Non-Normal Distributions

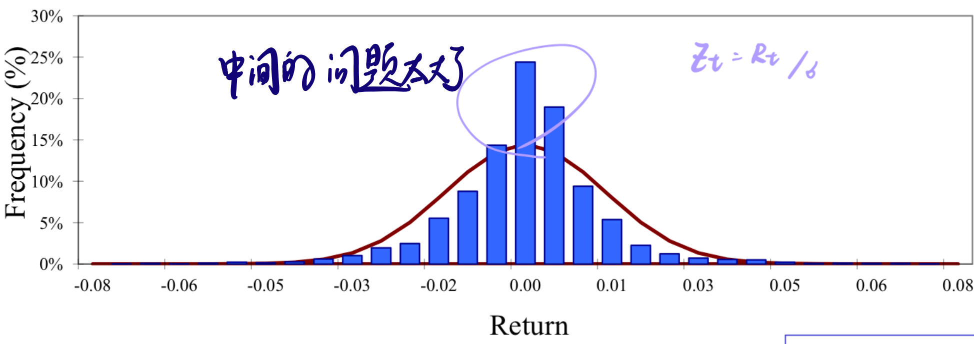
- The way we thinking about things

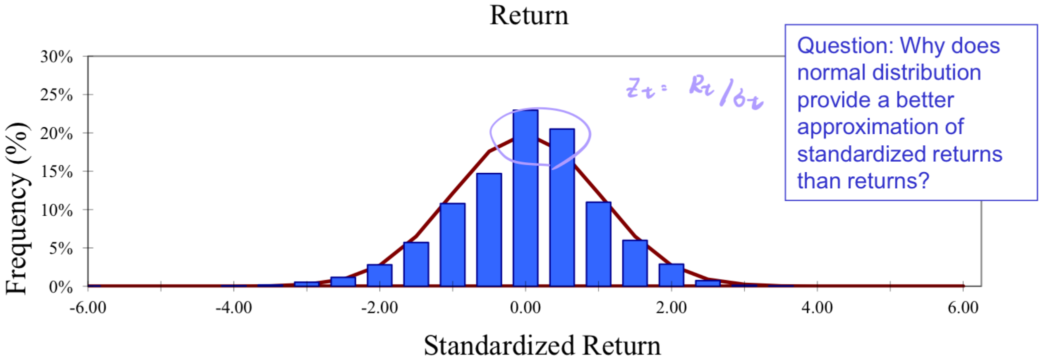
$$R_t = \sigma_t z_t \qquad Var(z_t) = 1$$

这样我们能计算 Standardized return $z_t = \frac{R_t}{\sigma_t}$

我们也能有另一种标准化的方式：用今天的 σ (GARCH 得到)

或是用 unconditional 的 σ： $z_t = \frac{R_t}{\sigma}$





为什么标准化这一过程这么重要呢？

$r_1 \sim N(0, 0.01^2)$ day 1's return

$r_2 \sim N(0, 0.015^2)$ next day's return

$r_3 \sim N(0, 0.02^2)$

$r_4 \sim N(0, 0.013^2)$

$\vdots$

每一个 return 都是服从正态分布的，但它们都有不同的方差 → 实际上数据往往是这样的

下面我们 take all returns and pool them together

The pooled distribution is not normal even if each return is conditionally Normal

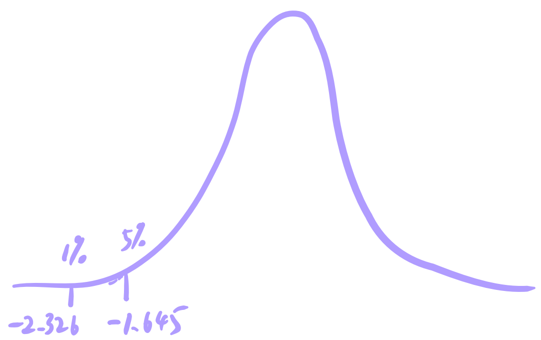
(on knowing σ_t^2)

