

已实现波动与日内价差条件下的 CVaR 估计^①

叶五一, 缪柏其

(中国科学技术大学统计与金融系, 合肥 230026)

摘要: 随着高频金融数据的获取, 已有很多基于高频数据的研究, 包括已实现波动率的估计及其分布特征分析等. 尝试结合日内高频数据和日收益率数据, 基于 Copula 方法分析了日收益率与“已实现”波动率以及日内价差之间的相依结构. 通过分象限对数据进行了 Copula 拟合, 给出了一类特殊数据的联合分布估计方法, 进而给出了已实现波动率和日内价差条件下的 CVaR 的估计方法. 最后基于中国股市上证综指和深证成指的高频收益率数据进行了实证分析, 并对两种条件下的 CVaR 方法进行了预测效果的比较, 实证结果表明已实现波动率条件下的 CVaR 预测效果更好.

关键词: Copula “已实现”波动率; 日内价差; CVaR

中图分类号: F830.59; O211.3 文献标识码: A 文章编号: 1007-9807(2012)08-0060-12

0 引言

风险管理在金融领域已经成了非常重要的课题, 而在险价值 VaR (value at risk) 为特定的投资组合提供了很好的风险度量, 准确的 VaR 估计方法非常重要, VaR 的估计方法有历史模拟方法、分析方法以及蒙特卡罗模拟方法等, Duffie 和 Pan^[1]、Dowd^[2] 等给出了关于 VaR 估计方法的系统的介绍. 关于 VaR 估计方法的改进有许多研究, Ait-Sahalia 和 Lo^[3] 给出了应用非参数方法进行 VaR 估计的方法, Wong 和 So^[4] 则给出了基于 GARCH 模型的参数估计方法, Fan 和 Gu^[5] 则给出了基于半参数方法的 VaR 估计方法, Engle 和 Manganelli^[6] 应用分位点回归的思想给出了 CAViaR 的估计方法. 国内学者也给出了 VaR 一些新的估计方法以及基于中国金融市场的实证分析. 王春峰等^[7] 应用 MCMC 方法给出了 VaR 一种新的估计方法, 并进行了相应的实证研究. 叶五一

等^[8] 则给出了应用 Bootstrap 方法的 VaR 估计方法. 余素红等^[9] 基于 GARCH 模型和 SV 模型对 VaR 进行了估计, 并对两个模型进行了预测效果的检验. 本文首先结合日收益率和日内高频收益率给出波动率的估计, 并在此基础上给出条件 VaR 的度量方法, 并进行预测效果的检验.

近年来, 由于存储技术和计算技术的进步, “高频”金融数据越来越容易获得, 许多学者就日内数据的基本特征进行了大量研究, Dacorogna^[10] 的专著给出了高频数据在金融领域的应用. 波动率的估计以及预测是金融风险度量和管理的基礎, 在金融领域有着非常重要的地位, 常用的有两种方法, 一种是低频数据波动率的估计方法, 包括 ARCH 类方法 (Engle^[11]) 和随机波动率 (SV) 模型^[9]. 另一种是基于高频数据的波动率估计方法, 主要是“已实现”波动率 (realized volatility) 估计方法. 该方法是针对高频时间序列而开发的全新的波动率度量方法. 在这种波动率度量方法中

① 收稿日期: 2009-07-23; 修订日期: 2009-11-28.

基金项目: 国家自然科学基金青年科学基金资助项目 (71001095; 70901067); 高等学校博士学科点专项科研基金资助项目 (20103402120010).

作者简介: 叶五一 (1979-), 男, 博士, 讲师. Email: wyue@ustc.edu.cn

避免了模型假定,而且计算简便,在金融研究领域和实际操作领域都有很广阔的应用前景。在估计“已实现”波动率时,需要确定抽样频率,Andersen 和 Bollerslev^[12]证实过去的外汇交易波动能很好地用 288 个每 5 min 取样的数据的变化平方和 (“已实现”波动率) 来估计。5 min 的抽样频率是在兼顾精确性和减少微观噪声之间的平衡,是用极大似然方法得到的理论最优频率。国内也有很多“已实现”波动率方面很好的研究,徐正国和张世英^[13]在“已实现”波动率的基础上提出更有效的调整“已实现”波动,针对调整“已实现”波动的长记忆性和“杠杆”效应建立 ARFIMAX 模型,并与 GARCH 模型以及 SV 模型比较了预测能力。施红俊等^[14]基于沪深股市中随机抽取的 30 支股票数据进行了实证分析。关于“已实现”波动率的分布性质,也已有很多相关的研究, Jacod 和 Protter^[15], Barndor - Nielsen 和 Shephard^[16] 的文章中研究了“已实现”波动率的渐近正态分布特性。Nielsen 和 Frederiksen^[17] 则研究了有限样本情况下对数“已实现”波动率的中心极限理论,指出在有限样本情况下,“已实现”波动率的对数具有渐近正态特性。然而,日收益率和“已实现”波动率之间的相依结构问题,在以往的文献中并没有涉及。通过对两者的相依结构进行分析,可以得到数据的联合分布以及条件 VaR 估计等。

描述市场之间相互关系的最直接简单的方法是线性相关系数,它主要描述的是服从多元正态分布或者椭圆分布的风险之间的关系。然而研究发现正态性假设与高频金融数据分布特点非常不相符,这就限制了其应用。另一种方法是秩相关系数,例如 Kendall 相关系数和 Spearman 相关系数。然而这些方法只能给出两个随机变量之间的某种相关关系,无法给出更深刻的相依关系。Copula 是利用样本数据和各种风险资产收益率的边缘分布来近似确定其联合分布的数学方法,是在构造多元联合分布以及随机变量间相关结构分析中常用的工具。Schweizer 与 Sklar^[18] 在 1983 年首先提出了 Copula 这个名词,近期 Copula 在金融领域中得到了很好的应用。Embrechts 等^[19] 把 Copula 引入到金融上以来,许多学者已经取得了较好的成果,

国内对 Copula 的研究刚刚开展起来,张尧庭^[20] 从理论上探讨了 Copula 在金融上应用的可行性,吴振翔等^[21] 探讨了 Copula 相依结构下资产的组合投资问题。李悦雷等^[22] 则基于极值相关分析方法对股指期货操纵防范进行了研究。叶五一等^[23] 应用 Copula 方法解决了价格日内波幅条件下的条件 VaR 估计问题。傅强和邢琳琳^[24] 则基于极值理论和 Copula 方法给出了条件 VaR 的新的估计方法。Ning 等^[25] 研究了收益率和已实现波动率之间的相依结构关系,并从尾部相依系数的角度进行了杠杆效应分析。本文将对“已实现”波动率条件下和日内价差条件下的 CVaR 进行估计,并比较和分析相应的预测效果。

