Note of Econophy

Q. X. Fang

2025-02-16

Contents

I Basic Concept

I.1 研究变量的选择

1. 价格变化：$Z(t) = Y(t + \Delta t) - Y(t)$

2. 考虑贴现的变化量:$Z_D(t) = (Y(t + \Delta t) - Y(t))D(t)$,其中$D(t)$是贴现因子,考虑了汇率等变化因素，但难以确定

3. 价格收益率：$R(t) = \frac{Y(t+\Delta t)-Y(t)}{Y(t)} = \frac{Z(t)}{Y(t)}$，对时间的变化过于敏感

4. 对数价格收益率：$r(t) = \ln\left(\frac{Y(t+\Delta t)}{Y(t)}\right) = \ln(1 + R(t))$，非线性的

当时间间隔很小时$\Delta t \ll \varepsilon, D(t) \approx 1$，则上述四个变量近似等价

以下选择对数收益率

I.2 时间尺度

1. 物理时间（包含未交易的 trash time）

2. 交易时间（无法包含休市时间的信息作用）

3. 交易次数（时间间隔的影响）

以下选择交易时间

I.3 统计分布

随机游走模型: 时间间隔$\Delta T = n\Delta t$的情况下, 游走位置是 n 个独立同分布的随机变量之和:

$$S_n = x_1 + ... + x_n$$

1. 二项分布：若随机变量的取值只有$x_i \in \{s, -s\}$，等概率情况，则:

$$E(S_n) = 0, \quad \sigma^2 = E(S_n^2) = ns^2 \propto \Delta T$$

方差是随步长线性增的:

$$\sigma = \sqrt{ns^2} \propto (\Delta T)^{\frac{1}{2}}$$

1. 正态分布：中心极限定理，当 n 足够大时，S_n近似正态分布:

$$P(S) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{S^2}{2\sigma^2}}$$

累积分布函数：对 S 积分

实证研究表明,对于金融资产收益率时间序列 S(t),其概率密度函数与正态分布曲线相比,呈现尖头胖尾的特征。这种尖头胖尾性可以简便地通过收益率时间序列的峰度(kurtosis)进行描述。

峰度,在统计学上,是指观测变量的四阶中心矩与标准差四次方的比值，定义为:

$$k = \frac{\sum_{i=1}^{N}\left[S(t_i)-\mu\right]^4}{\sigma^4}$$

平均值是μ，标准差是σ，N 是样本数。对于正态分布，峰度为 3，而对于尖头胖尾的分布，峰度大于 3。

1. 几何布朗运动模型:

$$\ln(P_T) \sim N\left[\ln P_0 + \left(\mu - \frac{\sigma^2}{2}\right)T, \sigma\sqrt{T}\right]$$

这是扩散过程

对数收益率为:

$$\ln P_T - \ln P_0 \sim N\left[\left(\mu - \frac{\sigma^2}{2}\right)T, \sigma\sqrt{T}\right]$$

服从正态分布

$N(\mu,\sigma)$是一个均值为μ，方差为σ的正态分布。

1. 学生 t 分布

2.3.1.4 学生 t 分布

为了完备起见, 这里还需要介绍一下学生 t 分布 (student's t-distribution)

$$P(z) = \frac{C_n}{(1+z^2/n)^{(n+1)/2}}$$

它是随机过程

$$z \equiv \frac{x\sqrt{n}}{\sqrt{y_1^2+\cdots+y_n^2}}$$

的分布。z 由从均值为 0、方差为 1 的正态独立随机变量 $y_1, y_2, \cdots, y_n$ 和 x 中产生。其中

$$C_n \equiv \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\Gamma(n/2)}$$

当 $n=1$ 时, $P(z)$ 为洛伦兹分布; 当 $n\to\infty$ 时, $P(z)$ 是高斯分布。一般地, 对 $k<n$, $P(z)$ 存在有限矩 $\mu_k \equiv \langle |z|^k \rangle$。因此, 符合学生 t 分布的随机过程可能同时具有无穷矩和有限矩。通过改变控制参数 n (它控制 k 阶矩的有限性), 人们可以以较好的精度来估计现实的价格对数收益率的分布。

