第16章 基于Copula的VaR度量与事后检验

清华大学经管学院 朱世武
Zhushw@sem.tsinghua.edu.cn
Resdat样本数据：www.resset.cn
SAS论坛： www.resset.cn

Copula函数

最常见的两种Copula函数是正态Copula函数和t-分布Copula函数。

$C_R(u_1,\cdots,u_n)$

正态Copula函数

正态分布随机变量 $X_1,\cdots,X_n$ 的均值分别为 $\mu_1,\cdots,\mu_n$ ，方差分别为 $\sigma_1,\cdots,\sigma_n$ ，相关矩阵为R，则随机变量 $U_i := \Phi(\frac{X_i-\mu_i}{\sigma_i}), i \in I$ 的分布函数 $C_R(u_1,\cdots,u_n)$ 为Copula函数，称为协方差矩阵为R的正态Copula函数，或称为Gauss Copula函数。（ Φ 为标准正态分布函数）。

$$C_R^{Ga}(u_1,\cdots,u_n) = \Phi_R(\Phi^{-1}(u_1),\cdots,\Phi^{-1}(u_n))$$

The Power to Know

t-分布Copula函数

t-分布Copula函数是正态Copula函数的变形。

正态分布随机变量 $X_1,\cdots,X_n$ 的均值分别为0，方差分别为1，相关矩阵为R。Y为χ^2 分布随机变量，自由度为υ，与 $(X_1,\cdots,X_n)'$ 独立。则随机变量 $U_i = t_\upsilon(\frac{\sqrt{\upsilon}}{\sqrt{Y}}X_i)$,的分布函数 $C_{\upsilon,R}(u_1,\cdots,u_n)$ 为Copula函数，称为自由度为 ，相关矩阵为R的t-分布Copula函数。

$$C^t_{v,R}(u_1,\cdots,u_n) = t_{v,R}(t_v^{-1}(u_1),\cdots,t_v^{-1}(u_n))$$

t-分布Copula函数继承了正态Copula函数的几乎全部性质。但在正态Copula函数模型中，极端事件的发生总是彼此独立的，即U_i 接近于0或1的可能性彼此独立。而在t-分布Copula函数中，极端事件是相关的。

The Power to Know.

联合分布模拟与收益率映射

假设有一个包括n个资产的组合，收益率为向量为 $(z_1, z_2, \dots, z_n)'$ 。设 $z_1, z_2, \dots, z_n$ 的边际分布分别为 $F_1, F_2, \dots, F_n$ ，$(z_1, z_2, \dots, z_n)'$ 的联合分布为 $F(z_1, z_2, \dots, z_n)$ 。

正态Copula模拟

1. 对R进行Cholesky分解 $\mathbf{R} = \mathbf{A}\mathbf{A}'$
2. 生成一个 $n \times 1$ 的服从独立标准正态分布的随机向量

$$\mathbf{Z} = (z_1, \cdots z_n)' \sim N\left(0, \begin{bmatrix} 1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & 1 \end{bmatrix}\right)$$

3. 令 $\mathbf{X} = \mathbf{A}\mathbf{Z}$，产生 $n \times 1$ 维向量，这时，

$$\mathbf{Corr}(\mathbf{X}) = \mathbf{Cov}(\mathbf{X}) = \mathbf{E}(\mathbf{X}\mathbf{X}') = \mathbf{E}(\mathbf{A}\mathbf{Z}\mathbf{Z}'\mathbf{A}') = \mathbf{A}\mathbf{E}(\mathbf{Z}\mathbf{Z}')\mathbf{A}' = \mathbf{A}\mathbf{I}\mathbf{A}' = \mathbf{A}\mathbf{A}' = \mathbf{R}$$

即X向量中的每一个随机变量均服从期望为0，方差为1的正态分布，并且它们相关矩阵正好为R。

4. 计算$u_i = F(x_i) \quad i = 1, \cdots, n$ ，得到

5. 用边际分布函数的反函数把$(u_1, \ldots, u_n)' \quad C_{\mathbf{R}}^{Ga}$ 映射到收益率

.

$$\mathbf{z} = (z_1, \ldots, z_n)' = (F_1^{-1}(u_1), \cdots, F_n^{-1}(u_n))'$$

若边际分布为正态分布，则 $\mathbf{z} = (z_1, \ldots, z_n)' = (\Phi_1^{-1}(u_1), \cdots, \Phi_n^{-1}(u_n))'$

