

文章编号: 1002-1566(2019)06-1104-15
DOI: 10.13860/j.cnki.sltj.20190617-001

资产组合非等间隔日内在险价值研究

鲁万波¹ 陈骋² 王建业¹

(1. 西南财经大学统计学院, 四川 成都 611130; 2. 伦敦政治经济学院统计系, 英国 伦敦 WC2A 2AE)

摘要: 当前对资产组合在险价值 (VaR) 的研究仅限于等间隔抽样数据的建模。本文提出资产组合的非等间隔日内在险价值 (Irregularly Spaced Intraday Value at Risk, ISIVaR) 研究方法, 克服资产组合逐笔交易数据非等间隔且不同步问题, 利用逐笔交易数据所包含的丰富市场微观结构信息对 VaR 进行估计。该方法基于更新时间方法将非同步的资产组合标值序列同步化; 运用 Copula 理论建立资产组合的非等间隔日内波动率模型, 并捕捉资产组合中各资产在截面上的相关关系; 最后利用这种截面相关关系, 使用蒙特卡洛模拟技术估计出资产组合的 ISIVaR。实证部分利用真实的逐笔交易数据验证了上述方法的有效性。

关键词: 高频数据; 资产组合; 非等间隔时间序列; 日内在险价值; Copula

中图分类号: C81, O212

文献标识码: A

A Study on the Irregularly Spaced Intraday Value at Risk for the Portfolio Selection

LU Wan-bo¹ CHEN Cheng² WANG Jian-ye¹

(1. School of Statistics, Southwestern University of Finance and Economics, Sichuan Chengdu 611130, China, 2. Department of Statistics, London School of Economics and Political Science, London, WC2A 2AE, UK)

Abstract: This research first proposed the method to estimate the Irregularly Spaced Intraday Value at Risk (ISIVaR) and solve the problem result from the irregularly spaced and asynchronous multivariate tick-by-tick data. Firstly, this research makes use of the auto correlation duration model to fit the price durations of each single asset in the portfolio. Then, based on the duration, the assets' intraday volatility and ISIVaR is estimated. Next, by Fresh Time method, this research synchronizes the price events sequences of the portfolio. Then, the Copula theory is used to model the irregularly spaced intraday volatility in order to capture the cross-sectional correlation information between the assets in the portfolio. Finally, based on the cross-section correlation, the ISIVaR of the portfolio is estimated by Monte Carlo simulation method. At the end of this research, an empirical study is presented to validate the feasibility of the proposed method.

Key words: high frequency data, portfolio selection, irregularly spaced time series, intraday VaR, Copula

收稿日期: 2018 年 2 月 5 日

收修改稿日期: 2019 年 4 月 5 日

基金项目: 本文获得国家自然科学基金面上项目 (71771187), 国家自然科学基金青年项目 (71101118), 教育部新世纪优秀人才支持计划项目 (NCET-13-0961) 的资助, 中央高校基本科研业务费专项资金项目 (JBK190602)。

0 引言

随着全球各金融市场软、硬件条件不断地成熟和完善,投资者可以获得的市場数据已经达到了逐笔交易 (tick-by-tick) 的水平。通过迅速捕捉由超高频数据信息传递出的投资机会从而获取更高收益成为可能。同时,由于高频交易单笔交易收益低但长期收益稳定,并且为市场带来更多的流动性,因而程序化交易受到了前所未有的重视,越来越多的程序化交易者参与到高频交易中。早在 2009 年,美国股市已经有 70% 以上的交易量由高频交易完成^[1]。这也促使越来越多的投资者和学者研究这种超高频的数据处理方法及其背后所隐含的信息。

高频交易具有电脑程序化交易,交易频繁、持仓时间短的特点,因而流动性的欠缺迫使投资者,尤其是从事高频交易的投资者以更高的交易成本来布置仓位,甚至会直接抹掉套利空间。因此,在高频交易背景下,流动性风险管理的重要性达到甚至超过了收益率波动风险,成为了投资者关注的重点。对于市场风险的衡量,自从 Morgan 于 1980 年提出 VaR 以来已经被广泛地运用到了风险管理中,现已成为业界的标准。

在市场微观结构领域,传统的在险价值估计方法存在诸多的弊端,一些学者进行了扩展研究^[2-3],但这些方法的应用还仅限于低频数据。市場上的逐笔交易数据不仅包含了量、价信息,更为宝贵的是成交的时间这一反映交易频繁程度,或者说流动性的信息被完整地记录下来。这些数据信息使得在超高频视角下研究流动性调整的波动率与在险价值成为可能。但是,这些数据本身的非等间隔记录特性却阻碍了传统方法在资产日内收益率和波动率上的直接应用。在此背景下,Engle 和 Russell (1998)^[4] 提出使用自回归条件持续时间 (Autoregression Conditional Duration, ACD) 模型来刻画股票交易的持续时间及其对收益率、波动率的影响,打开了利用非等间隔时间序列数据捕捉市场微观结构信息的大门。Colletaz, Hurlin 和 Tokpavi (2007)^[5] 针对非等间隔数据提出了一种定义在交易发生时刻基础上的市场风险测度以及后向检验方法,应用 ACD 模型和非参数分位数估计技术构建了一个半参数 ISIVaR 模型,并将该模型应用到美国纽约股票市场。随着市场逐笔交易数据的容易获得,当前关于日内波动率的研究不断涌现^[6-7]。

Gregor 和 Hendrik (2013)^[8] 开始对两个及以上资产进行研究,但分析的对象是等间隔数据或将非等间隔数据进行等间隔抽样。Engle 和 Russell (1998)^[4] 认为,若将非等间隔的数据以固定的频率抽样、汇总,使之成为等间隔数据,会产生两难的问题——若抽样频率过高,则会产生大量的没有新记录的观测点,从而在数据中人为地引入了异方差;相反,若频率过低,又会造成数据信息,尤其是市场微观结构信息的损失,这与使用超高频数据的初衷相背离。佐证之一是王峰和刘传哲 (2014)^[9] 对比了以高频数据为基础的日内 VaR (Intraday VaR, IVaR) 测度模型和使用等间隔高频数据的传统方法对上海燃料油期货市场数据进行研究,实证表明前者因为包含了更丰富的市场信息,在多头空头方都具有更好的一步向前预测能力。

总之,逐笔交易数据包含了丰富的市场微观结构信息,应当予以重视。因此,如何对多个资产非等间隔、非同步数据进行分析,目前还没有一个成熟的方法。针对这个问题,本文给出资产组合的非等间隔日内在险价值 (Irregular Spaced Intraday Value at Risk, ISIVaR) 的研究思路和解决办法。

1 单资产日内边际波动率及在险价值

1.1 价格持续时间模型

(1) 标值点过程。设 t 为日历时间, $\{t_i\}_{i \in \{1,2,\dots\}}$ 是某概率空间 $(\Omega, \mathcal{F}, P)$ 的非负随机到达

时间序列, 则序列 $\{t_i\}$ 为 $[0, +\infty)$ 上的一个点过程。用 x_i 表示两个连续事件的时间间隔:

$$x_i = \begin{cases} t_i - t_{i-1}, & i = 2, 3, \dots, \\ t_1, & i = 1, t_0 = 0, \end{cases} \quad (1)$$

则 $\{x_i\}_{i \in \{1, 2, \dots\}}$ 称为与点过程 $\{t_i\}$ 相关的持续时间过程。用 $\{z_i\}_{i \in \{1, 2, \dots\}}$ 表示与随机到达时间 $\{t_i\}_{i \in \{1, 2, \dots\}}$ 相对应的随机向量, 则 $\{z_i\}_{i \in \{1, 2, \dots\}}$ 称为标值序列, $\{t_i, z_i\}_{i \in \{1, 2, \dots\}}$ 形成一个标值点过程 (Market Point Process)。

某只股票的逐笔交易数据可视为由某标值点过程 $\{t_i, Q_{t_i}, p_{t_i}, z_i\}_{i \in \{1, 2, \dots, N\}}$ 生成, N 为样本容量。 t_i 为第 i 笔交易发生的时间, Q_{t_i} 为 t_i 上的成交股数, p_{t_i} 为成交价格。现实中, 由于交易系统记录的时间精度, 同一时点 t_i 上可能存多笔成交。 z_i 为一个 $k \times 1$ 维向量, 包括了其它标值, 如成交量、成交方向、买卖价差等。

