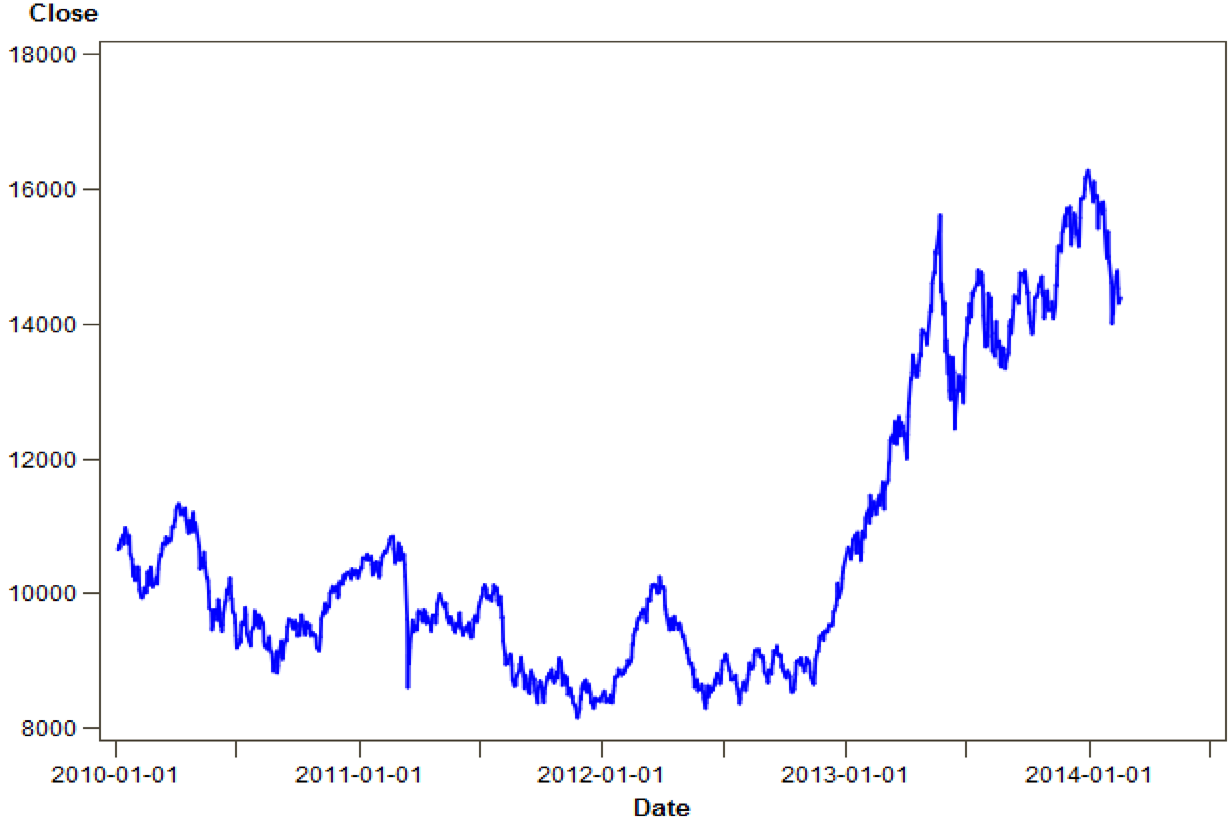
# Financial Time Series

Lecture 1

# What's this?


Close
18000
16000
14000
12000
10000
8000
2010-01-01
2011-01-01
2012-01-01
2013-01-01
2014-01-01
Date

# It’s

- The Nikkei 225
- Visualization of a typical financial time series
- What happened march 2011?
- Possible to predict?

# Financial return data

- Given the price of an asset $P_t$ at time $t$ we are often interested in modeling its ”returns”