1. 分形布朗运动

工具。分形布朗运动的表述方式很复杂 [McCauley, Gunaratne, and Bassler (2007)], 最初的定义就涉及复杂的数学语言:

设 (Ω, F^H, P_H) 为一概率空间, 常数 $H \in (0,1)$。具有 Hurst 参数 H 的分形布朗运动是一个满足下列条件的高斯过程 $\{B_H(t)\}_{t\in R^+}$:

(1) $B_H(0) = E_{P_H}[B_H(t)] = 0$, 对所有的 $t \in R^+$;

(2) $E_{P_H}[B_H(t)B_H(s)] = \frac{1}{2}\{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\}$, $s,t \in R^+$。

其中 $E_{P_H}(.)$ 表示在随机变量在概率侧度 P_H 下的期望, $F^H = \sigma\{B_H(s), s > 0\}$。$H$ 的选择将影响分形布朗运动的自相关性。

I.4 Levy 过程

I.4.1 特征函数

对概率密度函数做 Fourier 变换:

$$\varphi(q) = \int P(x)e^{iqx}\,\mathrm{d}x$$

I.4.2 稳定分布

随机变量之和的分布和 n 无关，服从变量分布

两个概率普渡函数的卷积:

$$P(S_2) = P(x_1) \otimes P(x_2)$$

但傅立叶变换后变成了直接乘积:

$$\mathcal{F}[P(S_2)] = \mathcal{F}[P(x_1)]\mathcal{F}[P(x_2)]$$

于是对于一个稳定分布应该是

I.4.3 Levy 稳定分布

形式如下

3. 列维稳定分布

稳定过程最一般的特征函数形式 [Levy (1925)]

$$\ln \varphi(q) = \begin{cases} \mathrm{i}\mu q - \gamma|q|^{\alpha}\left[1 - \mathrm{i}\beta \dfrac{q}{|q|}\tan\left(\dfrac{\pi}{2}\alpha\right)\right] & (\alpha \neq 1) \\ \mathrm{i}\mu q - \gamma|q|\left[1 + \mathrm{i}\beta \dfrac{q}{|q|}\dfrac{2}{\pi}\ln|q|\right] & (\alpha = 1) \end{cases}$$

这里 $0 < \alpha \leqslant 2$ (只有这个范围才对应稳定分布), γ 是一个正的尺度因子, μ 为任意实数, $\beta \in (-1, 1)$ 是一个非对称参数。

仅对很少几个 α 和 β 值, 列维稳定分布的密度函数存在解析形式:

- $\alpha = 1/2, \beta = 1$ 列维–斯米尔洛夫分布 (Levy-Smirnov);
- $\alpha = 1, \beta = 0$ 洛伦兹分布;
- $\alpha = 2$ 高斯分布。

2.3.2.2 截尾列维飞行分布

为了描述经验数据的有限方差和较大但有限的时间区间上的标度不变性, 1994 年, Mantegna 和 Stanley 提出了截尾列维分布 (truncated Levy flight, TLF) [Mantegna and Stanley (1994)] 过程, 其定义如下

$$P(x) \equiv \begin{cases} 0 & (x > l) \\ cP_L(x) & (-l \leqslant x \leqslant l) \\ 0 & (x < -l) \end{cases}$$

其中 $P_L(x)$ 是具有指数 α 和尺度因子 γ 的零均值对称列维分布, c 是标准化常数。虽然 TLF 能满足经验数据的标度特性和数字特征, 但这个模型也是唯象的。

I.5 幂律分布与标度律

特征函数

$$\varphi(q) = e^{-\eta\gamma\,|q|^{\alpha}}$$

则概率密度函数是:

$$P(S_n) = \frac{1}{\pi}\int e^{-\eta\gamma\,|q|^{\alpha}}\cos(qS_n)\,\mathrm{d}q$$

当$S_n = 0$，得到

$$P(S_n = 0) = \frac{\Gamma\left(\frac{1}{\alpha}\right)}{\pi\alpha(\gamma n)^{\frac{1}{\alpha}}}$$

