要点

- 分布
    1. GPD的依赖分布$P[X \leq x|X \geq V\tilde{a}R] \sim GPD(V\tilde{a}R, \hat{\xi}, \hat{\sigma} + \hat{\xi}(V\tilde{a}R - u))$
    2. Jarque-Bera Test检验tail的normality,但是数据多很容易reject
- semi-parametric
    - Threshold的选取要尽量大,但区域中shape ξ需要stable
    - 存在assumption即return必须是i.i.d(经常被violate)
- copula
    1. 计算copula找随机变量关系,再利用单调性回带
    2. 正态copula 没有tail dependence
    3. copula选择需要关注tail拟合
    4. 使用AIC/BIC来选择最优模型
    5. copula不可能是一个点,必定是线

1. 基本概念

收益率

今天减k天前的价格

- 收益率: $R_t = \frac{P_t}{P_{t-k}} - 1 = e^{r_t} - 1$
- 对数收益率: $r_t = \ln\frac{P_t}{P_{t-k}} = \ln(1 + R_t) = \ln P_t - \ln P_{t-k}$
- 分红收益率: $R_t = \frac{P_t + D_t}{P_{t-1}}$
- 对数分红收益率: $r_t = \ln\frac{P_t + D_t}{P_{t-1}}$

VaR

- 相对风险资产(收益率的负分位数): $P[R_t + V\tilde{a}R] = q$
- 风险资产(相对值乘以portfolio大小): 亏损超VaR的概率小于q, $VaR = P_t \cdot V\tilde{a}R_t$

Shortfall Distribution

- 给定亏损分布F,则控制亏损在x和VaR_q中的概率: $\Theta_q(x) = P[X \leq x|X > VaR_q] = \frac{F(x)-(1-q)}{q}$
- 亏损大于VaR情况下的期望值: $ES_q = E[X|X > VaR_q] = \frac{1}{q}\int_{x>VaR_q} xf(x)\mathrm{d}x$

收益率不同分布下的ES

正态分布	Dexp	GPD
$\mu + \sigma\frac{f(F^{-1}(\alpha))}{\alpha}$	$-\mu + \frac{1-\ln 2\alpha}{\lambda}$	$VaR_q + \frac{\sigma}{\xi q^{\xi}} = V\tilde{a}R_q + \frac{\hat{\sigma}+\xi(V\tilde{a}R_q - u)}{1-\hat{\xi}}$

时间序列数据

- 强平稳: 不同时间分布相同
- 弱平稳: 不同时间方差和均值相同
- 自方差: $COV(y_h) = \frac{1}{n}\sum_{t=1}^{n-h}(y_t - \mu)(y_{t+h} - \mu)$
- 自相关性: $\gamma_h = \frac{COV(y_h)}{\sigma^2}$

概率分布性质

- k阶中心矩: $\mu_k = E[(X - E[x])^k]$
- 偏度: $\frac{\mu_3}{\mu_2^{3/2}}$
- 峰度: $\frac{\mu_4}{\mu_2^2} - 3$
- 期望$E(X) = \int x f_X(x)\mathrm{d}x$
- 方差$\sigma = \int (x - E(X)) f_X(x)\mathrm{d}x$

- Tail probabilities描述当$x \to \infty$时$1 - F(x)$的变化情况,和指数分布比较: $\lim_{x\to\infty} e^{tx}\mathrm{d}F(x) = \infty$ 则heavy tail

2. 参数估计

Kernel Density Estimate

- 给定若干个离散样本x_i,构造平滑分布

$$\hat{f}_b(x) = \frac{1}{n}\sum_{i=1}^{n} K_b(x - x_i)$$

- K_b: kernel function,可以是任何density function,方差必须是1

- 本质上是样本的平滑分布,b越大,越平滑

Semi-parametric Estimation

- 当阈值u足够大时,任何tail都可以转化为GPD:

$$P[X > u + x | X > u] \to H_\xi(x)$$

- 估计流程
  1. Plot ECDF分布的Tail
  2. 使用shape plot画出GPD的$\hat{\xi}$和threshold u的关系
  3. threshold要尽量大,但是shape parameter要尽量稳定
  4. 通过Tail plot检验threshold选取的质量
  5. 计算$\tilde{V}aR_q = GPD^{-1}[1 - \frac{q}{1-ECDF(u)}]$
  6. 基于GPD期望的性质,可以得到$ES = \tilde{V}aR_q + \frac{\hat{\sigma}+\hat{\xi}(\tilde{V}aR_q - u)}{1-\hat{\xi}}$

