QF 604
ECONOMETRICS OF FINANCIAL MARKETS

LECTURE 7

Professor Lim Kian Guan

SMU Master of Science in
Quantitative Finance 2023-2024 T2

LECTURE OUTLINE

01 Maximum Likelihood Estimation

- Consider a Poisson (discrete) distribution with probability mass function:

$$f(x_t;\lambda) = \frac{e^{-\lambda}\lambda^{x_t}}{x_t!}$$

where (1) $x_t \geq 0$ is the number of independent events occuring each time period *t*.

(2) λ: the average number of events each time period. $\lambda = E(x_t) = Var(x_t)$.

- Suppose over *T* periods, observed x_t's are $\mathbf{x} = \{x_1, x_2, \ldots, x_T\}$. The likelihood function of the (observed) sample (or joint probability mass function) is

$$f(\mathbf{x};\lambda) = \prod_{t=1}^{T} f(x_t;\lambda) = \frac{e^{-T\lambda}\,\lambda^{\sum_{t=1}^{T} x_t}}{\prod_{t=1}^{T} x_t!}$$

01 Maximum Likelihood Estimation

- The log likelihood function of the sample is:

$$\ln L(\lambda) = \ln \prod_{t=1}^{T} f(x_t;\lambda) = -T\lambda + \left(\sum_{t=1}^{T} x_t\right)\ln\lambda - \sum_{t=1}^{T}\ln x_t!$$

- First order condition $\frac{d\ln L}{d\lambda}: -T + \frac{1}{\lambda}\left(\sum_{t=1}^{T} x_t\right) = 0$ (likelihood equation)

$$\hat{\lambda}_{ML} = \frac{\sum_{t=1}^{T} x_t}{T}$$

- Second order condition $\frac{d^2\ln L}{d\lambda^2} = -\frac{\sum_{t=1}^{T} x_t}{\lambda^2} < 0$ (at least one of the x_t's > 0)
- Another example: MLE of sampling from a normal distribution of $x_t \sim N(\mu,\sigma^2)$

$$\ln L(\mu,\sigma^2) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\sum_{t=1}^{T}\left[\frac{(x_t-\mu)^2}{\sigma^2}\right]$$

$$\hat{\mu}_{ML} = \frac{\sum_{t=1}^{T} x_t}{T} \equiv \bar{x} \quad \text{and} \quad \hat{\sigma}^2_{ML} = \frac{\sum_{t=1}^{T}(x_t-\bar{x})^2}{T}$$

- ML estimators are consistent, but they need not be unbiased in small sample.

01 Maximum Likelihood Estimation

- Estimate a and b in $Y_t = a + bX_t + e_t$ where $e_t \sim N(0, \sigma_e^2)$.
- Consider the MLE of sampling from a normal distribution of $e_t = Y_t - a - bX_t \sim N(0, \sigma_e^2)$.

$$\ln L(a, b, \sigma_e^2) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma_e^2 - \frac{1}{2}\sum_{t=1}^{T}\left[\frac{(y_t - a - bx_t)^2}{\sigma_e^2}\right]$$

- Then calculate $\frac{d\ln L}{da} = 0$, $\frac{d\ln L}{db} = 0$, and $\frac{d\ln L}{d\sigma_e^2} = 0$ respectively
- $\hat{b}_{ML} = \frac{\sum x_t y_t - T\bar{x}\bar{y}}{\sum x_t^2 - T\bar{x}^2}$, $\hat{a}_{ML} = \bar{y} - \hat{b}_{ML}\bar{x}$, and $\hat{\sigma}_{e,ML}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{a}_{ML} - \hat{b}_{ML}x_t)^2$
- Under normality, the ML estimators are also the OLS estimators for a and b.

  $\hat{\sigma}_{e,ML}^2$ is consistent but biased in small sample.

01 Properties of MLE

- Suppose RV Z_t is i.i.d., and takes sample values $\{z_1, z_2, \ldots, z_T\}$.
  Let the probability density function of RV Z_t in general be $f(z_t; \Lambda)$ with parameters $\Lambda_{n\times 1}$. Function $f(\cdot)$ is assumed to be continuous and smooth. $f(Z_t)$ are also i.i.d. Then, for any t

$$\int_{-\infty}^{\infty} f(z_t;\Lambda)dz_t = 1 \qquad \int_{-\infty}^{\infty} \frac{\partial f(z_t;\Lambda)}{\partial \Lambda} dz_t = 0 \qquad \int_{-\infty}^{\infty} \frac{\partial^2 f(z_t;\Lambda)}{\partial \Lambda \partial \Lambda^T} dz_t = 0$$

- $\frac{\partial \ln f(z_t;\Lambda)}{\partial \Lambda} f = \frac{\partial f(z_t;\Lambda)}{\partial \Lambda}$
- $\int_{-\infty}^{\infty} \frac{\partial f(z_t;\Lambda)}{\partial \Lambda} dz_t = 0 \Leftrightarrow \int_{-\infty}^{\infty} \frac{\partial \ln f(z_t;\Lambda)}{\partial \Lambda} f\, dz_t = 0 \Leftrightarrow \mathbf{E}\left[\frac{\boldsymbol{\partial} \boldsymbol{ln} \boldsymbol{f}(\boldsymbol{z_t};\boldsymbol{\Lambda})}{\boldsymbol{\partial\Lambda}}\right] = \mathbf{0}$
- $\frac{\partial^2 \ln f(z_t;\Lambda)}{\partial \Lambda \partial \Lambda^T} = \frac{\partial}{\partial \Lambda}\left(\frac{1}{f}\frac{\partial f(z_t;\Lambda)}{\partial \Lambda^T}\right) = \frac{1}{f}\frac{\partial^2 f}{\partial \Lambda \partial \Lambda^T} - \frac{1}{f^2}\frac{\partial f}{\partial \Lambda}\frac{\partial f}{\partial \Lambda^T} = \frac{1}{f}\frac{\partial^2 f}{\partial \Lambda \partial \Lambda^T} - \frac{\partial \ln f}{\partial \Lambda}\frac{\partial \ln f}{\partial \Lambda^T}$
- $\mathbf{E}\left[\frac{\boldsymbol{\partial}^2 \boldsymbol{ln} \boldsymbol{f}(\boldsymbol{z_t};\boldsymbol{\Lambda})}{\boldsymbol{\partial\Lambda\partial\Lambda^T}}\right] = -\mathbf{E}\left[\frac{\boldsymbol{\partial} \boldsymbol{ln} \boldsymbol{f}}{\boldsymbol{\partial\Lambda}}\frac{\boldsymbol{\partial} \boldsymbol{ln} \boldsymbol{f}}{\boldsymbol{\partial\Lambda^T}}\right]$
- $cov\left(\frac{\partial \ln f}{\partial \Lambda}\right) = I_1(\Lambda) = \mathrm{E}\left[\frac{\partial \ln f}{\partial \Lambda}\frac{\partial \ln f}{\partial \Lambda^T}\right]$

