Maximum Likelihood Estimation

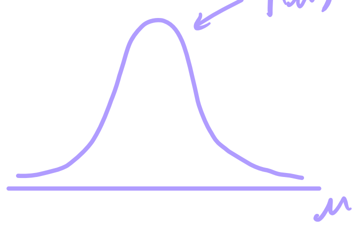
- main idea :

  1 observation $f(x,\mu) = \frac{1}{\sqrt{2\pi}\times 6} \times \exp\{-\frac{1}{2}(\frac{x-\mu}{6})^2\}$ ← only one parameter

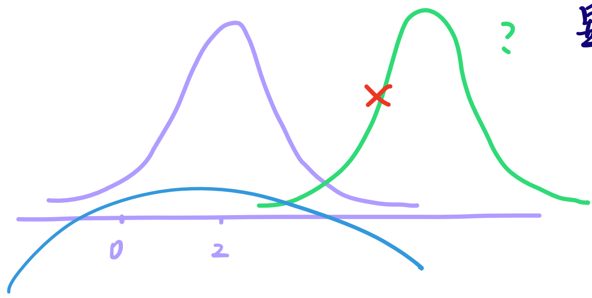
$f(x,(\mu)) \rightarrow f(x)$

我们现在想用 x 来估计 μ

⇓



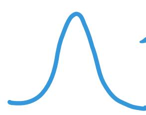
我们看到 $x=2$，那么 $\mu=2$？→ 是个好的estimate



显然 $\mu=2$ 概率最大

用 ln f 比较好，for two reasons { Theoretical Reason

practical reason : s.d.如果特别小

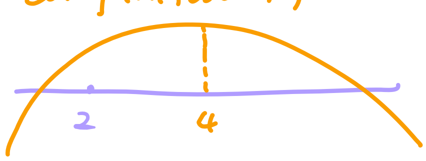


会有 floating point overflow problems

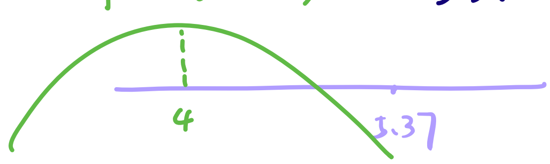
此外，我们现在只有一个 observation，如果有很多 observations 怎么办？需要算 joint distribution，即要把 density 乘起来 (if independent) $\ln ab = \ln a + \ln b$，这很好

假如我们有两个 observation：

$\ln f(x_1, \mu=4)$



$\ln f(x_2, \mu=4)$



其实不应该是4，而是 $\frac{2+5.37}{2}$

我们需要找一个 simple parameter factor maximize the sum of components
看起来不好找，但实际上很简单，求导就可以

把 lnf 加起来，找到最大值

$$f(x_3, x_2, x_1, \mu, \sigma) = f(x_3)f(x_2)f(x_1) \quad \text{(If independent)}$$

If not independent: $f(x_3|x_1,x_2)f(x_2|x_1)f(x_1) \rightarrow$ conditional density

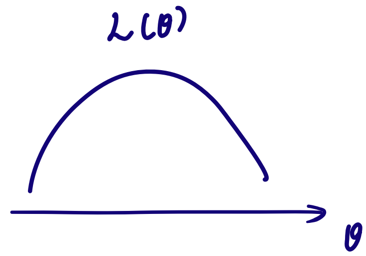
financial model 一般告诉你 conditional density 是什么

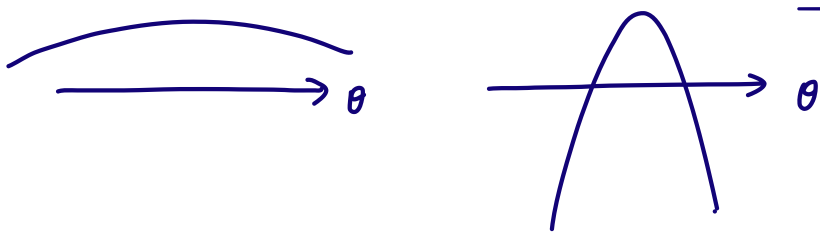
$x_2 = \alpha + \beta x_1 + \varepsilon_2$, $\varepsilon_2 \sim N(0, \sigma^2)$

$$\frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left\{-\frac{1}{2}\frac{[x_2-(\alpha+\beta x_1)]^2}{\sigma^2}\right\}$$

- log likelihood

one unknown parameter θ





×　　　　○

We care about curvature

GARCH Modeling

- GARCH — Generalized Autoregressive Conditional Heteroscedasticity (different variances)

  — 用于预测 volatility

  It's conditional because it depends on some past data

- $\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta \sigma_t^2$, $\alpha+\beta<1$ —— GARCH(1,1)

GARCH Model 有什么用呢?

