Financial Econometrics

Lecture 1: Financial data and their characteristics

Li Chen

WISE, Xiamen University

Financial Data

Commonly used financial data include

1. stock (future) market indices and prices
2. currency exchange rates
3. bond market rates
4. cryptocurrency exchange rates
5. resource prices (Gold, Silver, Crude Oil, etc.)
6. others...

Financial Data

Types of financial data include

- time series data
- cross-sectional data
- panel data

Can you give some examples for the above data types?

Financial Data

The main sources to obtain financial data include

- Websites:
  Yahoo Finance https://finance.yahoo.com
  Google Finance https://www.google.com/finance/
  ...
- Financial database:
  Wind https://www.wind.com.cn
  CSMAR https://cn.gtadata.com/
  ...
- Tranding softwares:
  Tonghuashun http://data.10jqka.com.cn
  ...

Financial Data Examples

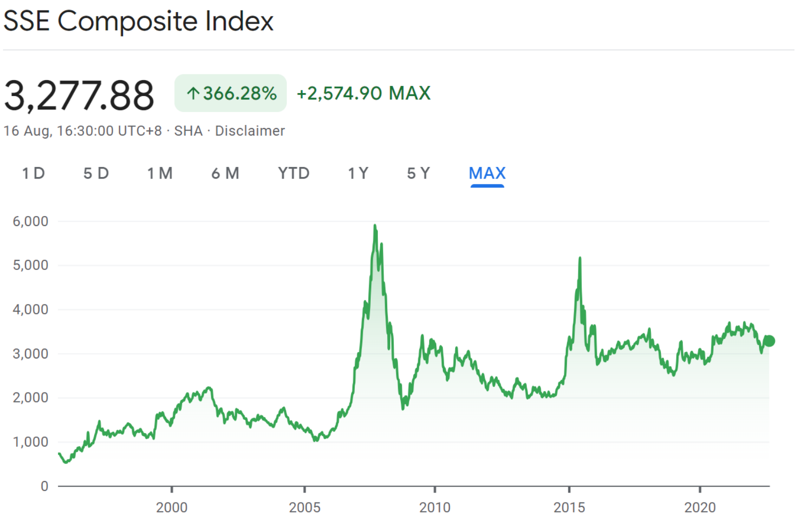
Financial Data Examples

Financial Data Examples

Financial Data Examples

Financial Data Examples

Financial Data Examples



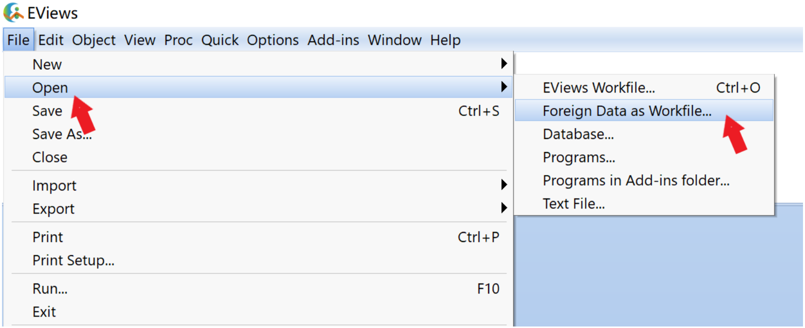
Financial Data Examples

Question:

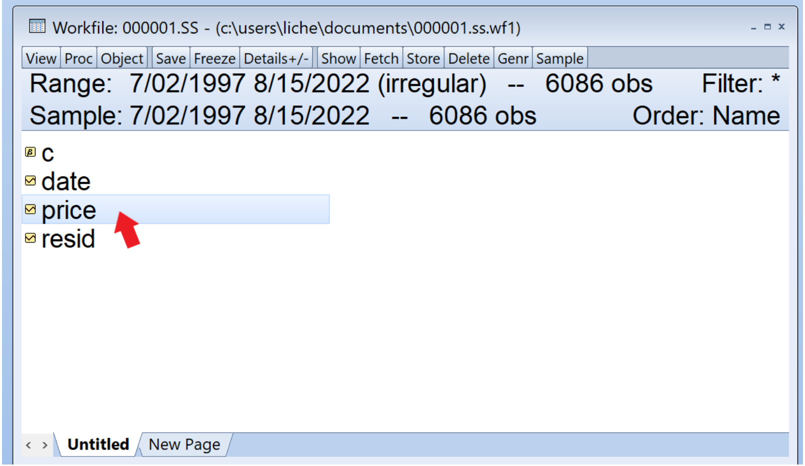
- Describe the main characteristics of the above financial data.
- What is the *frequency* of a sequence of time series data?