1 “已实现”波动率的估计

考虑简单的连续时间过程

$$d \ln P(t) = \sigma(t) dw(t)$$

其中 $\sigma(t)$ 的平方 $\sigma^2(t)$ 表示瞬时波动率,且假定和维纳过程 $w(t)$ 独立。积分波动率定义为

$$\langle IV \rangle_t = \int_t^{t+1} \sigma^2(s) ds$$

为了对日内数据进行建模,首先假定在 t 交易日,交易价格数据是离散的。用 $P_{t,i}, i = 1, \dots, n_t$ 表示第 t 交易日的第 i 个交易价格数据, n_t 是第 t 交易日所观测的数据总数。可以定义“已实现”波动率为金融资产日内高频收益率的平方和,由下式表出

$$\langle RV \rangle_t = \sum_{i=0}^{n_t} r_{t,i}^2 \quad (1)$$

其中 $r_{t,i} = \ln P_{t,i} - \ln P_{t,i-1}$ 为日内高频收益率。

为了得到条件 VaR 随着日内“已实现”波动率的变化情况,在本文的分析中,将对收益率为正时 ($r_{t,i}, \sqrt{\langle RV \rangle_t}$) 与收益率为负时 ($-r_{t,i}, \sqrt{\langle RV \rangle_t}$) 分别进行 Copula 相依结构分析,并通过推导得到全部数据的联合分布,其中 $-r_{t,i}$ 表示日收益率的相反数。

当在实际数据中存在微观噪声的影响时,“已实现”波动率并不是积分波动率的一致估计。当抽样频率过高时,微观噪声会引起“已实现”波

动率估计的偏差. Anderson 等^[26] 主张选定比较低的取样频率来减少偏差, 比如每 5 min 或者 15 min 取样 1 次, 而不是每秒取样. 这种方法称为稀疏抽样 (sparse sampling). 由于本文的重点不是抽样频率的确定, 在本文的分析中, 按照 Andersen 和 Bollerslev^[12] 的结论, 选取 5 min 的抽样数据进行“已实现”波动率的估计. 同时基于李胜歌和张世英^[27] 给出的抽样方法, 采用不同的频率进行“已实现”波动率的估计, 并进行相应的比较和分析.

2 基于 Copula 的 CVaR 估计

p 维总体 X 的联合分布可以唯一确定边缘分布, 反过来, 边缘分布不能唯一确定联合分布. Copula 就是由边缘分布来构造联合分布的一个 p 维函数, 其中联合分布可以有若干个参数. 设 C 为 Copula 函数, 则 $C(F_1(x_1), \dots, F_p(x_p))$ 就确定了 (X_1, \dots, X_p) 的联合分布, 它们的边缘分布为 $F_i, i = 1, 2, \dots, p$.

本文的研究目的是分析中国两个股票市场日收益率与“已实现”波动率之间的相依结构, 这就涉及到了二元 Copula 的应用. 阿基米德 (Archimedean) Copula 能够包含大多数常见的二元 Copula 结构, 因此在本文的分析中, 仅针对二元阿基米德 Copula 族进行分析.

2.1 阿基米德 Copula

设 $\varphi: [0, 1] \rightarrow [0, \infty]$ 为连续、严格递减的凸函数, $\varphi(0) = \infty, \varphi(1) = 0$, 且有连续、严格递减、凸的逆函数 $\varphi^{-1}: [0, \infty] \rightarrow [0, 1], \varphi^{-1}(\infty) = 0, \varphi^{-1}(0) = 1$, 则

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v))$$

称为由 $\varphi(\cdot)$ 生成的阿基米德 Copula, $\varphi(\cdot)$ 称为该 Copula 的生成函数.

$\varphi(t)$ 的不同形式决定了不同的阿基米德 Copula 族, 较常见的有 20 类, 对于每一类阿基米德 Copula, 可以根据样本数据来估计出参数 γ 的值. 在不同的 Copula 类中, Kendall 秩相关系数 ρ_τ 的取值范围是不一样的, 第一步根据不同的样本 Kendall 秩相关系数来确定阿基米德 Copula 候选

类. Genest 和 MacKay^[28] 给出 ρ_τ 和阿基米德 Copula 生成函数之间的关系为

$$\rho_\tau(X, Y) = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt$$

进一步, 由上式可以给出参数 γ 的矩估计. 而且可以应用文献 [30] 中给出的 3 种拟合优度检验方法在候选类中确定 1 个合适的类, 确定哪个 Copula 类能够最好地拟合样本数据.

2.2 一类特殊数据的联合分布估计

有一类特殊数据, 其散点图在第 1、2 象限具有一定的对称性, 全体数据之间的相关性并不显著, 而在两个象限内则相关性较大. 直接对全部数据应用 Copula 进行联合分布估计, 效果不是很好. 因此, 需要对两个象限的数据分别进行估计, 进而得到数据的联合分布.

假定二维随机变量 (X, Y) , 其中 $X \in R, Y \geq 0$, 其观测值为 $\{(X_i, Y_i)\}, i = 1, 2, \dots, n$. 将数据按照 X 分量的正负进行分类, 可以得到两组数据

$$\{(X_{1i}, Y_{1i})\}, i = 1, 2, \dots, n_1, X_{1i} \geq 0$$

$$\{(X_{2i}, Y_{2i})\}, i = 1, 2, \dots, n_2, X_{2i} < 0$$

为了分析方便, 对第 2 组数据的 X 分量求相反数, 得到变换后的数据为 $\{(-X_{2i}, Y_{2i})\}, i = 1, 2, \dots, n_2, X_{2i} < 0$. 此时两组数据具有类似的性质, 而且拟合效果要比直接拟合原来数据要好, 可以分别应用 Copula 方法对其联合分布进行拟合.

假定第 1 组数据, 经过拟合优度检验, 选择的最优阿基米德 Copula 的生成函数为 $\varphi_{1,\gamma}$, 则其联合分布可以表示为

$$\begin{aligned} P(X \leq x, Y \leq y | X \geq 0) &= F_1(x, y | x \geq 0) \\ &= C_1[G_1(x), H_1(y)] \\ &= \varphi_{1,\gamma}^{-1}[\varphi_{1,\gamma}(G_1(x)) + \varphi_{1,\gamma}(H_1(y))] \end{aligned}$$

其中 $G_1(x), H_1(y)$ 分别为第 1 组数据中随机变量 X 和 Y 的边缘分布函数.

