

Lecture Notes of Bus 41202 (Spring 2007)

Analysis of Financial Time Series

Ruey S. Tsay

What is financial time series (FTS) analysis?

Theory and practice of asset valuation over time.

Different from other T.S. analysis?

Not exactly, but with an added uncertainty.

For example, FTS must deal with the ever-changing business & economic environment and the fact that volatility is not directly observed.

Objective of the course

- to provide some basic knowledge of financial time series data
- to introduce some statistical tools & econometric models useful for analyzing these series.
- to gain empirical experience in analyzing FTS
- to study methods for assessing market risk
- to analyze high-dimensional asset returns.

Examples of financial time series

1. Daily log returns of GE stock

2. Quarterly earnings of Johnson & Johnson

Seasonal time series useful in

- earning forecasts
- pricing weather related derivatives (e.g. energy)
- modeling intraday behavior of asset returns

3. US monthly interest rates

Relations between the two series? Term structure of interest rates

4. Exchange rate between US Dollar vs Japanese Yen

Fixed income, hedging, carry trade

5. High-frequency financial data:

Tick-by-tick data of Boeing stock: December 5, 2005.

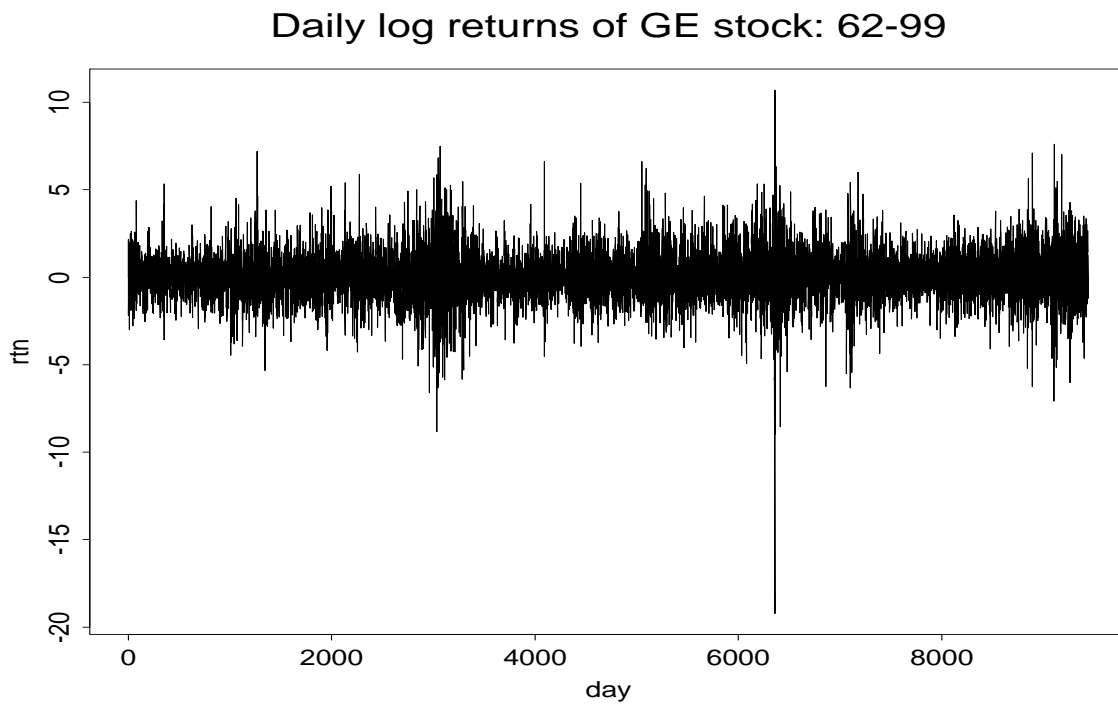


Figure 1: Daily log returns of GE stock

Quarterly earnings per share of Johnson & Johnson: 60-80

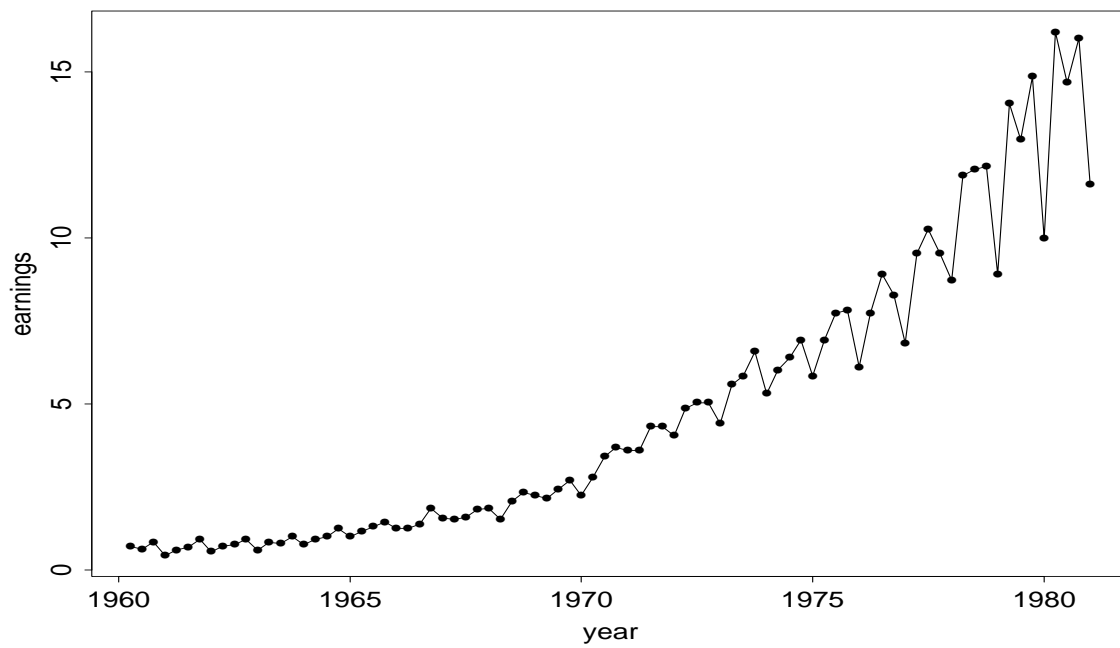


Figure 2: Quarterly earnings per share of Johnson and Johnson

Daily Exchange Rate: US-JP (1/3/94-2/28/01)

