

R 与 VaR 计算

Date: Jan 12, 2010; 09:47pm

Author: 邓一硕

R&数据艺术: (<http://yishuo.org>)

现代金融领域的计量工作大致分为两大块,即资产定价和风险度量。资产准确定价的最终也是为了避免投资过程中可能产生的风险。由此可见,风险度量工作实际上是现代金融计量领域的核心。目前,被公认的风险度量方法中 VaR 方法,因为原理简单、数理基础扎实而成为了一时翘楚。本文主要结合目前正在风靡的 R (<http://www.r-project.org>) 软件来介绍一下 VaR 的原理和实现途径,文章结构树如下:

- VaR 的定义
- VaR 的算法及其在 R 中的实现
- 历史模拟法 (PerformanceAnalytics 包)
- 蒙特卡洛模拟法 (PerformanceAnalytics 包)
- 方差-协方差法
- 指数加权平均法
- 混合正态模型方法
- 风险矩阵方法
- ARCH 类方法 (rgarch 包)
- 极值理论方法 (fExtremes 包)
- VaR 的检验
- VaR 的补充方法简介

1 VaR 的定义

VaR 是 Value at Risk 的简称,中文译作“在险价值”。它是指在市场正常波动情况下,某一金融资产组合,在未来一定时期内,在一定置信水平下所可能产生的最大损失。假设你拥有一个证券组合,该组合的初始价值为 p_0 , R 为特定期内的收益率,假设该组合的期末价值为 p , 则

$$p = (1 + r)p_0$$

若在某一置信水平下,该组合的最低价值为 p^* , 则

$$p^* = (1 + r^*)p_0$$

那么,根据 VaR 的定义得

$$VaR = p_0 - p^* = -p_0 r^*$$

由此可见,计算 VaR 值等价于计算一定置信水平下的资产组合的最低价值 p^* 或最低收益率 r^* , 而计算最低价值或收益率则需依赖于价值序列或收益率序列的分布或概率密度函数。换句话说,计算金融资产序列的分布函数或概率密度函数正是计算 VaR 值的核心之处。

2 VaR 的算法

清楚了 VaR 的定义,计算 VaR 的方法也就知道的差不多了。下面按照难以顺序依次介绍五种主要的 VaR 算法。

2.1 历史模拟法

历史模拟法是最简单的计算 VaR 的方法。历史模拟法的原理极其简单：收集一定量的数据，根据事先确定的置信水平，取相应的下分位数就可以了。

历史模拟法既可以计算单个金融资产的 VaR 值也可以用来计算资产组合的 VaR 值，两者区别不大。

列出来历史模拟法的计算步骤：

- (1) 收集金融资产的历史价格数据，数据量为 N ；
- (2) 计算金融资产的各期收益率，将收益率数据按大小排序；
- (3) 确定 VaR 的置信水平 α ，取出收益率序列升序排列的第 N 个数据。

举个例子，

例：某支股票连续 10 天的开盘价如下：

4.86	7.05	4.95	5.74	3.90	8.23	1.74
5.35	5.80	5.33	4.07	5.67	5.39	3.57
4.52	4.01	5.59	1.73	4.47	7.72	8.12

求未来一天的 95%置信水平下的 VaR 值。

一个手算的过程：

- (1) 计算收益率：

0.10	0.56	-0.31	0.08	0.03	-0.57	0.45	0.84	0.43	1.05
-0.16	-0.18	0.23	0.05	-0.29	1.25	-0.21	-0.35	-0.77	1.27

- (2) 排序

-0.77	-0.57	-0.35	-0.31	-0.29	-0.21	-0.18	-0.16	0.03	0.05
0.08	0.10	0.23	0.43	0.45	0.56	0.84	1.05	1.25	1.27

- (3) 计算 $(1-95\%) \times 20 = 1$ ，找到 (2) 中的第 1 个数为 -0.77。

- (4) 计算 $VaR = -8.12 \times (-0.77) = 6.2524$

因此，该金融资产在未来一天内，在 95%置信水平下的 VaR 值是 6.2524。

在 R 中的实现

```

P=read.table("/openprice.txt",head=T)      #读入历史数据
return=NULL                                #设定初始向量
for(i in 2:21) {return[i]=(diff(P)/P[i-1])}  #计算历史收益率
return                                     #返回历史收益率序列
sort(return)                              #对历史收益率序列进行排序
VaR=quantile(-return,95%,low=T)           #计算 VaR 值

library(PerformanceAnalytics)
VaR(R, p = 0.95, method="historical")

```

2.2 蒙特卡洛模拟法

蒙特卡洛模拟法就是根据历史的数据特征，模拟出一系列符合要求的伪随机数作为假想的未来数据。然后，以这些“假想的未来数据”作为研究对象应用类似历史模拟法的过程计算出 VaR 值。