一个原因是 each day conditional distribution 不一定是 normal 的. 另一个是 pooled distribution 不是正态分布

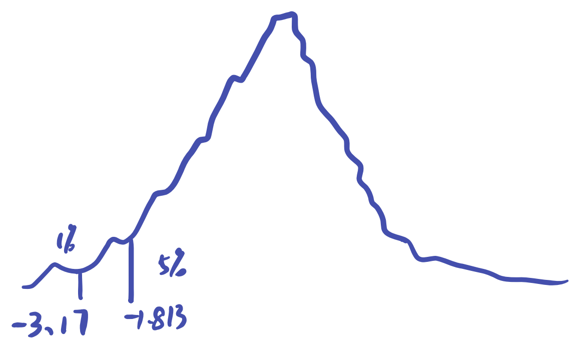
那我们怎么做呢?

首先用 QQ plot :

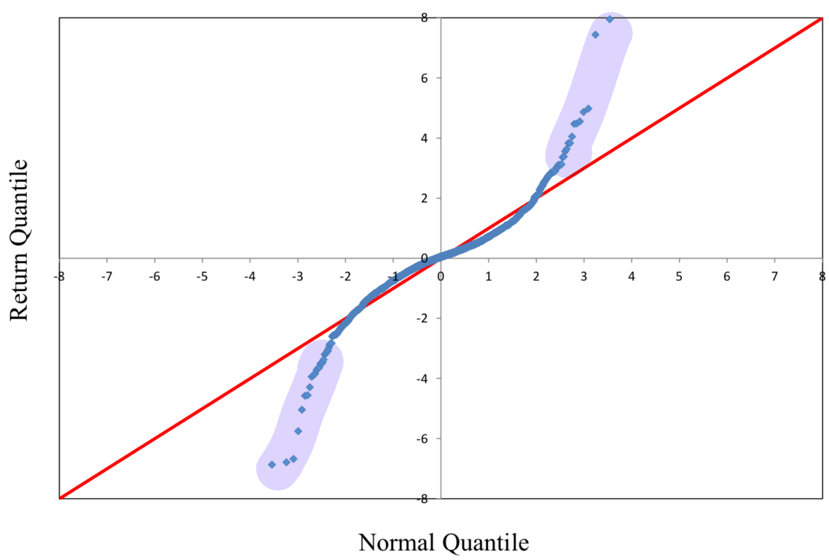
假如我们有一个正态分布 :



real distribution



如果 data 是正态分布的, 那么 quantile 应该是一样的



实际的 data 有 fat tails

我们试图将标准化后的 return 的方差定为 1

但我们用什么标准化呢？

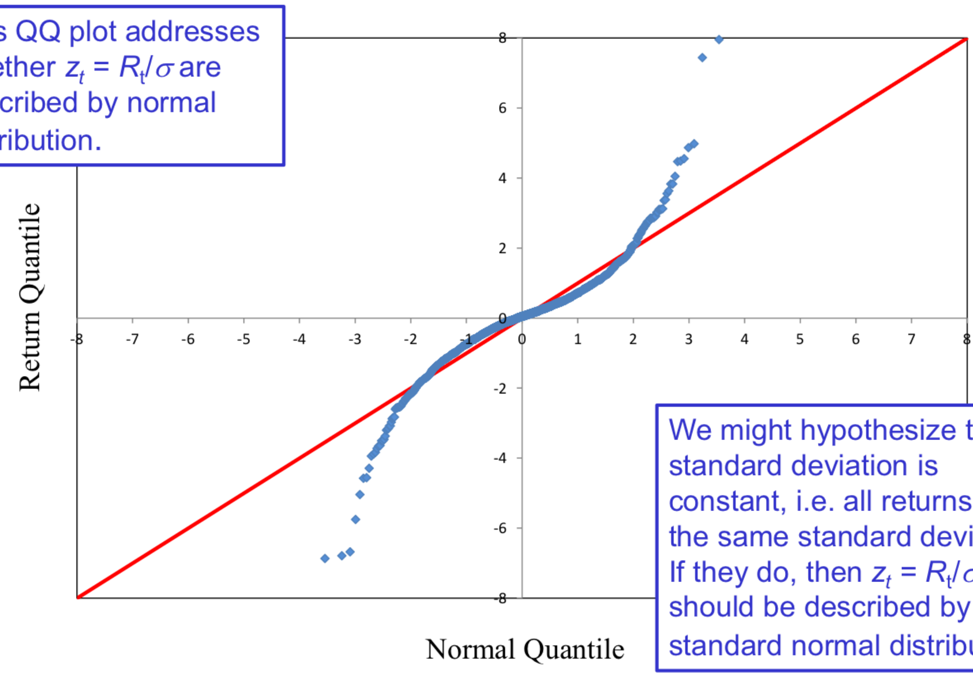
Now we consider two standardizations:

(a) Standardize by unconditional standard deviation σ, i.e. $z_t = R_t/\sigma$

(b) Standardize by conditional standard deviation σ_t from a variance forecasting model, e.g. a GARCH model. In this case $z_t = R_t/\sigma_t$

(a)：

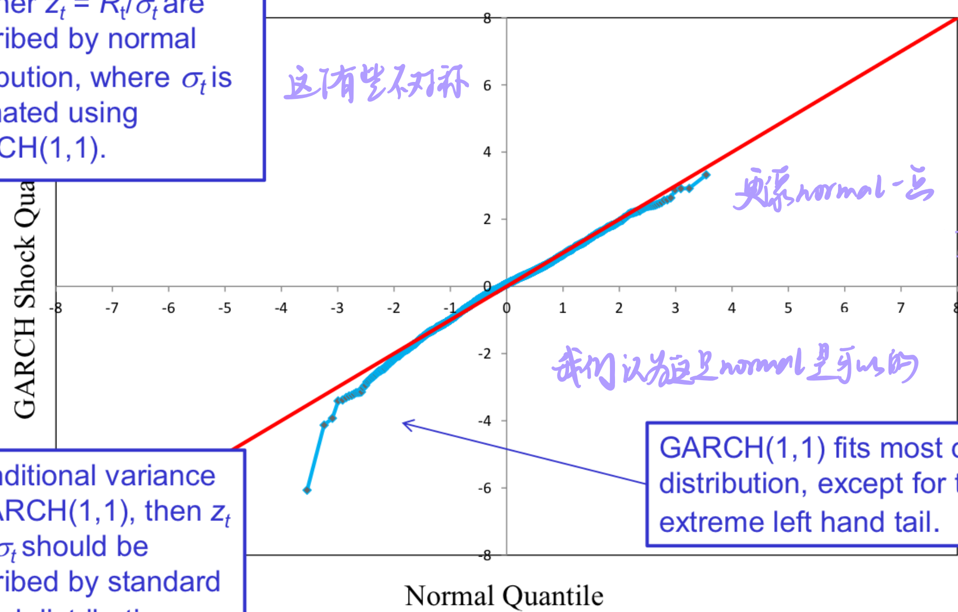
This QQ plot addresses whether $z_t = R_t/\sigma$ are described by normal distribution.



We might hypothesize that standard deviation is constant, i.e. all returns have the same standard deviation. If they do, then $z_t = R_t/\sigma$ should be described by standard normal distribution.

(b)：

This QQ plot addresses whether $z_t = R_t/\sigma_t$ are described by normal distribution, where σ_t is estimated using GARCH(1,1).



这行有些不对称

更像normal一点

我们认为这是normal是可以的

If conditional variance is GARCH(1,1), then $z_t = R_t/\sigma_t$ should be described by standard normal distribution.

GARCH(1,1) fits most of distribution, except for the extreme left hand tail.