为了避免买卖价格跳跃 (bid-ask bounce) 效应以及同一时点多笔成交带来的“零间隔”影响, 定义成交量加权价格变量。加权价格计算公式为:

$$p_{t_i} = \frac{Q'_{t_i} p_{t_i}}{Q'_{t_i} \mathbf{1}}. \quad (2)$$

本文从这一标值点过程 $\{t_i, p_{t_i}, z_{t_i}\}_{i \in \{1, 2, \dots\}}$ 中抽取价格出现了一定累积变化的子集。这里的 price 变化是指加权价 p_{t_i} 变化超过了预先设定的阈值 d_p , d_p 为价格绝对累积变化。由此产生了一个新的标值点过程 $\{t_j, p_{t_j}, z_{t_j}\}_{j \in \{1, 2, \dots, N_C\}}$ 。该过程具体的生成规则是: 第 1 个点为原序列第 1 个点; 若 $|p_{t_i} - p_{t_{j-1}}| > d_p$, i ($i \geq 2$), 则第 j ($j \geq 2$) 个点为原序列中第 i 个点, N_C 为新序列的样本量。定义相应的收益率序列: $r_{t_j} = \ln p_{t_j} - \ln p_{t_{j-1}}$ 。新生成的序列体现了价格信息变化所需要的时间, 是一个价格持续时间过程。

(2) 自回归持续时间 (ACD) 模型。对于持续时间数据, 学界广泛地采用 Engle 和 Russell 提出的 ACD 模型^[4]。模型的基本思路是将持续时间的条件均值动态参数化。定义 $\varphi_i \equiv \varphi_i(\theta) = E(x_i | \mathcal{F}_i; \theta)$ 为持续时间的条件均值, 其中 $\mathcal{F}_i$ 包含了 $\{(t_k, p_{t_k}, z_{t_k}), k = 1, 2, \dots, i-1\}$ 的全部信息。类似于 GARCH 模型, 设 $\varphi_i = \omega + j = 1, p\alpha_j x_{i-j+1} = 1, q\beta_l \varphi_{i-l}$, 则 ACD 模型定义为:

$$x_i = \varphi_i \cdot \nu_i, \quad (3)$$

其中, ν_i 为独立同分布的正值分布且 $E(\nu) = 1$ 。在此基础上, 对于持续时间自相关结构的设定和误差项 ν_i 分布的不同选择衍生出了 ACD 模型簇: 若 ν 服从指数分布, 则得到 EACD 模型; ν 服从威布尔分布得到 WACD 模型^[4]; ν 服从广义 Gamma 分布则得到 GACD 模型^[10]等。

1.2 非等间隔日内波动率模型和在险价值

Colletaz, Hurlin 和 Tokpavi (2007)^[5] 给出 ISIVaR 的定义: 对于给定的损失概率 α , 第 j 笔交易价格变化持续时间的期望 φ_j 和相应在险价值 $R_j(\alpha)$ 的信息对 $(\varphi_j, R_j(\alpha))$, 并满足:

$$\Pr(r_{t_j}(\alpha) < R_j(\alpha) | \mathcal{F}_j) = \alpha, \quad (4)$$

其中 $\mathcal{F}_i$ 包含了 $\{(t_k, p_{t_k}, z_{t_k}), k = 1, 2, \dots, j-1\}$, 即第 j 笔价格发生变化的交易之前的全部信息。

不失一般性, 设收益率 r_{t_j} 的条件均值为 0, 标准差为 σ_{t_j} , 则 r_{t_j} 可写为:

$$r_{t_j} = \mu_{t_j} + \sigma(t_j|\mathcal{F}_{t_j}) \cdot \varepsilon_{t_j}, \quad (5)$$

其中 μ_{t_j} 是收益率均值, σ 是 t_j 时刻价格变化的波动率, 而 ε_{t_j} 是独立同分布的新息过程, 其分布为零均值、单位方差。因此, 第 j 笔价格变化的在险价值:

$$R_j(\alpha) = F^{-1}(\alpha) \cdot \sigma(t_j|\mathcal{F}_{t_j}) + \mu_{t_j}, \quad (6)$$

其中 $F(\cdot)$ 为 ε_{t_j} 的累积分布函数。而风险函数

$$\lambda(t|\mathcal{F}_t, N(t)) = \lim_{\Delta \rightarrow 0} \frac{p(N(t+\Delta) > N(t)|\mathcal{F}_t, N(t))}{\Delta}, \quad (7)$$

其中 $N(t)$ 为一个计数过程, $\mathcal{F}_t = \{(t_j, p_{t_j}, z_{t_j}), t_j < t\}$ 为 $\mathcal{F}_j$ 在连续时间情形下的信息。对于 ACD 模型, 上式可简化为:

$$\lambda(t|\mathcal{F}_t, N(t)) = \lambda_\nu\left(\frac{x(t)}{\varphi_{N(t)+1}}\right) \cdot \frac{1}{\varphi_{N(t)+1}}, \quad (8)$$

其中, $\lambda_\nu(\cdot)$ 为误差项 ν 的风险函数, $x(g) = t - t_{N(t)}$ 。在 ν 服从威布尔分布的情况下,

$$\lambda(t|\mathcal{F}_t, N(t)) = \lambda_\nu\left(\frac{\Gamma(1 + \frac{1}{\gamma})}{\varphi_{N(t)+1}}\right) \cdot x(t)^{\gamma-1}\gamma, \quad (9)$$

其中, $\Gamma(\cdot)$ 为 Gamma 函数, γ 是威布尔分布的形状参数, 当 $\gamma = 1$ 时, Gamma 分布退化为指数分布, 此时价格风险为常数 $1/\varphi_{N(t)+1}$ 。Engle 和 Russell (1998)^[4] 给出了指数分布情况下瞬时波动率与价格风险函数的关系:

$$\tilde{\sigma}^2(t|\mathcal{F}_t) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} E\left(\left(\frac{p_{t+\Delta} - p_t}{p_t}\right)^2 \middle| \mathcal{F}_t\right). \quad (10)$$

借助于计数过程 $N(t)$, 可用价格风险函数来表示瞬时波动率:

$$\tilde{\sigma}^2(t|\mathcal{F}_t) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \Pr(|p_{t+\Delta} - p_t| \geq d_p | \mathcal{F}_t) \left(\frac{d_p}{p_t}\right)^2 = \lambda(t|\mathcal{F}_t, N(t)) \left(\frac{d_p}{p_t}\right)^2. \quad (11)$$

由此可见, 瞬时波动率可用 ACD 模型对价格持续期建模而得到。事实上, 根据 Giot (2005)^[11] 的证明, 由上式所定义的 $\tilde{\sigma}(t|\mathcal{F}_t)$ 在 $t \in [t_{j-1}, t_j]$ 之间均成立, 所以有 $\sigma(t|\mathcal{F}_t) = \tilde{\sigma}(t|\mathcal{F}_t)$, 因此第 j 笔价格时间的条件波动率:

$$\sigma^2(t_j|\mathcal{F}_{t_{j-1}}) = \lambda(t_{j-1}|\mathcal{F}_{t_{j-1}}, N(t_{j-1})) \left(\frac{d_p}{p_{t_{j-1}}}\right)^2. \quad (12)$$

在实际应用中, 价格持续时间具有明显的日内模式, 通常的做法是假设持续时间 $x_j = \tilde{x}_j \omega(t_j)$, 其中 $\tilde{x}_j$ 是剔除日内模式后的持续时间, $\omega(t_j)$ 是 t_j 时刻的日内模式。ACD 模型适用于 $\tilde{x}_j$, 因此, 经日内模式调整的波动率为:

$$\sigma^2(t_j|\mathcal{F}_{t_{j-1}}) = \lambda(t_{j-1}|\mathcal{F}_{t_{j-1}}, N(t_{j-1})) \left(\frac{d_p}{p_{t_{j-1}}}\right)^2 \frac{1}{\omega(t_j)}, \quad (13)$$

