统计量 |
|-------|-------------|---------------|-----------|
| Beta | 0.008266357 | 0.00009543792 | 23.877779 |
| Kesai | 0.2867022 | 0.09162471 | 3.129093 |

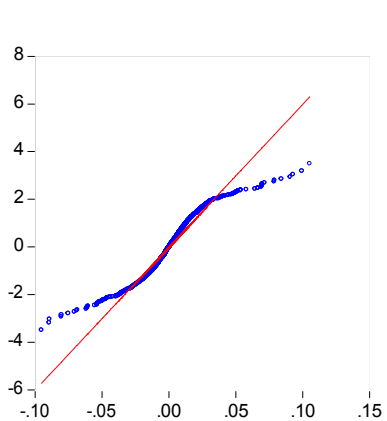


图 4.3 深证 Q-Q 图

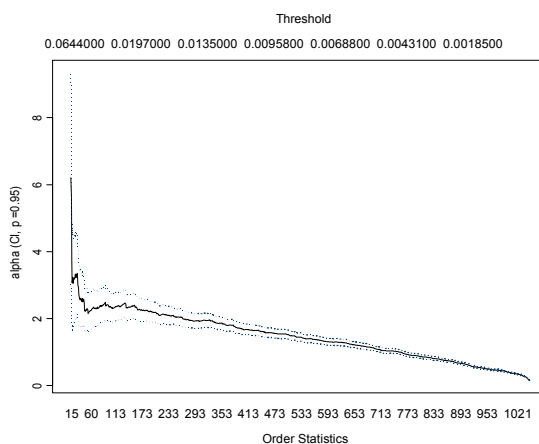


图 4.4 深证 Hill 图

表 4.2 参数估计结果

| 参数 | 估计值 | 标准差 | t 统计量 |
|----|-----|-----|-------|
|----|-----|-----|-------|

| | | | |
|-------|-----------|------------|--------|
| Beta | 0.0088427 | 0.00128764 | 6.8674 |
| Kesai | 0.3881023 | 0.1259446 | 3.0815 |

类似地, X, Y 选取的最大阈值为 0.020664, 得到上海和深圳超过这一阈值的数据分别为 131 个和 157 个, 两组数据超出阈值部分没有时间趋势, 故可以看作是独立同分布的样本。

下面比较当深圳股市出现最大损失时, 上海股市相应的条件分布。本文使用 S-Plus 软件对数据进行分析。

首先, 应用 Logistic 条件分布进行分析。假定边际分布满足 GEV 分布, 由极大似然估计得到参数向量的估计值, 见表 4.3。

表 4.3 X、Y 的边际分布参数估计

| | μ | σ | ξ |
|---|-----------|-----------|---------|
| X | 0.0243518 | 0.0043383 | 0.80927 |
| Y | 0.0247601 | 0.0049993 | 0.83570 |

其次, 由条件分布式 (4.2) 来研究深圳股市出现最大损失时, 上海股市也出现最大损失的概率, 也就是深证下跌幅度最大时, 上证也出现最大损失的 Logistic 条件分布。应注意的是, 我们的条件分布是计算当深证的最大损失为 $y = 0.105258$ 时, 计算上海同样出现下跌最大损失 $x = 0.099436$ 的概率。

选取上海和深圳股市中阈值最大的数 0.020664 作为积分下限, 积分上限取 1, 由式 (4.1) 和 (4.2) 计算得, 深圳股市出现最大损失时, 上海股市也出现最大损失的概率为 0.715445。

因此, 上海和深圳股市日收盘对数收益在极值处有很强的相关性, 即当一个股市出现极值时, 另一个股市也出现极值的趋势, 由此可以知道上海和深圳股市波动的总体趋势基本相同。

五、结 论

本文介绍了多变量极值分布和极值分布的 Copula 函数, 在此基础上给出了二元 POT 模型, 对联合分布的尾部进行估计。文章将多变量 EVT 和 Copula 相结合进行沪、深股市的相关性实证研究。通过运用 Logistic 模型研究得出, 在深市出现最大损失时, 沪市也出现最大损失的概率为 0.715445。

参考文献:

- [1] Fermanian, Jean-David, Scaillet O. Some statistical pitfalls in Copula modeling for financial applications. FAME Working Paper, 2004.
- [2] Fermanian, Scaillet O. Nonparametric estimation of copulas for time series. Risk, 2002, 4 (5): 25~54.
- [3] Fisher R A, Tippett L H C. Limiting forms of the frequency distributions of the largest of smallest member of a sample. Proceeding of Cambridge Philosophical Society, 1928, 24:180~190.
- [4] Frank M J. On the simultaneous associativity of $F(x, y)$ and $x+y- F(x, y)$. Aequationes Math, 1979, (19): 194~226.
- Frechet M. Sur la loi de probabili de l'ecart. maximum. Ann. Soc. Polonaise Math, Cracow, 1927, (6): 93~116.
- [5] Frees E, Valdez E. Understanding relationships using Copulas. North American Actuarial Journal. 1998, 2(1): 1~25.
- [6] Genest C, Rivest L P. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association, 1993, 88: 1034~1043.
- [7] Gnedenko. Sur la distribution limite du maximum d'une serie aleatoire. Ann. Math, 1943, 44:423~453.