- Mean-Variance optimization

  需要 mean (假设已知), covariance matrix (s.d. → 本课内容, correlation)

- Delta-normal & Monte Carlo VaR

  — 也用到了 covariance matrix

- Filtered Historical Simulation

  — 需要 rescaled returns $\frac{\sigma_{t+1}}{\sigma_t} \times R_t$. σ_{t+1} 和 σ_t 可用 GARCH 求

  — 为什么要用 GARCH 求呢?

  假如我们有 database of returns. Naive historical simulation 的一个局限就是一些数据是五年前的, 那时 volatility 更低 → not so relevant

  $\boxed{\underbrace{\frac{R_{t-\tau}}{\sigma_{t-\tau}}}_{var=1} \quad \sigma_{t+1}}$ → $var = \sigma_{t+1}^2$ ⇒ 这样我们就能用所有的数据了

- 假如 implied volatility 没有的话, 可以用 GARCH 的 volatility 来定价

  *如果有 implied volatility 那就一定用 implied volatility, 但有些国家就没有健全的期权机制

- 我们从 Simple Variance Forecasting 开始

$$R_{t+1} \equiv \ln\left(\frac{S_{t-1}}{S_t}\right)$$

$R_{t+1} = \sigma_{t+1} z_{t+1}$, $z_{t+1} \sim$ i.i.d. $N(0,1)$

$\Rightarrow \quad \sigma_{t+1}^2 = \sum_{\tau=1}^{m} \frac{1}{m} R_{t+1-\tau}^2$

- Risk Metrics

$\sigma_{t+1}^2 = \lambda \sigma_t^2 + (1-\lambda) R_t^2 \longrightarrow$ 这就和 GARCH 很像

$\Downarrow$

GARCH Model，看起来很灵活

$\sigma_{t+1}^2 = w + \alpha R_t^2 + \beta \sigma_t^2$, with $\alpha + \beta < 1$

$\Rightarrow$ GARCH(1, 1)

one lag of the return ↲ ↳ one lag of the conditional variance σ_t^2

$\sigma^2 \equiv E[\sigma_{t+1}^2] = w + \alpha E[R_t^2] + \beta E[\sigma_t^2]$

$= w + \alpha \sigma^2 + \beta \sigma^2$

$\Rightarrow \quad \sigma^2 = \frac{w}{1-\alpha-\beta}$ It's stationary

$\Rightarrow w = (1-\alpha-\beta)\sigma^2$

$\sigma_{t+1}^2 = (1-\alpha-\beta)\sigma^2 + \alpha R_t^2 + \beta \sigma_t^2 = \underline{\sigma^2} + \alpha\underline{(R_t^2 - \sigma^2)} + \beta\underline{(\sigma_t^2 - \sigma^2)}$

可以看作是 weighted average 对 "—"

$t \qquad t+1 \qquad \cdots \qquad t+k-1 \qquad t+k$

取决于之前的 date $\leftarrow \sigma_t^2 \xrightarrow{\text{决定}} \sigma_{t+1}^2 \longrightarrow \quad \cdots \quad \sigma_{t+k-1}^2 \xrightarrow{\text{决定}} \sigma_{t+k}^2$

$w + \alpha R_t^2 + \beta \sigma_t^2$

下面是很漂亮的数学：

$$E_t[\sigma_{t+k}^2] - \sigma^2 = \alpha E_t[R_{t+k-1}^2 - \sigma^2] + \beta E_t[\sigma_{t+k-1}^2 - \sigma^2]$$
$$= \alpha E_t[\sigma_{t+k-1}^2 Z_{t+k-1}^2 - \sigma^2] + \beta E_t[\sigma_{t+k-1}^2 - \sigma^2]$$
$$= (\alpha+\beta)(E_t[\sigma_{t+k-1}^2] - \sigma^2)$$

很容易找到通项公式

$$E_t[\sigma_{t+k}^2] - \sigma^2 = (\alpha+\beta)^{k-1}(\sigma_{t+1}^2 - \sigma^2)$$

→ standard deviation per day

注意 Returns are uncorrelated, 但不是 independent

* $\alpha+\beta<1$, variance will be mean-reverting → stationary

但时常 $\alpha+\beta = 0.99 / 0.98$

但是 GARCH model 的一些 parameters 需要 MLE

data	variance
r_1	σ_1^2
r_2	$w+\alpha r_1^2+\beta\sigma_1^2$
r_3	$w+\alpha r_2^2+\beta\sigma_2^2$