对于变换后的第 2 组数据, 假定最优阿基米德 Copula 的生成函数为 $\varphi_{2,\eta}$, 则按照类似的方法, 其联合分布可以表示为

$$\begin{aligned} P(-X \leq x, Y \leq y | X < 0) &= F_2(x, y | X < 0) \\ &= C_2[G_2(x), H_2(y)] \\ &= \varphi_{2,\eta}^{-1}[\varphi_{2,\eta}(G_2(x)) + \varphi_{2,\eta}(H_2(y))] \end{aligned}$$

其中 $G_2(x)$ 、 $H_2(y)$ 分别为变换后第 2 组数据中随机变量 $-X$ 和 Y 的边缘分布函数。

经过简单变换可得

$$\begin{aligned} P(X \leq x, Y \leq y | X < 0) &= P(-X \geq -x, Y \leq y | X < 0) \\ &= P(Y \leq y | X < 0) - \\ &P(-X < -x, Y \leq y | X < 0) \\ &= H_2(y) - F_2(-x, y | X < 0) \\ &= H_2(y) - C_2(G_2(-x), H_2(y)) \end{aligned}$$

综合上述分析,由全概率公式,全部数据的联合分布可以表示为

$$\begin{aligned} P(X \leq x, Y \leq y) &= P(X \geq 0)P(X \leq x, Y \leq y | X \geq 0) + \\ &P(X < 0)P(X \leq x, Y \leq y | X < 0) \\ &= P(X \geq 0)F_1(x, y | X \geq 0) + \\ &P(X < 0)[H_2(y) - F_2(-x, y | X < 0)] \end{aligned}$$

即

$$F(x, y) = pC_1(G_1(x), H_1(y)) + (1-p)[H_2(y) - C_2(G_2(-x), H_2(y))] \quad (2)$$

其中 $p = P(X \geq 0)$; $G_1(x)$ 、 $H_1(y)$ 、 $G_2(x)$ 、 $H_2(y)$ 的定义如前,且由前面的分析可知,随机变量 Y 的分布在两种情况下相同,即 $H_1(y) = H_2(y)$,后面统一以 $H(y)$ 表示。

2.3 条件 VaR 的估计

VaR 的含义是处于风险中的价值,“VaR 是指在市场的正常波动下,在给定的置信水平下,某一金融资产或者证券投资组合在未来的特定的一段时间内的最大的可能的损失。”

假定 Y_t 是某债券或者投资组合的价格过程, X_t 是状态过程或者是信息向量,在实际应用中,包括股票指数、债券收益率、流动性指标以及其他的能够影响价格过程的因素。

在时期 $[t, t+h)$ 中,组合的对数收益率为 $r_t^h = \ln Y_{t+h} - \ln Y_t$ 。按照 VaR 的定义,置信水平为 $p(0 < p < 1)$ 的 VaR 为

$$\text{VaR}(p) = \inf_V \{ P(r_t^h \leq V) \geq 1 - p \}$$

假定 r_t^h 在 X_t 条件下的分布函数为 $F(\cdot | X_t)$, 那么条件 VaR 被定义为

$$\text{CVaR}(p) = F^{-1}(1 - p | X_t)$$

式中 $F^{-1}(\cdot | X_t)$ 为 $F(\cdot | X_t)$ 的反函数或者称为条件分位数函数; CVaR 表示条件 VaR。由定义可见 VaR 以及 CVaR 实际上分别为分位点和条件分位点的值。

由上面的分析可以得到随机向量 (X, Y) 的联合分布函数为

$$F(x, y) = pC_1(G_1(x), H(y)) + (1-p)[H(y) - C_2(G_2(-x), H(y))]$$

经过推导可以得到联合密度

$$\begin{aligned} f(x, y) &= \frac{\partial^2 F(x, y)}{\partial x \partial y} \\ &= p \frac{\partial^2 C_1(G_1(x), H(y))}{\partial G_1 \partial H} g_1(x) h(y) + \\ &(1-p) \frac{\partial^2 C_2(G_2(-x), H(y))}{\partial G_2 \partial H} \times \\ &g_2(-x) h(y) \end{aligned}$$

其中 $g_1(\cdot)$ 、 $g_2(\cdot)$ 、 $h(\cdot)$ 为相应随机变量的密度函数。

由 $X | Y = y$ 的条件密度函数

$$f_{X|Y}(x | Y = y) = \frac{f(x, y)}{h(y)}$$

可以得到 $X | Y = y$ 的条件分布函数为

$$\begin{aligned} F_{X|Y}(x, y) &= \int_{-\infty}^x \frac{f(t, y)}{h(y)} dt \\ &= p \frac{\partial C_1(G_1(x), H(y))}{\partial H} + \\ &(1-p) \left[1 - \frac{\partial C_2(G_2(-x), H(y))}{\partial H} \right] \end{aligned} \quad (3)$$

首先由全部数据得到参数 p 的经验估计,然后由两部分数据可以得到最优 Copula 函数,此时便可以确定联合分布函数以及条件分布函数,然后可以根据其反函数估计一定条件下的 CVaR。

2.4 CVaR 预测效果检验

得到 CVaR 的估计值以后,需要对其预测效果进行检验。常用的方法有 Kupiec^[29] 提出的似然比检验方法。

假定估计 CVaR 的置信水平为 α ,实际考察天数为 T ,失败天数为 N ,那么失败概率 p 的频率估计为 N/T ,准确性的 VaR 模型的失败概率 p 不应该显著不同于 p^* ($p^* = 1 - \alpha$) ,可以检验原假设 $H_0 : p = p^*$. Kupiec 提出了对零假设最合适的似然比检验 ,在零假设成立的条件下近似地成立

$$\langle LR \rangle = 2 \ln [(1 - N/T)^{T-N} (N/T)^N] - 2 \ln [(1 - p^*)^{T-N} (p^*)^N] \sim \chi_1^2 \quad (4)$$

本文还将同时采用 Clements 等^[30] 提出的检验方法进行 VaR 预测效果的检验 ,定义

$$L_{1,\alpha} = E[\alpha - I(r_t < VaR_t)] ,$$

$$L_{2,\alpha} = E[\alpha - I(r_t < VaR_t)](r_t - VaR_t)$$

假定用 T 天的数据进行预测 ,则上式的估计式分别为

$$\hat{L}_{1,\alpha} = \frac{1}{T} \sum_{t=1}^T [\alpha - I(r_t < VaR_t)] ,$$

$$\hat{L}_{2,\alpha} = \frac{1}{T} \sum_{t=1}^T [\alpha - I(r_t < VaR_t)] \times (r_t - VaR_t) \quad (5)$$

上面的两个损失函数中 $L_{1,\alpha}$ 表示实际失败率与期望失败率的差 $L_{2,\alpha}$ 则同时考虑了实际损失与 VaR 之间的差距 .上述统计量可以比较不同 VaR 估计方法的预测效果 .