其中, $\lambda(\cdot)$ 为价格时间的风险函数。因而第 j 笔价格事件的 VaR 为:

$$R_j(\alpha) = \mu_{t_j} + F^{-1}(\alpha)\sigma(t_j|\mathcal{F}_{t_j}) = \mu_{t_j} + F^{-1}(\alpha)\lambda(t_{j-1}|\mathcal{F}_{t_{j-1}}, N(t_{j-1}))^{\frac{1}{2}} \frac{d_p}{p_{t_{j-1}}} \left(\frac{1}{\omega(t_j)} \right)^{\frac{1}{2}}. \quad (14)$$

2 资产组合非等间隔日内在险价值

基于上述单个资产非等间隔日内波动率及在险价值的研究思路, 本节将单资产扩展为多资产, 给出多个资产非等间隔日内在险价值的研究方法。

2.1 非同步交易同步化

多个资产之间在时序和截面上的相关关系一直是经济、金融领域研究的重点与热点。但非同步交易的多元时间序列会直接阻碍截面相关关系的研究。在本文中, 价格事件以非等间隔、非同步的形式传递出新的价格。为了同时利用市场价格时间在时序和截面上的信息来估计各资产的波动以及在险价值, 首先需要将非同步的时间序列同步化。而目前使用较多的同步化方法是“更新时间 (Fresh Time)”方法。Barndorff 等 (2010)^[12] 给出更新时间的完整定义及其算法:

设组合中共有 d 个资产, 其中第 k 个资产价格事件在前文为计数过程 $N^{(k)}(t)$, 即第 k 个资产的价格事件出现在 $t_1^{(k)}, t_2^{(k)}, \dots, t_n^{(k)}, \dots$ 。定义更新时间为在单位时间 $t \in [0, 1]$ 内, 第一次更新时间为 $\tau_1 = \max\{t_1^{(1)}, t_1^{(2)}, \dots, t_1^{(d)}\}$, 接下来的更新时间为:

$$\tau_{j+1} = \max\left\{t_{N_{\tau_j}^{(1)}+1}^{(1)}, t_{N_{\tau_j}^{(2)}+1}^{(2)}, \dots, t_{N_{\tau_j}^{(d)}+1}^{(d)}\right\}. \quad (15)$$

由此生成的更新时间样本量为 n_f 。在这个定义下, τ_1 是所有资产的价格事件都发生的时点, 也就是说所有的价格已经是更新的, τ_2 是所有资产的价格都再次更新的时点, 以此类推。

更新时间方法将非同步时间序列同步化问题转化为如何由各资产价格事件发生的时间 $t_1^{(k)}, t_2^{(k)}, \dots, t_n^{(k)}, \dots$ 来生成同步化的价格事件时间 $\tau_1, \tau_2, \dots, \tau_{n_f}$ 。新序列对于原序列样本的保留程度由新样本量与原有样本量的比值来描述, 即 $d \cdot n_f / \sum_{k=1}^d n^{(k)}$ 。

由上述定义可知, $\{\tau_i\}_{i \in \{1, 2, \dots, n_f\}}$ 为 $[0, +\infty)$ 上的一个新的随机点过程。用 $\{Z_i^{(k)}\}_{i \in \{1, 2, \dots, n_f\}}$, $k = 1, 2, \dots, d$ 表示与随机到达时间 $\{\tau_i\}_{i \in \{1, 2, \dots, n_f\}}$ 相对应的标值序列, $\{\tau_i, Z_i^{(k)}\}_{i \in \{1, 2, \dots, n_f\}}$ 形成一个新的标值点过程。在这个定义下, 某个资产组合的更新时间过程可视为由标值点过程生成: $\{\tau_i, p_{\tau_i}^{(k)}, \sigma_{\tau_i}^{2(k)}, Z_i^{(k)}\}_{i \in \{1, 2, \dots, n_f\}}$, 其中 $p_{\tau_i}^{(k)}$ 为第 k 只股票在该更新时点对应的价格, 即 $p_{\tau_i}^{(k)} = p_{t_j: \{\max(t_j \leq \tau_i)\}}^{(k)}$, 进而可以得到相应的收益率 $r_{\tau_j}^{(k)} = \ln p_{\tau_i}^{(k)} - \ln p_{\tau_{i-1}}^{(k)}$ 。 $\sigma_{\tau_i}^{2(k)}$ 为第 k 只股票在该更新时点对应的收益率方差, 是在这个持续期期间的平均方差, 即 $\sigma_{\tau_i}^{2(k)} = \frac{1}{n_{\tau_i}^{(k)}} \sum \sigma_{t_j: \{\tau_{i-1} < t_j \leq \tau_i\}}^{2(k)}$, 其中 $n_{\tau_i}^{(k)}$ 是 τ_{i-1} 至 τ_i 期间的价格事件次数。 $z_i^{(k)}$ 包括了其它相关标值变量。由于需要计算收益率, 更新时间序列的第一个观测将损失, 相应的脚标向前一期平移即可。

2.2 多元波动率模型

对于资产组合收益率及其波动的建模, 常用的方法包括 EWMA 模型、BEKK 模型、DCC 模型等。但在本文中, 时间序列是非等间隔发生的, 不能直接使用这些模型。对于 d 个资产, 经过更新时间方法同步化后的资产组合向量可表示如下:

$$r_{\tau_j} = \mu_{\tau_j} + \varepsilon_{\tau_j} \cdot \sigma_{\tau_j}, \quad (16)$$

其中,

$$\mathbf{r}_{\tau_j} = \begin{pmatrix} r_{\tau_j}^{(1)} \\ \vdots \\ r_{\tau_j}^{(d)} \end{pmatrix}, \quad \boldsymbol{\mu}_{\tau_j} = \begin{pmatrix} \mu_{\tau_j}^{(1)} \\ \vdots \\ \mu_{\tau_j}^{(d)} \end{pmatrix}, \quad \boldsymbol{\sigma}_{\tau_j} = \begin{pmatrix} \sigma_{\tau_j}^{(1)} \\ \vdots \\ \sigma_{\tau_j}^{(d)} \end{pmatrix}, \quad \boldsymbol{\varepsilon}_{\tau_j} = \begin{pmatrix} \varepsilon_{\tau_j}^{(1,1)} & \cdots & \varepsilon_{\tau_j}^{(1,d)} \\ \vdots & \ddots & \vdots \\ \varepsilon_{\tau_j}^{(d,1)} & \cdots & \varepsilon_{\tau_j}^{(d,d)} \end{pmatrix},$$

$\mu_{\tau_j}^{(k)}$ 是 τ_j 时刻第 k 只股票的收益率均值, $\sigma^{(k)}(\tau_j|\mathcal{F}_{\tau_j})$ 是经价格持续时间调整的条件波动率。假设 d 个资产按固定权重 $\boldsymbol{\omega} = (\omega^{(1)}, \dots, \omega^{(d)})^T$ 组合且 $\boldsymbol{\omega}^T \cdot \mathbf{1} = 1$, 则组合的收益率为 $\gamma_{\tau_j} = \boldsymbol{\omega}^T \cdot \mathbf{r}_{\tau_j}$ 。这里 $\sigma_{\tau_j}^{(k)}$ 包含了各资产波动率的自相关信息, 而 $\boldsymbol{\varepsilon}_{\tau_j}$ 中也保留了资产组合在横截面上的相关信息。

Andréas 和 Heinen (2007)^[13] 研究发现将单变量的 ACD 模型拓展到向量 ACD 模型非常困难。鲁万波和李会会 (2014)^[14] 利用 Copula 理论建立多元 Copula-ACD 模型, 一定程度上能够规避向量 ACD 模型参数估计的困难。但经过更新时间同步化的多元时间序列波动率是否仍具有前文所述的单变量价格事件持续时间与及时波动率的关系, 或应该怎样进行修改和调整, 当前仍不清楚。因此, 本文采用 Copula 方法直接研究 $\boldsymbol{\varepsilon}_{\tau_j}$ 在横截面上的相关关系。假设 F_k 为 $\varepsilon_{\tau_j}^{(k)}$ 的边缘分布, 即 $\varepsilon_{\tau_j}^{(k)} \sim F_k(\varepsilon_{\tau_j}^{(k)})$ 。根据 Sklar 定理^[15], 存在一个函数 $C(\cdot)$ 使得 $\varepsilon_{\tau_j}^{(k)}$ 的联合分布可表示为边缘分布的函数:

$$F(\varepsilon_{\tau_j}^{(1)}, \varepsilon_{\tau_j}^{(2)}, \dots, \varepsilon_{\tau_j}^{(d)}) = C(F_1(\varepsilon_{\tau_j}^{(1)}), F_2(\varepsilon_{\tau_j}^{(2)}), \dots, F_d(\varepsilon_{\tau_j}^{(d)})). \quad (17)$$