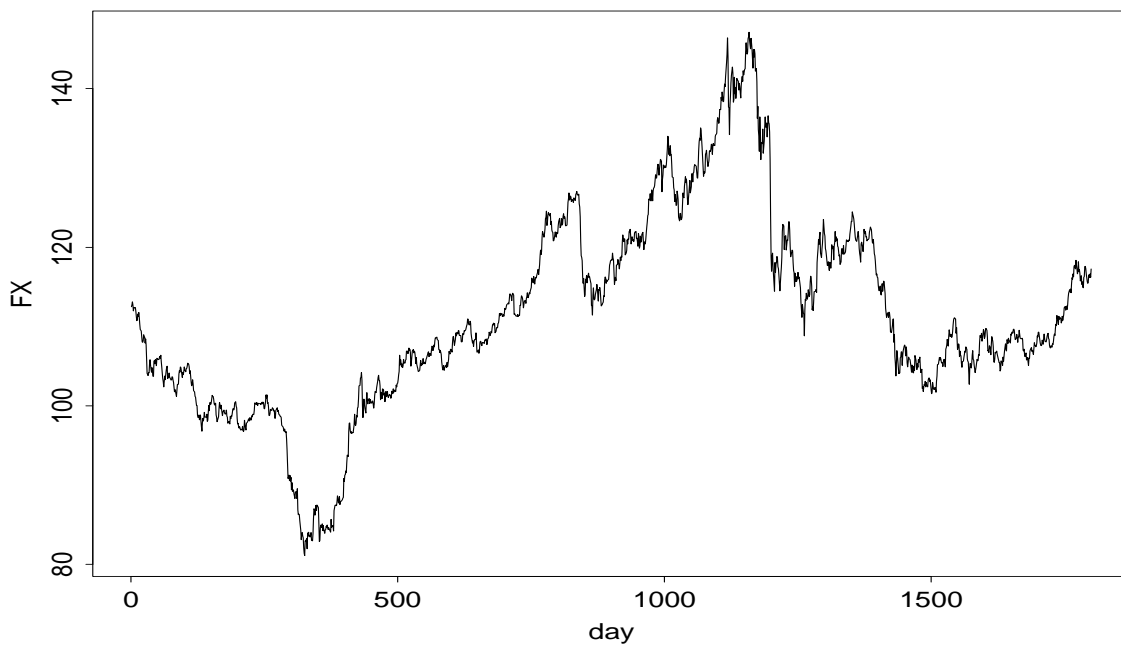


Figure 3: Daily Exchange Rate: Dollar vs Yen

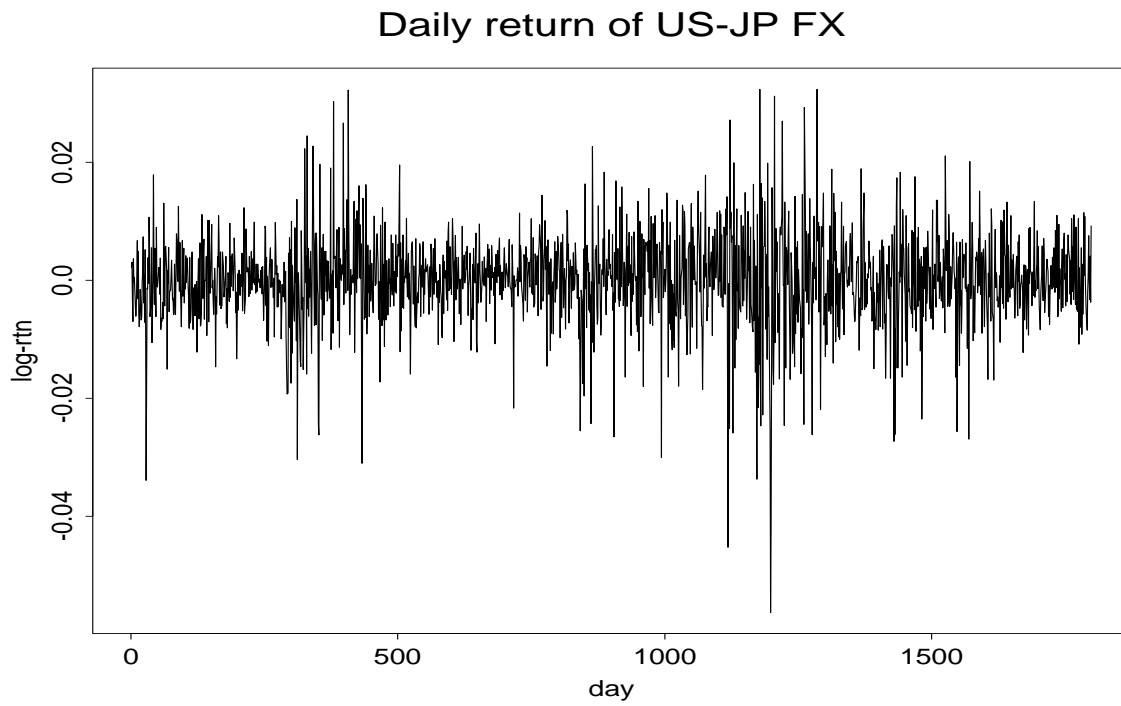
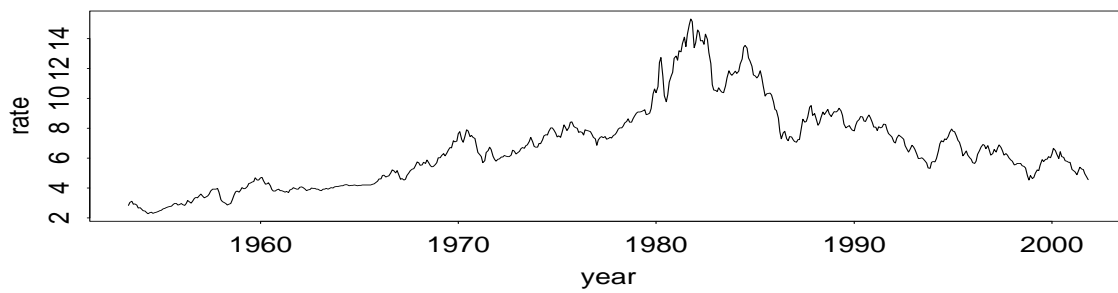