01 Maximum Likelihood Estimation

- The joint pdf of $\{z_1, z_2, \dots, z_T\}$ (or Z) is likelihood function

$$L(Z;\Lambda) = \prod_{t=1}^{T} f(z_t;\Lambda)$$

Log likelihood function $\ln L(Z;\Lambda) = \sum_{t=1}^{T} \ln f(z_t;\Lambda)$

- $\mathrm{E}\left[\frac{\partial \ln L(z_t;\Lambda)}{\partial \Lambda}\right] = \mathrm{E}\left[\sum_{t=1}^{T} \frac{\partial \ln f(z_t;\Lambda)}{\partial \Lambda}\right] = 0$
- $\mathrm{E}\left[\frac{\partial^2 \ln L(z_t;\Lambda)}{\partial \Lambda \partial \Lambda^T}\right] = \mathrm{E}\left[\sum_{t=1}^{T} \frac{\partial^2 \ln f(z_t;\Lambda)}{\partial \Lambda \partial \Lambda^T}\right] = -E\left[\sum_{t=1}^{T} \frac{\partial \ln f}{\partial \Lambda} \frac{\partial \ln f}{\partial \Lambda^T}\right] = -E\left[\frac{\partial \ln L}{\partial \Lambda} \frac{\partial \ln L}{\partial \Lambda^T}\right]$
- $cov\left(\frac{\partial \ln L(z_t;\Lambda)}{\partial \Lambda}\right) = E\left[\frac{\partial \ln L}{\partial \Lambda} \frac{\partial \ln L}{\partial \Lambda^T}\right] = E\left[\sum_{t=1}^{T} \frac{\partial \ln f}{\partial \Lambda} \frac{\partial \ln f}{\partial \Lambda^T}\right] = T \times I_1(\Lambda) = I_T(\Lambda)$
- $E\left[\frac{\partial \ln L}{\partial \Lambda} \frac{\partial \ln L}{\partial \Lambda^T}\right]$ is called **Fisher's information matrix** of order (sample size) T.

01 Maximum Likelihood Estimation

- Recall MLE estimator $\widehat{\Lambda}_{ML}$ maximizes $\ln L(Z;\Lambda) \Rightarrow$ FOC

  Sets $\frac{\partial \ln L(Z;\Lambda)}{\partial\Lambda} = 0 \;\Rightarrow\; \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \ln f(z_t;\Lambda)}{\partial\Lambda}|_{\Lambda=\widehat{\Lambda}_{ML}} = 0$

- Recall the GMM. By the LLN, the sample moment converges to the population moment:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \ln f(z_i;\Lambda)}{\partial\Lambda}|_{\Lambda=\widehat{\Lambda}_{ML}} \to E\left[\frac{\partial \ln f\left(Z_t;\widehat{\Lambda}_{ML}\right)}{\partial\widehat{\Lambda}_{ML}}\right]$$

- But $E\left[\frac{\partial \ln f(Z_t;\Lambda)}{\partial\Lambda}\right] = 0$, then $\widehat{\Lambda}_{ML} \to \Lambda$, i.e. ML estimator is consistent.

- Therefore, $\sqrt{T}(\widehat{\Lambda}_{ML} - \Lambda) \to N(0, I_1(\Lambda)^{-1})$.

  See **Proof** in pages 255 – 256 of book.

  The maximum likelihood estimator $\widehat{\Lambda}_{ML}$ is asymptotically normal.

01 Cramer-Rao Lower Bound

The Cramer-Rao lower bound or inequality is an important result in maximum likelihood theory. Suppose **h(Z)** is any unbiased n × 1 vector estimate of Λ, then the variances of **h(Z)** are bounded from below by the diagonal elements of n × n matrix $I_T^{-1}(\Lambda)$.

$$I_T^{-1} = \begin{bmatrix} r_{11} & \cdots & & \\ \vdots & r_{22} & & \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix}$$

An estimator that attains the lower bound is said to be a minimum variance unbiased or efficient estimator.

ML estimator is consistent and asymptotically normal ~ $N(0, I_1(\Lambda)^{-1})$. It is also asymptotically unbiased. Thus, ML estimator variances are bounded below by I_T^{-1}.

Therefore, ML estimators are asymptotically efficient, attaining the Cramer-Rao lower bound.

See **Proof** in pages 257 – 258 of book.