↓

for each observation we can write the variance

conditional density

$$\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_2^2(\sigma_1^2, w, \alpha, \beta)}\cdot \exp\left(-\frac{1}{2}\frac{r_2^2}{\sigma_2^2(\sigma_1^2, w, \alpha, \beta)}\right)$$

↓

$\sigma_2^2 = w+\alpha r_1^2+\beta\sigma_1^2$

→ log conditional density function and add them up

GARCH Modeling

$$E_t[\sigma_{t+k}^2] - \sigma^2 = (\alpha+\beta)^{k-1}(\sigma_{t+1}^2 - \sigma^2)$$ → use for forcasting

在 GPD 中我们做什么了呢？我们得到了所有 observation 的 density. GARCH 也是一样的

- 现在我们假设 return 是符合正态分布的：

$$R_t = \sigma_t z_t, \quad z_t \sim i.i.d.\ N(0,1)$$

Probability of R_t is

$$l_t = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{R_t^2}{2\sigma_t^2}\right)$$

那么 joint likelihood 就是

$$L = \prod_{t=1}^{T} l_t = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{R_t^2}{2\sigma_t^2}\right)$$

注意：我们这里把 likelihood 乘起来并不意味着 return 是相互独立的

$$\sigma_t^2 = \sigma^2(1-\alpha-\beta) + \alpha R_{t-1}^2 + \beta\sigma_{t-1}^2$$

在用 MLE 时，我们会 maximize σ^2, α, β

$$= \omega \quad + \alpha R_{t-1}^2 + \beta\sigma_{t-1}^2$$

这个时候就是 maximize ω, α, β

$$\text{Max} \ln L = \text{Max} \sum_{t=1}^{T} \ln(l_t) = \text{Max} \sum_{t=1}^{T} \left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{1}{2}\frac{R_t^2}{\sigma_t^2}\right]$$

可以看作 maximize $\left[-\frac{1}{2}\left(\sum_{t=1}^{T} \ln\sigma_t^2 + \frac{R_t^2}{\sigma_t^2}\right)\right]$

MLE 是很好的，随着 sample size $T \to \infty$，parameter estimates → true values

estimates 的方差也最小

注意！就算 conditional distribution is not normal，MLE 的 estimate 还是会收敛于真实值

我们在做MLE时，需要用到第四个 parameter：

$$\sigma_1^2 = \omega + \alpha R_0^2 + \beta \sigma_0^2$$

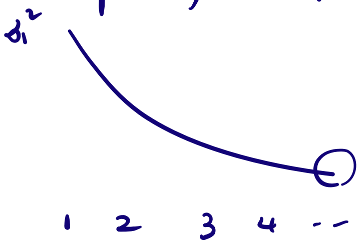
—— 假如我们的数据是从 $t=1$ 开始的，R_0 和 σ_0 根本不存在，那该怎么办呢？
我们就需要引入 σ_1^2

σ_1 是哪来的呢？ $l_1 = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{R_1^2}{2\sigma_1^2}\right)$

用MLE 多估一个 parameter 就可以了

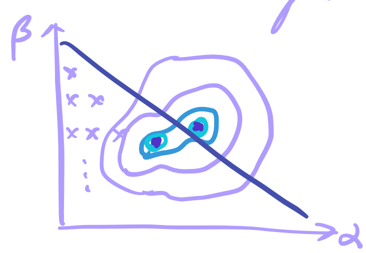
而 long-run variance 可以看作 sample variance

$\hat{\sigma}_1^2 = \frac{1}{N}\sum_{t=1}^{N} R_t^2$ → 用 sample variance 估计，但其实是错的，而很多人是这样做的。σ_1^2 其实不影响后来的方差



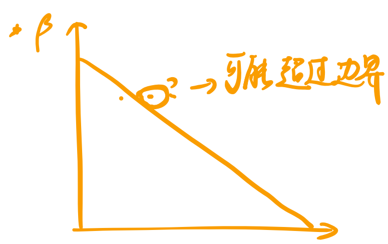
* 注意！Optimizer 是需要我们猜值的，在 Bloomberg 中是还是不能乱猜的。