Figure 4: Daily log returns of FX (Dollar vs Yen)

(a) Monthly US interest rates: 10-year maturity



(b) 1-year maturity

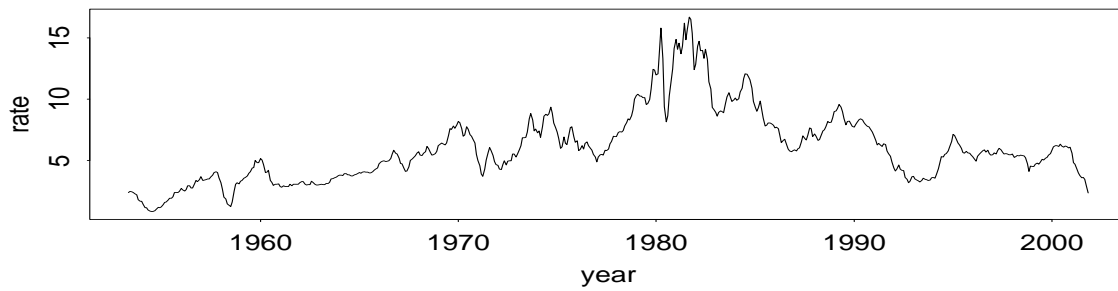


Figure 5: Monthly US interest rates

symbol	date	time	price	size
BA	05DEC2005	9:31:10	69.4500	60700
BA	05DEC2005	9:31:11	69.4500	100
BA	05DEC2005	9:31:11	69.4500	200
.....				
BA	05DEC2005	9:31:11	69.4500	2500
BA	05DEC2005	9:31:11	69.4500	100
BA	05DEC2005	9:31:11	69.4500	100
BA	05DEC2005	9:31:12	69.4500	1600
BA	05DEC2005	9:31:12	69.4500	1500
BA	05DEC2005	9:31:12	69.4500	1700
BA	05DEC2005	9:31:12	69.4500	100
BA	05DEC2005	9:31:13	69.4500	100
BA	05DEC2005	9:31:15	69.4500	100
BA	05DEC2005	9:31:18	69.4700	100
.....				
BA	05DEC2005	9:31:18	69.4500	100
BA	05DEC2005	9:31:19	69.4500	100
.....				
BA	05DEC2005	9:31:19	69.4500	100
BA	05DEC2005	9:31:27	69.4500	100
BA	05DEC2005	9:31:53	69.4700	1700
BA	05DEC2005	9:31:56	69.4800	700
BA	05DEC2005	9:32:03	69.4900	200
BA	05DEC2005	9:32:06	69.4900	100
BA	05DEC2005	9:32:07	69.4800	400
BA	05DEC2005	9:32:08	69.4900	100

Outline of the course

- Returns & their characteristics: empirical analysis
- Simple linear time series models
- Univariate volatility modeling
- Nonlinearity in level and volatility
- Neural network
- High-frequency financial data and market micro-structure
- Continuous-time models and derivative pricing
- Value at Risk and extreme value theory
- Multivariate models: factor models, dynamic and cross dependence

Asset Returns

Let P_t be the price of an asset at time t , and assume no dividend.

One-period simple return: Gross return

$$1 + R_t = \frac{P_t}{P_{t-1}} \quad \text{or} \quad P_t = P_{t-1}(1 + R_t)$$

Simple return:

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

Multiperiod simple return: Gross return

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

The k -period simple net return is $R_t(k) = \frac{P_t}{P_{t-k}} - 1$.

Example: Suppose the daily closing prices of a stock are

Day	1	2	3	4	5
Price	37.84	38.49	37.12	37.60	36.30

1. What is the simple return from day 1 to day 2?

$$\text{Ans: } R_2 = \frac{38.49 - 37.84}{37.84} = 0.017.$$

2. What is the simple return from day 1 to day 5?

$$\text{Ans: } R_5(4) = \frac{36.30 - 37.84}{37.84} = -0.041.$$

3. Verify that $1 + R_5(4) = (1 + R_2)(1 + R_3) \cdots (1 + R_5)$.

Time interval is important! Default is one year.