What's the best way to compute VaR

最好的是 Filtered Historical Simulation，因为可以忽略数据的分布，（正态和t都是对称的）

rescale the return

$\frac{r_{t-\tau}}{\sigma_{t-\tau}}\sigma_t$ 这样 std.dev = σ_t

Natural choice from GARCH Model

这样能捕捉数据的分布

我们假设 correlation 是恒定的

补充：教材中 standardized ≠ standard

Standard t distribution

$$f_{t(d)}(x;d)=\frac{\Gamma((d+1)/2)}{\Gamma(d/2)\sqrt{d\pi}}(1+x^2/d)^{-(1+d)/2},\ \text{for } d>0$$

$E[x]=0$, when $d>1$ $\quad Var[x]=\frac{d}{d-2}$ when $d>2$

对我们来说这个性质实不怎么样. 我们的模型中 $Var(z_t)=1$

这样我们要 rescale t-distribution

$$z=\frac{x-E[x]}{\sqrt{Var[x]}}=\frac{x}{\sqrt{d/(d-2)}}$$

注意！在用 t-distribution 中 MLE 要先 log 再加. 因为Γ↑非常快，可能会非常大

这样 $\mu=1$ $\quad \sigma^2=1$ $\quad \zeta_1=0$ $\quad \zeta_2=\frac{6}{d-4}$

用这个方法可以求d，但不如MLE好用

$$\ln L_1=\sum_{t=1}^{T}\ln(f(z_t;d))$$

$$=T\{\ln(\Gamma((d+1)/2))-\ln(\Gamma(d/2))-\ln(\pi)/2-\ln(d-2)/2\}$$

$$-\frac{1}{2}\sum_{t=1}^{T}(1+d)\ln(1+(R_{PF,t}/\sigma_{PF,t})^2/(d-2))$$

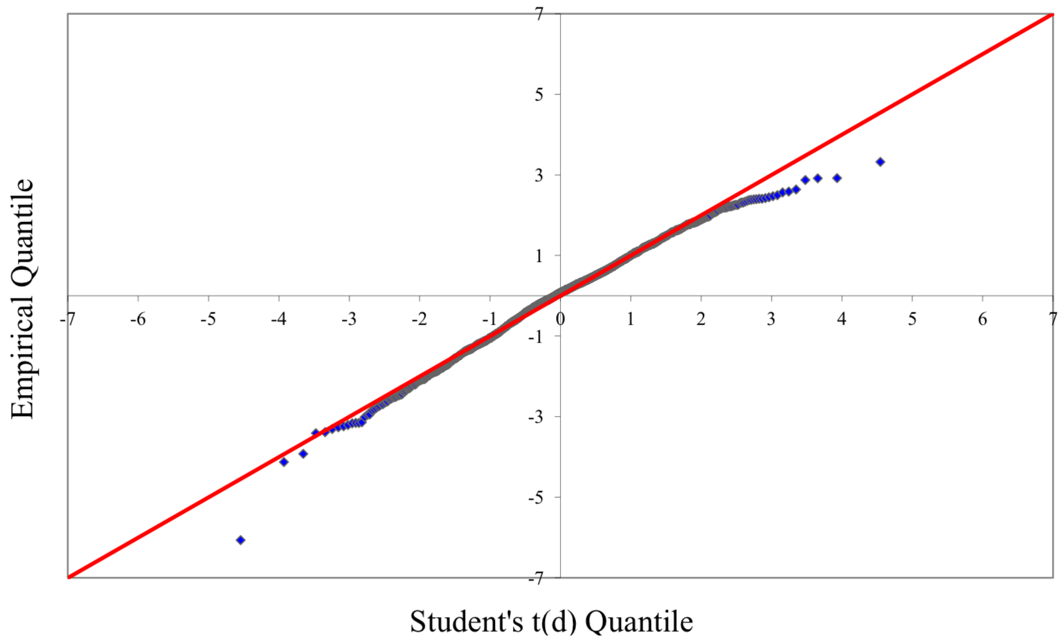
t distribution 的 MLE

另一种方法是estimate everything in one step using density function

1. estimate the GARCH(1,1) using Normal Distribution

2. construct $z_t = \frac{R_t}{\sigma_t}$

$\hookrightarrow$ satisfies standardized t distribution



用 t-distribution 看起来不错，但它不是完美的．因为 t-distribution 是对称的 left hand tail 决定了 df，这样 right hand tail 就不准了