3. Portfolio基本性质

双资产管理

给定双资产组成的Portfolio $V = n_1P_1 + n_2P_2$

- 权重向量$w = \begin{bmatrix} \frac{n_1P_1}{n_1P_1' + n_2P_2'} & \frac{n_2P_2}{n_1P_1' + n_2P_2'} \end{bmatrix}$
- 收益率$R = \frac{V'}{V} = \frac{n_1P_1' + n_2P_2'}{n_1P_1 + n_2P_2}$,通过$R$来计算$VaR$和Expected Shortfall
- 对于双资产Portfolio $V = wR_1 + (1-w)R_2$,则通过分散投资不相关资产来最小化风险

$$w = \frac{\sigma_2^2 - \rho_{12}\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2}$$

<u>Copula模型应用</u>

1. 对每个asset估计marginal distribution $\hat{F}_P$
2. 使用QQ plot判断$\hat{F}_p$的估计准确度,注意tail
3. 计算asset之间的spearman相关系数
4. 用$\hat{F}_p$生成概率并带入copula模型
5. 绘制copula的等高线,分析估计质量
6. 从copula生成模拟数据,并获得不同权重时的VaR

风险资产模型定义

资产收益率向量	资产收益率期望	收益率协方差矩阵
$R=[R_1\cdots R_N]^T$	$E[R]=[\mu_1\cdots\mu_N]^T$	$\mathrm{Cov}(R)=\Sigma$

已知上述条件,找到最合适的权重向量$W=[w_1\cdots w_N]^T$在收益率最大的情况下最小化风险

- 最小化标准差: $\min w^T\Sigma w$
- 最大化收益率: $w^T\mu=\mu^*$
- 权重和为1: $w^TI=1$
- 可能还有分散投资,不能做空等附加条件
- 生成的结果为<u>有效边界: 在固定资产的方差σ的情况下,可以获得的最大收益μ</u>

无风险资产混合投资

- 在资产组合中引入了收益率为μ_f的无风险资产
- 需要最大化夏普比率:描述了额外收益$E(R)-\mu_f$和风险σ的比值,<u>最大化夏普比率等于寻找切线</u>

$$\frac{E(R)-\mu_f}{\sigma}$$

- 令风险资产权重为w则有
  - 收益率$\mu=\mu_f+w(\mu_p-\mu_f)$
  - 方差$\sigma=w\sigma_p$

最大化夏普比率

- 对于双资产混合无风险模型,令$v_1=\mu_1-\mu_f,v_2=\mu_2-\mu_f$,则第一个资产的权重有

$$w_T=\frac{v_1\sigma_2^2-v_2\rho_{12}\sigma_1\sigma_2}{v_1\sigma_2^2+v_2\sigma_1^2-(v_1+v_2)\rho_{12}\sigma_1\sigma_2}$$

- 在最大化夏普比率的基础上,根据资产模型的方差限制,获得无风险资产的配置

1. 根据w_T获得风险资产标准差σ_T,μ_T
2. 根据给定的风险目标,获得风险资产总占比权重 $w=\frac{\sigma_{target}}{\sigma_T}$
3. 或者根据给定的收益率目标,获得风险资产总比权重$w=\frac{\mu_{target}-\mu_f}{\mu_T-\mu_f}$
4. 每个风险资产的比重为$w\cdot w_T$

4. 单变量分布

正态分布 N(μ, σ^2)

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

指数分布 Exp(λ)

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 \end{cases}$$
$$F(x) = 1 - e^{-\lambda x}$$

- $\mu = \sigma = \frac{1}{\lambda}$

双指数分布 Dexp(μ, λ)

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x-\mu|}$$
$$F(x) = \begin{cases} \frac{1}{2}e^{-\lambda|x-\mu|} & x < \mu \\ 1 - \frac{1}{2}e^{-\lambda|x-\mu|} & x \geq \mu \end{cases}$$

- $\sigma = \frac{\sqrt{2}}{\lambda}$
- 峰度为3

t分布

- ν越大,tail越小, 峰度$\frac{\sigma}{\nu-4}$
- $\sigma = \lambda\sqrt{\frac{\nu}{\nu-2}}$
- $t_\nu(\mu, \lambda^2)$是$Y = \mu + \lambda t_\nu$的表示,标准差不是λ

Pareto Distribution

$$f(x) = \frac{\alpha\mu^\alpha}{x^{\alpha+1}}$$
$$F(x) = 1 - \frac{\mu}{x}^\alpha \quad x \geq \mu$$

GPD(ξ, σ)

- 多项式级的tail

$$f(x) = \frac{1}{\sigma}(1 + \frac{\xi(x-\mu)}{\sigma})^{-\frac{1}{\xi}-1}$$
$$F(x) = 1 - (1 + \frac{\xi(x-\mu)}{\sigma})^{-\frac{1}{\xi}}$$