Financial Data Analysis: EViews

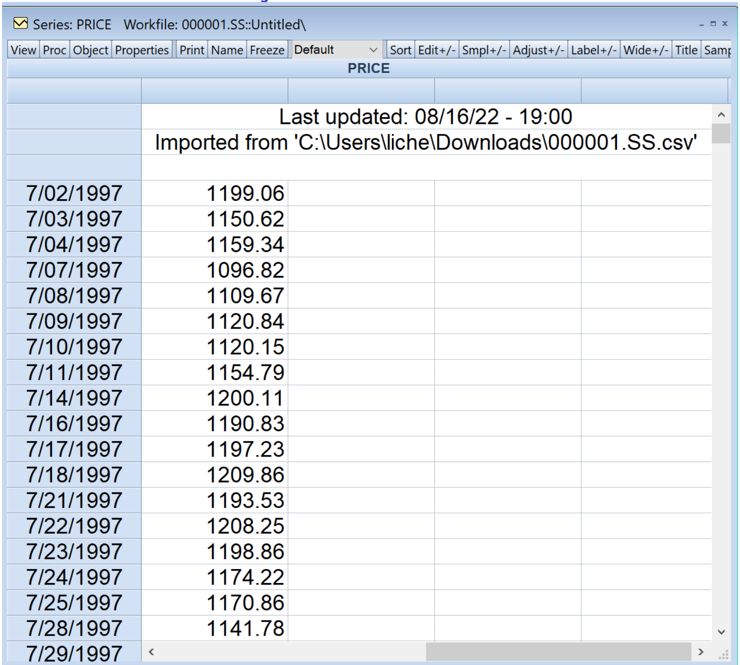
1. Download *Shanghai Stock Exchange Composite Index* data 1997-7-2 to 2022-8-15 and save them as .csv file
2. Import the data into EViews.