Annualized (average) return:

$$\text{Annualized}[R_t(k)] = \left[\prod_{j=0}^{k-1} (1 + R_{t-j}) \right]^{1/k} - 1.$$

An approximation:

$$\text{Annualized}[R_t(k)] \approx \frac{1}{k} \sum_{j=0}^{k-1} R_{t-j}.$$

Continuously compounding: Illustration of the power of compounding (int. rate 10% per annum)

Type	#(payment)	Int.	Net
Annual	1	0.1	\$1.10000
Semi-Annual	2	0.05	\$1.10250
Quarterly	4	0.025	\$1.10381
Monthly	12	0.0083	\$1.10471
Weekly	52	$\frac{0.1}{52}$	\$1.10506
Daily	365	$\frac{0.1}{365}$	\$1.10516
Continuously	∞		\$1.10517

$$A = C \exp[r \times n]$$

where r is the interest rate per annum, C is the initial capital, n is the number of years, and \exp is the exponential function.

Present value:

$$C = A \exp[-r \times n]$$

Continuously compounded (or log) return

$$r_t = \ln(1 + R_t) = \ln \frac{P_t}{P_{t-1}} = p_t - p_{t-1},$$

where $p_t = \ln(P_t)$.

Multiperiod log return:

$$\begin{aligned} r_t(k) &= \ln[1 + R_t(k)] \\ &= \ln[(1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})] \\ &= \ln(1 + R_t) + \ln(1 + R_{t-1}) + \cdots + \ln(1 + R_{t-k+1}) \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1}. \end{aligned}$$

Example (continued). Use the previous daily prices.

1. What is the log return from day 1 to day 2?

$$\text{A: } r_2 = \ln(38.49) - \ln(37.84) = 0.017.$$

2. What is the log return from day 1 to day 5?

$$\text{A: } r_5(4) = \ln(36.3) - \ln(37.84) = -0.042.$$

3. It is easy to verify $r_5(4) = r_2 + \dots + r_5$.

Portfolio return: N assets

$$R_{p,t} = \sum_{i=1}^N w_i R_{it}$$

Example: An investor holds stocks of IBM, Microsoft and Citi-Group. Assume that her capital allocation is 30%, 30% and 40%. Use the monthly simple returns in Table 1.2. What is the mean simple return of her stock portfolio?

Answer: $E(R_t) = 0.3 \times 1.42 + 0.3 \times 3.37 + 0.4 \times 2.20 = 2.32$.

Dividend payment:

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1, \quad r_t = \ln(P_t + D_t) - \ln(P_{t-1}).$$

Excess return: (adjusting for risk)

$$Z_t = R_t - R_{0t}, \quad z_t = r_t - r_{0t}$$

where r_{0t} denotes the log return of a reference asset (e.g. risk-free interest rate).

Relationship:

$$r_t = \ln(1 + R_t), \quad R_t = e^{r_t} - 1.$$

If the returns are in **percentage**, then

$$r_t = 100 \times \ln\left(1 + \frac{R_t}{100}\right), \quad R_t = [\exp(r_t/100) - 1] \times 100.$$

Temporal aggregation of the returns produces

$$\begin{aligned} 1 + R_t(k) &= (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}), \\ r_t(k) &= r_t + r_{t-1} + \cdots + r_{t-k+1}. \end{aligned}$$

These two relations are important in practice, e.g. obtain annual returns from monthly returns.

Example: If the monthly log returns of an asset are 4.46%, -7.34% and 10.77%, then what is the corresponding quarterly log return?

Answer: $4.46 - 7.34 + 10.77 = 7.89\%$.

Example: If the monthly simple returns of an asset are 4.46%, -7.34% and 10.77%, then what is the corresponding quarterly simple return?

Answer: $R = (1 + 0.0446)(1 - 0.0734)(1 + 0.1077) - 1 = 1.0721 - 1 = 0.0721 = 7.21\%$

Distributional properties of returns

Key: What is the distribution of

$\{r_{it}; i = 1, \dots, N; t = 1, \dots, T\}$?

Some theoretical properties:

Moments of a random variable X with density $f(x)$: ℓ -th moment

$$m'_\ell = E(X^\ell) = \int_{-\infty}^{\infty} x^\ell f(x) dx$$

First moment: mean or expectation of X .

ℓ -th central moment

$$m_\ell = E[(X - \mu_x)^\ell] = \int_{-\infty}^{\infty} (x - \mu_x)^\ell f(x) dx,$$

2nd c.m.: Variance of X .

Skewness (symmetry) and kurtosis (fat-tails)

$$S(x) = E \left[\frac{(X - \mu_x)^3}{\sigma_x^3} \right], \quad K(x) = E \left[\frac{(X - \mu_x)^4}{\sigma_x^4} \right].$$

$K(x) - 3$: *Excess kurtosis*.

Why are mean and variance of returns important?

They are concerned with long-term return and risk, respectively.

Why is symmetry of interest in financial study?

Symmetry has important implications in holding short or long financial positions and in risk management.

Why is kurtosis important?

Related to volatility forecasting, efficiency in estimation and tests, etc.

High kurtosis implies heavy (or long) tails in distribution.

Estimation:

Data: $\{x_1, \dots, x_T\}$

- sample mean:

$$\hat{\mu}_x = \frac{1}{T} \sum_{t=1}^T x_t,$$

- sample variance:

$$\hat{\sigma}_x^2 = \frac{1}{T-1} \sum_{t=1}^T (x_t - \hat{\mu}_x)^2,$$

- sample skewness:

$$\hat{S}(x) = \frac{1}{(T-1)\hat{\sigma}_x^3} \sum_{t=1}^T (x_t - \hat{\mu}_x)^3,$$

- sample kurtosis:

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}_x^4} \sum_{t=1}^T (x_t - \hat{\mu}_x)^4.$$

Under normality assumption,

$$\hat{S}(x) \sim N\left(0, \frac{6}{T}\right), \quad \hat{K}(x) - 3 \sim N\left(0, \frac{24}{T}\right).$$

Some simple tests for normality (for large T).

1. Test for symmetry:

$$S^* = \frac{\hat{S}(x)}{\sqrt{6/T}} \sim N(0, 1)$$

if normality holds.