- [8] Gouriéroux C, Monfort A, Tognon C. Pseudo-maximum likelihood methods: theory. *Econometrica*, 1984, 52:681~700.
- [9] Gumbel E J. Bivariate exponential distributions. *J. Amer. Statist. Assoc.* 1960, (55): 549~555.
- [10] Hall White. Maximum likelihood estimation of misspecified models. *Econometrica*, 1982, 50: 1~25.
- [11] Hann L D. On regular variation and its application to the weak convergence of sample extremes. *Mathematical Centre Tracts*, 1970, 32: 523~528.
- [12] Jenkinson A F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 1955, 87: 158~171.
- [13] Jobson J D, Korkie B. Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, 1980, 75: 544~554.
- [14] Jordan J V & Mackay R J. Assessing Value at Risk for equity portfolios: implementing alternative techniques. Working Paper, George Washington University, Washington, D.C., 1995.
- [15] Jose A, Lopse. Regulatory evaluation of value at risk models. Wharton Financial Institute Center, working paper, 1996, 10.
- [16] J.P. Morgan. RiskMetrics- Risk Metrics monitor. Second Quarter, 1996.
- [17] Kearns P, Paqan A. Estimating the density tail index for financial time series. *The Review of Economics and Statistics*, 1997, 79(5):171~175.
- [18] Klugman S A, Parsa R. Fitting bivariate loss distributions with Copulas. *Insurance: Mathematics & Economics*, 1999, 24: 139~148.
- [19] Lindskog F. Modelling dependence with Copulas and applications to risk management. Working paper, 2000.
- [20] Malavergne Y, Sornette D. Testing the Gaussian Copula hypothesis for financial asset dependencies. Working paper, Institute of Geophysics and Planetary Physics, UCLA, California, 2001.
- [21] Patton A J. Modeling: time-varying exchange rate dependence using the conditional Copula. Working Paper of London School of Economics & Political Science, 2001.
- [22] Rank D J. Copulas in financial risk management, in a thesis submitted for the diploma in mathematical finance, D.f.C.E.U.o. Oxford, Editor, 2000.
- [23] Roberto De Matteis. Fitting Copula to data. Diploma thesis, IMU, Zurich, 2001.
- [24] 陈守东, 胡铮洋, 孔繁利. Copula 函数度量风险价值的 Monte Carlo 模拟. *吉林大学社会科学学报*, 2006, 46 (2): 85~91.
- [25] 史道济. 二元极值分布参数的最大似然估计与分布估计. *天津大学学报*, 1994, 27(3): 294~299.
- [26] 张尧庭. 连接函数 (Copula) 技术与金融风险分析. *统计研究*, 2002, (4): 48~51.

The Correlation Analysis of Shanghai and Shenzhen stock markets Based on the Copula binary Extreme Risk model

Chen Shou-dong Kong Fan-li Chi Xian-liang

(Center for Quantitative Economic of Jilin University, Business School of Jilin University, Changchun, 130012)

Abstract: This paper presents a multivariable extreme distribution and the distribution of Copulas extreme, and this paper introduces the binary POT model, estimating the Joint Distribution tail. The article will combine

multivariable EVT and Copula study the correlation of Shanghai and Shenzhen stock markets. Through the Logistic models, we find when Shenzhen markets have the largest loss, Shanghai stock market has also experienced the largest losses with a probability of 0.715445.

Key words: Multivariable EVT; Copula model; Binary POT model; Tail estimates

收稿日期: 2007 年 8 月 19 日

基金项目: 本文得到“吉林大学‘985 工程’项目”吉林大学经济分析与预测创新基地、04 年教育部重大项目(05JJD790005)资助、06 年国家社会科学基金项目(06JY010)资助。

作者简介: 陈守东, 男, 1955 年 1 月生, 汉族, 天津市蓟县人, 吉林大学数量经济研究中心、商学院财务系主任、教授, 博士生导师 博士学位 研究方向: 数量经济学, 通讯地址: 长春市前进大街 2699 号吉林大学商学院, 130012, 联系电话: 0431-5166334(办)13331666307(手机) 0431-5168766(传真), 电子邮件: chensd@jlu.edu.cn.
孔繁利(1973 —) 男 内蒙古通辽市人 吉林大学商学院数量经济学专业博士研究生。胡铮洋(1980 —) 男 吉林省长春市人 吉林大学商学院数量经济学专业博士研究生。