Financial Data Analysis: EViews

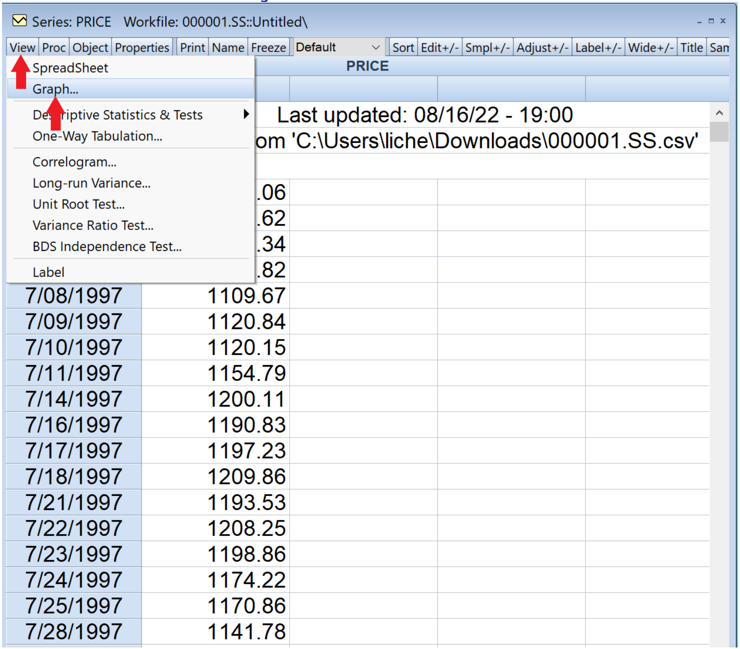


Financial Data Analysis: EViews

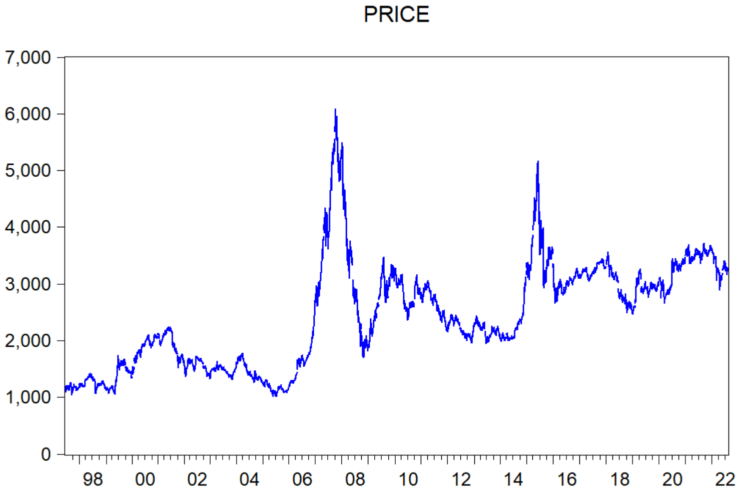


	PRICE
7/02/1997	1199.06
7/03/1997	1150.62
7/04/1997	1159.34
7/07/1997	1096.82
7/08/1997	1109.67
7/09/1997	1120.84
7/10/1997	1120.15
7/11/1997	1154.79
7/14/1997	1200.11
7/16/1997	1190.83
7/17/1997	1197.23
7/18/1997	1209.86
7/21/1997	1193.53
7/22/1997	1208.25
7/23/1997	1198.86
7/24/1997	1174.22
7/25/1997	1170.86
7/28/1997	1141.78
7/29/1997	

Financial Data Analysis: EViews



Financial Data Analysis: EViews



Financial Data Analysis

Question:

- Investigate and describe what happened for the *Shanghai Stock Exchange Composite Index* since 1997.
- Are the variations in the historical data at different time periods comparable?
- If not, then how can we adjust the index sequence to make them comparable over time?

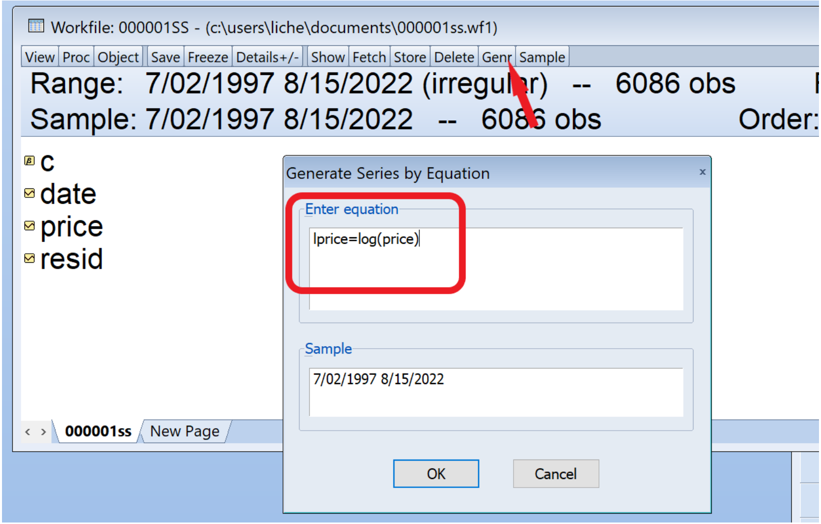
Log-transformations

- We denote the stock index or stock price at time t by P_t.
- The natural logarithm of P_t is called the **log price** that