01 Likelihood Ratio Test

- Expand using Taylor series:

$$\ln L(z;\Lambda) = \ln L(z;\widehat{\Lambda}_{ML}) + (\Lambda - \widehat{\Lambda}_{ML})\frac{\partial \ln L(z;\widehat{\Lambda}_{ML})}{\partial \Lambda} + \frac{1}{2}(\Lambda - \widehat{\Lambda}_{ML})^T \frac{\partial^2 \ln L(z_t;\widehat{\Lambda}_{ML})}{\partial \Lambda \partial \Lambda^T}(\Lambda - \widehat{\Lambda}_{ML}) + \mathrm{o}(T)$$

$$\to \ln L(z;\widehat{\Lambda}_{ML}) + 0 + \frac{1}{2}\left[\sqrt{T}(\Lambda - \widehat{\Lambda}_{ML})^T \left(\frac{1}{T}\frac{\partial^2 \ln L(z_t;\widehat{\Lambda}_{ML})}{\partial \Lambda \partial \Lambda^T}\right)\sqrt{T}(\Lambda - \widehat{\Lambda}_{ML})\right] \text{ as } T\uparrow\infty$$

Or, $2\,[\ln L(z;\Lambda) - \ln L(z;\widehat{\Lambda}_{ML})] \to N(0, I_1(\Lambda)^{-1}) \times [-I_1(\Lambda)] \times N(0, I_1(\Lambda)^{-1})$

Or, $-2\ln\left[\frac{L(z;\Lambda)}{L(z;\widehat{\Lambda}_{ML})}\right] \to N(0, I_1(\Lambda)^{-1}) \times [I_1(\Lambda)] \times N(0, I_1(\Lambda)^{-1}) \sim \chi^2_n$

i.e. converges to a chi-square RV with *n* degrees of freedom, *n* being the dimension of Λ.

- Suppose we want to test $H_0\!: \Lambda = \Theta$, i.e., testing if true Λ is a set of values given by Θ , then if the null hypothesis is true, asymptotically

$$-2\ln\left[\frac{L(z;\Theta)}{L(z;\widehat{\Lambda}_{ML})}\right] = T(\widehat{\Lambda}_{ML} - \Theta)^T I_1(\Theta)(\widehat{\Lambda}_{ML} - \Theta) \to \chi^2_n$$

where $I_1(\Theta)$ is estimated by $-\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \ln f(\Theta)}{\partial \Lambda \partial \Lambda^T}$. Reject null hypothesis if the test statistic occurs in the right tail of the χ^2_n distribution at a p-value smaller than the significance level of the test.

02 Conditional Heteroskedasticity

- For ARMA(1,1) which includes AR(1) and MA(1) processes,

$$y_t = \theta + \lambda y_{t-1} + u_t + \alpha u_{t-1},$$

  - The conditional variance $Var(y_t|y_{t-1}) = (1 + a^2)\sigma_u^2$ is constant.
  - The conditional mean $\mathrm{E}(y_t|y_{t-1}) = \theta + \lambda y_{t-1}$ changes.

- Similarly, we can show that for: $y_t = \theta + \lambda x_t + u_t$
  - $E(u_t) = 0, Cov(u_t, u_{t-k}) = 0, k \neq 0$, and x_t, u_t are stochastically stationary and independent
  - $Var(y_t|x_t) = Var(u_t) = \sigma_u^2$ is constant.

- Suppose we model a process on the variance of u_t (not u_t itself) such that:

$$Var(u_t) = \alpha_0 + \alpha_1 u_{t-1}^2$$

  - This is an autoregressive conditional heteroskedasticity or ARCH(1) model in disturbance u_t.

02 Conditional Heteroskedasticity

- We can also write the equation in terms of a u_t process as follows:

$$u_t = e_t\sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}$$

where $e_t \sim N(0,1)$.

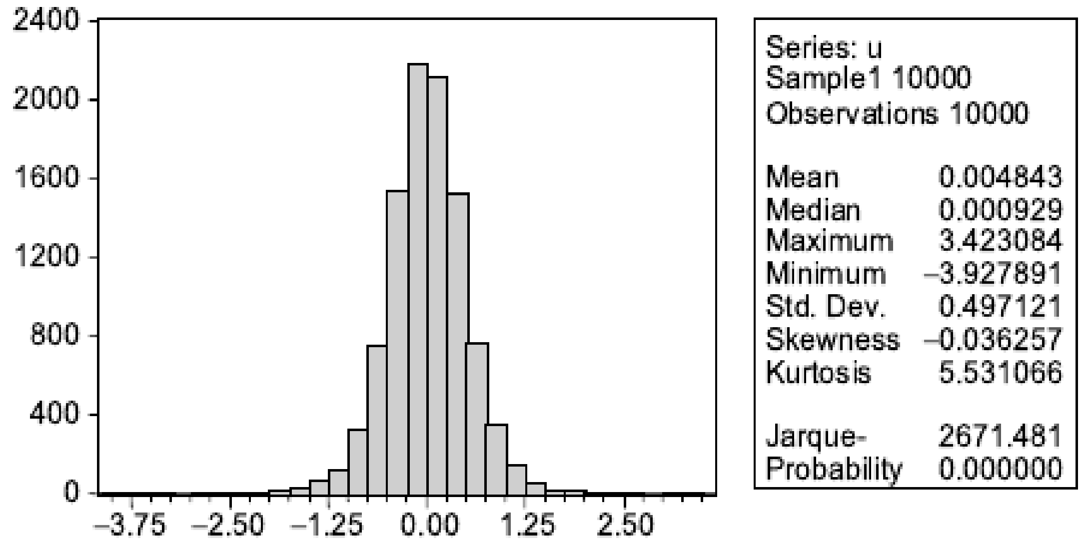


Figure 9.1: Monte Carlo Simulation of Errors $u_t = e_t\sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}$

- Conditional on u_{t-1}, u_t is normally distributed. Using Monte Carlo simulation with a sample size of 10,000, a histogram of the distribution of u_t is produced as shown in Figure 9.1.
- Clearly, unlike unit normal e_t, unconditional u_t's empirical distribution as shown in Fig. 9.1 has a larger kurtosis (>3) than a normal random variable. The Jarque-Bera test statistic rejects the null of normal distribution for u_t.
- When $\text{Var}(u_t) = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_{q-1} u_{t-q-1}^2 + \alpha_q u_{t-q}^2$, we call the conditional variance of u_t above an ARCH(q) process.

02 Conditional Heteroskedasticity

- Another model of changing conditional variance is,

$$Var(u_t) = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 Var(u_{t-1})$$

  This is generalized autoregressive conditional heteroskedasticity or the GARCH(1,1) model in u_t.