由此, 资产组合的收益率表示为:

$$\mathbf{r}_{\tau_j} = \boldsymbol{\mu}_{\tau_j} + \boldsymbol{\sigma}_{\tau_j} \odot \boldsymbol{\varepsilon}_{\tau_j}, \quad \boldsymbol{\varepsilon}_{\tau_j} = \begin{pmatrix} \varepsilon_{\tau_j}^{(1)} \\ \vdots \\ \varepsilon_{\tau_j}^{(d)} \end{pmatrix} \sim F(\varepsilon_{\tau_j}^{(1)}, \varepsilon_{\tau_j}^{(2)}, \dots, \varepsilon_{\tau_j}^{(d)}), \quad (18)$$

“ $\odot$ ”表示矩阵 Hadamard 积, 即对应元素相乘。

在金融风险管理领域, 常见的连接函数包括椭圆 Copula 簇中的正态 Copula 和 t-Copula 以及阿基米德 Copula 簇中的 Frank-Copula、Clayton-Copula 和 Gumbel-Copula。这五种连接函数具有简单的形式且分布特征具有代表性, 被广泛运用于构建多元联合分布。根据 Nelsen 和 Roger (2007)^[16], 正态 Copula 定义为:

$$C^{Norm}(\mathbf{u}) = \Phi(u_1, u_2, \dots, u_d),$$

其中 Φ 为多元正态分布函数; t-Copula 定义为:

$$C^t(\mathbf{u}) = \int_{-\infty}^{t_{\nu}^{-1}(u_1)} \cdots \int_{-\infty}^{t_{\nu}^{-1}(u_d)} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu\pi)^d|\mathbf{P}|}} \left(1 + \frac{\mathbf{x}'\mathbf{P}^{-1}\mathbf{x}}{\nu}\right)^{-\frac{\nu+d}{2}} dx,$$

Frank-Copula 定义为:

$$C^{Frank}(\mathbf{u}) = \frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1) \cdots (e^{-\theta u_d} - 1)}{(e^{-\theta} - 1)^{d-1}}\right), \quad \theta > 0,$$

Gumbel-Copula 定义为:

$$C^{Gumbel}(\mathbf{u}) = \exp \left(- \left((-\ln u_1)^{\theta} + (-\ln u_2)^{\theta} + \cdots + (-\ln u_d)^{\theta} \right)^{1/\theta} \right), \quad \theta > 1,$$

Clayton-Copula 定义为:

$$C^{Clayton}(\mathbf{u}) = (u_1^{-\theta} + u_2^{-\theta} + \cdots + u_d^{-\theta})^{1/\theta}, \quad \theta > 0.$$

张金清和李徐 (2008)^[17] 归纳指出: 正态 Copula、t-Copula 和 Frank-Copula 为对称分布, Gumbel-Copula、Clayton-Copula 为非对称分布, t-Copula 具有一定的厚尾特征, Frank-Copula 具有较大方差且上下尾都较厚, Gumbel-Copula 具有较厚的下尾部, 而 Clayton-Copula 上尾部较厚。

2.3 资产组合非等间隔日内在险价值及检验

不失一般性, 假设 d 个资产按固定权重 $\boldsymbol{\omega} = (\omega^{(1)} \cdots \omega^{(d)})^T$ 组合且 $\boldsymbol{\omega}^T \cdot \mathbf{1} = 1$, 则组合的收益率为 $\gamma_{\tau_i} = \boldsymbol{\omega}^T \mathbf{r}_{\tau_j}$ 。组合收益率的 $1 - \alpha$ 水平在险价值 $R_{\tau_i}(\alpha)$ 定义为: 使组合收益率 γ_{τ_i} 低于 $R_{\tau_i}(\alpha)$ 的概率为 α , 即:

$$\begin{aligned} \alpha &= P(\gamma_{\tau_i} < R_{\tau_i}(\alpha) | \mathcal{F}_{\tau_{i-1}}) = P(\boldsymbol{\omega}^T \mathbf{r}_{\tau_j} < R_{\tau_i}(\alpha) | \mathcal{F}_{\tau_{i-1}}) \\ &= P(\boldsymbol{\omega}^T (\boldsymbol{\mu}_{\tau_j} + \boldsymbol{\sigma}_{\tau_j} \odot \boldsymbol{\varepsilon}_{\tau_j}) < R_{\tau_i}(\alpha) | \mathcal{F}_{\tau_{i-1}}). \end{aligned} \quad (19)$$

在给定历史信息集 $\{\mathcal{F}_{\tau_{i-1}}\}$ 的情况下 $\boldsymbol{\sigma}_{\tau_j}^T$ 、 $\boldsymbol{\mu}_{\tau_j}$ 可单独估计, $\boldsymbol{\varepsilon}_{\tau_j}$ 的联合分布由 Copula 给出。进而求解式 (19) 即可得到 $R_{\tau_i}(\alpha)$ 。进一步展开, 可以得到:

$$\begin{aligned} \alpha &= P(\boldsymbol{\omega}^T (\boldsymbol{\sigma}_{\tau_j} \odot \boldsymbol{\varepsilon}_{\tau_j}) < (R_{\tau_i}(\alpha) - \boldsymbol{\omega}^T \boldsymbol{\mu}_{\tau_j}) | \mathcal{F}_{\tau_{i-1}}) \\ &= P((\boldsymbol{\omega} \odot \boldsymbol{\sigma}_{\tau_j})^T \boldsymbol{\varepsilon}_{\tau_j} < (R_{\tau_i}(\alpha) - \boldsymbol{\omega}^T \boldsymbol{\mu}_{\tau_j}) | \mathcal{F}_{\tau_{i-1}}), \end{aligned} \quad (20)$$

其中 $(\boldsymbol{\omega} \odot \boldsymbol{\sigma}_{\tau_j})^T \boldsymbol{\varepsilon}_{\tau_j}$ 是多个随机变量 $\boldsymbol{\varepsilon}_{\tau_j}$ 的加权求和, 理论上在给定单个元素的边缘分布以及其联合分布的情况下使用卷积公式可以得到。但是在复杂的边缘、联合分布下, 式 (20) 的解析解将非常繁琐, 因此考虑使用蒙特卡洛模拟的方法进行模拟求解。基本思路是先利用样本信息估计 $\boldsymbol{\varepsilon}_{\tau_j}$ 的连接函数, 即其联合分布。由该分布生成大量的随机数 $\tilde{\boldsymbol{\varepsilon}}_{\tau_j}$, 将随机数带回 $(\boldsymbol{\omega} \odot \boldsymbol{\sigma}_{\tau_j})^T \boldsymbol{\varepsilon}_{\tau_j}$ 并找到其 α 分位点 $R_{\tau_i}^*(\alpha)$, 进而求出资产组合的日内非等间隔在险价值:

$$R_{\tau_i}(\alpha) = \boldsymbol{\omega}^T \boldsymbol{\mu}_{\tau_j} + R_{\tau_i}^*(\alpha). \quad (21)$$

关于这个在险价值的检验, 同样可使用 Kupiec (1995)^[18] 提出的 LR 统计量。定义碰撞序列:

$$h_j = \begin{cases} 1, & r_{t_j} < -R_j(\alpha), \\ 0, & r_{t_j} \geq -R_j(\alpha). \end{cases}$$

检验 VaR 的估计是否准确变换为检验如下原假设:

$$H_0: h_j \sim \text{i.i.d. Bernoulli}(\alpha), \quad \forall j = 1, 2, \cdots, N_c.$$