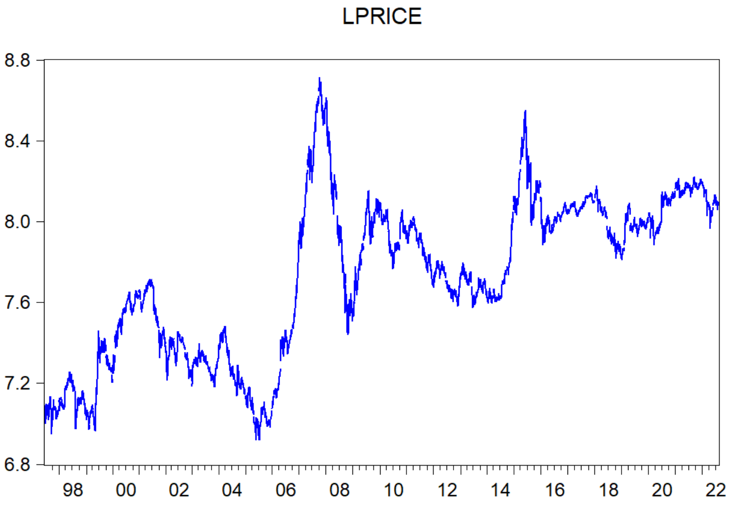
$$p_t = \log P_t.$$

- The changes in p_t are comparable regardless of the level of P_t.
- That is, same distances between log prices represent equal percentage changes.

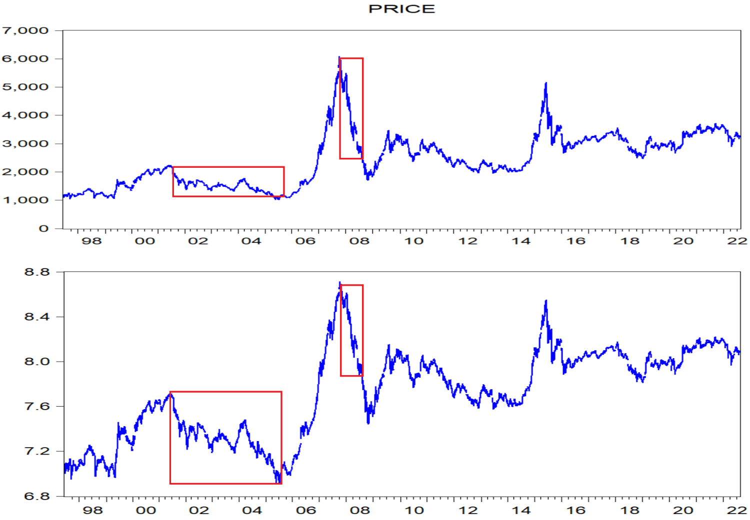
Log-transformations



Log-transformations



Log-transformations



Financial Returns

- The first-order difference of p_t is the so-called **log return** that

$$r_t = p_t - p_{t-1} = \log(P_t / P_{t-1}).$$

- The proportional change

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

  is called the **simple return**.
- By simple algebra, we have

$$r_t = \log(1 + R_t) \text{ or } R_t = e^{r_t} - 1.$$

- r_t is also called continuously compounded return because

$$\lim_{m \to \infty} \left(1 + \frac{r}{m}\right)^m = e^r.$$

Question: Which one is larger, simple return or log return? Prove your answers.

Financial Returns

When r_t and R_t are close to 0, we apply the Taylor series expansion around 0. Since $r_t = \log(1 + R_t)$, we have

$$r_t = \log(1 + R_t) = R_t \underbrace{- \frac{R_t^2}{2} + \frac{R_t^3}{3}}_{<0} \underbrace{- \frac{R_t^4}{4} + \frac{R_t^5}{5}}_{<0} - \cdots \tag{1}$$

