Financial Risk Management

Spring 2016

Dr. Ehud Peleg

Maximum Likelihood Estimation

UCLA Anderson
THINK IN THE NEXT

Agenda

- Likelihood function
- Maximum Likelihood Estimation
- Applying MLE to Volatility Models
- Confidence Intervals
- Likelihood Ratio Tests

Likelihood Function - Example

- Suppose we draw one number from a normal distribution, what is the probability density function, if we know μ=1 and σ=3?

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\cdot 3} e^{-\frac{(x-1)^2}{2\cdot 3^2}}$$

- Suppose we don't know μ and σ, but the number we draw is 1. What is the likelihood?

$$L(\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(1-\mu)^2}{2\sigma^2}}$$

Maximum Likelihood Estimator

- Suppose our data: $X = (x_1, x_2, \cdots, x_n)$ are a realization from a joint probability density . $f(X;\theta)$

  θ is the vector of parameters of the density function.
- The likelihood is: $L(\theta;X) = f(X;\theta)$

  i.e. a function of θ where the observed data are known.
- In maximum likelihood methods, we find the parameters that maximize the likelihood of the observed sample.

Log Likelihood for iid Data

- Typically, we prefer to maximize log-likelihood rather than likelihood.
- If the data are i.i.d. then we have:

$$l(\theta;X)=\log L(\theta;X)=\log\prod_{i=1}^{n}f(x_i;\theta)=\sum_{i=1}^{n}\log L(\theta;x_i)$$

- The parameters that maximize this function can be shown to be good estimators of the true parameter.

Likelihood for iid data example

- Suppose we draw 3 numbers from a Normal distribution: 1, -2, 3.
- Suppose we know mean = 0. The free parameter is v, the variance.
- What is the log likelihood?

$$L(\nu;x_i)=\frac{1}{\sqrt{2\pi\nu}}e^{-\frac{(x_i-0)^2}{2\nu}}\Rightarrow l(\nu;x_i)=-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\left[\ln(\nu)+\frac{x_i^2}{\nu}\right]$$

$$l(\nu;x_1,x_2,x_3)=\sum l(\nu)=-\frac{3}{2}\ln(2\pi)-\frac{3}{2}\ln(\nu)-\frac{1}{2}\left[\frac{(1)^2}{\nu}+\frac{(-2)^2}{\nu}+\frac{(3)^2}{\nu}\right]$$

- MLE will be *v* that maximizes this expression.

Simple MLE Example

- We observe that a coin falls on heads one time in ten trials. What is our estimate of the probability, *p*, of a coin falling on heads?
- The likelihood of the outcome is:

$$L\left(p\right) = 10\,p(1-p)^9$$

- Let's look at the first order condition to find the maximum.

$$\frac{\partial L}{\partial p} \propto \left(1-p\right)^9 - 9p\left(1-p\right)^8 = 0$$

$$1 - p = 9p$$

$$p = \tfrac{1}{10}$$

Simple MLE Example (cont.)

- Suppose we observed two heads in 10 tosses, what is the MLE?
- The likelihood of the outcome is:

$$L(p) = \binom{10}{2} p^2 (1-p)^8$$

- Let's look at the first order condition to find the maximum.

$$\frac{\partial L}{\partial p} \propto 2p(1-p)^8 - 8p^2(1-p)^7 = 0$$

$$(1-p) - 4p = 0$$

$$p = 1/5$$

MLE for N(0, ν)

Estimate the variance, *v*, from *n* observations, $u_1 \ldots u_n$, drawn from a normal distribution with mean zero:

Likelihood: $$\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi v}}\exp\left(\frac{-u_i^2}{2v}\right)\right]=\left[\frac{1}{2\pi v}\right]^{\frac{n}{2}}\cdot\prod_{i=1}^{n}\left[\exp\left(\frac{-u_i^2}{2v}\right)\right]$$