为了两两比较估计方法 ,可以对 A、B 两种估计方法的损失函数 $L_{2,\alpha}$ 之差进行分析 ,定义

$$d_{t,\alpha} = [\alpha - I(r_t < VaR_{\alpha}^A)] [r_t - VaR_{\alpha}^A] - [\alpha - I(r_t < VaR_{\alpha}^B)] [r_t - VaR_{\alpha}^B]$$

可以在上述损失差 $d_{t,\alpha}$ 服从正态分布的假定下 ,对 A、B 两种方法的预测效果进行检验 ,即原假设为 $H_0 : d_{t,\alpha} = 0$.

定义统计量

$$Z = \frac{\bar{d}}{\sqrt{S^2/T}} \quad (6)$$

其中 \bar{d} 与 S^2 分别表示 $d_{t,\alpha}$, $t = 1, \dots, T$ 的样本均值和样本方差 ,则 $Z \sim t(T-1)$,即统计量 Z 服从自由的为 $T-1$ 的学生 t 分布 .基于 Z 统计量可以对原假设 $H_0 : d_{t,\alpha} = 0$ 进行假设检验 .

3 我国股票市场的实证分析

本文采用的研究数据如下: 1) 研究对象为中

国股票市场的上证综指与深证成指的高频数据 .

2) 数据样本的时间范围是 2004-1-2 ~ 2008-6-16 . 3) 数据样本的采样频率为 5 min . 4) 考虑到隔夜收益可能造成数据的伪自相关 ,所以剔除了每天记录的第 1 个观察值 . 本文将对我国上述两个股票市场指数的高频数据进行实证研究 ,分析日收益率与“已实现”波动率之间的相依结构 .

3.1 高频收益率、“已实现”波动率以及日内价差的描述性特征

在估计“已实现”波动率时 ,为了同时考虑估计的精确性和微观噪声的影响 ,采用 5 min 高频收益率进行估计 ,按照式(1)可以得到“已实现”波动率的计算结果 .

也可以用日内价差(price range) 度量波动率(叶五一等^[23]) ,为了使其标准化 ,可以定义

$$\langle PR \rangle_t = \frac{2(P_{\max,t} - P_{\min,t})}{(P_{\max,t} + P_{\min,t})}$$

这里 $P_{\max,t}$ 和 $P_{\min,t}$ 分别表示第 t 交易日的最高价和最低价 .

为了进行比较 ,也会相应给出日内价差与收益率的 Copula 相依结构以及日内价差条件下的 CVaR 估计结果 .

在对日收益率与“已实现”波动率以及日内价差之间的相依结构进行分析时 ,为了方便 ,一般取收益性的相反数进行分析 .首先在图 1 中给出中国股市两种指数 $(-r_t, \sqrt{\langle RV \rangle_t})$ 与 $(-r_t, \langle PR \rangle_t)$ 的散点图 .本文中市场风险用收益率 r_t 的相反数 $y_t = -r_t$ 来描述 ,即以 y_t 表示损失 , y_t 越大 ,损失越大 .

由图 1 可以看出 ,两种指数的散点图具有相似的特征 ,在两个象限内两者之间具有明显的相关特征 ,然而总体来看 ,相关性不大 .本文分别对第 1 象限和第 2 象限的数据分别应用 Copula 进行拟合 ,从而得到全部数据更加准确的联合分布 .由于两个象限的 Copula 结构相似 ,本文只对损失部分 ,即第 1 象限的数据进行重点分析 .

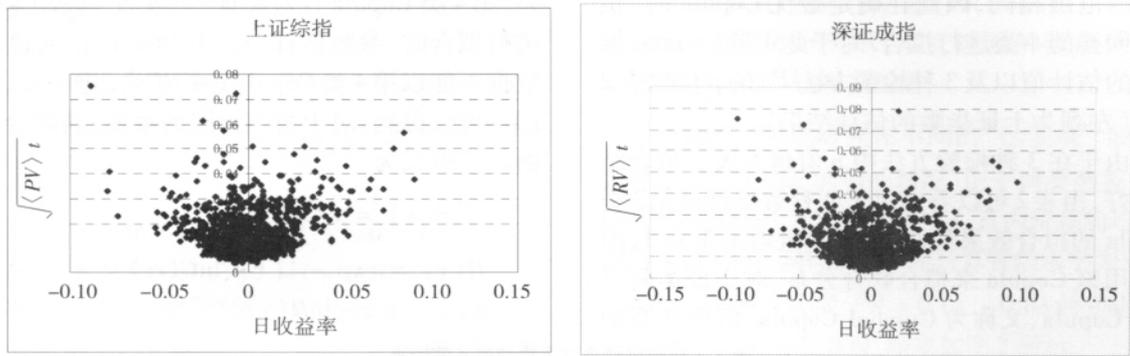


图 1a $(-r_t, \sqrt{\langle RV \rangle}_t)$ 的散点图

Fig. 1a Scatter figure of $(-r_t, \sqrt{\langle RV \rangle}_t)$

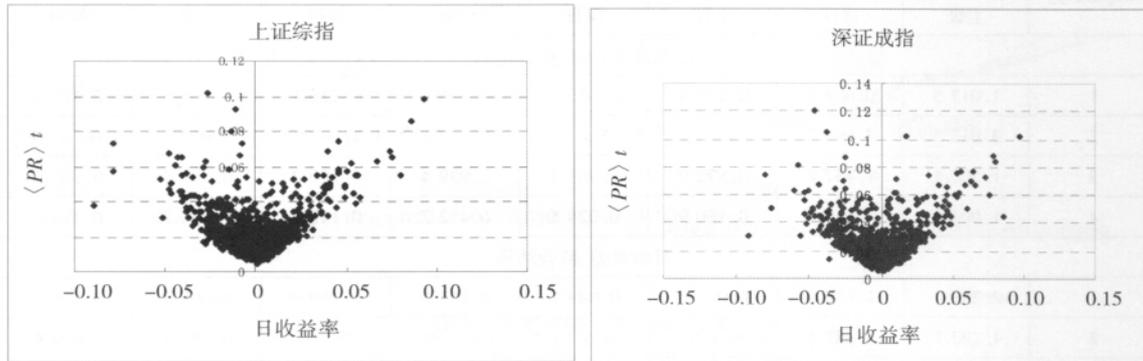


图 1b $(-r_t, \langle PR \rangle_t)$ 的散点图

Fig. 1b Scatter figure of $(-r_t, \langle PR \rangle_t)$

表 1 给出了两种指数第 1 象限中“已实现”波动率、日内价差的描述性特征以及与日收益率 (损失) 的 Kendall 秩相关系数, 以及 $p = P(X \geq 0)$ 的经验估计 \hat{p} .