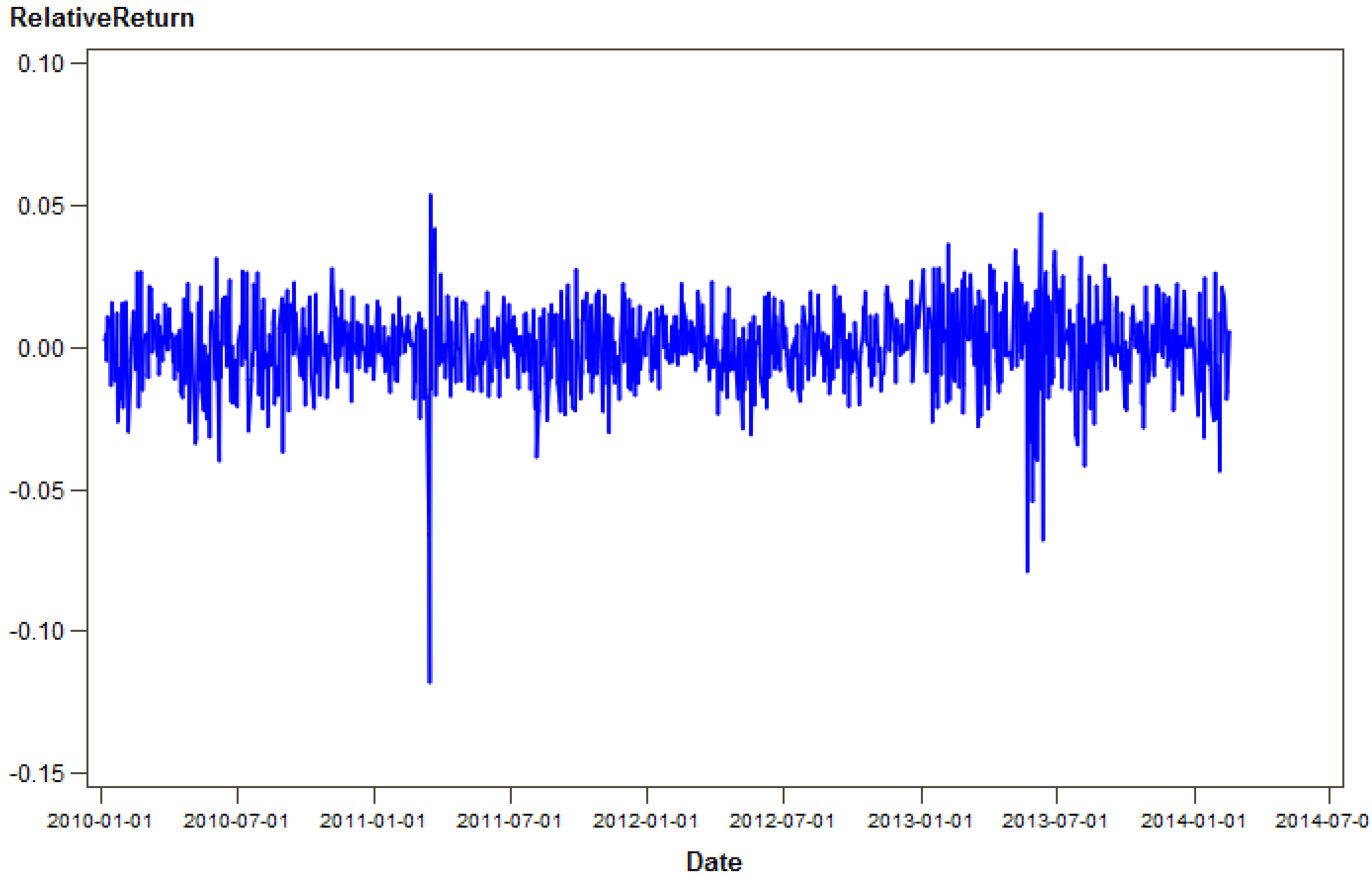
- We may consider ”absolute returns”

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

- Or ”relative (or simple) returns”

$$\tilde{R}_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

# ^N225 Relative Returns



# Financial return data

- Or ”log-returns”

$$r_t = ln\frac{P_t}{P_{t-1}}$$

- Log-returns are commonly used since it is often assumed (e.g., Black-Scholes) that the price process is of the type

$$P_t = e^{X_t}$$

# Black-Scholes

- In the asset price model used in the Black-Scholes option pricing theory it is assumed that

$$P_t = P_0 exp\{\mu t + \sigma W_t\}$$

- Here $\{W_t\}$ is a standard Wienerprocess, so that

$$r_t \sim N(\mu, \sigma)$$

# Stylized facts of log-returns

- However, in practice the distribution of log-returns

- Has more observations close to zero
- Has heavier tails (leptokurtosis)
- Is skewed, typically to the left

# Ways to examine deviations from normality

- Histogram
- Normality tests (Anderson-Darling, Kolmogorov-Smirnov)
- Testing for skewness
- Testing for (excess) kurtosis