Log Likelihood: $$\frac{n}{2}\ln\left(\frac{1}{2\pi}\right)-\frac{n}{2}\ln(v)-\sum_{i=1}^{n}\left[\frac{u_i^2}{2v}\right]$$

FOC: $$-\frac{n}{2v}+\frac{1}{2v^2}\sum_{i=1}^{n}u_i^2=0$$

MLE: $$v=\frac{1}{n}\sum_{i=1}^{n}u_i^2$$

MLE and Time Varying Volatility

- Models like GARCH(1,1) and EWMA assume that daily returns are Normal with mean zero and volatility, v_i.
  - Note, this is the distribution conditional on all previous daily returns
  - Later, we will relax the Normal assumption with wider tail distributions
- The likelihood is now:

$$\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi v_i}}\exp\left(\frac{-u_i^2}{2v_i}\right)\right]$$

MLE for N(0, ν_i)

- We choose parameters that maximize, where now volatility is changing each day:

$$\sum_{i=1}^{n}\left[-\ln(v_i)-\frac{u_i^2}{v_i}\right]$$

- For the models we previously discussed:

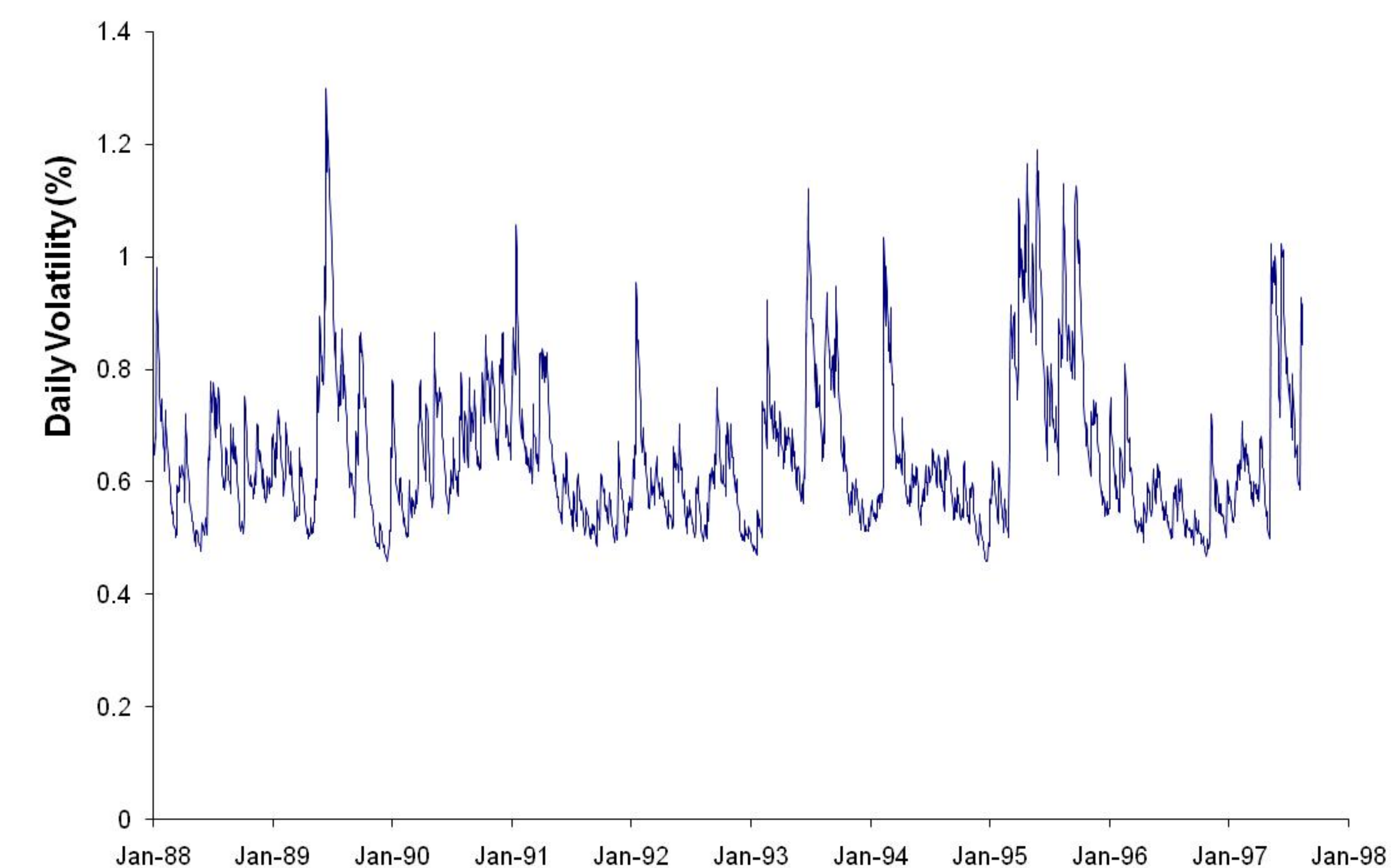
$$EWMA \qquad\qquad : \nu_i = \lambda\nu_{i-1} + \left(1-\lambda\right)u_{i-1}^2$$