- Suppose a y_t process contains a disturbance random error that behaves according to GARCH in the above equation. Then $Var(y_t|\Phi_t) = Var(u_t|x_t, u_{t-1}, \sigma_{t-1}^2) = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 \sigma_{t-1}^2$ is no longer constant, but instead, changes with t or more precisely, the information available at t, Φ_t, even as u_{t-1} and also $Var(u_{t-1})$ change over time.

- Process such as y_t or u_t itself is said to exhibit conditional heteroskedasticity or dynamic volatility.

02 Conditional Heteroskedasticity

- Simulate sample paths of a process $\{y_t\}_{t=1,2,\ldots,200}$, weakly stationary without conditional heteroskedasticity:

$$y_t = \theta + \lambda x_t + u_t,$$

$$x_t \sim N(1, 0.4), u_t \sim N(0,2), \theta = \frac{1}{2}, \lambda = \frac{1}{2}$$

- Let's chart the path of y_t process.
- The plot of the time-path of y_t is shown in Fig. 9.2. It is seen that y_t behaves like a random series with a mean at 1 and the two dotted lines are the two standard deviations away from the mean. In this case, they are $1 \pm 2\sqrt{2.1}$ (about 4 and -2, respectively), with 2.275% probability of exceeding each way the region between the two dotted lines.

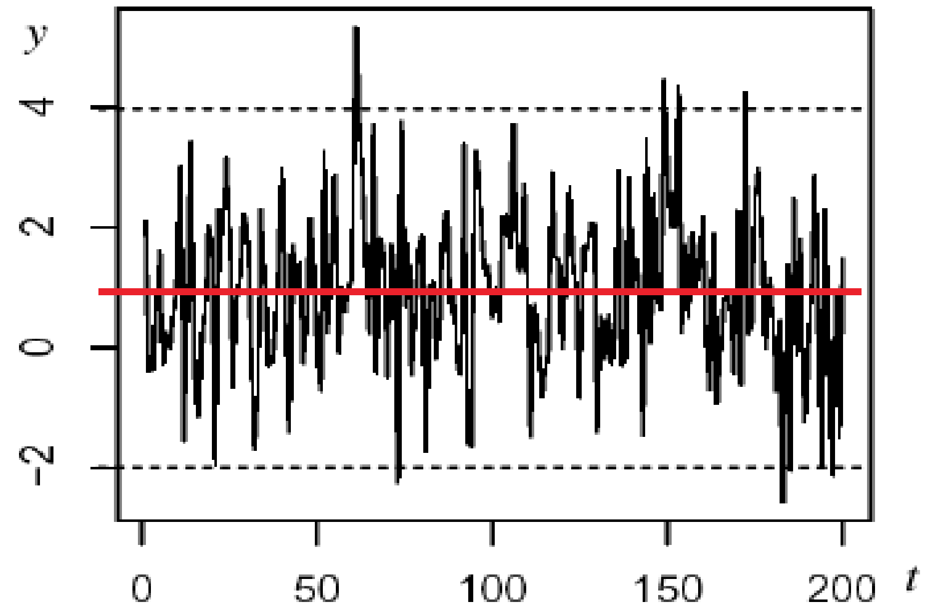


Figure 9.2: Stationary Process $y_t \sim N(1, 2.1)$

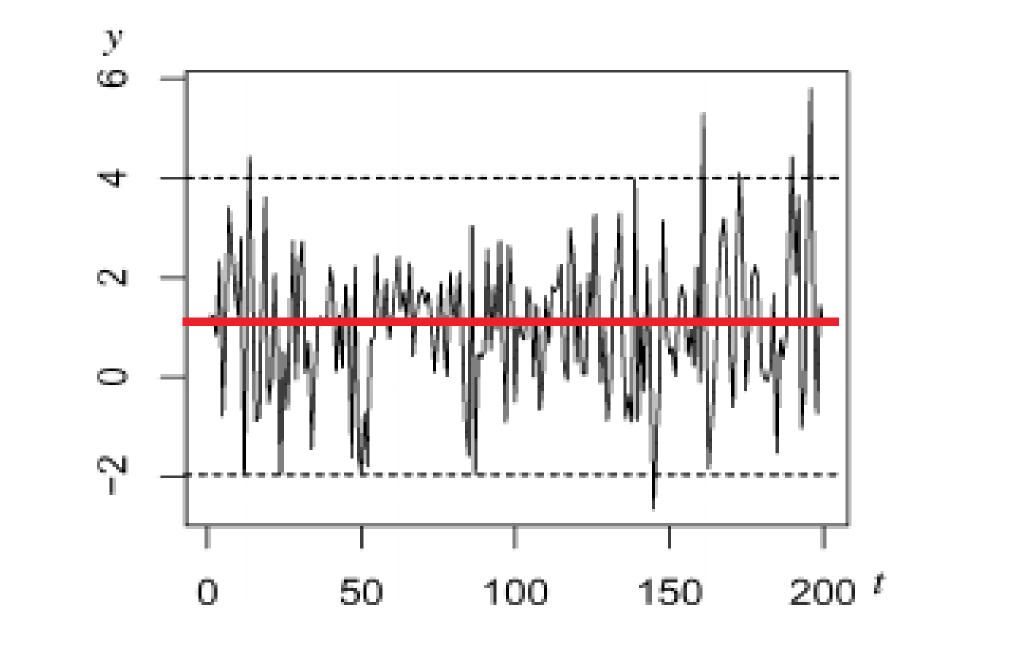
02 Conditional Heteroskedasticity

- Simulate sample paths of another process $\{y_t\}_{t=1,2,\ldots,200}$:

$$y_t = \theta + \lambda x_t + u_t, \qquad x_t \sim N(1, 0.4),$$

$$Var(u_t) = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 Var(u_{t-1})$$

- We set initial $Var(u_0) = \sigma_0^2 = 2$.
- $\theta = \frac{1}{2}, \lambda = \frac{1}{2}, \alpha_0 = \frac{1}{2}, \alpha_1 = \frac{1}{4}$ and $\gamma_1 = \frac{1}{2}$. $y_0 = \frac{1}{2} + \frac{1}{2}x_0 + u_0$, $u_0 \sim N(0,2)$
- Next simulate $u_1 = e_1\sqrt{\alpha_0 + \alpha_1 u_0^2 + \gamma_1 Var(u_0)}$ for $e_1 \sim N(0,1)$.
- Then simulate $u_2 = e_2\sqrt{\alpha_0 + \alpha_1 u_1^2 + \gamma_1 Var(u_1)}$ for $e_2 \sim N(0,1)$.
- The figure shows a 'similar' y_t process as in Fig.9.2, with $y_t = \frac{1}{2} + \frac{1}{2}x_t + u_t$. Its variance follows the GARCH error process: $Var(u_t) = 0.5 + 0.25u_{t-1}^2 + 0.5Var(u_{t-1})$.