Decision rule: Reject H_0 of a symmetric distribution if $|S^*| > Z_{\alpha/2}$ or p-value is less than α .

2. Test for tail thickness:

$$K^* = \frac{\hat{K}(x) - 3}{\sqrt{24/T}} \sim N(0, 1)$$

if normality holds.

Decision rule: Reject H_o of normal tails if $|K^*| > Z_{\alpha/2}$ or p-value is less than α .

3. A joint test (Jarque-Bera test):

$$JB = (K^*)^2 + (S^*)^2 \sim \chi_2^2$$

if normality holds, where χ_2^2 denotes a chi-squared distribution with 2 degrees of freedom.

Decision rule: Reject H_o of normality if $JB > \chi_2^2(\alpha)$ or p-value is less than α .

Empirical properties of returns

Data sources:

- Course web:

<http://faculty.chicagogsb.edu/ruey.tsay/teaching/bs41202/sp2007/>

- CRSP: Center for Research in Security Prices (via Wharton WRDS)

<http://wrdsx.wharton.upenn.edu/>

- Various web sites, e.g. Federal Reserve Bank at St. Louis

<http://research.stlouisfed.org/fred2/>

- Data sets of the textbook:

<http://faculty.chicagogsb.edu/ruey.tsay/teaching/fts2/>

See Figures and Tables of Chapter 1 for summary, including comparison between empirical dist and normal dist

Empirical dist of asset returns tends to be skewed to the left with heavy tails and has a higher peak than normal dist.

Demonstration of Data Analysis

R demonstration: Use monthly IBM stock returns rom 1926 to 2004.

```
**** Task: (a) Set the working directory
           (b) Load the library ‘fSeries’
           (c) Compute summary statistics
           (d) Perform Jarque-Bera test for normality

> setwd("C:/teaching/bs41202") % Set the working directory
> library(fSeries) % Load the package fSeries.
Loading required package: mgcv
This is mgcv 1.3-22
Loading required package: nnet
Loading required package: fCalendar
Loading required package: fEcofin

Rmetrics, (C) 1999-2006, Diethelm Wuertz, GPL
fCalendar: Time, Date and Calendar Tools
Loading required package: fBasics
Loading required package: MASS

Rmetrics, (C) 1999-2005, Diethelm Wuertz, GPL
fBasics: Markets and Basic Statistics

Rmetrics, (C) 1999-2006, Diethelm Wuertz, GPL
fSeries: The Dynamical Process Behind Financial Markets

> da=read.table("m-ibm2604.txt",header=T) %Load data into R workspace
> dim(da) %Find the dimension of the data table
[1] 948 2
```

```

> ibm=da[,2] %First column is the ‘date’.
> basicStats(ibm) % Compute the descriptive statistics
      round.ans..digits...6.
nobs          948.000000
NAs           0.000000
Minimum      -0.261900
Maximum       0.353800
1. Quartile  -0.027610
3. Quartile   0.052248
Mean          0.014085
Median        0.011550
Sum           13.352570
SE Mean       0.002298
LCL Mean      0.009575
UCL Mean      0.018595
Variance      0.005006
Stdev         0.070754
Skewness      0.271671
Kurtosis      2.181691
> jbTest(ibm) %Jarque-Bera Normality test

```

Title:

Jarque - Bera Normality Test

Test Results:

PARAMETER:

Sample Size: 948

STATISTIC:

LM: 201.601

ALM: 205.816

P VALUE:

LM p-value: < 2.2e-16

ALM p-value: < 2.2e-16

Asymptotic: < 2.2e-16

Description:

Wed Mar 21 10:35:11 2007

Warning messages:

1: p-value greater/smaller than printed p-value in: .pTable(X, STAT, N, digits)

2: p-value greater/smaller than printed p-value in: .pTable(X, STAT, N, digits)

>q() % quit R.

Spplus demonstration: Use daily IBM stock returns

*** Task: (a) Load Finmetrics module

- (b) Load data into Splus
- (c) Compute summary statistics
- (d) Perform normality test

\$ Splus

```
> module(finmetrics)           % load FinMetrics Module
(click on 'File', then click on 'Load module' on a window version to
select the 'FinMetrics'.)
```

```
> x=matrix(scan(file='d-ibmvew6202.txt'),4) % Load data into S-plus
(click on 'File', then 'Import data', then 'From File' to obtain
a pop-up window to browse the data file.)
```

```
> ibm=x[2,]           % Select IBM simple returns
> pibm=ibm*100        % Percentage simple return
> rt=log(ibm+1)       % transform into log returns
> summaryStats(pibm) % Compute summary statistics
```

Sample Quantiles:

```
   min   1Q median   3Q   max
-22.96 -0.84    0 0.881 13.16
```

Sample Moments:

```
   mean   std skewness kurtosis   % Kurtosis (Not excess)
0.05152 1.652  0.07828   13.34
```

Number of Observations: 10194

```
> normalTest(pibm,method='jb') % JB normality test
```