重标度化:

$$\tilde{P}\left(\tilde{S}_n\right) = P(S_n)n^{\frac{1}{\alpha}}$$

标准化:

$$\int_{\{-\infty\}}^{\{\infty\}} \tilde{P}\left(\tilde{S}_n\right) \mathrm{d}\tilde{S}_n = 1$$

给出S_n标度变化的形式:

$$\tilde{S} = S_n n^{-\frac{1}{\alpha}}$$

4.4.2　有限系统中的幂率

今天，幂率分布被用于描述开放系统。然而，被观察到的标度特性通常会受到规模的有限性或系统内在的一些其他限制的影响。临界现象［147］就是一个很好的实例，它既体现了幂率应用的成果，也反映了在应用幂率中存在的困难。在无限系统的临界状态中我们观测到了幂率相关函数，但如果系统是有限的，那么有限性就限制了观测幂率行为的范围。尽管存在这些限制，引入与使用标度概念——它与幂率的相关性本质有密切关系——对我们准确理解临界现象仍非常关键，即算我们所考虑的是有限系统［59］。

I.6 时间相关类研究

I.6.1 簇集性

波动率:

$$V = \sigma_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[S(t_i) - \mu]^2}$$

或者:

$$V = \sigma_t = \frac{1}{N}\sum_{i=1}^{N}[S(t_i) - \mu]$$

波动率的簇集性: 大的价格变化会引起更大的价格变化，小的价格变化会引起更小的价格变化

I.6.2 持续性

当$H=0.5$时,时间序列是一个无偏的随机游走过程;当$0.5<h\le 1$时,是一个持续性的随机游走,具有长期记忆性;当$0\le H<0.5$时,是一个反持续性的随机游走,即均值回复过程。

在更一般的情形下,若$\sigma \propto (\Delta T)^{\alpha}$,将随机过程比拟为扩散过程,则$\alpha<0.5,\alpha=0.5,\alpha>0.5$的情形,分别可定义所谓的次扩散过程(subdiffusion)、扩散过程 (diffusion process) 和超扩散过程(super-diffusion)

I.6.3 相关性

1. 自相关函数: 考察一个收益率时间序列$S(t)$,其自协方差$Cov(\Delta t)$定义为时间序列$S(t)$与其经过时间平移后的序列$S(t-\Delta t)$之间的协方差。自协方差通过方差进行标准化之后,即成为自相关函数$\rho(\Delta t)$。

   协方差计算:

$$Cov(\Delta t)=E\{[S(t)-\mu][S(t-\Delta t)-\mu]\}$$

   自相关函数:

$$\rho(\Delta t)=\frac{Cov(\Delta t)}{\sigma^2}$$

2. 互相关函数: 两个时间序列$S_1(t)$和$S_2(t)$的互协方差$Cov(S_1,S_2)$定义为两个时间序列之间的协方差。互协方差通过两个时间序列的标准差进行标准化之后,即成为互相关函数$\rho(S_1,S_2)$。

   协方差计算:

$$Cov(S_1,S_2)=E\{[S_1(t)-\mu_1][S_2(t)-\mu_2]\}$$

   互相关函数:

$$\rho(S_1,S_2)=\frac{Cov(S_1,S_2)}{\sigma_1\sigma_2}$$

I.6.4 度量空间

考虑一个标准化的收益率:

$$\tilde{S}_i = \frac{S_i - \langle S_i \rangle}{\sqrt{\langle S_i^2 \rangle - \langle S_i \rangle^2}}$$

取同个时间内的N组值,因此收益率是一个N维向量,自然可以定义向量的距离:

$$d_{(ij)_=} \sqrt{\left(S_i - S_j\right)^2} = \sqrt{\sum_{\{k=1\}}^{N} \left(\tilde{S}_i^2 + \tilde{S}_j^2 - 2\tilde{S}_{ik}\tilde{S}_{jk}\right)}$$