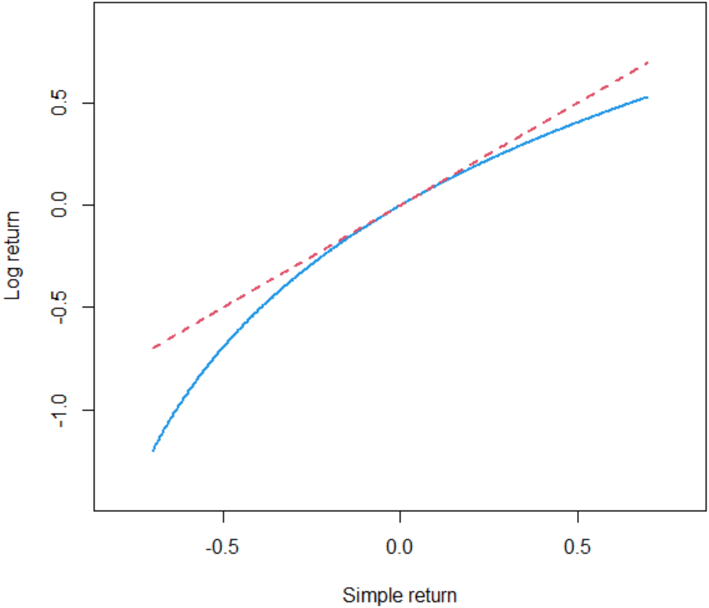
Alternatively, since $R_t = e^{r_t} - 1$, we have

$$R_t = r_t + \underbrace{\frac{r_t^2}{2!} + \frac{r_t^3}{3!}}_{>0} + \underbrace{\frac{r_t^4}{4!} + \frac{r_t^5}{5!}}_{>0} + \cdots \tag{2}$$

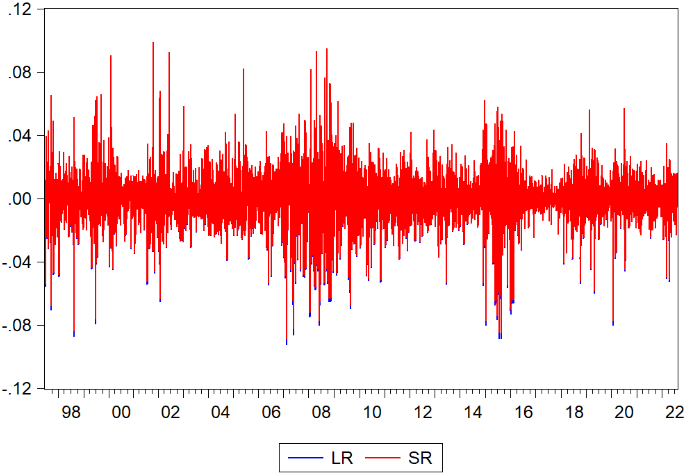
Therefore,

$$r_t \leq R_t. \tag{3}$$

Financial Returns



Financial Returns



Financial Returns

We consider the stock price from period t to $t+k$. The simple return from t to $t+k$ is

$$
\begin{aligned}
1 + R_t(k) =& \frac{P_{t+k}}{P_t} = \frac{P_{t+k}}{P_{t+k-1}} \frac{P_{t+k-1}}{P_{t+k-2}} \cdots \frac{P_{t+1}}{P_t} &(4)\\
=&(1 + R_{t+k})(1 + R_{t+k-1}) \cdots (1 + R_{t+1}). &(5)
\end{aligned}
$$

While the log return is

$$
\begin{aligned}
r_t(k) =& \log\left(\frac{P_{t+k}}{P_t}\right) = \log\left(\frac{P_{t+k}}{P_{t+k-1}} \frac{P_{t+k-1}}{P_{t+k-2}} \cdots \frac{P_{t+1}}{P_t}\right) &(6)\\
=& r_{t+1} + r_{t+2} + \cdots + r_{t+k}. &(7)
\end{aligned}
$$

Financial Returns - Descriptive statistics

Population version

- Mean $\mu = E(r_t)$,
- Variance $\sigma^2 = E(r_t - \mu)^2$,
- Skewness $sk = \frac{E(r_t-\mu)^3}{\sigma^3}$,
- Kurtosis $K = \frac{E(r_t-\mu)^4}{\sigma^4}$.

Sample version

- Sample mean $\bar{r} = T^{-1} \sum_{t=1}^{T} r_t$,
- Sample variance $\hat{\sigma}^2 = (T-1)^{-1} \sum_{t=1}^{T}(r_t - \bar{r})^2$,
- Sample skewness $\widehat{sk} = \frac{T^{-1} \sum_{t=1}^{T}(r_t-\bar{r})^3}{\hat{\sigma}^3}$,
- Sample kurtosis $\hat{K} = \frac{T^{-1} \sum_{t=1}^{T}(r_t-\bar{r})^4}{\hat{\sigma}^4}$,

where r_t is a time series of financial returns for $t = 1, 2, ..., T$.