Test for Normality: Jarque-Bera

Null Hypothesis: data is normally distributed

Test Statistics:

```
Test Stat 45445.62
p.value    0.00
```

Dist. under Null: chi-square with 2 degrees of freedom

Total Observ.: 10194

```
> q() % Exit Splus.
```

Use of SCA: An illustration.

```
*** Task: (a) load daily returns of IBM, VW, EW from the ASCII file
***         "d-ibmvew6202.txt" into SCA workspace,
***         (b) obtain the summary statistics for the simple
***         returns of IBM.
***         (c) perform skewness, kurtosis and normality tests.
```

```

***
*** Commands and output will be discussed in class.
--
input yrmdd,ibm,vw,ew. file 'd-ibmvwew6202.txt' % Load data

YRMMDD , A10194 BY 1 VARIABLE, IS STORED IN THE WORKSPACE
IBM , A10194 BY 1 VARIABLE, IS STORED IN THE WORKSPACE
VW , A10194 BY 1 VARIABLE, IS STORED IN THE WORKSPACE
EW , A10194 BY 1 VARIABLE, IS STORED IN THE WORKSPACE
--
rt=ln(ibm+1) % transform to log returns
--
pibm=ibm*100 % percentage simple returns
--
desc pibm % Calculate descriptive statistics

VARIABLE NAME IS PIBM
NUMBER OF OBSERVATIONS 10194
NUMBER OF MISSING VALUES 0

STATISTIC STD. ERROR STATISTIC/S.E.
MEAN 0.0515 0.0164 3.1480
VARIANCE 2.7299
STD DEVIATION 1.6522
C. V. 32.0728
SKEWNESS 0.0783 0.0243
KURTOSIS 10.3400 0.0485 % Excess kurtosis

QUARTILE
MINIMUM -22.9630
1ST QUARTILE -0.8400
MEDIAN 0.0000
3RD QUARTILE 0.8810
MAXIMUM 13.1640

RANGE
MAX - MIN 36.1270
Q3 - Q1 1.7210
--
skew=0.0783/.0243 % t-ratio for skewness
--
kur=10.34/.0485 % t-ratio for kurtosis
--
jb=skew*skew+kur*kur % JB-statistic
--
pv=2*(1-cdfn(abs(skew))) % Compute p-value of Skewness test

```

```

--
print skew, pv.

VARIABLE      SKEW      PV
COLUMN-->      1        1
  ROW
    1          3.222    .001
--
pv=2*(1-cdfn(kur))  % Compute p-value for Kurtosis test
--
print kur, pv

VARIABLE      KUR      PV
COLUMN-->      1        1
  ROW
    1          213.196  .000
--
pv=1-cdfc(jb,2)    % Compute p-value for JB statistic
--
print jb,pv

VARIABLE      JB      PV
COLUMN-->      1        1
  ROW
    1          45462.863 .000
--
stop           % Exit SCA

```

Normal and lognormal dists

Y is lognormal if $X = \ln(Y)$ is normal.

If $X \sim N(\mu, \sigma^2)$, then $Y = \exp(X)$ is lognormal with mean and variance

$$E(Y) = \exp\left(\mu + \frac{\sigma^2}{2}\right), \quad V(Y) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1].$$

Conversely, if Y is lognormal with mean μ_y and variance σ_y^2 , then $X = \ln(Y)$ is normal with mean and variance

$$E(X) = \ln \left[\frac{\mu_y}{\sqrt{1 + \frac{\sigma_y^2}{\mu_y^2}}} \right], \quad V(X) = \ln \left[1 + \frac{\sigma_y^2}{\mu_y^2} \right].$$

Application: If the log return of an asset is normally distributed with mean 0.0119 and standard deviation 0.0663, then what is the mean and standard deviation of its simple return?

Answer: Solve this problem in two steps.

Step 1: Based on the prior results, the mean and variance of $Y_t = \exp(r_t)$ are

$$E(Y) = \exp \left[0.0119 + \frac{0.0663^2}{2} \right] = 1.014$$

$$V(Y) = \exp(2 \times 0.0119 + 0.0663^2)[\exp(0.0663^2) - 1] = 0.0045$$

Step 2: Simple return is $R_t = \exp(r_t) - 1 = Y_t - 1$. Therefore,

$$E(R) = E(Y) - 1 = 0.014$$

$$V(R) = V(Y) = 0.0045, \quad \text{standard dev} = \sqrt{V(R)} = 0.067$$

Remark: See the monthly IBM stock returns in Table 1.2.

Processes considered

- return series (e.g., ch. 1, 2, 5)
- volatility processes (e.g., ch. 3, 4, 10, 12)
- continuous-time processes (ch. 6)
- extreme events (ch. 7)
- multivariate series (ch. 8, 9, 10)

Likelihood function (for self study)

Finally, it pays to study the likelihood function of returns $\{r_1, \dots, r_T\}$ discussed in Chapter 1.

Basic concept:

Joint dist = Conditional dist \times Marginal dist, i.e.

$$f(x, y) = f(x|y)f(y)$$

For two consecutive returns r_1 and r_2 , we have

$$f(r_2, r_1) = f(r_2|r_1)f(r_1).$$

For three returns r_1 , r_2 and r_3 , by repeated application,

$$\begin{aligned} f(r_3, r_2, r_1) &= f(r_3|r_2, r_1)f(r_2, r_1) \\ &= f(r_3|r_2, r_1)f(r_2|r_1)f(r_1). \end{aligned}$$

In general, we have

$$\begin{aligned}
 & f(r_T, r_{T-1}, \dots, r_2, r_1) \\
 = & f(r_T | r_{T-1}, \dots, r_1) f(r_{T-1}, \dots, r_1) \\
 = & f(r_T | r_{T-1}, \dots, r_1) f(r_{T-1} | r_{T-2}, \dots, r_1) f(r_{T-2}, \dots, r_1) \\
 = & \vdots \\
 = & \left[\prod_{t=2}^T f(r_t | r_{t-1}, \dots, r_1) \right] f(r_1),
 \end{aligned}$$

where $\prod_{t=2}^T$ denotes product.

If $r_t | r_{t-1}, \dots, r_1$ is normal with mean μ_t and variance σ_t^2 , then likelihood function becomes

$$f(r_T, r_{T-1}, \dots, r_1) = \prod_{t=2}^T \frac{1}{\sqrt{2\pi}\sigma_t} \exp \left[\frac{-(r_t - \mu_t)^2}{2\sigma_t^2} \right] f(r_1).$$

For simplicity, if $f(r_1)$ is ignored, then the likelihood function becomes

$$f(r_T, r_{T-1}, \dots, r_1) = \prod_{t=2}^T \frac{1}{\sqrt{2\pi}\sigma_t} \exp \left[\frac{-(r_t - \mu_t)^2}{2\sigma_t^2} \right].$$

This is the *conditional* likelihood function of the returns under normality.