譬如 2.1 中的数据，用蒙特卡洛模拟法计算 VaR 值的过程时是这样的。首先，观察历史数据的特征，比如绘制直方图：

```
hist(r)
```

观察到历史数据的分布特征十分类似于正态分布，因此，可以设想未来的收益率数据很有可能也服从正态分布，所以我们就可以用随机数发生器产生大量的（比如 1000 个）数据来模拟未来的情形。

```
rnorm(10000)
```

把产生的数据作为研究对象，用 2.1 中的方法计算 VaR 值即可。

蒙特卡洛模拟法在应用时，有两个比较核心的地方。一个是你模拟的市场因子是什么？是收益率还是价格？一般而言，模拟的都是收益率，这是因为价格本身没有很好的模拟基础。另外一个，你模拟的市场因子服从何种分布？目前比较常用的有混合正态分布（双峰分布），t 分布，广义误差分布以及加入偏度的偏正态分布、偏 t 分布、偏广义误差分布，这些在 R 软件中都可以做到。

（混合正态分布图）

（t 分布图）

（广义误差分布图）

（偏正态分布分布图）

（偏 t 分布图）

（偏广义误差分布图）

2.3 方差-协方差法

方差-协方差方法有不同的算法，下面分别进行简单的介绍。

2.3.1 指数加权移动平均法¹

指数加权移动平均法将收益率的方差定义为：

$$\sigma_{t+1|t}^2 = \sum_{k=0}^t \lambda^k (1 - \lambda) (r_{t-k} - \bar{r})^2$$

风险矩阵在单变量情况下的形式为：

$$r_t = \sigma_t \epsilon_t$$

$$\sigma_t = (1 - \lambda) r_{t-1}^2 + \lambda \sigma_{t|t-1}^2$$

$$\text{则 } E(r_t) = 0, \text{Var} R_t = \sigma_t = z_\alpha \sigma_{t+1|t} \sqrt{k}$$

其中， z_α 为 α 对应的正态分布的分位数。 K 为测度期限。 λ 决定了不同历史时期数据在估计中的相对权重，成为衰减因子。关于衰减因子的选择：在前一期 t 基础上预测的 $t+1$ 期回报的方差为 $E_t(r_{t+1}^2) = \sigma_{t+1|t}^2$ ，定义方差预测误差为 $\epsilon_{t+1|t} = r_{t+1}^2 - \sigma_{t+1|t}^2$ ，它满足预测误差的期望值为0，即 $E(\epsilon_{t+1|t}) = E(r_{t+1}^2) - \sigma_{t+1|t}^2 = 0$ 。每日均值平方根误差(RMSE)为

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_0^T (r_{t+1}^2 - \sigma_{t+1|t}^2)}$$

选择可使RMSE达到最小的 λ 作为最优衰减因子。目前，这个方法已经很少用了。

2.3.2 混合正态模型方法²

混合正态模型是为了解决金融时间序列的尖峰厚尾现象而提出来的。假设某个分布是两个正态分布混合而成，密度形式如下：

$$g(x; \theta) = p f_1(x; \mu_1, \sigma_1) + (1 - p) f_2(x; \mu_2, \sigma_2)$$

其中， $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, p)$ 为待估参数。 f_1, f_2 为两个不同正态分布的密度函数。混合正态分布的核心在于参数估计，一般用极大似然法来估计。

基于混合正态模型构造的VaR如下：

$$r_t / \sigma_t = p f_1(x; \mu_1, \sigma_1) + (1 - p) f_2(x; \mu_2, \sigma_2)$$

$$\sigma_{t+1|t}^2 = (1 - \lambda) r_t^2 + \lambda \sigma_{t|t-1}^2$$

$$\text{同理, } \text{Var} R_t = \sigma_t = z_\alpha \sigma_{t+1|t} \sqrt{k}$$

2.3.3 ARCH类方法

这是目前应用最广泛的方法。ARCH模型的基本形式为：

ARCH模型的基本变种是GARCH模型。GARCH模型的基本形式为：

改变GARCH模型的分布假设可以得到GARCH-t模型、GARCH-GED模型、GARCH-SGED

¹张国勇、杨宝臣《VaR计算方法综述》

² 张国勇、杨宝臣《VaR计算方法综述》

模型等；改变 GARCH 模型的主方程可以得到 EGARCH 模型、IGARCH 模型等。当然也可以衍生出 EGARCH-GED 等模型。

library(rgarch)

2.4 极值理论 (Extrem Value Theory) 方法

虽然很多人把极值理论计算 VaR 标榜为新兴的方法。然而，这一种算法实际上也出现很久了。

极值理论是上世纪 70 年代正式发展起来的一门理论，它针对极端事件建模。由于直方图中极端事件位于尾部，因此，有时候又说，极值理论主要是针对尾部建模。极值理论依托于两类重要的分布：极值分布和超限分布。