因为只估计 α 和 β，不需要 optimizer



我们要施加限制条件

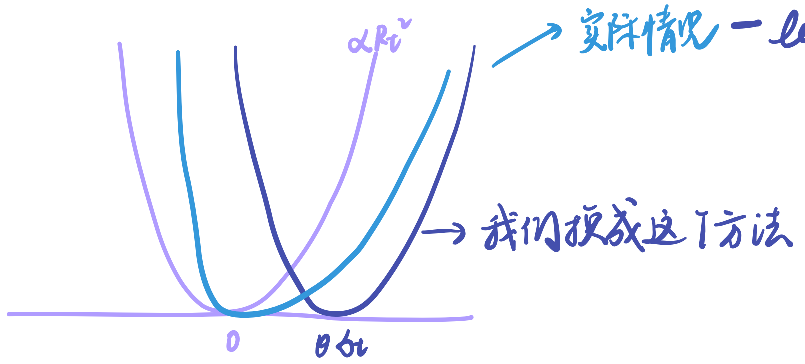
$\alpha, \beta \geq 0$，因为 σ 不能 <0，$\alpha + \beta < 1$



More on GARCH Modeling

· 之前我们一直在用 GARCH (1, 1). 它有什么问题呢?

It's SYMMETRIC · $\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta \sigma_t^2$



实际情况有什么特征呢? αR_t^2 is symmetric around 0

Return 同样变化 5%, -5% 带来的 variance 增长 比 5% 带来的 Variance 增长大 ⇒ 这个叫做 leverage effect 为了捕捉这一特征, 我们引入 θ

$$\sigma_{t+1}^2 = \omega + \alpha (R_t - \theta\sigma_t)^2 + \beta\sigma_t^2 = \omega + \alpha\sigma_t^2 (z_t - \theta)^2 + \beta\sigma_t^2$$ → NGARCH model

capture the asymmetry

这是 first approach → 和 GARCH (1, 1) 一样简单

* $E[\sigma_{t+1}^2] = \omega + (\alpha(1+\theta^2) + \beta) E[\sigma_t^2]$

* 为什么叫做 leverage effect?

当 stock price ↓, companies are more leveraged

A more natural way to do is to introduce a dummy variable I allows the slope to be different to the right and left of 0.

这样 Model 就变成了 $\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \alpha\theta I_t R_t^2 + \beta\sigma_t^2$

上面的这些方法都有较强的操作性．为什么呢？

$$\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sigma_t}\cdot\exp\left\{-\frac{1}{2}\frac{R_t^2}{\underline{\sigma_t^2}}\right\}$$

↳ 对于 almost any function 来说，它都是 stationary 的 → 有的可能不是 stationary
只是 specification 不一样

θ 捕捉了 leveraged effect — GJR-GARCH model.

接下来我们放宽条件，α, β 不一定 >0 → $\ln\sigma_{t+1}^2$ 可以 <0

$$\ln\sigma_{t+1}^2 = \omega + \alpha\left(\phi R_t + \gamma\left[|R_t| - E|R_t|\right]\right) + \beta\ln\sigma_t^2$$

$\alpha\phi<0$ → leverage effect

当然我们也可以取更多的 lag ： GARCH (p, q)

$$\sigma_{t+1}^2 = \omega + \sum_{i=1}^{p}\alpha_i R_{t+1-i}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t+1-j}^2$$ 但一般取 lag = 2

estimation is annoying．昨天的 variance 和今天的 variance 有很强的相关性，我们也很难解释

让我们回到 GARCH(1、1) model 中来 可能 long-run variance 是随机过程

* 右侧可以加任意 Explanatory variables．比如可以加一个 Calendar Effect 的 dummy variable ，比如是否是周一、是否在 Dec 15 ~ Jan 1 之间

→ 这就涉及到了 Model Selection 的问题：选哪个 Model 好呢？

举一个例子：

GARCH(1,1)：α、β、σ^2、σ_1^2 $\theta=0$ $\ln L_0$

NGARCH(1,1)：α、β、σ^2、σ_1^2、θ $\ln L_1$

$\ln L_1$ 更大，因为在GARCH(1,1)可以看作 $\theta=0$，unrestricted 一定大于等于 restricted 的 log likelihood

它们的差值 $2(\ln L_1 - \ln L_0) \sim \chi_1^2 = LR$

df = # of constraints

under the H0：added parameters in model 1 are significant

如果 LR is large（> χ_1 critical value，我们就应该选更加灵活的model）

有什么 Limitation 呢？Always reject the less flexible model

因为什么呢？mean return 的 confidence interval 是很大的，so volatile that doesn't contain too much information.

我们不知道 CAPM，FF-3 factor model 是否准确 Average return 不能用来预测 expected return

而对于 variation 来说，data is so informative，模型会永远选择包括更多 parameter 的模型

那我们怎么办呢？ad hoc thing，如果你有一个好模型（good variance forecast），$R_{t+1}^2 = b_0 + b_1 \sigma_{t+1}^2 + e_{t+1}$，$b_0=0$，$b_1=1$

我们需要针对具体数据具体分析