相应的 LR 统计量为:

$$LR = -2 \ln \frac{\alpha^{n_h} (1 - \alpha)^{n - n_h}}{(n_h/n)^{n_h} (1 - n_h/n)^{n - n_h}}. \quad (22)$$

在 H_0 为真, 即估计的 VaR 准确的情况下, 该统计量服从自由度为 1 的卡方分布。

3 实证研究

3.1 数据选取及基本统计特征

实证分析选用民生银行 (600016)、招商银行 (600036) 和兴业银行 (601166) 2014 年 8 月 1 日至 29 日期间所有交易日的逐笔成交数据。数据来源于国泰安 CSMAR 高频数据库。选择三只银行股, 主要考虑是银行股体量较大, 能够在一定程度上抵御非市场因素的扰动, 因而本文未选择国有“五大行”, 而是选择了上述市场化程度相对较高的三家银行。民生银行、招商银行和兴业银行这 21 个交易日的原始样本分别为 513296, 358026 和 538709 个, 日均样本量分别为 24442.67, 17048.86 和 25620.29 个。原始数据中包含了股票代码、成交日期、成交时间点 (精确到毫秒)、成交价格、成交股数、成交量等信息。上证交易所连续交易时间为 9:30–11:30、13:00–15:00。本文剔除了 9:30 之前、11:30–13:00 以及 15:00 之后的数据。

价格持续时间是价格信息变化的时间间隔。在高频数据中, 价格往往会短暂的变化并很快回到之前的价格。Engle 和 Russell (1998)^[4] 指出, 市场中的交易者会有报价错误或者是对仓位控制。为了剔除这种影响, 需设定一个阈值“ dp ”来筛选市场价格变化大于这个阈值的观测。由于上证市场报价的最小单位为 0.01 元, 实证部分以 0.01 作为阈值考察价格绝对变化与持续时间的关系。

三只股票价格持续时间数据的描述统计结果如表 1 所示。其中, N 为本量, Min 为最小值, Mean 为均值, Med 为中位数, Max 为最大值, SD 为标准差, CV 为变异系数, Skew 为偏度, Kurt 为峰度, LB(30) 为滞后 30 期的 LB 检验统计量。由表 1 可见: (1) 各股票价格持续时间的标准差均大于均值, 变异系数大于 1, 表现出明显的过度分散性; (2) 各股票价格持续时间的偏度远大于 0, 峰度远大于 3, 说明价格持续时间的分布存在明显的正偏和尖峰特性; (3) 价格持续时间的 LB 检验统计量远大于 5% 显著性水平下的临界值 $\chi^2_{0.05}(30) = 43.77$, 说明价格持续期序列存在强烈的自相关性。

表 1 价格持续期描述性统计量

	N	Min	Mean	Med	Max	SD	CV	Skew	Kurt	LB(30)
民生银行	21770	0.01	13.89	2.45	5258.56	77.7	5975.6	26.4	1269.6	859.2
招商银行	16084	0.01	18.80	4.10	1121.71	56.5	812.8	7.9	87.7	1779.6
兴业银行	18236	0.01	16.57	2.51	1834.90	59.8	2265.3	10.0	163.1	3434.3

由于价格持续时间表现出明显的日内模式^[4], 原始的持续时间不能反映持续时间的特性, 需要将日内模式剔除后再对价格持续时间进行 ACD 建模。本文采用三次样条光滑的方法, 以 30 分钟为单位的分组 (全天交易时间 4 小时, 共 8 组) 价格持续时间的组内均值为节点进行拟合。

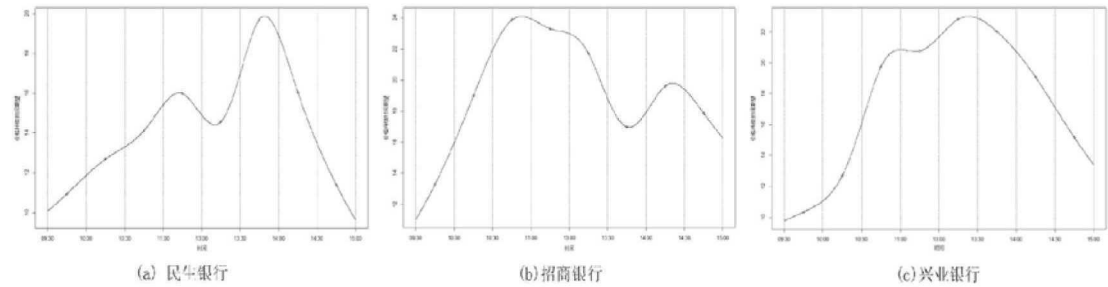


图 1 价格持续时间日内模式

图 1 为三只银行股票价格持续时间的日内模式样条拟合。可见价格事件平均持续时间在早上开盘时最短,随后持续上升,在下午开盘时有一定减小,之后回升至下午盘中达到最大后在盘尾再次回落。与之相对应的是市场参与者在每日交易时段之初和结尾交易较为频繁。这种日内模式与之前包括 Engle 和 Russell (1998)^[4] 在内的研究发现一致。

表 2 给出了剔除日内模式后的价格持续时间 $\tilde{x}_j$ 的描述性统计结果。其中各指标意义与表 1 相同。对比表 2 与表 1 可以发现,经过日内模式调整以后的价格持续时间的均值、标准差、变异系数、偏度均小于未剔除的情况, LB 统计量也有所减小,说明剔除日内模式后,价格持续时间的自相关性有所降低。然而 LB 统计量仍然远大于 5% 显著性水平下的临界值 $\chi_{0.05}^2(30) = 43.77$, 说明价格持续时间序列仍然存在较强的自相关性。

表 2 剔除日内模式后的价格持续期描述性统计量

	N	Min	Mean	Med	Max	SD	CV	Skew	Kurt	LB(30)
民生银行	21770	0.00	1.01	0.18	496.28	5.92	5.87	36.48	2475.89	686.24
招商银行	16084	0.00	1.01	0.22	95.04	3.00	2.98	8.60	125.12	1503.35
兴业银行	18236	0.00	1.00	0.16	86.47	3.45	3.44	8.82	118.15	3119.85

3.2 非等间隔非同步数据同步化

利用前文所述的更新时间方法,可以得到一个同步化后的资产组合非等间隔的时间序列。图 2 给出了三只股票经过时间更新后的部分结果。经同步化后的序列共计 2541 个样本,样本保留度为 13.59%。经更新时间方法同步化后的新序列同样可用值点过程 $\{\tau_i, p_{\tau_i}^{(k)}, \sigma_{\tau_i}^{2(k)}, z_i^{(k)}\}_{i \in \{1, 2, \dots, n_f\}}$ 描述,进而可得到相应的收益率 $r_{\tau_i}^{(k)} = \ln p_{\tau_i}^{(k)} - \ln p_{\tau_{i-1}}^{(k)}$ 。同样采用三次样条光滑,图 3 给出了三只股票组成的资产组合更新时间序列的价格持续时间日内模式。由图 1 和图 3 可以看出,同步化后组合序列的价格持续时间仍然呈现出与单只股票相似的日内模式,从另一面说明了该方法的合理性。