若边际分布为t分布，则 $\mathbf{z} = (z_1, \ldots, z_n)' = (t_{v_1}^{-1}(u_1), \cdots, t_{v_n}^{-1}(u_n))'$

其中 $v_1, \cdots, v_n$ 分别是相应t分布的自由度。

t-分布Copula模拟

1. 对R进行Cholesky分解 $\mathbf{R} = \mathbf{A}\mathbf{A}'$
2. 生成一个 $n \times 1$ 的服从独立标准正态分布的随机向量

$$\mathbf{Z} = (z_1, \cdots z_n)' \sim N\left(0, \begin{bmatrix} 1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & 1 \end{bmatrix}\right)$$

3. 令 $\mathrm{Y} = \mathbf{A}\mathbf{Z}$，产生 $n \times 1$ 维向量，这时，

$$\mathbf{CorrY} = \mathbf{Cov(X)} = \mathbf{E}\left(\mathbf{XX}'\right) = \mathbf{E}\left(\mathbf{AZZ'A}'\right) = \mathbf{AE}\left(\mathbf{ZZ}'\right)\mathbf{A}' = \mathbf{AIA}' = \mathbf{AA}' = \mathbf{R}$$

即Y向量中的每一个随机变量均服从期望为0，方差为1的正态分布，并且它们相关矩阵正好为R。

4. 产生与Y相互独立的变量S，服从χ_{ν}^{2} 分布。

5. 令 $\mathbf{X}=\dfrac{\sqrt{\nu}}{\sqrt{S}}\mathbf{Y}$ ，则X服从自由度为ν 的t分布。

6. 计算$u_i = t_\nu(x_i) \quad i=1,\ldots,n$ ，得到$(u_1,\ldots,u_n)' \sim C_{\nu,R}^{n}$

7. 用边际分布函数的反函数把$(u_1,\ldots,u_n)' \sim C_{\nu,R}^{t}$ 映射到收益率

$$\mathbf{z}=(z_1,\ldots,z_n)'=(F_1^{-1}(u_1),\cdots,F_n^{-1}(u_n))'$$

若边际分布为正态分布，则 $\mathbf{z}=(z_1,\ldots,z_n)'=(\Phi_1^{-1}(u_1),\cdots,\Phi_n^{-1}(u_n))'$

若边际分布为t分布，则 $\mathbf{z}=(z_1,\ldots,z_n)'=(t_{v_1}^{-1}(u_1),\cdots,t_{v_n}^{-1}(u_n))'$

