
Portfolio Construction and Analytics 读书笔记

(第二章)

目录	1
Contents	1
Contents	1
1 资产管理的介绍	1
2 随机变量、概率分布和重要的统计概念	1
3 常见的分布函数介绍	1
4 统计学模型	1
4.1 经典收益模型	1
4.2 回归分析	1
4.2.1 一个简单的例子	1
4.2.2 回归分析在投资中的应用	2
4.3 因子分析	2
4.4 主成分分析	2
4.5 自回归条件方差模型	3

1 资产管理的介绍

2 随机变量、概率分布和重要的统计概念

3 常见的分布函数介绍

4 统计学模型

4.1 经典收益模型

最为经典的线性回归模型可以表示为：

$$r_i = \alpha_i + f_1\beta_{i1} + \cdots + f_p\beta_{ip} + \varepsilon_i \quad i = 1, \dots, n,$$

r_i :第*i*份资产第收益率

f_k : 影响因子

β :回归系数（灵敏度）

α :常数

ε_i 随机扰动

我们之所以给出上述的模型，是假设收益率可以通过一些可观测的因子线性表出。上述的模型的衍生是多种多样的。例如：上述的模型是静态的，我们可以在不同的时间下观测数据，从而建立模型：

$$r_{i,t+1} = \alpha_i + f_{1t}\beta_{i1} + \cdots + f_{pt}\beta_{ip} + \varepsilon_{it} \quad i = 1, \dots, n,$$

4.2 回归分析

4.2.1 一个简单的例子

假设我们想研究保洁公司股票的收益率和标普500指数的关系，我们可以建立如下的回归模型：

$$r_{P\&G} = \alpha + \beta r_{S\&P500} + \varepsilon$$

其中：

$r_{P\&G}$: 保洁公司股票的收益率。

$r_{P\&G500}$:标普500指数收益率。

我们选取了过去78个月份的数据进行回归分析，得到如下的回归方程：

$$r_{P\&G} = 0.0021 + 0.4617r_{S\&P500}$$

经过P值检验， β 系数显著不为0，经过F检验回归方程显著。为了确保回归模型的有效性，我们还需要对残差进行如下的检验：

- 1.对 ε 进行正态性检验
- 2.对 ε 进行方差齐性检验
- 3.对 ε 进行自相关检验

4.2.2回归分析在投资中的应用

回归分析在投资中的应用主要在于以下四个方面：

- 1.建立投资策略
- 2.选择投资策略
- 3.选择标的资产
- 4.评估策略表现

4.3 因子分析

一个因子分析模型和回归分析模型十分相似，例如资产回报率模型可以写成：

$$r = \alpha + B \cdot f + \varepsilon$$

其中 α 是N维列向量，表示资产的平均回报率。 f 是K维因子向量， B 是 $N \times K$ 维因子载荷矩阵。

因子分析存在的问题：尽管我们可以计算出具体的因子，但是有时候我们难以对因子给出合理的解释。

4.4 主成分分析

主成分分析（Principal Component Analysis, PCA），是一种统计方法。通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量，转换后的这组变量叫主成分。

设是数据的协方差矩阵，PCA的主要步骤：

求解矩阵 \mathbf{W}^T 满足如下的条件。

$$\text{Max } \mathbf{W}^T \Sigma \mathbf{W}$$

$$\text{s.t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}$$

作为一个例子，我们考虑10支股票（AXP, T, BA, CAT, CVX, CSCO, KO, DD, XOM, GE）构成的投资组合在过去78个月内的收益率。我们计算得到各个主成分如下：

EXHIBIT 4.4 Principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
AXP	-0.3125	-0.3965	-0.0053	-0.1065	0.2329	-0.3591	0.7197	-0.0855	-0.0244	0.1563
T	-0.262	0.2583	-0.5274	-0.5126	0.5049	0.2171	-0.0773	0.0167	0.0072	-0.1233
BA	-0.319	-0.1623	0.3003	0.4218	0.51	0.3489	-0.2128	-0.2478	-0.3287	0.0829
CAT	-0.3592	-0.0749	0.0022	-0.3208	-0.3152	-0.3924	-0.4024	-0.092	-0.5604	0.1562
CVX	-0.316	0.4582	0.2953	-0.0732	0.0018	-0.0931	-0.0679	-0.1833	0.4842	0.5622
CSCO	-0.3206	-0.1258	0.1474	-0.2682	-0.4821	0.6971	0.2597	0.0245	-0.0325	-0.0057
KO	-0.2625	0.2633	-0.5851	0.5335	-0.3049	-0.0072	0.1629	-0.3324	-0.0654	-0.0178
DD	-0.3654	-0.3085	0.081	-0.0091	-0.0691	-0.162	-0.2745	-0.2502	0.5104	-0.5788
XOM	-0.2559	0.5627	0.3717	0.0788	0.0302	-0.1638	0.256	0.2926	-0.2266	-0.4931
GE	-0.3634	-0.1923	-0.1883	0.2768	-0.0081	-0.0166	-0.1704	0.7934	0.1595	0.1854

4-1 主成分

EXHIBIT 4.5 Standard deviations of the 10 principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	2.3003	1.1724	0.9020	0.7931	0.7036	0.6943	0.5649	0.4933	0.4500	0.3865
Proportion of variance	0.5291	0.1374	0.0814	0.0629	0.0495	0.0482	0.0319	0.0243	0.0203	0.0149
Cumulative proportion	0.5291	0.6666	0.7480	0.8109	0.8604	0.9086	0.9405	0.9648	0.9851	1.0000

4-2 各个主成分占比

从中我们可以看到，前三个主成了解释了总体74 %的方差。通常我们选取占比较高的前几个成分作为我们的主成分，一旦我们确定了主成分我们就可以采用如下的公式计算得分了：

$$x_k = \sum_{i=1}^N \beta_{ik} r_i$$

4.5 自回归条件方差模型

传统的回归模型假设残差的方差是相同的，然而这样的假设存在如下的问题：

1.金融资产的回报率的振幅随时间变化，并且在某一时期，巨大的资产振幅往往预示着未来一段时间的振幅也会比较大。这也是我们所说的波动集群效应。

2.高频金融数据往往具有厚尾性，这与传统的数据服从正态分布的假设相矛盾。

基于上述的问题，我们引入了ARCH和GARCH模型。

ARCH(q):

以 ε_t 表示收益或者收益残差，假设 $\varepsilon_t = \sigma_t z_t$ ，此处 $z_t \sim i.i.d N(0, 1)$ （即独立同分布，均符合期望为0，方差为1 的正态分布）。在此条件下 σ_t^2 可以写成：

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2,$$

其中 $\alpha_i \geq 0, i > 0$.

GARCH(p,q):

$$y_t = x_t' b + \epsilon_t$$

$$\epsilon_t | \psi_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$$

对于ARCH和GARCH模型，我们容易计算他们的VaR和CVaR：

$$VaR_{1-\varepsilon} = V_t(-u_r + q_{1-\varepsilon} \sigma_r)$$

其中 $q_{1-\varepsilon}$ 为标准正态分布 $100(1 - \varepsilon)$ 分位数。同时：

$$CVaR(1 - \varepsilon) = V_t(-u_r + \frac{\phi(q_{1-\varepsilon})}{\varepsilon} \sigma_r)$$