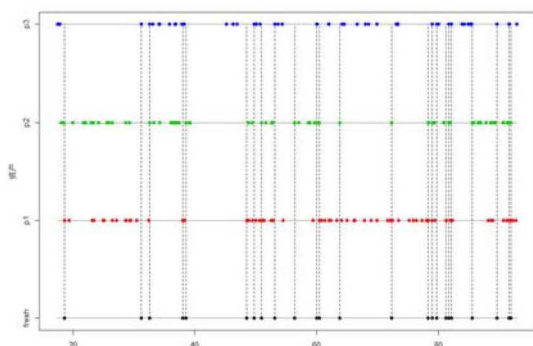


图 2 部分更新结果

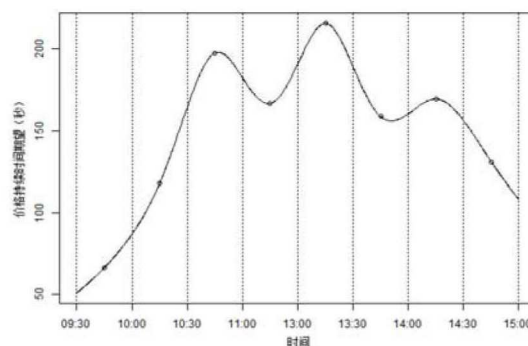


图 3 资产组合价格持续时间日内模式

3.3 资产截面相关项的边缘分布

将原始数中前 18 个交易日的数据作为训练样本 (共 1896 个观测), 后 3 个交易日数据作为验证样本 (共 149 个观测)。根据式 (7), 在给定历史信息集 $\{\mathcal{F}_{\tau_{i-1}}\}$ 的情况下, 第 k 只股票收益率条件均值的估计可由简单的均值模型给出。历史研究发现, 股票的日内收益率自相关及偏相关很微弱, 几乎等同于白噪声, 所以这里仅借鉴 AR(1) 的形式给出 $\mu_{\tau_j}^{(k)}$ 的朴素 (naive) 估计。

各股票在同步化后,抹去了 τ_{i-1} 至 τ_i 的价格信息而直接记录 τ_{i-1} 和 τ_i 时点上的价格,由此计算的收益率可以视作 τ_{i-1} 至 τ_i 的平均收益率。因此在给定 $\{\mathcal{F}_{\tau_{i-1}}\}$, 更新序列的日内波动率一步向前预测也采用该股票在 τ_{i-1} 至 τ_i 期间的平均波动率 $\hat{\sigma}_{\tau_i}^{2(k)} = \frac{1}{n_{\tau_i}^{(k)}} \sum \sigma_{t_j: \{\tau_{i-1} < t_j \leq \tau_i\}}^{2(k)}$, 其中 $n_{\tau_i}^{(k)}$ 是 τ_{i-1} 至 τ_i 期间该股的价格事件次数。

在获得了 $\hat{\sigma}_{\tau_j}^{2(k)}$ 和 $\hat{\mu}_{\tau_j}^{(k)}$ 后,可计算包含截面相关信息的扰动项 $\hat{\varepsilon}_{\tau_j}^{(k)} = \frac{r_{\tau_j}^{(k)} - \hat{\mu}_{\tau_j}^{(k)}}{\hat{\sigma}_{\tau_j}^{(k)}}$ 。截面相关项 $\hat{\varepsilon}_{\tau_j}^{(k)}$ 的边缘分布对于构建 Copula 函数非常重要。Cochrane (2000)^[19] 研究发现,随着抽样频率的减少,即从日内的 1 分钟、5 分钟、半小时到日度、周、月度收益率,其分布由尖峰厚尾分布逐渐趋近于正态。因此对日度及更高频收益率的描述通常采用 Student's t 分布。但本文所研究的经日内模式调整后的非等间隔日内收益率的边缘分布应该采用什么分布,需要进一步验证。常见的边缘分布包括正态 (Normal) 分布、学生 t 分布。但拟合的结果显示学生 t 分布并不能很好地反映更新序列资产收益率的分布情况,通过对数据的观察和试验发现,双指数 (Double Exponential) 分布能够更好地拟合更新序列的资产收益率。双指数分布的密度函数为:

$$f_{X \sim \text{dexp}}(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}, \quad -\infty < x < +\infty, \quad (23)$$

其中 μ 为位置参数, b 为形状参数。

图 4 展示了民生银行标准化后的截面相关项核密度估计的参数拟合。图中粗虚线为核密度估计,细虚线为学生 t 分布拟合,细点线为正态分布拟合,细实线为双指数分布拟合。从图中可以看出,各股票真实的边缘分布峰部较为“扁圆”,而尾部厚度介于正态和学生 t 分布之间。经过反复试验,发现双指数分布能够更好地拟合各股票,尤其是其尾部分布。

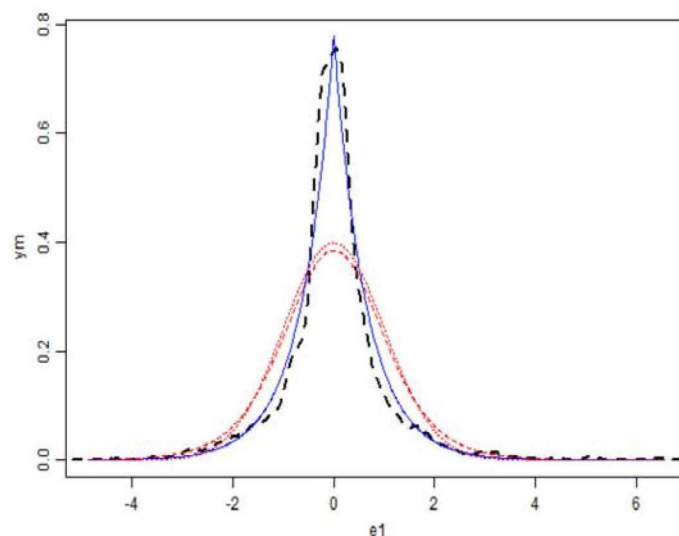


图 4 民生银行截面相关项边缘分布核密度估计及参数拟合

表 3 资产的截面相关项相关系数

相关系数		ρ_{12}	ρ_{13}	ρ_{23}
Kendall	τ	0.1343	0.1700	0.1895
	z	8.7663	11.0920	12.3690
Pearson	r	0.1990	0.2471	0.2749
	t	8.8388	11.0980	12.4450

3.4 资产组合非等间隔日内在险价值的计算和检验

(1) 截面相关系数。接下来使用 Copula 建立 $\varepsilon_{\tau_j}^{(k)}$ 的联合分布。表 3 给出了三个资产间的两两相关系数。其中皮尔逊相关系数 (Pearson) 用于描述线性相关, 但各资产截面相关项之间的关系可能并非线性。因此采用更加稳健的 Kendall 秩相关系数 (Kendall)。表 3 中两个相关系数靠上一栏为相关系数, 靠下一栏为相应的检验统计量。Kendall 秩相关系数的检验统计量 (τ) 服从正态分布, 皮尔逊相关系数 (r) 的检验统计量服从自由度为样本量减去 2 的 t 分布。从表 3 可以看出, 无论是基于皮尔逊相关系数还是 Kendall 相关系数, 三个资产的截面相关项均存在显著的正相关。

图 5 为截面相关项在三维空间的联合分布图。可以看出截面相关项呈现出“椭球”形状, 并且这个“椭球”是“斜向上”置于三维空间中, 展示出明显的正相关关系。

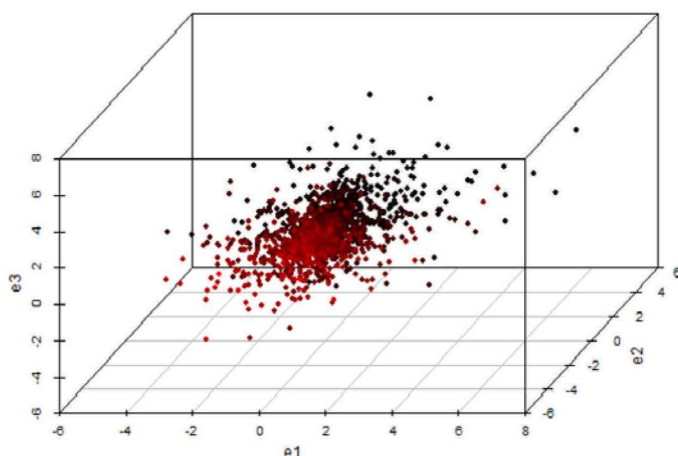


图 5 截面相关项三维关系图

(2) 资产组合非等间隔日内在险价值算法。本文尝试了边缘分布为双指数分布、学生 t 分布和正态分布, 以及连接函数为正态、学生 t 、Gumbel、Frank 和 Clayton 分布构成的 15 种组合, 并根据训练数据的表现选择相应的模型对验证数据作样本外一步向前预测。参考张金清和李徐 (2008)^[17]、柏满迎和孙禄杰 (2007)^[20] 介绍的蒙特卡洛模拟思路, 基于连接函数的资产组合的 VaR 计算步骤为:

步骤 1 利用训练数据的 $\varepsilon_{\tau_j}^{(k)}$, 采用极大似然法估计其边缘分布 $F_k(\cdot)$ 的相关参数;

步骤 2 利用边缘分布计算截面相关项的分位数 $\mu_j^{(k)} = F_x(\varepsilon_{\tau_j}^{(k)})$ 并带入连接函数 $C(u^{(1)}, u^{(2)}, u^{(3)})$ 利用极大似然法估计连接函数的相关参数;

步骤 3 循环, 令 $j = 0$,

步骤 3.1 用步骤 2 得到的连接函数 $C(\cdot)$ 生成 5000 个三维随机变量 ε , 其中 $\varepsilon_l = (\varepsilon_l^{(1)}, \varepsilon_l^{(2)}, \varepsilon_l^{(3)})$, $l = 1, 2, \dots, 5000$;

步骤 3.2 计算 $V = \varepsilon \cdot (\omega \odot \sigma_{\tau_j})$, V 是一个长度为 5000 的向量;

步骤 3.3 找到 V 的 α 分位点, 作为 $R_{\tau_j}^*(\alpha)$ 的一次估计;

将步骤 3.1 至步骤 3.3 重复 200 次, 取 α 分位点的平均值作为 $R_{\tau_j}^*(\alpha)$ 的估计。则资产组合在 τ_j 时点的 ISIVaR 为:

$$\hat{R}_{\tau_j}(\alpha) = \hat{\omega}^T \hat{\mu}_{\tau_j} + R_{\tau_j}^*(\alpha), \quad (24)$$

步骤 3.4 让 $j = j + 1$ 。若 $j \leq n_f$, 回到步骤 3.1, 否则跳出循环;

步骤 4 分别对训练数据和验证数据部分统计资产组合收益率 r_{τ_j} 向下突破 $\hat{R}_{\tau_j}(\alpha)$ 的次数, 并进行检验。

利用上述算法步骤对前文所述的三种边缘分布和五类连接函数共 15 种组合, 在 95%、99% 和 99.5% 置信水平下的 ISIVaR 进行计算。

(3) 连接函数估计结果。采用极大似然法, 步骤 2 中估计各 Copula 的参数。其中, 正态和 Student's t-Copula 需要估计截面相关项的相关系数矩阵, 估计时使用表 3 中估计的 Kendall 相关系数作为初始值。

表 4 各 Copula 的参数估计

边缘分布	Copula	Par	Est	Sd	Z	P	LH
双指数分布	Frank	θ	1.1347	0.0827	13.7200	0.0000	-109.6
	Gumbel	θ	1.1204	0.0109	103.0000	0.0000	-106.2
	Clayton	θ	0.1845	0.0169	10.9000	0.0000	-84.4
	Normal	ρ_{12}	0.1900	0.0196	9.6717	0.0000	-135.5
		ρ_{13}	0.1916	0.0194	9.8916	0.0000	
		ρ_{23}	0.2375	0.0196	12.1357	0.0000	
	Student's t	ρ_{12}	0.1921	0.0212	9.0547	0.0000	-153.0
		ρ_{13}	0.1958	0.0209	9.3841	0.0000	
		ρ_{23}	0.2341	0.0208	11.2341	0.0000	
		df	15.6188	2.9722	5.2550	0.0000	
正态分布	Frank	θ	1.7232	0.1107	15.5700	0.0000	-131.2
	Gumbel	θ	1.1003	0.0118	93.0300	0.0000	-71.21
	Clayton	θ	0.1134	0.0130	8.7080	0.0000	-59.73
	Normal	ρ_{12}	0.1908	0.0198	9.6257	0.0000	-126.0
		ρ_{13}	0.1719	0.0200	8.5854	0.0000	
		ρ_{23}	0.2325	0.0193	12.0460	0.0000	
	Student's t	ρ_{12}	0.1956	0.0219	8.9245	0.0000	-140.1
		ρ_{13}	0.1804	0.0223	8.0924	0.0000	
		ρ_{23}	0.2346	0.0206	11.3801	0.0000	
		df	33.7352	7.2090	4.6796	0.0000	
学生 t	Frank	θ	2.0001	0.1177	16.9900	0.0000	-144.9
	Gumbel	θ	1.2079	0.0163	74.3500	0.0000	-127.7
	Clayton	θ	0.2566	0.0224	11.4800	0.0000	-87.92
	Normal	ρ_{12}	0.2936	0.0237	12.4126	0.0000	-146.9
		ρ_{13}	0.2954	0.0248	11.9145	0.0000	
		ρ_{23}	0.3145	0.0211	14.9150	0.0000	
	Student's t	ρ_{12}	0.2929	0.0264	11.1018	0.0000	-177.7
		ρ_{13}	0.3015	0.0274	10.9856	0.0000	
		ρ_{23}	0.2994	0.0236	12.6811	0.0000	
		df	10.2196	1.4753	6.9272	0.0000	

注: LH 为极大似然值, Par 为 Copula 待估参数, Est 为参数估计值, Sd 为标准差, Z 为相应检验统计量, P 为其 P 值, df 为自由度。

表 4 给出了边缘分布为双指数分布、正态分布、学生 t 分布, 以及连接函数为正态、学生 t、Gumbel、Frank 和 Clayton 分布时各 Copula 的参数估计情况。从表 4 的结果来看, 相应的 Z 统计量和 P 值显示各 Copula 的参数系数均显著异于零, 可以认为上述 Copula 对于训练数据的截面相关性描述在统计上均有一定的解释力。分别来看, 边缘分布为双指数分布的连

接函数中, Clayton-Copula 具有最大的似然函数值, 其次 Frank-Copula 和 Gumbel-Copula 似然函数值相差不大。边缘分布为正态的连接函数中, Clayton-Copula 似然函数值最大, 其次是 Gumbel-Copula。而边缘分布为学生 t 分布的 Clayton-Copula 似然函数值明显高于组内其它连接函数。总的来看, 边缘分布为正态和双指数分布的连接函数似然函数明显高于边缘分布为 Student's t 的连接函数, 因此推测更新序列中各资产收益率波动的分布特征, 可能并不是之前研究在日度及更高频率的等间隔时间序列中呈现的厚尾特征。相反, 其尾部应该为图 4 所展示的那样, 介于 Student's t 分布和正态分布之间。同时本文发现 Clayton-Copula 在各组中均具有最大的似然函数值, 因此推测更新序列资产组合中单个资产间波动率的横截面相关性可能存在三维空间上的非对称且上侧尾部较厚的特征。

(4) 资产组合非等间隔日内在险价值估计。由算法步骤 3 和步骤 4, 利用蒙特卡洛模拟, 使用估计出的联合分布生成大量三维随机样本点。将这些样本点带入假定的数据生成过程, 即多元波动率模型, 根据 α 分位点计算出各边缘、联合分布假设下的 ISIVaR。

不同 α 水平下训练样本一步向前预测的估计结果如表 5 所示。其中每栏第一行为“碰撞事件”的频率, 第二行为 Kupiec 的 LR 检验统计量的 P 值。由表 5 可以看出, 边缘分布为学生 t 分布下的 Copula 碰撞事件频率远低于预设的在险价值各水平, 检验 P 值均低于 0.01, 认为估计的 VaR 并不准确, 严重低估了 VaR。其原因在于学生 t 分布尾部相较于真实值过厚, 从而在模拟计算时错误地模拟出了过多的极端事件。由此说明学生 t 分布并不适用于描述更新序列的尾部特征。而对 VaR 的估计更看重的是对于样本分布尾部的拟合情况。而且在边缘分布与连接函数的联合作用下, 估计出的联合分布的尾部特征与单独观察边缘分布的特征也存在差距。本文中的三维联合分布以及更高维的联合分布与边缘分布的交互影响下, 样本特征、边缘分布特征和联合分布特征, 三者之间的关系不会像图 5 那样直观。