Unconditional variance of u_t (via law of iterated expectations) is :

$E(var(u_t|\Phi_{t-1})) = \sigma_u^2$

Now,

$E(var(u_t|\Phi_{t-1})) = \alpha_0 + \alpha_1 E(u_{t-1}^2) + \gamma_1 E(Var(u_{t-1}|\Phi_{t-2}))$

Or $\sigma_u^2 = \alpha_0 + \alpha_1\sigma_u^2 + \gamma_1\sigma_u^2$

Or $\sigma_u^2 = \alpha_0/(1 - \alpha_1 - \gamma_1) = 2$.

02 Conditional Heteroskedasticity

- Another simulation using the same $y_t = \frac{1}{2} + \frac{1}{2}x_t + u_t$ with unconditional mean and variance of y_t at the same 1 and 2.1 values, respectively.
- Now its variance follows the GARCH error process:

$$Var(u_t) = 0.1 + 0.25u_{t-1}^2 + 0.7Var(u_{t-1})$$

- Fig.9.4 shows the persistent and much higher volatility with y_t's exceeding +15 and falling below -15 in the observations from 100 to 150.
- Suppose $y_t = \theta + \lambda x_t + \gamma\sigma_t^2 + u_t$, where $\sigma_t^2 = Var(y_t|x_t, u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2$
- Then the y_t process is an ARCH-in-mean or ARCH-M model and can produce similar high deviations of y_t seen in Fig. 9.4.

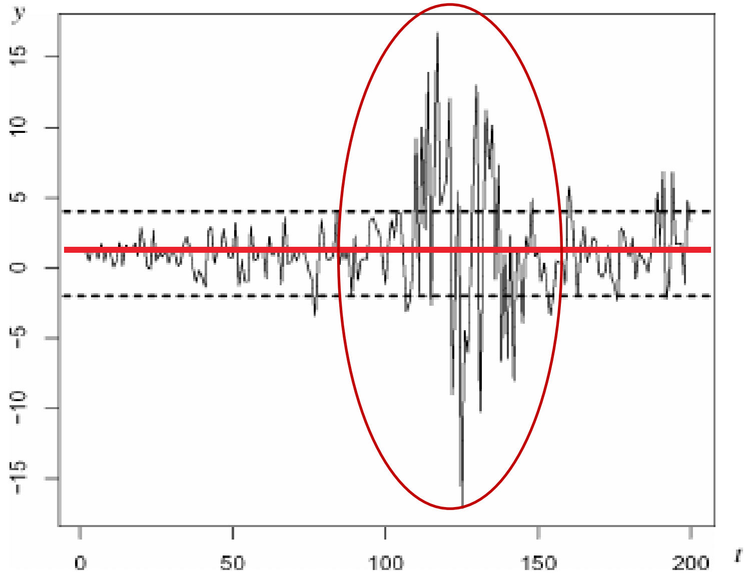


Figure 9.4: GARCH Error Process: Unconditional y_t has mean, variance 1, 2.1. $\mathrm{var}(u_t) = 0.1 + 0.25u_{t-1}^2 + 0.7\,\mathrm{var}(u_{t-1})$

02 Stationarity Conditions

- While GARCH processes are conditionally non-stationary with changing variances, they are still unconditionally stationary processes.
- Expand $Var(u_t) = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 Var(u_{t-1})$:

$$Var(u_t) = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 Var(u_{t-1})$$
$$= \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1[\alpha_0 + \alpha_1 u_{t-2}^2 + \gamma_1 Var(u_{t-2})\,]$$
$$= \alpha_0(1 + \gamma_1 + \gamma_1^2 + \cdots) + \alpha_1(u_{t-1}^2 + \gamma_1 u_{t-2}^2 + \gamma_1^2 u_{t-3}^2 + \cdots)$$

- Taking unconditional expectation on both sides, and assuming there exists stationarity so that $\sigma_u^2 = E(u_{t-1}^2) = E(u_{t-2}^2) = E(u_{t-3}^2) = \cdots$, then $\sigma_u^2 = \frac{\alpha_0}{1-\gamma_1} + \alpha_1(\sigma_u^2 + \gamma_1\sigma_u^2 + \gamma_1^2\sigma_u^2 + \cdots) = \frac{\alpha_0}{1-\gamma_1} + \frac{\alpha_1\sigma_u^2}{1-\gamma_1} = \frac{\alpha_0+\alpha_1\sigma_u^2}{1-\gamma_1}$, supposing $|\gamma_1| < 1$.
- Then, $\sigma_u^2 = \alpha_0/(1 - \gamma_1 - \alpha_1)$, supposing $\alpha_0 > 0$ and $|\gamma_1 + \alpha_1| < 1$.

03 Estimating GARCH using MLE

Suppose we have process:

$$Y_t = \theta + \lambda X_t + u_t$$

where u_t follows GARCH(1,1) process,

$$u_t = e_t \sigma_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 \sigma_{t-1}^2$$

e_t is i.i.d. $N(0,1)$ and $\sigma_t^2 = Var(u_t)$.

We first consider estimating using ML the GARCH (1,1) process u in the above equations. The procedure can be similarly applied to GARCH (q, p).