表 1 $\sqrt{\langle RV \rangle}_t$ 与 $\langle PR \rangle_t$ 的描述性特征及 Kendall 相关系数

Table 1 Kendall correlation coefficients and descriptive characteristics of $\sqrt{\langle RV \rangle}_t$ and $\langle PR \rangle_t$

		\hat{p}	\hat{p}	均值	标准差	偏度	峰度	Kendall 相关系数
上证	$\sqrt{\langle RV \rangle}_t$	0.459 63	0.891 217	0.015 4	0.009 2	1.362 0	4.772 9	0.336 3 (-0.266 4)
深证		0.450 75	0.889 074	0.016 3	0.010 3	1.554 7	6.329 4	0.311 4 (-0.297 4)
上证	$\langle PR \rangle_t$	0.453 62	0.889 776	0.021 9	0.012 7	1.951 4	8.557 1	0.534 1(-0.435 4)
深证		0.464 75	0.892 415	0.024 1	0.014 5	1.941 9	8.579 3	0.484 9(-0.494 9)

注: 括号内的结果为正收益率与 $\sqrt{\langle RV \rangle}_t$ 以及 $\langle PR \rangle_t$ 之间(第 2 象限数据) Kendall 相关系数

3.2 Copula 相依结构分析

对于所研究的已实现波动率与日收益率数据, 第 1 象限的样本容量分别为 528 和 508. 经过计算, 所选样本已实现波动率与日损失数据之间的 Kendall 秩相关系数分别为

$$\hat{\rho}_\tau = \frac{1}{C_n^2} \sum_{i < j} \text{sign} [(x_i - x_j)(y_i - y_j)]$$

$$= 0.336 3; 0.311 4$$

由 Kendall 秩相关系数范围, 可以选择多种

Copula 类进行拟合, 由于多数 Copula 类的拟合效果都很差, 限于篇幅, 本文没有给出全部结果, 只给出下面 4 类阿基米德 Copula 的拟合结果, 其生成函数分别为

- 1) $\varphi_\gamma(t) = (t^{-\gamma} - 1) / \gamma$
- 2) $\varphi_\gamma(t) = (1 - t)^\gamma$
- 3) $\varphi_\gamma(t) = (-\ln t)^\gamma$
- 4) $\varphi_\gamma(t) = (1/t - 1)^\gamma$

日内价差与日收益率数据的 Kendall 秩相关

系数的范围相同,因此在确定最优 Copula 时,也选择同样的 4 类进行拟合. 对于此 4 类 Copula,参数 γ 的估计值以及 3 种检验方法^[25]的 p 值如表 2 所示(左列为上证指数的估计结果).

由于在 3 种检验方法中 p 值越大表示拟合结果越好,由表 2 可以看出对于两种数据都是第 3 类 Copula 的拟合效果最好,而且 p 值均大于 0.3,因此就用该 Copula 来拟合联合分布. 对于该类阿基米德 Copula,又称为 Gumbel Copula. 值得注意的

是,第 4 类 Copula 在对深证“已实现”波动率数据进行拟合时,参数估计值为 0.968 1,说明该部分数据不能以第 4 类 Copula 进行拟合. 进一步,由式 (3) 可以得到,对于第 1 象限的数据,其联合分布的具体形式为

$$F(rv, y) = C(G(rv)),$$

$$H(y) = \exp - \{ [(-\ln G(rv))^\gamma + (-\ln H(y))^\gamma]^{(1/\gamma)} \} \quad (7)$$

表 2 参数估计值以及拟合的 p 值结果

Table 2 Estimator of parameters and p -value of goodness of fit

Copula 类	γ		p (非参数方法)		p (Copula 分布方法)		p (Klugman-Parsa 方法)	
	上证	深证	上证	深证	上证	深证	上证	深证
已实现波动率 拟合结果								
1	1.013 5	0.904 3	0.180 9	0.040 7	0.425 6	0.216 1	0.908 5	0.165 2
2	3.013 5	2.904 3	4.16×10^{-27}	5.12×10^{-29}	3.54×10^{-24}	3.42×10^{-32}	0.012 2	9.00×10^{-6}
3	1.500 4	1.452 2	0.870 9	0.963 3	0.939 4	0.780 2	0.289 4	0.307 6
4	1.004 5	0.968 1	0.180 9	0.028 5	0.472 2	0.143 6	0.780 2	0.393 6
日内价差 拟合结果								
1	2.292 7	1.882 4	0.071 4	0.039 1	0.533 2	0.359 9	0.069 7	0.106 3
2	4.292 7	3.882 4	5.16×10^{-15}	1.01×10^{-17}	5.64×10^{-11}	1.79×10^{-8}	0.084 2	0.032 6
3	2.146 4	1.941 2	0.888 6	0.976 2	0.994 6	0.897 0	0.206 8	0.143 6
4	1.430 9	1.294 1	0.345 2	0.077 4	0.305 8	0.143 6	0.433 6	0.106 3

由表 2 中的拟合结果,可以看出日内价差与日损失数据之间的最优 Copula 也是常见的 Gumbel Copula,这和文献 [23] 中的结果相一致.

3.3 条件 VaR 估计

为了得到 CVaR 准确的估计,首先需要全部数据的联合分布函数进行估计. 按照第 3 部分的分析,首先分别得到两个象限数据的联合分布,

进而由全概率公式便可以得到全部数据的联合分布. 其中在分析时,边缘分布为经验分布函数. 经过与 3.2 部分类似的检验可知,其它的几类数据也是 Gumbel Copula 的拟合效果最好,只是参数估计结果有所不同. 下面首先给出当收益率分别为正和为负时, Gumbel Copula 的参数的估计结果.

表 3 不同象限数据的 Copula 参数估计

Table 3 Estimator of Copula parameters from different quadrants

	$(-r_t, \sqrt{\langle RV \rangle}_t)$	$(r_t, \sqrt{\langle RV \rangle}_t)$	$(-r_t, \langle PR \rangle_t)$	$(r_t, \langle PR \rangle_t)$
上证	1.500 370 5	1.363 141	2.146 383	1.771 165
深证	1.452 221 9	1.423 284 9	1.941 370 6	1.979 806

得到 Gumbel Copula 参数的估计以后,便可以得到各部分数据的联合分布函数,加上参数 p 的经验估计,进而可以得到全部数据的联合分布函数.