2.4.1 广义极值分布：

如果随机时间序列 x_n 独立同分布于分布函数 $F(x)$ ，其均值为 μ ，方差为 σ^2 ，无论样本数据的初始分布是什么样子，极值的渐进分布只有三种分布：Gumbel 分布、Frechet 分布、Weibull 分布，而他们又可以被统一为广义极值分布 (Generalized Extrem Value Distribution)。

$$f(x; \xi) = \begin{cases} (1 + \xi x)^{-1/\xi-1} \exp[-(1 + \xi x)^{-1/\xi}] & \text{if } \xi \neq 0 \\ \exp[-x - \exp(-x)] & \text{if } \xi = 0 \end{cases}$$

其中 ξ 是形状参数， $\alpha = 1/\xi$ 为尾部指数。

$\xi > 0$ 时， $-\infty < x < +\infty$ ； $\xi < 0$ 时， $x < -1/\xi$ ； $\xi = 0$ 时， $x > -1/\xi$

通过加入位置参数 μ 和规模参数 σ^2 可以得到完整的 GEV 分布，其密度函数如下：

$$f(x; \mu, \sigma, \xi) = \begin{cases} \frac{1}{\sigma} [1 + \xi (\frac{x-\mu}{\sigma})]^{-1/\xi-1} \exp[-(1 + \xi (\frac{x-\mu}{\sigma}))^{-1/\xi}] & \text{if } \xi \neq 0 \\ \frac{1}{\sigma} \exp[-\frac{x-\mu}{\sigma} - \exp(-\frac{x-\mu}{\sigma})] & \text{if } \xi = 0 \end{cases}$$

计算出分布的分位数公式，可进一步推出相应的 VaR 公式为：

$$VaR_p = \begin{cases} \mu - \frac{\sigma}{\xi} \{1 - [-n \ln(1-p)]^{-\xi}\} & \text{if } \xi \neq 0 \\ \mu - \sigma \ln[-n \ln(1-p)] & \text{if } \xi = 0 \end{cases}$$

2.4.2 广义帕累托分布³：

研究极值时还有另一种方法：首先选取一个足够大的门限值（阈值） u ，然后，对 x_n 中所有超过 u 的值进行建模。这种方法称为 POT (Peaks Over Threshold) 法。 x_n 中所有大于 u 的样本 x_i 称为超限值。 $x_i - u$ 称为超出量。根据条件概率，容易算出来超出量的分布 $F_u(x)$ 如下：

$$F_u(x) = P((X - u) \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}$$

³ 郑振龙、王保和《基于极值理论的风险价值度量》

等式两端对 x 求导得到超出量分布的密度函数：

$$f_u(x) = \frac{f(x+u)}{1-F(u)}$$

对于下面的分布：

$$F_{(u)}(x) = P(X \leq x | X > u) = \frac{P(u < X \leq x)}{P(X > u)} = \frac{F(x) - F(u)}{1 - F(u)}$$

称 $F_{(u)}(x)$ 为超限分布。等式两端关于 x 求导得到超限分布的密度函数为：

$$f_{(u)}(x) = \frac{f(x)}{1 - F(u)}$$

应用过程中一般用广义帕累托分布（GPD: Generalized Paroto Distribution）来逼近超限分布，广义帕累托分布的密度函数如下：

$$f(x; \xi, \sigma, \mu) = \begin{cases} \frac{1}{\sigma} [1 + \frac{\xi(x-\mu)}{\sigma}]^{-1/\xi-1} & \text{if } \xi \neq 0 \\ \frac{1}{\sigma} \exp[-\frac{x-\mu}{\sigma}] & \text{if } \xi = 0 \end{cases}$$

其中 μ 为位置（location）参数， σ 为尺度（scale）参数， ξ 为形状（shape）参数。在广义帕累托分布的假设下，VaR 的计算公式：

$$VaR_p = \begin{cases} u - \frac{\hat{\sigma}}{\hat{\xi}} \{1 - [\frac{N}{N_u}(1-p)]^{-\hat{\xi}}\} & \xi \neq 0 \\ u - \hat{\sigma} \ln[\frac{N}{N_u}(1-p)] & \xi = 0 \end{cases}$$

其中 u 为阈值， $\hat{\sigma}, \hat{\xi}$ 分别是 σ, ξ 的估计值， N 为总样本数目， N_u 为超限值的数目。

```
library(fExtremes)
tailRisk()
```

3 VaR 的检验

关于厚尾，本文引用 Ramazan Gancay 的定义，即：如果尾部密度函数是幂指数衰减的，就称为厚尾；如果尾部是指数衰减或有有限的终点的，称为细尾。

4 附录

VaR 相关 R 包有 VaR 包、fExtremes 包、ghyp 包、PerformanceAnalytics 包、fAssets 包、actuar 包；极值理论相关 R 包有 evir 包、evdbayes 包、evd 包、POT 包；GARCH 模型相关 R 包有 fGarch 包、goGarch 包、rgarch 包等。