其中 $v_1,\cdots,v_n$ 分别是相应t分布的自由度。

投资组合VaR度量

假设一资产组合中包含20只股票，利用前面介绍的方法估计该资产组合特定置信水平下的日风险值（VaR）。

计算环境

2004年12月底以前的历史数据，计算2005年第一个交易日的VaR。假设投资总额为100万元，构造股票投资组合，针对投资组合做相应计算。

类似地，可以计算2005年全年交易日的VaR，并作相应的事后检验。

数据集：

个股数据集ResDat.Qttndist；

计算步骤

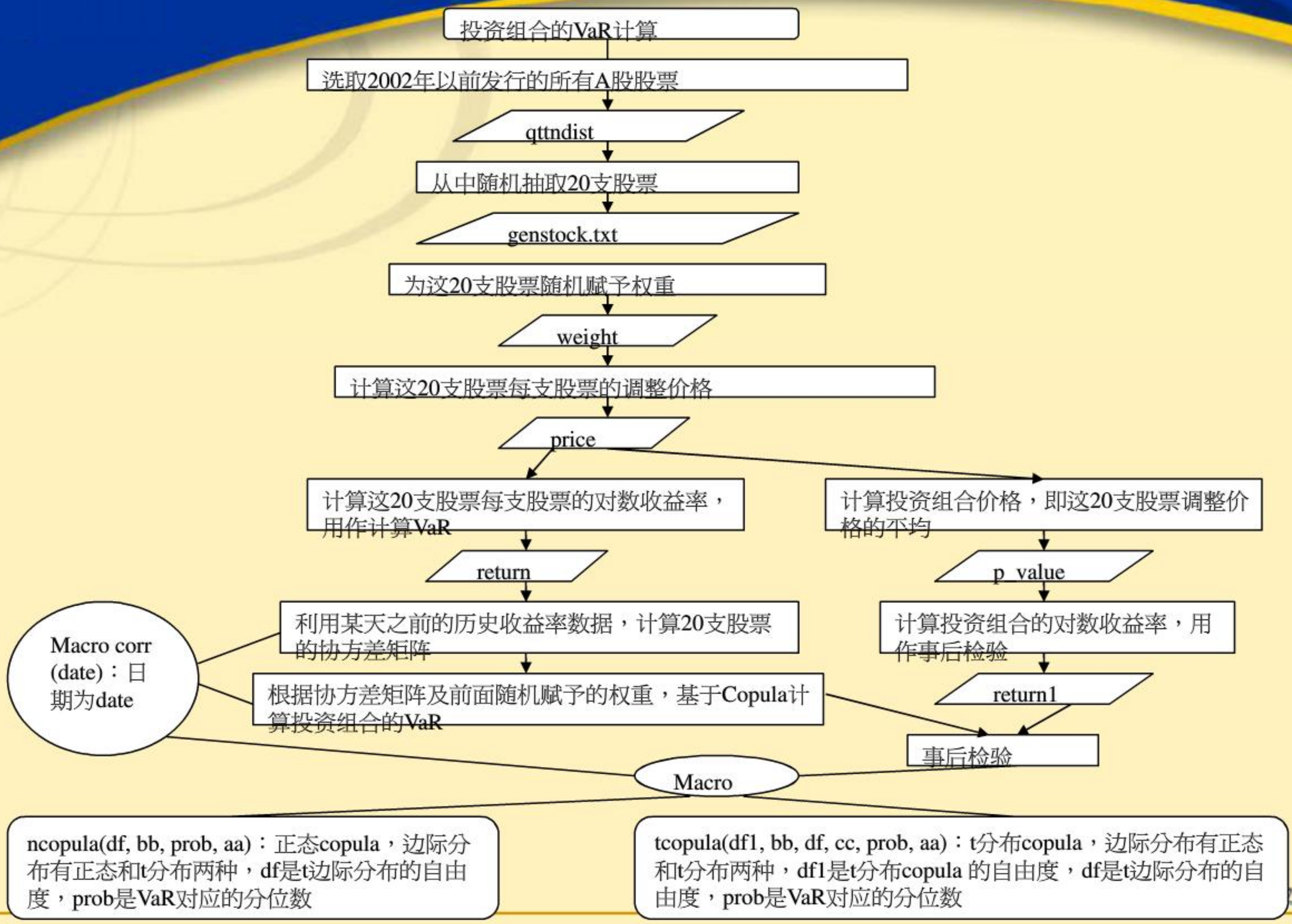
第一步：确定权重，构造投资组合；

第二步：对每只股票的收益率进行模拟；

第三步：利用前两步所得，计算组合的的收益率；

第四步：利用第三步所得的组合的收益率，计算投资组合的VaR；

第五步：对投资组合的VaR作相应的事后检验。


投资组合的VaR计算
选取2002年以前发行的所有A股股票
qttndist
从中随机抽取20支股票
genstock.txt
为这20支股票随机赋予权重
weight
计算这20支股票每支股票的调整价格
price
计算这20支股票每支股票的对数收益率，用作计算VaR
计算投资组合价格，即这20支股票调整价格的平均
return
p_value
利用某天之前的历史收益率数据，计算20支股票的协方差矩阵
计算投资组合的对数收益率，用作事后检验
Macro corr (date)：日期为date
根据协方差矩阵及前面随机赋予的权重，基于Copula计算投资组合的VaR
return1
事后检验
Macro
ncopula(df, bb, prob, aa)：正态copula，边际分布有正态和t分布两种，df是t边际分布的自由度，prob是VaR对应的分位数
tcopula(df1, bb, df, cc, prob, aa)：t分布copula，边际分布有正态和t分布两种，df1是t分布copula 的自由度，df是t边际分布的自由度，prob是VaR对应的分位数