Other dists, e.g. Student- t , can be used to handle heavy tails.

Model specification

- μ_t : discussed in Chapter 2
- σ_t^2 : Chapetrs 3 and 4.

Takeaway

1. Understand the summary statistics of asset returns
2. Understand various definitions of returns & their relationships
3. Learn basic characteristics of FTS.

Linear Time Series (TS) Models

Financial TS: collection of a financial measurement over time

Example: log return r_t

Data: $\{r_1, r_2, \dots, r_T\}$ (T data points)

Purpose: What information contained in $\{r_t\}$?

Basic concepts

- Stationarity:
 - Strict: distributions are time-invariant
 - Weak: first 2 moments are time-invariant

What does weak stationarity mean in practice?

Past: time plot of $\{r_t\}$ varies around a fixed level within a finite range!

Future: the first 2 moments of future r_t are the same as those of the data so that meaningful inferences can be made.

- Mean (or expectation) of returns:

$$\mu = E(r_t)$$

- Variance (variability) of returns:

$$\text{Var}(r_t) = E[(r_t - \mu)^2]$$

- Sample mean and sample variance are used to estimate the mean and variance of returns.

$$\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t \quad \& \quad \text{Var}(r_t) = \frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{r})^2$$

- Test $H_o : \mu = 0$ vs $H_a : \mu \neq 0$. Compute

$$t = \frac{\bar{r}}{\text{std}(\bar{r})} = \frac{\bar{r}}{\sqrt{\text{Var}(r_t)/T}}$$

Compare t ratio with $N(0, 1)$ dist.

Decision rule: Reject H_o of zero mean if $|t| > Z_{\alpha/2}$ or p-value is less than α .

- Lag- k autocovariance:

$$\gamma_k = \text{Cov}(r_t, r_{t-k}) = E[(r_t - \mu)(r_{t-k} - \mu)].$$

- Serial (or auto-) correlations:

$$\rho_\ell = \frac{\text{cov}(r_t, r_{t-\ell})}{\text{var}(r_t)}$$

Note: $\rho_0 = 1$ and $\rho_k = \rho_{-k}$ for $k \neq 0$. Why?

Existence of serial correlations implies that the return is predictable, indicating market inefficiency.

- Sample autocorrelation function (ACF)

$$\hat{\rho}_\ell = \frac{\sum_{t=1}^{T-\ell} (r_t - \bar{r})(r_{t+\ell} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2},$$

where \bar{r} is the sample mean & T is the sample size.

- Test zero serial correlations (market efficiency)

– Individual test: for example,

$$H_o : \rho_1 = 0 \text{ vs } H_a : \rho_1 \neq 0$$

$$t = \frac{\hat{\rho}_1}{\sqrt{1/T}} = \sqrt{T} \hat{\rho}_1$$

Asym. $N(0, 1)$.

Decision rule: Reject H_o if $|t| > Z_{\alpha/2}$ or p-value less than α .

– Joint test (Ljung-Box statistics):

$$H_o : \rho_1 = \dots = \rho_m = 0 \text{ vs } H_a : \rho_i \neq 0$$

$$Q(m) = T(T + 2) \sum_{\ell=1}^m \frac{\hat{\rho}_\ell^2}{T - \ell}$$

Asym. chi-squared dist with m degrees of freedom.

Decision rule: Reject H_o if $Q(m) > \chi_m^2(\alpha)$ or p-value is less than α .

- Sources of serial correlations in financial TS

– Nonsynchronous trading (ch. 5)

– Bid-ask bounce (ch. 5)

– Risk premium, etc. (ch. 3)

Thus, significant sample ACF does not necessarily imply market inefficiency.

Example: Monthly returns of IBM stock from 1926 to 1997.

- R_t : $Q(5) = 5.4(0.37)$ and $Q(10) = 14.1(0.17)$
- r_t : $Q(5) = 5.8(0.33)$ and $Q(10) = 13.7(0.19)$

Remark: What is p-value? How to use it?

Implication: Monthly IBM stock returns do not have significant serial correlations.

Example: Monthly returns of CRSP value-weighted index from 1926 to 1997.

- R_t : $Q(5) = 27.8$ and $Q(10) = 36.0$
- r_t : $Q(5) = 26.9$ and $Q(10) = 32.7$

All highly significant. Implication: there exist significant serial correlations in the equal-weighted index returns. (Nonsynchronous trading might explain the existence of the serial correlations, among other reasons.)

R demonstration: Monthly IBM returns from 1926 to 1997.

```
> library('fSeries')
> ibm=read.table("m-ibm2697.txt")
> acf(ibm,lag.max=15)
> x1=acf(ibm,lag.max=15)
> names(x1)
[1] "acf"      "type"     "n.used"  "lag"     "series"  "snames"
> x1$acf
, , 1
      [,1]
[1,]  1.000000000
[2,]  0.074250386
[3,]  0.010515948
[4,] -0.023832213
[5,] -0.006133517
.... (editted)
```

```

[13,] 0.001143822
[14,] -0.057881530
[15,] -0.056303333
[16,] -0.021747045

> x2=pacf(ibm,lag.max=15)
> names(x2)
[1] "acf"      "type"      "n.used"    "lag"       "series"    "snames"
> x2$acf
, , 1
      [,1]
[1,] 0.0742503857
[2,] 0.0050305627
[3,] -0.0251217508
[4,] -0.0025929619
[5,] -0.0061669432
... (edited)
[13,] -0.0562812983
[14,] -0.0467710983
[15,] -0.0124448122

> Box.test(ibm,lag=5,type="Ljung")

      Box-Ljung test

data:  ibm
X-squared = 5.4474, df = 5, p-value = 0.3638

> Box.test(log(ibm+1),lag=5,type="Ljung")

      Box-Ljung test

data:  log(ibm + 1)
X-squared = 5.7731, df = 5, p-value = 0.3289