From the Equation of u_t, conditional RV $u_{t+1}|\sigma_{t+1} \sim \mathrm{N}(0, {\sigma_{t+1}}^2)$ is independent of RV $u_t|\sigma_t \sim \mathrm{N}(0, {\sigma_t}^2)$, i.e., $e_{t+1}$is independent of e_t

03 Estimating GARCH using MLE

- The joint density or likelihood function of sample values $\{u_1, u_2, \ldots, u_T\}$ (if they are observed or if their estimates $\hat{u}_t$ are observed), is expressed as product of conditional and marginal densities:

$$\begin{aligned} f(u_1, u_2, \ldots, u_T) &= f(u_T|u_{T-1}, u_{T-2}, \ldots, u_0)\, f(u_0, u_1, \ldots, u_{T-1}) \\ &= f(u_T|u_{T-1}, u_{T-2}, \ldots, u_0)\, f(u_{T-1}|u_{T-2}, u_{T-3}, \ldots, u_0) \\ &\quad \times f(u_0, u_1, \ldots, u_{T-2}) \\ &= f(u_T|u_{T-1}, u_{T-2}, \ldots, u_0)\, f(u_{T-1}|u_{T-2}, u_{T-3}, \ldots, u_0) \\ &\quad \times \ldots \times f(u_1|u_0) \times f(u_0)\,. \end{aligned}$$

- We shall assume the initial values u_0, σ_0 are given. These may be approximated using $u_0 = 0$ (since $E(u_0) = 0$), and $\sigma_0 = \sqrt{\frac{1}{T}\sum_{t=1}^{T} u_t^2}$.
- The log likelihood function of the sample data is then

$$\ln f(u_1, u_2, \ldots, u_T) = \left[\sum_{t=1}^{T} \ln f(u_t|u_{t-1}, \ldots, u_0)\right] + \ln f(u_0)\,.$$

- Since each $u_t \sim N(0, \sigma_t^2)$,

$$\ln f(u_t|u_{t-1}, \ldots, u_0) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma_t^2 - \frac{1}{2}\frac{u_t^2}{\sigma_t^2}\,, \qquad (9.17)$$

and each σ_t^2 at t is a pre-determined linear function of $u_0^2, u_1^2, \ldots, u_{t-1}^2$ and σ_0.

03 Estimating GARCH using MLE

- If we let parameter vector $\omega = (\alpha_0, \alpha_1, \gamma_1)$, the maximum likelihood first order conditions are to set

$$\sum_{t=1}^{T} \frac{\partial \ln f}{\partial \omega} = \sum_{t=1}^{T} \frac{1}{2} \sigma_t^{-2} \frac{\partial \sigma_t^2}{\partial \omega} \left(\frac{u_t^2}{\sigma_t^2} - 1 \right) = 0$$

- However, the solution for ω_{ML} in the equation above is not straightforward since the equation is not analytical in ω as there is an iteration on $\frac{\partial \sigma_t^2}{\partial \omega}$ and σ_t^2 $(\forall_t)$ in the above equation, is also a linear function of lagged u_{t-i}^2's with coefficients in ω. Therefore, an iterative numerical technique is called for.
- The maximum likelihood estimator of $\beta = (\theta, \lambda)^T$ can also be found.

See **more details** of above in pages 268 – 270 of book.

03 Quasi MLE

- Suppose the assumption of normality in e_t is incorrectly specified. The true i.i.d. e_t distribution is non-normal.
- Using asymptotic theory, it can be shown that the ML estimators $\hat{\omega}_{ML}$ and $\hat{\beta}_{ML}$ derived under the assumption of conditional normality are generally still consistent and follow asymptotic normal distributions. These estimators are called quasi maximum likelihood:

$$\sqrt{T}\left(\hat{\Lambda}_{QML} - \Lambda\right) \xrightarrow{d} N(0, H^{-1}SH^{-1})$$

- $H = -\frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial^2 \ln f}{\partial\Lambda\partial\Lambda^T}\right]$
- $S = \frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial \ln f}{\partial\Lambda}\frac{\partial \ln f}{\partial\Lambda^T}\right]$

- The population moments are replaced by the sample moments:

$$\hat{H} = -\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial^2 \ln f}{\partial\Lambda\partial\Lambda^T}\right] \quad \text{and} \quad \hat{S} = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial \ln f}{\partial\Lambda}\frac{\partial \ln f}{\partial\Lambda^T}\right]$$

for independent Z_t in the argument of f.

03 Diagnostic for ARCH-GARCH

- ARCH (q) is modelled by $E(u_t^2) = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_{q-1} u_{t-q+1}^2 + \alpha_q u_{t-q}^2$, assuming $E(u_t) = 0$. Then this heuristically as a regression of u_t^2 on its lags up to lag q, adding a white noise e_t.

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_{q-1} u_{t-q+1}^2 + \alpha_q u_{t-q}^2 + e_t$$

- GARCH (p,q) for $p \neq 0$ can also heuristically, be expressed as follows for an arbitrarily large number of lags N, where we would set u_{t-N-1}^2 equal to some constant. The c_j's are constants: $u_t^2 = c_0 + c_1 u_{t-1}^2 + c_2 u_{t-2}^2 + \cdots + c_q u_{t-q}^2 + \cdots + c_N u_{t-N}^2 + e_t$.

- Suppose we estimate via OLS and then obtain the estimated residuals: $\hat{u}_t = y_t - \hat{\theta} - \hat{\lambda} x_t$.

- Then use the Ljung and Box Q-test on autocorrelations on $\hat{u}_t^2$ to test if autocorrelations are present or not, viz. H_0: $\rho(1) = \rho(2) = \rho(3) = \cdots = \rho(q) = 0$. If H_0 is rejected, then ARCH (q) or GARCH(p,q) are plausible.

04 Euler Equation Revisited under MLE

- The stochastic Euler equation for asset pricing under power utility is:

$$E_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{\gamma-1} R_{t+1}\right] = \frac{1}{\rho}$$

where ρ is the time discount factor, $1-\gamma$ is the constant relative risk aversion coefficient, C_t is per capita consumption at time t, and R_{t+1} is the return of an asset, e.g., a stock or the market portfolio, over period (t, t+1].