对于拟合最优的第 3 类 Gumbel Copula,由式 (4) 可以得出,全部数据的联合分布函数为

$$F(x, y) = p \exp\{ - [(-\ln G_1(x))^\gamma + (-\ln H(y))^\gamma]^{(1/\gamma)} \} + (1-p) \times [H(y) - \exp\{ - [(-\ln G_2(-x))^\gamma + (-\ln H(y))^\gamma]^{(1/\gamma)} \}] \quad (8)$$

由式 (6) 以及式 (8),可以得到条件分布函数为

$$F_{X|Y}(x|y) = p \frac{[-\ln H(y)]^{\gamma-1}}{H(y)} \times \exp(-A^{1/\gamma}) A^{\frac{1}{\gamma-1}} + (1-p) \times \left\{ 1 - \frac{(-\ln H(y))^{\eta-1}}{H(y)} \exp(-B^{1/\eta}) B^{\frac{1}{\eta-1}} \right\} \quad (9)$$

其中

$$A = [(-\ln G_1(x))^\gamma + (-\ln H(y))^\gamma],$$

$$B = [(-\ln G_2(x))^\eta + (-\ln H(y))^\eta]$$

条件分布式(9)没有解析表达的反函数,因此不能给出 CVaR 的显式表达式. 给定置信水平

p , 以及条件 $\langle RV \rangle = \langle rv \rangle$ 或 $\langle PR \rangle = \langle pr \rangle$, 可以通过数值方法得到 CVaR, 本文涉及的数值计算均采用 Matlab 工具编写完成.

下面本文首先研究在不同的“已实现”波动率条件下, 条件 VaR 的值的情况. 由于两类数据 \sqrt{RV} 的均值分别为 0.015 4 和 0.016 3, 日内价差的均值分别为 0.219 和 0.241. 本文首先计算置信水平为 95%, $\sqrt{RV} = 0.006 \sim 0.02$, $PR = 0.01 \sim 0.03$ 的变化范围内的条件 VaR 变化情况. 结果如图 2 所示.

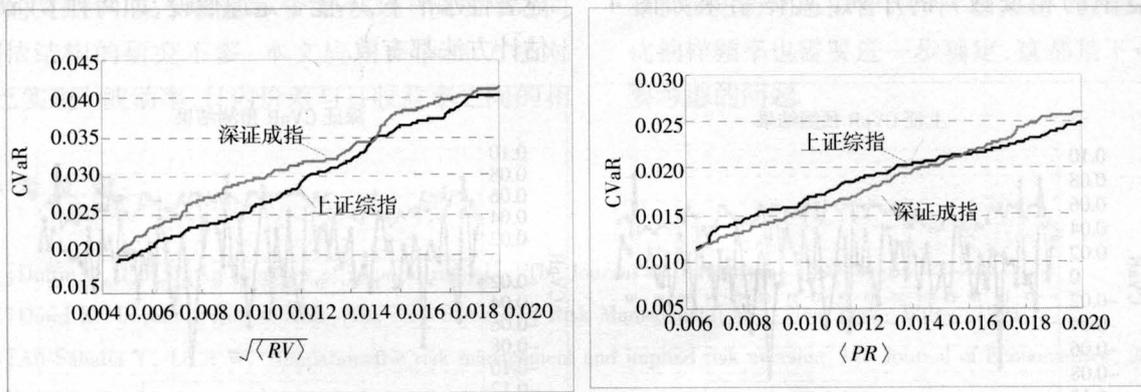


图 2 不同波动下 95% CVaR 值

Fig. 2 95% CVaR value under different volatility

图 2 给出了 95% 置信水平下 CVaR 值随着已实现波动率和日内价差的变化而变化的趋势, 可以看出在两种条件下 CVaR 的变化都具有相同的上升趋势, 即随着已实现波动率和日内价差的增大, CVaR 值也增大, 这和市场的实际情况相一致. 有一点不同的是, 已实现波动率条件下的 CVaR 变化趋势并不满足线性, 而日内价差的 CVaR 变化趋势则具有一定的线性现象.

为了找出在相同条件下在不同置信水平下的

CVaR 值的变化情况, 本文分别固定 $\sqrt{\langle RV \rangle} = 0.015 4, 0.016 3$, $\langle PR \rangle = 0.021 9, 0.024 1$ 给出置信水平由 0.9 到 0.999 时 CVaR 值. 结果如图 3 所示. 由图 3 可以看出, 随着置信水平的增大, CVaR 值逐渐增大, 这和一般的理论相一致. 开始阶段 CVaR 变化相对平缓, 但是随着置信水平 p 的增大, CVaR 的变化较快. 从图 3 可以看出, 对于两种条件下的 CVaR, 都是深证成指的 CVaR 较大.

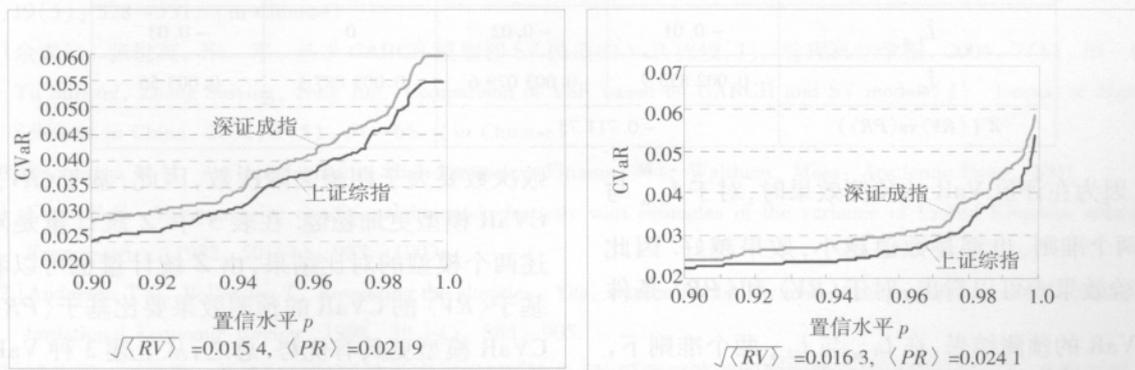


图 3 不同置信水平下 CVaR 值

Fig. 3 CVaR value under different confidence levels

投资者在进行组合投资时,如果能综合考虑资产的日收益率和“已实现”波动风险,则可以更准确地度量组合所暴露的市场风险. 这为投资者在进行风险分析以及构建投资组合时提供了很有用的信息.

3.4 VaR 预测结果

按照 CVaR 的计算方法,对最后 100 天的数据,给出了条件分别为 $\langle RV \rangle$ 和 $\langle PR \rangle$ 的 CVaR 预测结果. 为了进行比较,在图中同时给出了日损失(收益率的相反数)的序列,具体结果如图 4 所示.

由图 4 可以看出,在 $\langle RV \rangle$ 和 $\langle PR \rangle$ 两种条件下的 CVaR 预测结果变化趋势基本相同. $\langle RV \rangle$ 条件下的 CVaR 结果比 $\langle PR \rangle$ 条件下的大,而且波动性较小.

对于 Kupiec 提出的 LR 统计量的检验结果, $\langle LR \rangle$ 统计量的值越小,拟合效果越好. 由表 2 可以看出,对于两类 CVaR 估计方法,似然比统计量 $\langle LR \rangle$ 都不是很大,深证指数 PR 条件下的 CVaR 的预测结果 $\langle LR \rangle$ 最大,为 1.615 8,说明在 0.05 显著性水平下,不能否定原假设,即两种 CVaR 的估计方法都有效.