$$GARCH\left(1,1\right) : \nu_i = \gamma V_L + \alpha u_{i-1}^2 + \beta\nu_{i-1}$$

MLE for N(0, ν_i) – cont.

- For EWMA – We estimate λ.
- For GARCH(1,1) - We can:
  - Estimate three parameters (α, β, $\omega=\gamma V_L$)
  - Or, assume the the long-run average volatility (V_L) equals to the sample variance and estimate only two parameters using MLE. This is called: Variance Targeting

Daily Volatility of Yen: 1988-1997



Estimating GARCH(1,1)

Day	S_i	u_i	$v_i = \sigma_i^2$	$-\ln v_i - u_i^2/v_i$
1	0.007728			
2	0.007779	0.006599		
3	0.007746	-0.004242	0.00004355	9.6283
4	0.007816	0.009037	0.00004198	8.1329
5	0.007837	0.002687	0.00004455	9.8568
….				
2423	0.008495	0.000144	0.00008417	9.3824
				22,063.5833

1. Assume estimates for alpha, beta and omega.
2. Compute $\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$
3. Compute likelihood and sum across all days.
4. Let optimizer find the parameters to maximize likelihood

Programing MLE Estimator

- Write likelihood function:
  - Takes parameters as inputs
  - Computes the likelihood given the data
- Call a minimization/maximization algorithm such as *optim* in R or *fminunc* in Matlab

$$\boxed{l = \frac{n}{2}\ln\left(\frac{1}{2\pi}\right) + \frac{1}{2}\sum_{i=1}^{n}\left[-\ln\left(v_i\right) - \frac{u_i^2}{v_i}\right]}$$

Asymptotic Properties of MLE Estimator

- It can be shown that: $\hat{\theta} \xrightarrow{d} N\left(\theta, \frac{1}{n} I(\theta)^{-1}\right)$
- I(θ) is called the Fisher Information:

$$I(\theta) = E\left[\left(\frac{\partial}{\partial \theta} \ln L(\theta; X)\right)^2\right] = -E\left(\frac{\partial^2}{\partial \theta^2} \ln L(\theta; X)\right)$$

- We use a sample estimate of the Fisher Information:

$$\overline{I}\left(\hat{\theta}\right) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} \ln L\left(\hat{\theta}; x_i\right)$$

Confidence Interval for MLE Estimator

- This allows us to test hypotheses and build confidence interval around our estimate, if *n* is large enough.
- The standard error of the estimator is:

$$se\left(\hat{\theta}\right) = \sqrt{\frac{1}{n}\bar{I}\left(\hat{\theta}\right)^{-1}}$$

- The *(1-α)* confidence interval for θ is:

$$\hat{\theta} \pm se\left(\hat{\theta}\right)\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$$

Hypothesis Testing using MLE

- We can test the hypothesis at level α:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

- By forming a standard normal: $Z = \dfrac{\hat{\theta} - \theta_0}{se\left(\hat{\theta}\right)} \sim N(0,1)$

- Reject *H_0* if: $|Z| \geq \Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right)$

Normal Distribution Example

Suppose that $x_1,\ldots,x_n$ are $i.i.d.\ N\left(\mu,\sigma^2\right)$ with σ^2 known.