GED($\mu, \lambda^2\nu$)

$$f(x) = \kappa_\nu \exp\left(-\frac{1}{2}|\frac{x}{\lambda_\nu}|^\nu\right)$$

- ν越小,tail越heavy
- $\sigma = 1$
- 转换后$Y = \mu + \lambda X, \sigma = \lambda$

Gamma 分布 $\Gamma(\lambda, \alpha)$

描述了α个独立指数分布/Gamma分布的和

- Gamma Function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x$$

$$f_{\alpha,\lambda}(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 \end{cases}$$

  对于正整数α,$\Gamma(\alpha) = (\alpha - 1)!$

- $\mu = \frac{\alpha}{\lambda}$
- $\sigma = \frac{\sqrt{\alpha}}{\lambda}$
- Gamma分布可加性 $Gamma(\gamma, \alpha_1,) + Gamma(\gamma, \alpha_2) = Gamma(\gamma, \alpha_1 + \alpha_2)$

5. 多变量分布

多变量正态分布

参数	随机变量 X	PDF	期望	方差
μ, Σ	p维向量X	$f_X(x) = \frac{1}{\sqrt{(2\pi)^p \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$	μ	Σ

多变量T分布

参数	随机变量X	期望	方差
μ, Λ, ν	$T_n = \mu + \sqrt{\frac{\nu}{W}} X, (X \sim N(0, \Lambda))$ (多变量正态分布除以方差后转换得到)	μ	$\frac{\nu}{\nu-2}\Lambda$

- 所有变量的marginal distribution都是自由度ν的T分布,因此tail的性质相同
- 所有变量之间uncorrelated但是dependent,但是当ν很大时,可以假设independent

Copulas

定义

- 将随机变量转化到uniform distribution,只考虑他们出现在p-quantile的概率 $U_p = F_{X_p}(X_p)$

$$C(u_1, u_2, \cdots, u_p) = P[U_1 \leq u_1, \cdots, U_p \leq u_p]$$

$$F_{X_1, \cdots X_p}(x_1, \cdots, x_p) = C(F_{X_1}(x_1), \cdots, F_{X_p}(x_p))$$

- 通过Copula可以得到变量的联合分布: $f(x, y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y)$

Copula和相关性

- 对于独立变量 $C(u_1, u_2, \cdots, u_p) = \prod_{i=1}^p u_i$
- 对于完全互相依赖的变量,$C(u_1, u_2, \cdots, u_p) = \min\{c_i\}$

正态分布 Copula

- 当变量服从联合正态分布的时Copula
- 只和变量之间的相关系数ρ有关

$$\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{F_X^{-1}(u_1)}\int_{-\infty}^{F_Y^{-1}(u_2)}\exp\left(-\frac{1}{2}\frac{s^2-2\rho st+t^2}{1-\rho^2}\right)\mathrm{d}s\mathrm{d}t$$

T-Copuls

- 当变量服从联合T-分布时的Copula
- Copula和μ无关,因为不会影响rank

建模流程

1. 估计每个变量的边际分布 $\hat{F}_p$
2. 回带每个变量的边际分布$\hat{F}_p$
3. 获得Copula的表达式估计值$\hat{C}$

Archimedean-type Copula定义

- $C(u_1,u_2,\cdots,u_p)=\phi^{-1}(\phi(u_1)+\cdots\phi(u_d))$
- ϕ是严格递减的凸函数
- $\phi(0)=\infty,\phi(1)=0$
- 通过估计不同模型的θ来找到合理的copula分布

Gumbel (logistic Copula)

$$C(u_1,u_2,\cdots,u_p)=\exp\left(-[(-\log u_1)^\theta+\cdots(-\log u_d)^\theta]^{\frac{1}{\theta}}\right)$$

- $\phi(u)=-\log u$
- 适合Heavy tail分布,当$\theta=1$的时候是独立分布,θ越大相关性越强

Clayton Copula

$$C(u_1,u_2,\cdots,u_p)=(u_1^{-\theta}+\cdots u_d^{-\theta}-d+1)^{-\frac{1}{\theta}}$$

- $\phi(u)=u^{-\theta}-1$
- 当$\theta\to 0$时,变成独立分布

Frank Copula

$$C(u_1,u_2,\cdots,u_p)=-\frac{1}{\theta}\log\left(1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{\exp(-\theta)-1}\right)$$

- $\phi(u)=-\ln\frac{\exp(-\theta u)-1}{\exp(-\theta)-1}$
- 当$\theta\to 0$时,变成独立分布