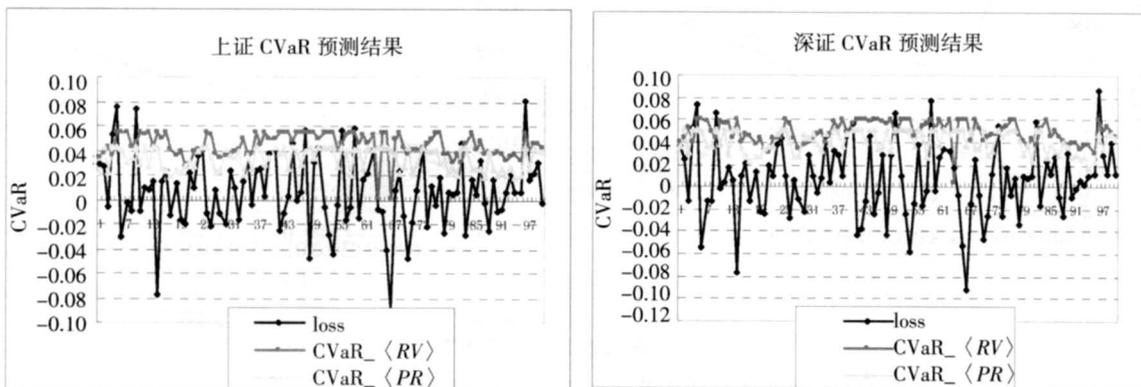


图 4 95% $\langle RV \rangle$ 与 $\langle PR \rangle$ 条件下 CVaR 预测结果

Fig. 4 Forecasting results of 95% CVaR conditioned on $\langle RV \rangle$ and $\langle PR \rangle$

表 4 VaR 预测效果检验以及两两比较结果表

Table 4 Testing of VaR forecasting effect and comparative results between $\langle RV \rangle$ and $\langle PR \rangle$

	上证		深证	
	CVaR_⟨RV⟩	CVaR_⟨PR⟩	CVaR_⟨RV⟩	CVaR_⟨PR⟩
失败次数	6	7	5	8
$\langle LR \rangle$	0.198 4	0.753 0	0	1.615 8
$\hat{L}_{1,\alpha}$	-0.01	-0.02	0	-0.03
$\hat{L}_{2,\alpha}$	0.002 823 9	0.003 079 6	0.002 887 1	0.003 56
$Z(\langle RV \rangle \text{ vs } \langle PR \rangle)$	-0.711 75		-2.17	

因为在评价 VaR 的预测效果时,对于 $\hat{L}_{1,\alpha}$ 与 $\hat{L}_{2,\alpha}$ 两个准则,也都是数值越小,效果越好. 因此由检验效果表可以看出:对于 $\langle RV \rangle$ 和 $\langle PR \rangle$ 条件下 CVaR 的预测结果,在 $\hat{L}_{1,\alpha}$ 与 $\hat{L}_{2,\alpha}$ 两个准则下,都是 $\langle RV \rangle$ 条件下的 CVaR 模型的预测效果要好. 本质上 $\hat{L}_{1,\alpha}$ 与失败次数是相同的, $\hat{L}_{1,\alpha}$ 为负说明失

败次数要大于期望失败次数,因此,基于 $\langle RV \rangle$ 的 CVaR 模型更加稳健. 在表 3 中 Z 统计量是对上述两个模型的对比结果,由 Z 统计量也可以看出基于 $\langle RV \rangle$ 的 CVaR 的预测效果要比基于 $\langle PR \rangle$ 的 CVaR 模型更具有优势. 总之,从上述 3 种 VaR 预测效果检验标准来看,基于 $\langle RV \rangle$ 的 CVaR 估计方法都具有一定的优势. 本文同时也按照李胜歌和

张世英^[27] 给出的频率确定方法,对“已实现”波动率进行了估计,然后给出条件 VaR 的估计. 实证结果发现,采用 1 min、2 min 抽样频率得到的结果与 5 min 的结果没有本质的区别.

4 结束语

关于“已实现”波动率本身的研究已有很多,包括抽样频率的选择、波动率的分布特征等. 在已有的文献中,关于收益率与“已实现”波动率之间相依结构的研究不多. 本文应用 Copula 方法对“已实现”波动率、日内价差与日收益率之间的相

依结构进行了分析. 同时针对一类特殊的数据,本文首先对数据进行分组,对两组数据分别应用 Copula 对联合分布进行估计,进而由全概率公式得到了全部数据的联合分布以及条件分布,并给出了 CVaR 的估计. 同时对中国股市数据进行了实证分析,经过 CVaR 预测效果检验,得出已实现波动率条件下的 CVaR 预测效果更好. 当然,本文的研究还存在很多不足,例如,需要考虑动态 Copula 函数的应用,需要对“已实现”波动率的分布特征进行动态拟合,另外“已实现”波动率的最优抽样频率也需要进一步确定,这都是下一步需要考虑的问题.

参考文献:

- [1] Duffie D, Pan J. An overview of value at risk [J]. *The Journal of Derivatives*, 1997, 4(3): 7-49.
- [2] Dowd K. *Beyond Value at Risk: The New Science of Risk Management* [M]. New York: Wiley, 1998.
- [3] Ait-Sahalia Y, Lo A W. Nonparametric risk management and implied risk aversion [J]. *Journal of Econometrics*, 2000, 94(1/2): 9-51.
- [4] Wong C M, So M K P. On conditional moments of GARCH models, with applications to multiple period value [J]. *Statistica Sinica*, 2003, 13(3): 1015-1044.
- [5] Fan J, Gu J. Semiparametric estimation of value at risk [J]. *Econometrics Journal*, 2003, 6(2): 261-290.
- [6] Engle R F, Manganelli S. CAViaR: Conditional autoregressive value at risk by regression quantiles [J]. *Journal of Business and Economic Statistics*, 2004, 22(4): 367-381.
- [7] 王春峰, 万海辉, 李刚. 基于 MCMC 的金融市场风险 VaR 的估计 [J]. *管理科学学报*, 2000, 3(2): 54-61.
Wang Chunfeng, Wan Haihui, Li Gang. Estimation of value-at-risk using MCMC [J]. *Journal of Management Sciences in China*, 2000, 3(2): 54-61. (in Chinese)
- [8] 叶五一, 缪柏其, 吴振翔. 基于 Bootstrap 方法的 VaR 计算 [J]. *系统工程学报*, 2004, 19(5): 528-531.
Ye Wuyi, Miao Baiqi, Wu Zhenxiang. Bootstrap method based evaluating VaR [J]. *Journal of Systems Engineering*, 2004, 19(5): 528-531. (in Chinese)
- [9] 余素红, 张世英, 宋军. 基于 GARCH 模型和 SV 模型的 VaR 比较 [J]. *管理科学学报*, 2004, 7(5): 61-65.
Yu Suhong, Zhang Shiyong, Song Jun. Comparison of VaR based on GARCH and SV models [J]. *Journal of Management Sciences in China*, 2004, 7(5): 61-65. (in Chinese)
- [10] Dacorogna M M. *An Introduction to High-Frequency Finance* [M]. Waltham, Mass: Academic Press, 2001.
- [11] Engle R F. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation [J]. *Econometrica*, 1982, 50(4): 987-1007.
- [12] Andersen T G, Bollerslev T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts [J]. *International Economic Review*, 1998, 39(4): 885-905.
- [13] 徐正国, 张世英. 调整“已实现”波动率与 GARCH 及 SV 模型对波动的预测能力的比较研究 [J]. *系统工程*, 2004, (8): 60-63.
Xu Zhengguo, Zhang Shiyong. The comparative on volatility prediction ability of adjusted realized volatility, GARCH mod-