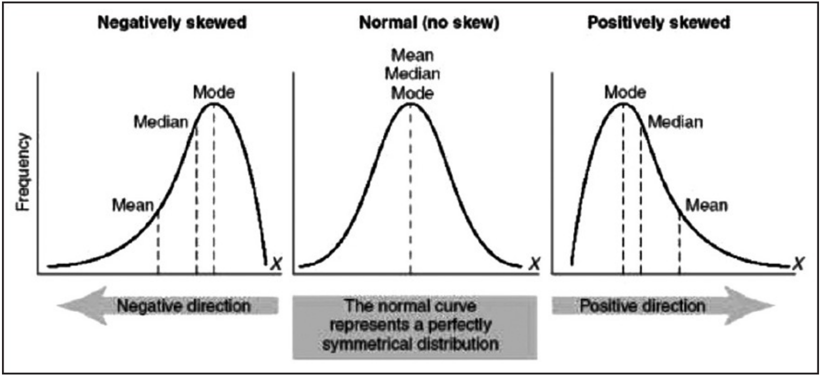
Financial Returns - Descriptive statistics

For a normal distribution, $sk = 0$ and $K = 3$.

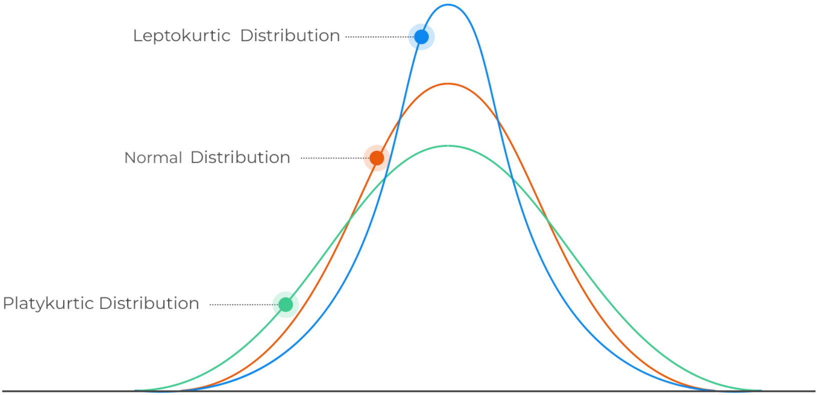
- $sk = 0$: symmetric (but not necessarily);
- $sk < 0$: negatively skewed or left skewed;
- $sk > 0$: positively skewed or right skewed;
- $K = 3$: Mesokurtic;
- $K > 3$: Leptokurtic or heavy tail;
- $K < 3$: Platykurtic or thin tail.

Excess kurtosis: $excess\, K = K - 3$.

Financial Returns - Descriptive statistics



Financial Returns - Descriptive statistics



Question: What is the implication for financial returns with a heavy-tailed distribution?

Financial Returns - Descriptive statistics

To test whether r_t follows a *Normal Distribution*, we apply the Jarque-Bera test

$$\mathbb{H}_0 : r_t \text{ follows Normal distribution} \tag{8}$$

$$\mathbb{H}_1 : r_t \text{ does not follow Normal distribution} \tag{9}$$

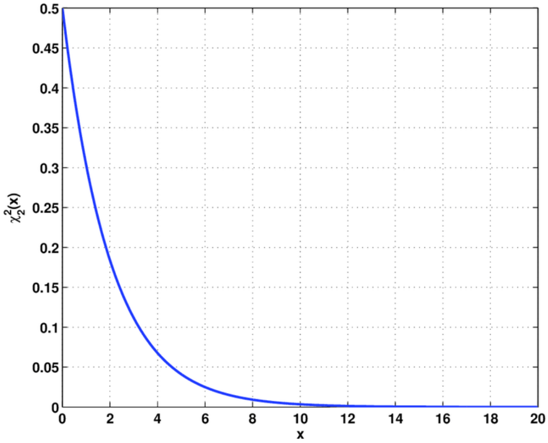
The test statistic

$$JB = \frac{T}{6}\left(\widehat{sk}^2 + \frac{(\hat{K}-3)^2}{4}\right), \tag{10}$$