- Let $G_{t+1} = C_{t+1}/C_t$, $g_{t+1} = \ln G_{t+1}$, and $r_{t+1} = \ln R_{t+1}$. Then the Euler equation can be expressed as:

$$E_t\left[\exp\left((\gamma-1)g_{t+1} + r_{t+1}\right)\right] = \exp(-\ln\rho).$$

04 Euler Equation Revisited under MLE

- Assume g_{t+1} and r_{t+1} to be normally distributed conditional on information at t that are restricted to lagged values of g_t and r_t only. Then,

$$\exp\left[E_t\left((\gamma-1)g_{t+1}+r_{t+1}\right)+\tfrac{1}{2}Var_t\left((\gamma-1)g_{t+1}+r_{t+1}\right)\right]=\exp(-\ln\rho)$$

$$\text{or}\quad E_t\left((\gamma-1)g_{t+1}+r_{t+1}\right)+\tfrac{1}{2}Var_t\left((\gamma-1)g_{t+1}+r_{t+1}\right)=-\ln\rho$$

- If we further assume $Var_t\left((\gamma-1)g_{t+1}+r_{t+1}\right)=\sigma^2$, i.e., a constant $\forall t$, then

$$E_t(r_{t+1})=-(\gamma-1)E_t(g_{t+1})-\ln\rho-\frac{1}{2}\sigma^2$$

04 Euler Equation Revisited under MLE

- Use the conditional expectations to perform the MLE, i.e., let

$$y_t = (\gamma - 1)g_{t+1} + r_{t+1} + \ln\rho + \frac{1}{2}\sigma^2$$

be normally distributed with mean zero and variance σ^2, i.e. $N(0, \sigma^2)$.

- The density function of y_t is

$$f(y_t) = (2\pi)^{-1/2}(\sigma^2)^{-1/2}e^{-\left[(\gamma-1)g_{t+1}+r_{t+1}+\ln\rho+\frac{1}{2}\sigma^2\right]/2\sigma^2}$$

- The likelihood function of sample $\{y_1, y_2, \ldots, y_T\}$ is

$$L(\gamma, \rho, \sigma^2) = \prod_{t=1}^{T} f(y_t; \gamma, \rho, \sigma^2).$$

- The log likelihood function:

$$\ln L(\gamma, \rho, \sigma^2) = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{\left[(\gamma-1)g_{t+1}+r_{t+1}+\ln\rho+\frac{1}{2}\sigma^2\right]^2}{\sigma^2}$$

04 Euler Equation Revisited under MLE

- The log likelihood function:

$$\ln L(\gamma,\rho,\sigma^2) = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{\left[(\gamma-1)g_{t+1}+r_{t+1}+\ln\rho+\frac{1}{2}\sigma^2\right]^2}{\sigma^2}$$

- Solve for the maximum or first order conditions in the above log equation to derive the ML estimators $\hat{\Lambda}_{ML} \equiv (\hat{\gamma}_{ML},\hat{\rho}_{ML},\hat{\sigma}_{ML})^T$.
- Obtain the asymptotically efficient covariance matrix of the estimators as $\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\ln L}{\partial\Lambda}\frac{\partial\ln L}{\partial\Lambda^T}\right)^{-1} = I_1(\Theta)^{-1}$.
- Apply the asymptotic likelihood ratio test, $T\left(\hat{\Lambda}_{ML}-\Theta\right)^T I_1(\Theta)(\hat{\Lambda}_{ML}-\Theta)\sim\chi_3^2$, where the null hypothesis is: $H_0\!:\Lambda=\Theta$.

04 Time Series Model Estimation

Example: AR(1)

- AR(1): $Y_t = \lambda Y_{t-1} + u_t$ for t=0,1,2,….,T
  where $u_t \sim N(0, \sigma^2)$ and $Y_0 \sim N(0, \sigma^2/(1-\lambda^2))$
- log likelihood of $\{Y_T, Y_{T-1}, Y_{T-2},\dots,Y_1,Y_0\}$ is
  $L = \ln f(Y_T|Y_{T-1},\dots,Y_1,Y_0)\, f(Y_{T-1}|Y_{T-2},\dots,Y_1,Y_0)\dots f(Y_2|Y_1,Y_0)\, f(Y_1|Y_0)\, f(Y_0)$
- But pdf $f(Y_0) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\lambda^2)}} e^{-\frac{1}{2}Y_0^2/[\sigma^2/(1-\lambda^2)]}$
  so $\ln f(Y_0) = -\frac{1}{2}\ln[2\pi\sigma^2] + \frac{1}{2}\ln(1-\lambda^2) - \frac{1}{2\sigma^2}Y_0^2(1-\lambda^2)$
- pdf $f(Y_{T-j}|Y_{T-j-1},\dots,Y_1,Y_0)$ = pdf $f(Y_{T-j} - \lambda Y_{T-j-1})$ = pdf $f(u_{T-j}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}u_{T-j}^2/\sigma^2}$
  for j = 0,1,2,….,T-1
  so $\ln f(Y_{T-j} - \lambda Y_{T-j-1}) = -\frac{1}{2}\ln[2\pi\sigma^2] - \frac{1}{2\sigma^2}\left(Y_{T-j} - \lambda Y_{T-j-1}\right)^2$
- Thus $L = -\frac{T+1}{2}\ln[2\pi\sigma^2] + \frac{1}{2}\ln(1-\lambda^2) - \frac{1}{2\sigma^2}\left[Y_0^2(1-\lambda^2) + \sum_{j=0}^{T-1}\left(Y_{T-j} - \lambda Y_{T-j-1}\right)^2\right]$

04 Time Series Model Estimation

Example: AR(1)

- $\max\limits_{\lambda,\sigma^2} L = -\frac{T+1}{2}\ln[2\pi\sigma^2] + \frac{1}{2}\ln(1-\lambda^2) - \frac{1}{2\sigma^2}\left[Y_0^2(1-\lambda^2) + \sum_{j=0}^{T-1}\left(Y_{T-j} - \lambda Y_{T-j-1}\right)^2\right]$