The log likelihood: $$\ln L\left(\mu;X\right) = -\frac{n}{2}\left[\ln\left(2\pi\right)+\ln\left(\sigma^2\right)\right]-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

FOC: $$\frac{\partial \ln L\left(\mu;X\right)}{\partial\mu} = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}x_i - n\mu\right) = 0$$

MLE Estimator: $$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{X}$$

Fisher Information: $$\bar{I}\left(\hat{\theta}\right) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln L\left(\theta;x_i\right)$$ With: $$\frac{\partial^2 \ln L\left(\mu;x_i\right)}{\partial\mu^2} = -\frac{1}{\sigma^2}$$

$$\Rightarrow \bar{I}\left(\hat{\mu}\right) = -\frac{1}{n}\cdot n\cdot\frac{-1}{\sigma^2} = \frac{1}{\sigma^2}$$

Standard Error: $$se_{\hat{\mu}} = \sqrt{\frac{1}{n}\bar{I}\left(\hat{\mu}\right)^{-1}} = \frac{\sigma}{\sqrt{n}}$$

Normal Distribution – Cont.

$95\% C.I.$ for the mean is :

$$\bar{x} \pm \frac{\sigma}{\sqrt{n}} \Phi^{-1}(0.975) = \bar{x} \pm \frac{\sigma}{\sqrt{n}} 1.96$$

A 2-sided test for the mean different from zero :

$$Z = \frac{\bar{x}}{\sigma / \sqrt{n}}$$

Reject if: $|Z| \geq 1.96$

Homework :

Suppose that $X_1, \ldots, X_n$ are $i.i.d.$ $N\left(0, S = \sigma^2\right)$.

Estimate $\hat{S}$. *What is* $se_{\hat{S}}$?

Likelihood Ratio Test

- Suppose we want to test the hypothesis that some constraints on the parameters hold:
  - For example, μ = 0.
- We can compute the Maximum Likelihood given the constraints, and compare it to the Maximum Likelihood without constraints.
  - What gain in Maximum Likelihood are we giving up by the constraints?
- The following asymptotically holds:
  - *m* is the number of effective constraints.

$$2\left[\max\left\{\ln L(\theta)\right\} - \max\left\{\ln L(\theta_c)\right\}\right] \sim \chi_m^2$$

Normal Distribution – L.R. Test

$$T\text{est for the mean different from zero, when } \sigma \text{ is known :}$$

Unconstrained: $$\max\left\{\ln L(\theta)\right\} = -\frac{n}{2}\left[\ln(2\pi) + \ln\left(\sigma^2\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$

Constrained: $$\max\left\{\ln L(\theta_c)\right\} = -\frac{n}{2}\left[\ln(2\pi) + \ln\left(\sigma^2\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - 0\right)^2$$

$$\Downarrow$$

$$LR = 2\left[\max\left\{\ln L(\theta)\right\} - \max\left\{\ln L(\theta_c)\right\}\right] = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}\left(x_i\right)^2 - \sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right] =$$

$$= \frac{1}{\sigma^2}\left[2\overline{x}\sum_{i=1}^{n}x_i - n\overline{x}^2\right] = \frac{n\overline{x}^2}{\sigma^2} \sim \chi_1^2$$

- Reject at 5% if LR>3.841
- Other rejection boundaries: 6.635 at 1%, 2.706 at 10%

Homework

- Daily EUR data from 2/5/14 – 2/5/15
- Estimate EWMA parameter λ using MLE
- Test whether λ is different from 0.96 using Likelihood Ratio Test.

Thanks