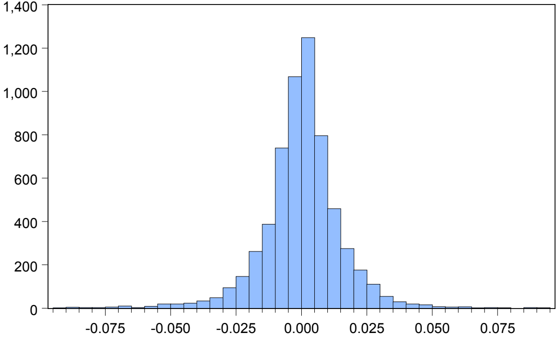
where $\widehat{sk}$ and $\hat{K}$ are the sample skewness and sample kurtosis, respectively.

Financial Returns - Descriptive statistics

- Under $\mathbb{H}_0$, $JB \sim \chi_2^2$ when $T \to \infty$.
- Under 5% significance level, we reject $\mathbb{H}_0$ when $JB > 5.99$.



Histogram and discriptive statistics



Series: R
Sample 7/02/1997 8/15/2022
Observations 6085

Mean	0.000165
Median	0.000565
Maximum	0.094008
Minimum	-0.092562
Std. Dev.	0.015293
Skewness	-0.355217
Kurtosis	8.079492
Jarque-Bera	6669.656
Probability	0.000000

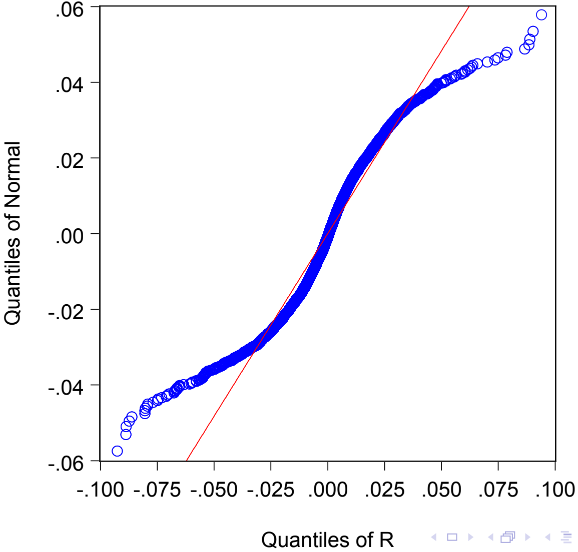
The QQ plot

A simpler way to examine Normality: Quantile-Quantile (QQ) plot.

- We compare the quantiles of the log return data with the normal distribution that has the same mean and variance.
- If log return follows normal distribution, then they should have the same quantiles (on the 45 degree line).

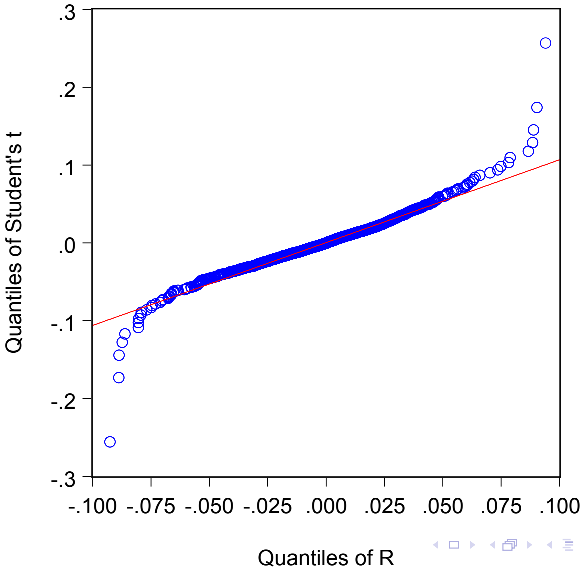
The QQ plot

Compare the distribution of LR with Normal distribution.