- el and SV model[J]. *Systems Engineering*, 2004, (8): 60–63. (in Chinese)
- [14] 施红俊, 马玉林, 陈伟忠. 实际波动率理论及实证综述[J]. *山东科技大学学报(自然科学版)*, 2003, (3): 104–108.
- Shi Hongjun, Ma Yulin, Chen Weizhong. Summarization about the theory and empirical evidences of realized volatility[J]. *Journal of Shandong University of Science and Technology(Natural Science)*, 2003, (3): 104–108. (in Chinese)
- [15] Jacod J, Protter P. Asymptotic error distributions for the Euler method for stochastic differential equations[J]. *Annals of Probability*, 1998, 26(1): 267–307.
- [16] Barndorff-Nielsen O E, Shephard N. Econometric analysis of realised volatility and its use in estimating stochastic volatility models[J]. *Journal of the Royal Statistical Society B*, 2002, 64(2): 253–280.
- [17] Nielsen M Ø, Frederiksen P H. Finite sample accuracy and choice of sampling frequency in integrated volatility estimation[J]. *Journal of Empirical Finance*, 2008, 15(2): 265–286.
- [18] Schweizer B, Sklar A. *Probabilistic Metric Spaces*[M]. New York: North-Holland/Elsevier, 1983.
- [19] Embrechts P, McNeil A J, Straumann D. Correlation and dependence in risk management: Properties and pitfalls[M]// Dempster M A H. (Ed.). *Risk Management: Value at Risk and Beyond*, Cambridge: Cambridge University Press, 1999: 176–223.
- [20] 张尧庭. 连接函数(Copula)技术与金融风险分析[J]. *统计研究*, 2002, (4): 48–51.
- Zhang Raoting. Copula technology and analysis of financial risk[J]. *Statistical Research*, 2002, (4): 48–51. (in Chinese)
- [21] 吴振翔, 叶五一, 缪柏其. 基于 Copula 的外汇投资组合风险分析[J]. *中国管理科学*, 2004, 12(4): 1–5.
- Wu Zhenxiang, Ye Wuyi, Miao Baiqi. Risk analysis of foreign exchange markets by Copula[J]. *Chinese Journal of Management Science*, 2004, 12(4): 1–5. (in Chinese)
- [22] 李悦雷, 张 维, 熊 熊, 等. 基于极值相关分析方法的股指期货操纵防范研究[J]. *管理科学学报*, 2010, 13(1): 104–111.
- Li Yuelei, Zhang Wei, Xiong Xiong, et al. Investigation on prevention of manipulation in the stock index future markets based on method of extreme correlation[J]. *Journal of Management Sciences in China*, 2010, 13(1): 104–111. (in Chinese)
- [23] 叶五一, 缪柏其, 吴振翔. 基于 Copula 方法的条件 VaR 估计[J]. *中国科学技术大学学报*, 2006, 36(9): 917–922.
- Ye Wuyi, Miao Baiqi, Wu Zhenxiang. Estimating conditional VaR based on Copula method[J]. *Journal of University of Science and Technology of China*, 2006, 36(9): 917–922. (in Chinese)
- [24] 傅 强, 邢琳琳. 基于极值理论和 Copula 函数的条件 VaR 计算[J]. *系统工程学报*, 2009, 24(5): 531–537.
- Fu Qiang, Xing Linlin. Calculate the condition VaR by extreme value theory and Copula function[J]. *Journal of Systems Engineering*, 2009, 24(5): 531–537. (in Chinese)
- [25] Ning C, Xu D H, Wirjanto T S. Modeling the leverage effect with copulas and realized volatility[J]. *Finance Research Letters*, 2008, 5(4): 221–227.
- [26] Andersen T G, Bollerslev T, Diebold F X, et al. Modeling and forecasting realized volatility[J]. *Econometrica*, 2003, 71(2): 579–625.
- [27] 李胜歌, 张世英. 金融高频数据的最优抽样频率研究[J]. *管理学报*, 2008, 5(6): 801–808.
- Li Shengge, Zhang Shiyong. Optimal sampling frequency for high-frequency financial data[J]. *Chinese Journal of Management*, 2008, 5(6): 801–808. (in Chinese)
- [28] Genest C, Mackay J. The joy of Copulas: Bivariate distributions with uniform marginals[J]. *American Statistician*, 1986,

40 (2) : 280 - 283.

[29] Kupiec P. Techniques for verifying the accuracy of risk measurement models [J]. *Journal of Derivatives*, 1995, 3(2) : 73 - 84.

[30] Clements M P, Galvão A B, Kim J H. Quantile forecasts of daily exchange rate returns from forecasts of realized volatility [J]. *Journal of Empirical Finance*, 2008, 15 (4) : 729 - 750.

[31] De Matteis R. Fitting Copulas to data [D]. Institute of Mathematics University of Zurich, 2001.

Estimating of CVaR with consideration of realized volatility and price range

YE Wu-yi, MIAO Bai-qi

University of Science and Technology of China, Hefei 230026, China

Abstract: There are lots of studies based on high-frequency data, including the estimating method and distribution character of realized volatility and so on. In this paper, the dependence structure between daily return and realized volatility and that between daily return and price range are analyzed based on Copula method from intraday high-frequency data and daily returns. Joint distribution of one special type of data is estimated based on distributions of different quadrants, and CVaR with consideration of realized volatility and price range is also calculated. An empirical analysis of high-frequency data from Shanghai and Shenzhen Stock Markets is presented. By comparing the forecasting results, the conclusion is obtained that the CVaR with consideration of realized volatility forecasts market risks better.

Key words: Copula; realized volatility; price range; conditional value at risk