- $\partial \mathrm{L}/\partial\sigma^2 = -\frac{T+1}{2\sigma^2} + \frac{1}{2\sigma^4}\left[Y_0^2(1-\lambda^2) + \sum_{j=0}^{T-1}\left(Y_{T-j} - \lambda Y_{T-j-1}\right)^2\right]$

  $\partial \mathrm{L}/\partial\lambda = -\frac{\lambda}{1-\lambda^2} + \frac{1}{\sigma^2}\left[\lambda Y_0^2 + \sum_{j=0}^{T-1}\left(Y_{T-j} - \lambda Y_{T-j-1}\right) Y_{T-j-1}\right]$

- Setting the derivatives or FOC to zeros and solving:

$$\hat{\sigma}_{ML}^2 = \frac{\left[Y_0^2(1-\lambda^2)+\sum_{j=0}^{T-1}\left(Y_{T-j}-\lambda Y_{T-j-1}\right)^2\right]}{T+1}$$

and $\left(\frac{\hat{\lambda}_{ML}}{1-\hat{\lambda}_{ML}^2}\right)\hat{\sigma}_{ML}^2 - \hat{\lambda}_{ML} Y_0^2 = \sum_{j=0}^{T-1}\left(Y_{T-j} - \lambda Y_{T-j-1}\right) Y_{T-j-1}$

- Solve iteratively to obtain $\hat{\sigma}_{ML}^2$ and $\hat{\lambda}_{ML}$

04 Volatility Clustering and Exchange Margins

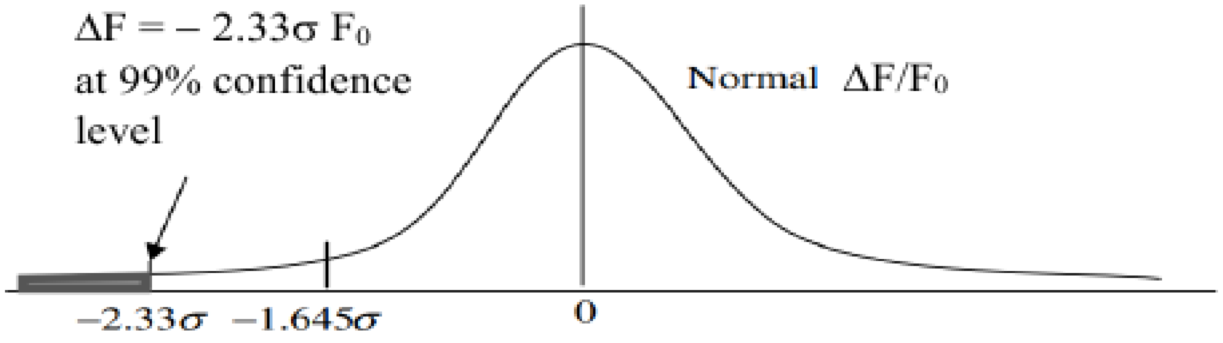


Figure 9.5: Value-at-Risk

The optimal setting of margins is closely related to the Value-at-Risk (VaR). Given a historical time series of daily futures price $\{F_t\}$ and its changes $\{\Delta F_t\}$, the pdf of the daily futures "return" or rate of change may be assumed to be normal, so $\frac{\Delta F_{t+1}}{F_t} \sim N(0, \sigma_{t+1}^2)$, and is depicted in Figure 9.5.

In Figure 9.5, the futures price at end of the previous trading day close is F_0. The uncertain or random next day price is F_1. The change is $\Delta F = F_1 - F_0$. For a normal distribution $\Delta F/F_0$, there is 1% probability or chance that $\Delta F/F_0 < -2.33\sigma$ where σ is anticipated volatility of next day futures "return". See shaded area in Figure - at 99% confidence level (or 99% probability), loss $\Delta F = F_1 - F_0$ would be at most $-2.33\sigma F_0$.

04 Volatility Clustering and Exchange Margins

- Suppose for the following day, $\frac{\Delta F}{F_0} \sim N(0, \sigma^2)$, and volatility is forecast as $\hat{\sigma}$ in order to estimate the VaR of a long N225 futures contract position.
- Daily VaR at 99% confidence level is such that $\text{Prob}(F_1 - F_0 < -2.33\hat{\sigma}F_0) = 1\%$ or $\text{Prob}(F_1 - F_0 \geq -2.33\hat{\sigma}F_0) = 99\%$. The VaR is $2.33\hat{\sigma}F_0$, assuming F_t index is denominated in \$.
- Each day t before the next trading day $t + 1$, an Exchange has to decide the level of maintenance margin per contract, $\$x_t$, so that within the next day, chances of the Exchange taking risk of a loss before top-up by the trader, i.e. when event $\{F_{t+1} - F_t < -x_t\}$ or loss exceeding maintenance margin happens, is 1% or less.

Then, x_t is set by the Exchange to be large enough, i.e. set $x_t \geq 2.33\hat{\sigma}_{t+1}F_t$.

04 Volatility Clustering and Exchange Margins

SMU SINGAPORE MANAGEMENT UNIVERSITY | Lee Kong Chian School of Business

Thus at t, forecasting or estimating σ_{t+1} is an important task for setting Exchange maintenance margin for the following day $t+1$. We can model the conditional variance of daily rates of the futures price change $\Delta F_{t+1}/F_t$ as a GARCH(1,1) process. Assume $E[\Delta F_{t+1}/F_t] = 0$ over a day.

Let $\Delta F_{t+1}/F_t = u_{t+1}$, $u_{t+1} = e_{t+1}\sigma_{t+1}$, and

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 u_t^2 + \gamma_1 \sigma_t^2$$

where $e_{t+1} \sim N(0,1)$. We estimate the parameters $\{\alpha_0, \alpha_1, \gamma_1\}$ using the MLE method for GARCH (1,1) in sub-section 9.6.2, and then use them to forecast the following day's volatility $\text{var}(u_{t+1})$ or σ_{t+1}^2 given observed u_t, u_{t-1}, u_{t-2}. and so on.

Practice Exercises (not graded)

+S&P_GARCH.ipynb
S&P500_Yahoo_Finance.csv