计算结果

表16.1 正态Copula相关结果

标识	置信水平	t边际分布下的VaR（百万元）	正态边际分布下的VaR（百万元）	实际失效天数（t边际分布）	实际失效天数（正态边际分布）	理论失效天数
ncopula+02+5	95%	0.030049	0.026622	4	9	12
ncopula+02+1	99%	0.046011	0.040755	1	2	2
ncopula+04+5	95%	0.028293	0.026592	7	10	12
ncopula+04+1	99%	0.041623	0.039117	1	2	2
ncopula+10+5	95%	0.025758	0.025122	9	9	12
ncopula+10+1	99%	0.038854	0.037894	2	2	2
ncopula+20+5	95%	0.026524	0.026194	7	8	12
ncopula+20+1	99%	0.038349	0.037871	2	2	2
ncopula+40+5	95%	0.027491	0.02732	10	10	12
ncopula+40+1	99%	0.037167	0.036935	2	2	2

表16.2 t-分布Copula相关结果

标识	置信水平	t边际分布下的VaR（百万元）	正态边际分布下的VaR（百万元）	实际失效天数（t边际分布）	实际失效天数（正态边际分布）	理论失效天数
tcopula+02+02+5	95%	0.046792	0.041422	1	1	12
tcopula+02+02+1	99%	0.132916	0.112787	0	0	2
tcopula+02+04+5	95%	0.042522	0.039959	1	2	12
tcopula+02+04+1	99%	0.151929	0.14079	0	0	2
tcopula+02+10+5	95%	0.042293	0.041247	2	2	12
tcopula+02+10+1	99%	0.097435	0.094938	0	0	2
tcopula+02+20+5	95%	0.045533	0.044946	2	2	12
tcopula+02+20+1	99%	0.095744	0.094541	0	0	2
tcopula+02+40+5	95%	0.047528	0.04723	1	1	12
tcopula+02+40+1	99%	0.085071	0.084532	0	0	2
……		……		……		

以上结果表明，采用Copula计算的VaR比上章只用正分布的VaR大，所以，事后检验效果更好。

极大似然法拟合t分布

从事后检验结果可以看出，不同自由度下，t分布的检验效果均优于正态边际分布。下面给出确定t分布自由度n的极大似然法。

t分布概率密度函数为，

$$t(x;\quad n)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}(1+\frac{x^{2}}{n})^{-(n+1)/2}$$

其中n为自由度。t分布拟合就是估计该密度函数的自由度参数n。

通过最大化样本的对数似然函数，

$$Max\sum_{i=1}^{m}\log t(x_i;\quad n)$$

就可以得到参数n的估计值。

$Max\sum_{i=1}^{m}\log t(x_i;\quad n)$ 对迭代方法与初值都有很多要求，此处采用一种比较直观的方法。即通过令n取一列值，分别计算对数似然函数 $\sum_{i=1}^{m}\log t(x_i;\quad n)$ 的值，取其最大值所对应的 $\hat{n}$值为自由度参数n的估计值。

```
/*求t边际分布的自由度*/
options nodate nosource nonotes;
proc means data=return1 noprint;
var return1;
output out=b mean=m  std=std;
data b;
set b;
call symput('mean', m);/*求组合收益率的均值和方差*/
call symput('std', std);
data result;
delete;
%macro a(x);
%do i=2 %to  &x ;
data t;
set return1;
n=&i;
t=(Gamma((n+1)/2)*(1+((return1-&mean)/(&std))**2/n)**(-
(n+1)/2))/(gamma(n/2)*(3.14159*n)**0.5);
lnt=log(t);
```

```
proc means data=t  noprint;
var lnt;
output out=t1  sum=sum ;
data t1;
set t1;
n=&i;
data result;
set result t1;
%end;
%mend a;
%a(40);
run;
proc sort data=result;
by descending sum;
data result;
set result;
if _n_=1;
call symput('n', n);
run;
```

计算结果：

Obs	_TYPE_	_FREQ_	sum	n
1	0	961	-1337.83	10
2	0	961	-1337.93	11
3	0	961	-1337.99	9
4	0	961	-1338.20	12
5	0	961	-1338.57	8
6	0	961	-1338.57	13
7	0	961	-1339.01	14
8	0	961	-1339.48	15
9	0	961	-1339.84	7
10	0	961	-1339.98	16
11	0	961	-1340.48	17
12	0	961	-1340.98	18
13	0	961	-1341.48	19
……		……		……

从结果可以看出，当自由度取10时，似然函数取得极大值。