有什么解决办法呢？可以用两个 t-distribution 拼在一起

$$f_{asyt}(z; d_1, d_2) =$$

$$\begin{cases} BC\left[1+(Bz+A)^2/\left((1-d_2)^2(d_1-2)\right)\right]^{-(1+d_1)/2}, & \text{if } z < -A/B \\ BC\left[1+(Bz+A)^2/\left((1+d_2)^2(d_1-2)\right)\right]^{-(1+d_1)/2}, & \text{if } z \geq -A/B \end{cases}$$

where

$$A = 4d_2C\frac{d_1-2}{d_1-1}, \quad B = \sqrt{1+3d_2^2-A^2}, \quad C = \frac{\Gamma((d_1+1)/2)}{\Gamma(d_1/2)\sqrt{\pi(d_1-2)}}$$

Wednesday Material

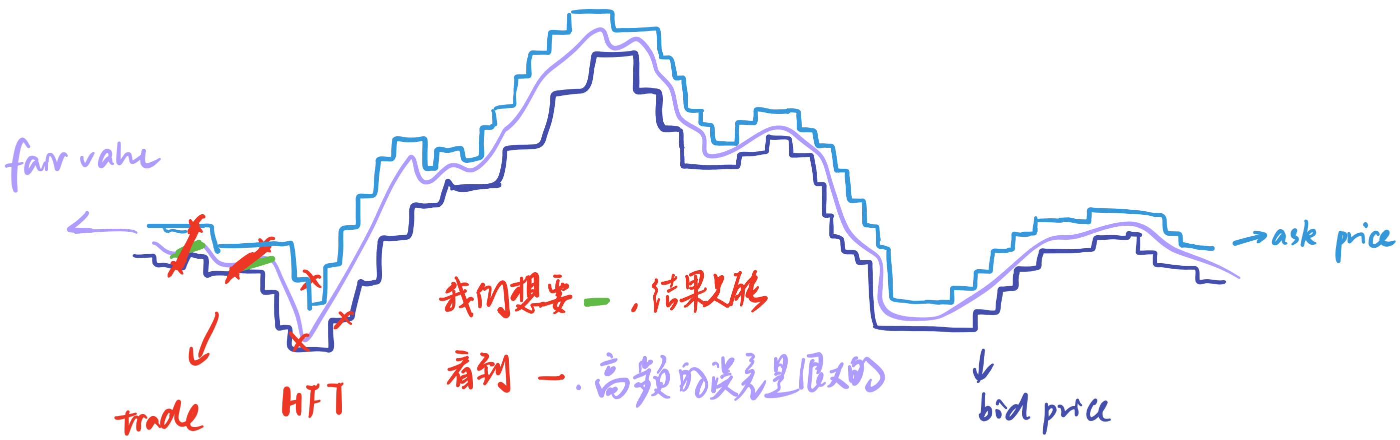
为什么用 daily return 呢? $\ln S_t - \cancel{\ln S_{t-1}}$ 和frequency 没啥关系.

为什么不用小时/分钟的数据: $\cancel{\ln S_{t-1}} - \cancel{\ln S_{t-2}}$ 反正都取首尾的值

$\cancel{\ln S_{t-2}} - \cancel{\ln S_{t-3}}$

$\cancel{\ln S_{t-3}} - \ln S_{t-4}$

在 Ito's Process中, 如果你知道 path 的话, 如果很高频的数据, 就能很好的估计 variance. 但是如果我们用高频数据



谁提供了quote? 如果有公司想买 100,000 股某股票

Goldman & Sachs.

+ brokerage firm execution algorithm

我们观测到的是每过一会就有一个 trade

但是 daily return 好在哪儿呢

但是在 data 中

bid − ask 是不对称的