表 5 训练样本一步向前资产组合 ISIVaR 检验

Copula	双指数			正态			学生 t		
	5%	1%	0.5%	5%	1%	0.5%	5%	1%	0.5%
Frank	0.0332	0.0105	0.0047	0.0253	0.0111	0.0084	0.0185	0.0058	0.0021
	0.0004	0.8119	0.8747	0.0000	0.6435	0.0534	0.0000	0.0461	0.0436
Gumbel	0.0364	0.0111	0.0053	0.0338	0.0142	0.0105	0.0216	0.0063	0.0021
	0.0044	0.6435	0.8667	0.0006	0.0810	0.0029	0.0000	0.0849	0.0436
Clayton	0.0274	0.0047	0.0005	0.0290	0.0105	0.0063	0.0185	0.0016	0.0005
	0.0000	0.0104	0.0004	0.0000	0.8119	0.4308	0.0000	0.0000	0.0004
Normal	0.0274	0.0069	0.0016	0.0258	0.0105	0.0063	0.0179	0.0026	0.0016
	0.0000	0.1447	0.0137	0.0000	0.8119	0.4308	0.0000	0.0001	0.0137
Student's t	0.0380	0.0021	0.0005	0.0269	0.0063	0.0016	0.0211	0.0016	0.0005
	0.0123	0.0000	0.0004	0.0000	0.0849	0.0137	0.0000	0.0000	0.0004

利用验证数据进行样本外一步向前预测, 其检验情况如表 6 所示。如前分析, 边缘分布为学生 t 分布的连接函数对于 ISIVaR 的估计误差比较严重, 因此在该方法样本外验证过程中不予考虑。从表 6 可以发现, 样本外一步向前预测中, 边缘分布为双指数分布的 Clayton-Copula 和 T-Copula 在 95%、99% 和 99.5% 置信水平下的 ISIVaR 估计情况良好, P 值均大于或接近 0.05。然而边缘分布为正态情况时, 对于除 95% 置信水平以外的 ISIVaR 估计则并不理想, P 值均小于 0.01。正如前文所分析的, 正态分布的尾部较双指数分布更薄, 因此没能够模拟出足够多的“极端事件”, 导致其明显高估了 ISIVaR。

对比表 5 和表 6 发现, 样本内和样本外一步向前预测的结果比不完全一致。其原因可能在于验证数据的样本量 (149 个观测) 远小于训练数据样本量 (1896 个观测), 由此导致样本中随机因素对于估计和检验的影响增大。但从另一个角度来看, 实证结果中验证部分表现并不比训练部分差, 说明本文提出的基于 Copula 理论的资产组合 ISIVaR 的计算方法不存在过拟合, 是一种比较稳健的方法。

表 6 验证样本一步向前资产组合 ISIVaR 检验

Copula	双指数			正态		
	5%	1%	0.5%	5%	1%	0.5%
Frank	0.0671	0.0470	0.0336	0.0604	0.0470	0.0403
	0.3612	0.0010	0.0011	0.5721	0.0010	0.0001
Gumbel	0.0738	0.0470	0.0336	0.0738	0.0470	0.0470
	0.2113	0.0010	0.0011	0.2113	0.0010	0.0000
Clayton	0.0671	0.0268	0.0201	0.0604	0.0470	0.0336
	0.3612	0.0873	0.0488	0.5721	0.0010	0.0011
Normal	0.0671	0.0403	0.0268	0.0604	0.0470	0.0403
	0.3612	0.0051	0.0081	0.5721	0.0010	0.0001
Student's t	0.0738	0.0268	0.0201	0.0604	0.0403	0.0201
	0.2113	0.0873	0.0488	0.5721	0.0051	0.0488

4 结论

本文基于价格持续时间的 ACD 模型, 通过更新时间方法将非同步的资产组合标值序列同步化, 运用 Copula 理论建立资产组合的非等间隔日内波动模型, 捕捉了组合中资产在截面上的相关性, 并基于这种相关关系, 通过蒙特卡洛模拟的方法估计出资产组合的 ISIVaR。

研究发现: (1) 无论是同步化的单个资产还是同步化的资产组合, 其价格持续时间均存在明显的日内模式, 同步化的单资产与资产组合的日内模式相吻合, 说明通过更新时间在对组合中各资产的非等间隔时间序列数据进行同步化的同时, 比较客观地保留了原先单个资产的市场微观结构信息; (2) 对于三只银行股而言, 多元波动率模型中包含截面相关信息的扰动项尾部厚度介于同方差的正态分布和学生 t 分布之间, 利用双指数分布拟合相对较好, 正态分布其次, 但学生 t 分布会产生严重偏差。在此基础上, 边缘分布为双指数分布的 Clayton-Copula 或 t-Copula 对资产组合 ISIVaR 估计结果比较准确。(3) 采用的同步化资产组合的横截面相关关系建模的思路切实可行, 训练数据与验证数据表现基本一致, 说明该方法不存在过拟合的问题, 是一种比较稳健的方法。

[参考文献]

- [1] Narang, Rishi K, 郭剑光. 打开量化投资的黑箱 [M]. 北京: 机械工业出版社, 2012.
- [2] 许启发, 徐金菊, 蒋翠侠. 基于神经网络分位数回归的多期 CVaR 风险测度 [J]. 数理统计与管理, 2017, 36(4): 715-730.
- [3] 赵丽丽, 张波. 基于 ICA 模型的投资组合稳健 VaR 方法研究 [J]. 数理统计与管理, 2019, 38(2): 367-380.
- [4] Engle R F, Russell R. Autoregressive conditional duration: A new model for irregularly spaced transaction data [J]. Econometrica, 1998, 66(5): 1127-1162.
- [5] Colletaz G, Hurlin C, Tokpavi S. Irregularly spaced intraday value at risk (ISIVaR) models forecasting and predictive abilities [R]. Orleans: University of Orleans, 2007.

- [6] Dionne G, Duchesne P, Pacurar M. Intraday value at risk (IVaR) using tick-by-tick data with application to the Toronto Stock Exchange [J]. *Journal of Empirical Finance*, 2009, 16(5): 777–792.
- [7] 鲁万波, 王卫东. 基于价格持续时间的中国股市日内风险价值预测 [J]. *数理统计与管理*, 2012, 31(3): 527–536.
- [8] Gregor N F W, Hendrik S. Forecasting liquidity-adjusted intraday Value-at-Risk with vine copulas [J]. *Journal of Banking and Finance*, 2013, 37(9): 3334–3350.
- [9] 王峰, 刘传哲. 期货市场日内 VaR 测度模型与应用 [J]. *数理统计与管理*, 2014, 33(6): 1091–1100.
- [10] Lunde A. A generalized Gamma autoregressive conditional duration model [R]. Aalborg: Aalborg University, 1999.
- [11] Giot P. Market risk models for intraday data [J]. *The European Journal of Finance*, 2005, 11(4): 309–324.
- [12] Barndorff N O E, Hansen P R, Lunde A, et al. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading [J]. *Journal of Econometrics*, 2010, 162(2): 149–169.
- [13] Andréas Heinen, Erick Rengifo. Multivariate autoregressive modeling of time series count data using copulas [J]. *Journal of Empirical Finance*, 2007, 14(4): 564–583.
- [14] 鲁万波, 李会会. 多元 Copula-ACD 模型及其应用 [J]. *中国管理科学*, 2014, 22: 293–298.
- [15] Sklar A. Fonctions de repartition à n dimensions et leurs marges [J]. *Publication de l'Institut de Statistique l'Universite Paris*, 1959, 8: 229–231.
- [16] Nelsen R B. An Introduction to Copulas [M]. Springer Science & Business Media, 2007.
- [17] 张金清, 李徐. 资产组合的集成风险度量及其应用 — 基于最优拟合 Copula 函数的 VaR 方法 [J]. *系统工程理论与实践*, 2008, (6): 14–21.
- [18] Kupiec P H. Techniques for verifying the accuracy of risk measurement models [J]. *The Journal of Derivatives*, 1995, 3(2): 73–84.
- [19] Cochrane J H. New facts in finance [J]. *Economic Perspectives*, Federal Reserve Bank of Chicago, 2000, 30(2): 36–58.
- [20] 柏满迎, 孙禄杰. 三种 Copula-VaR 计算方法与传统 VaR 方法的比较 [J]. *数量经济技术经济研究*, 2007, (2): 154–160.