The QQ plot

How about t-distribution?



Measuring serial autocorrelation

We define k-th order autocovariance as

$$\gamma(k) = Cov(r_t, r_{t+k}) = E[r_t r_{t-k}] - E[r_t]E[r_{t-k}], \tag{11}$$

and k-th order serial autocorrelation (AC) as

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}. \tag{12}$$

We estimate $\gamma(k)$ by

$$\widehat{\gamma}(k) = T^{-1} \sum_{t=1}^{T-k} (r_t - \bar{r})(r_{t+k} - \bar{r}), \tag{13}$$

where $\bar{r} = T^{-1} \sum_{t=1}^{n} r_t$. Therefore, $\widehat{\rho}(k) = \widehat{\gamma}(k)/\widehat{\gamma}(0)$.

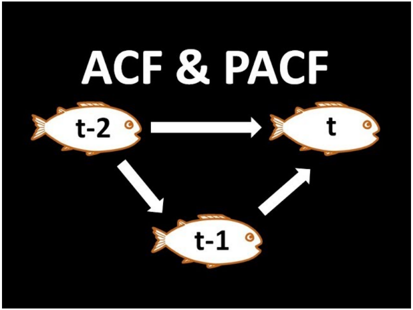
Stationary

A time series is called weakly stationary (or second-order stationary), if its mean, variance, and autocovariances are time-invariant. That is, μ, σ^2, and $\gamma(k)$ do not change over time.

- Stock prices (indices) are usually non-stationary;
- Stock returns are usually stationary.

Partial autocorrelation

- The correlation between r_t and r_{t-2} may be **indirect**.
- In other words, r_{t-2} affect r_t because r_{t-2} first affects r_{t-1} and then r_{t-1} affects r_t.



Partial autocorrelation

- We measure the direct correlation between r_{t-k} and r_t by the partial autocorrelation (PAC).
- The k^{th}-order PAC is usually estimated by the linear regression model of AR(k)

$$r_t = c_0 + c_1 r_{t-1} + c_2 r_{t-2} + \cdots + c_k r_{t-k} + e_t,$$

  where c_k is the value of the k^{th}-order PAC.
- By definition, the first-order AC and PAC are the same.

The LM test

To test whether the first k autocorrelations are zeros or not, we employ the Lagrange Multiplier (LM) test.

$$\mathbb{H}_0 : \rho_1 = \rho_2 = ... = \rho_k = 0, \tag{14}$$

$$\mathbb{H}_1 : \text{at least one of } \rho_1, ..., \rho_k \text{ is not zero.} \tag{15}$$

If the p-value of the test is smaller than the significance level (say 5%), we reject the null hypothesis and conclude that r_t is autocorrelated.

Correlogram

Date: 09/15/22 Time: 21:07
Sample: 7/02/1997 8/15/2022
Included observations: 6085

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.019	0.019	2.2468	0.134
		2	-0.024	-0.025	5.8227	0.054
		3	0.033	0.034	12.539	0.006
		4	0.042	0.040	23.046	0.000
		5	-0.002	-0.002	23.084	0.000
		6	-0.053	-0.053	40.381	0.000
		7	0.028	0.028	45.259	0.000
		8	0.004	-0.001	45.380	0.000
		9	-0.005	0.000	45.518	0.000
		10	-0.003	-0.001	45.578	0.000
		11	0.016	0.014	47.148	0.000
		12	0.030	0.027	52.805	0.000
		13	0.030	0.033	58.315	0.000
		14	-0.027	-0.029	62.770	0.000
		15	0.047	0.047	76.503	0.000

Summary

- We usually take the logarithm of financial prices (indices) before analyzing them.
- Financial returns are usually
    - leptokurtic with heavy tail;
    - negatively skewed;
    - not Normally distribution;
    - not autocorrelated or very weakly correlated;
    - ...

Exercise: Analyze a time series sequence of financial data, and try to discover their characteristics.