```

Plus demonstration

```

> ibm=scan(file='m-ibm2697.txt') % Load data
> autocorTest(ibm,lag=5)          % Perform Q(5) test

```

Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation

Test Statistics:

```
Test Stat 5.4474
p.value 0.3638
```

```
Dist. under Null: chi-square with 5 degrees of freedom
Total Observ.: 864
```

```
> ibm=log(ibm+1)      % Convert into log returns
> autocorTest(ibm,lag=5)
```

```
Test Statistics:
Test Stat 5.7731
p.value 0.3289
```

```
Dist. under Null: chi-square with 5 degrees of freedom
```

SCA Demonstration: Output edited.

```
input ibm. file 'm-ibm2697.txt'  % Load data
--
acf ibm. maxl 10.  % Compute 10 lags of ACF.
```

NAME OF THE SERIES	IBM
TIME PERIOD ANALYZED	1 TO 864
MEAN OF THE (DIFFERENCED) SERIES . . .	0.0142
STANDARD DEVIATION OF THE SERIES . . .	0.0670
T-VALUE OF MEAN (AGAINST ZERO)	6.2246

AUTOCORRELATIONS

1- 10	.07	.01	-.02	-.01	-.01	-.01	-.00	.07	.05	.04	% ACF
ST.E.	.03	.03	.03	.03	.03	.03	.03	.03	.03	.03	% Stan. error
Q	4.8	4.9	5.4	5.4	5.4	5.5	5.5	10.2	12.6	14.1	% Ljung-Box Q

```
--
p=1-cdfc(5.4,5)  % Calculate p-value
--
print p          % Print p-value
```

.369

Back-shift (lag) operator

A useful notation in TS analysis.

- Definition: $Br_t = r_{t-1}$ or $Lr_t = r_{t-1}$
- $B^2r_t = B(Br_t) = Br_{t-1} = r_{t-2}$.

B (or L) means time shift! Br_t is the value of the series at time $t - 1$.

Suppose that the daily log returns are

Day	1	2	3	4
r_t	0.017	-0.005	-0.014	0.021

Answer the following questions:

- $r_2 =$
- $Br_3 =$
- $B^2r_5 =$

Question: What is B^2 ?

What are the important statistics in practice?

Conditional quantities, not unconditional

A proper perspective: at a time point t

- Available data: $\{r_1, r_2, \dots, r_{t-1}\} \equiv F_{t-1}$

- The return is decomposed into two parts as

$$\begin{aligned} r_t &= \text{predictable part} + \text{not predic. part} \\ &= \text{function of elements of } F_{t-1} + a_t \end{aligned}$$

In other words, given information F_{t-1}

$$\begin{aligned} r_t &= \mu_t + a_t \\ &= E(r_t | F_{t-1}) + \sigma_t \epsilon_t \end{aligned}$$

- μ_t : conditional mean of r_t
- a_t : shock or innovation at time t
- ϵ_t : an iid sequence with mean zero and variance 1
- σ_t : conditional standard deviation (commonly called volatility in finance)

Traditional TS modeling is concerned with μ_t :

Model for μ_t : **mean equation**

Volatility modeling concerns σ_t .

Model for σ_t^2 : **volatility equation**

Univariate TS analysis serves two purposes

- a model for μ_t
- understanding models for σ_t^2 : properties, forecasting, etc.

Linear time series: r_t is linear if

- the predictable part is a linear function of F_{t-1}
- $\{a_t\}$ are indep. and have the same dist. (iid)

Mathematically, it means r_t can be written as

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i},$$

where μ is a constant, $\psi_0 = 1$ and $\{a_t\}$ is an iid sequence with mean zero and well-defined distribution.

In the economic literature, a_t is the *shock* (or *innovation*) at time t and $\{\psi_i\}$ are the *impulse* responses of r_t .

White noise: iid sequence (with finite variance), which is the building block of linear TS models.

White noise is not predictable, but has zero mean and finite variance.

Univariate linear time series models

1. autoregressive (AR) models
2. moving-average (MA) models
3. mixed ARMA models
4. seasonal models
5. regression models with time series errors
6. fractionally differenced models (long-memory)

Example Quarterly growth rate of U.S. real gross national product (GNP), seasonally adjusted, from the second quarter of 1947 to the first quarter of 1991.

An AR(3) model for the data is

$$r_t = 0.005 + 0.35r_{t-1} + 0.18r_{t-2} - 0.14r_{t-3} + a_t, \quad \hat{\sigma}_a = 0.01,$$

where $\{a_t\}$ denotes a white noise with variance σ_a^2 . Given r_n, r_{n-1} & r_{n-2} , we can predict r_{n+1} as

$$\hat{r}_{n+1} = 0.005 + 0.35r_n + 0.18r_{n-1} - 0.14r_{n-2}.$$

Other implications of the model?

Example: Monthly simple return of CRSP equal-weighted index

$$R_t = 0.013 + a_t + 0.178a_{t-1} - 0.13a_{t-3} + 0.135a_{t-9}, \quad \hat{\sigma}_a = 0.073$$

Checking: $Q(10) = 11.4(0.122)$ for the residual series a_t .

Implications of the model?

Important properties of a model

- Stationarity condition
- Basic properties: mean, variance, serial dependence
- Empirical model building: specification, estimation, & checking
- Forecasting