标准化的向量是归一化的，于是自然得到度量的结果是:

$$d_{ij} = \sqrt{2\left(1 - \sum_{\{k=1\}}^{N} \tilde{S}_{ik}\tilde{S}_{jk}\right)} = \sqrt{2\left[1 - \rho\left(S_i, S_j\right)\right]}$$

这样就构造了收益率空间上的度量结构，可以证明这个度量满足度量空间条件:

1. 非负性:

    $$d_{ij} \geq 0$$

2. 同一性:

    $$d_{ij} = 0$$

    当且仅当

    $$i = j$$

3. 对称性:

    $$d_{ij} = d_{ji}$$

4. 三角不等式:

    $$d_{ij} \leq d_{ik} + d_{kj}$$

此外，有研究指出，收益率空间上的度量结构是一个超度量空间，即满足超三角不等式:

$$d_{ij} \leq \max\left(d_{ik}, d_{kj}\right)$$

超度量空间与分层方法相关，其可以描述复杂系统，例如 spin glasses 系统。

度量空间的某种分类可以获得超度量空间的集合，这成为亚超度量空间，可以通过关联分层的最小生成树方法获得。

II 时间相关类研究

II.1 ARCH 模型

定义σ_t是$t-1$时刻所估计的金融产品价格在t时刻的波动率，σ_t^2为所估计的t时刻价格的方差，设金融产品的价格序列为$\{Y_t\}$，则其对数收益率序列为$\{S_t\}$，则：

$$S_t = \ln(Y_t) - \ln(Y_{t-1})$$

t时刻前 m 长度的对数收益序列来估计t时刻的方差：

$$\sigma_t^2 = \frac{1}{m-1} \sum_{\{i=1\}}^{m} \left(S_{t-i} - \langle S \rangle\right)^2$$

其中均值的定义是：

$$\langle u \rangle = \frac{1}{m} \sum_{\{i=1\}}^{m} S_{t-i}$$

期望值远小于标准差，因此可近似得到：

$$\sigma_t^2 = \frac{1}{m-1} \sum_{\{i=1\}}^{m} S_{t-i}^2$$

自回归的 ARCH(m)模型是：

$$\sigma_t^2 = \alpha_0 + \sum_{\{i=1\}}^{m} \alpha_i S_{t-i}^2$$

其中$\alpha_0 = \gamma V_L$，V_L是长期平均方差，γ和α_i都是权重，u_{t-i}是对数收益率的序列，数据要更新更好，因此权重的设置上自然有

$$0 < \alpha_0 < 1, \alpha_i > \alpha_j, i < j$$

归一化的要求是：$\gamma + \sum_i \alpha_i = 1$

II.1.1 例：ARCH(1)模型

$$\sigma_t^2 = \alpha_0 + \alpha_1 S_{t-1}^2$$

如果随机变量是具有高斯分布的随机变量，那么可以计算无条件方差：

$$\sigma^2 = \frac{\alpha_0}{1-\alpha_1}$$

其中$1-\alpha_1 \neq 0$

峰度是:

$$k = \frac{\langle x^4 \rangle}{\langle x^2 \rangle^2} = 3 + \frac{6\alpha_1^2}{1-3\alpha_1^2}$$

II.2 GARCH 模型

GARCH 模型是在 ARCH 模型的基础上引入了波动率的自回归项，即:

$$\sigma_t^2 = \alpha_0 + \sum_{\{i=1\}}^{m} \alpha_i S_{t-i}^2 + \sum_{\{j=1\}}^{n} \beta_j \sigma_{t-j}^2$$

归一有限条件是: $\gamma + \sum_i \alpha_i + \sum_j \beta_j = 1$

II.2.1 例: GARCH(1,1)模型

$$\sigma_t^2 = \alpha_0 + \alpha_1 S_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

这是一个递推公式，因此可以写成

$$\sigma_t^2 = \left(\sum_{i=0}^{t-2} \beta^i\right) \alpha_0 + \alpha \sum_{i=0}^{t-2} (\beta^i u_{t-1-i}^2) + \beta^{t-1} \sigma_1^2$$