# Sample statistics

- We have observed $x_1, \ldots, x_T$
- Sample mean

$$\hat{\mu}_x = \frac{1}{T} \sum_{i=1}^{T} x_i$$

- Sample variance

$$\hat{\sigma}^2{}_x = \frac{1}{T-1} \sum_{i=1}^{T} (x_i - \hat{\mu}_x)^2$$

# Sample statistics

- Sample skewness

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

- Sample kurtosis

$$\widehat{K}(x) = \frac{1}{(T-1)\hat{\sigma}^4{}_x} \sum_{i=1}^{T} (x_i - \hat{\mu}_x)^4$$

# Excess kurtosis and tests

- The normal distribution has kurtosis 3 so by ”excess kurtosis” we mean $K(x) - 3$

- To test for skewness and kurtosis we may use that $\hat{S}(x)$ and $\widehat{K}(x) - 3$ are asymptotically normal with mean zero and variances $6/T$ and $24/T$,respectively under the assumption of normal data

# Jarque-Bera test

- We may test both skewness and kurtosis at once, using that

$$\frac{\hat{S}^2(x)}{6/T} + \frac{\left(\widehat{K}(x) - 3\right)^2}{24/T}$$

is asymptotically ${\chi^2}_2$ under the assumption of normal data

**Table 1.2. Descriptive Statistics for Daily and Monthly Simple and Log Returns of Selected Indexes and Stocks[a]**

| Security | Start | Size | Mean | Standard Deviation | Skewness | Excess Kurtosis | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | *Daily Simple Returns (%)* | | | | |
| SP | 62/7/3 | 10446 | 0.033 | 0.945 | −0.95 | 25.76 | −20.47 | 9.10 |
| VW | 62/7/3 | 10446 | 0.045 | 0.794 | −0.76 | 18.32 | −17.14 | 8.66 |
| EW | 62/7/3 | 10446 | 0.085 | 0.726 | −0.89 | 13.42 | −10.39 | 6.95 |
| IBM | 62/7/3 | 10446 | 0.052 | 1.648 | −0.08 | 10.21 | −22.96 | 13.16 |
| Intel | 72/12/15 | 7828 | 0.131 | 2.998 | −0.16 | 5.85 | −29.57 | 26.38 |
| 3M | 62/7/3 | 10446 | 0.054 | 1.465 | −0.28 | 12.87 | −25.98 | 11.54 |
| Microsoft | 86/3/14 | 4493 | 0.157 | 2.505 | −0.25 | 8.75 | −30.12 | 19.57 |
| Citi-Group | 86/10/30 | 4333 | 0.110 | 2.289 | −0.10 | 6.79 | −21.74 | 20.76 |
| | | | | *Daily Log Returns (%)* | | | | |
| SP | 62/7/3 | 10446 | 0.029 | 0.951 | −1.41 | 36.91 | −22.90 | 8.71 |
| VW | 62/7/3 | 10446 | 0.041 | 0.895 | −1.06 | 23.91 | −18.80 | 8.31 |
| EW | 62/7/3 | 10446 | 0.082 | 0.728 | −1.29 | 14.70 | −10.97 | 6.72 |
| IBM | 62/7/3 | 10446 | 0.039 | 1.649 | −0.25 | 12.60 | −26.09 | 12.37 |
| Intel | 72/12/15 | 7828 | 0.086 | 3.013 | −0.54 | 7.54 | −35.06 | 23.41 |
| 3M | 62/7/3 | 10446 | 0.044 | 1.469 | −0.69 | 20.06 | −30.08 | 10.92 |
| Microsoft | 86/3/14 | 4493 | 0.126 | 2.518 | −0.73 | 13.23 | −35.83 | 17.87 |
| Citi-Group | 86/10/30 | 4333 | 0.084 | 2.289 | −0.21 | 7.47 | −24.51 | 18.86 |

# Testing for skewness, example

- If we consider the descriptives found for Intel, we see that for $T = 7828$ observations of daily simple returns the skewness is $\hat{S}(r) = -0.16$

- The observed value of the test statistic is

$$\frac{-0.16}{\sqrt{6/7828}} \approx -5.78$$

- The p-value for the two-sided test is 0.0000… so clearly there is (negative) skewness

# Testing for excess kurtosis, example

- If we consider the descriptives found for Citi-Group, we see that for $T = 4333$ observations of daily simple returns the excess kurtosis is $\widehat{K}(r) - 3 = 6.79$

- The observed value of the test statistic is

$$\frac{6.79}{\sqrt{24/4333}} \approx 91.23$$

- The p-value for the one-sided test for excess kurtsosis is 0.0000… so clearly there is excess kurtosis

# Distributions for returns

- In many models it is assumed that returns follow normal distributions

- This of course makes the models mathematically tractable

- But, as seen above, returns are typically not normal

# Distributions for returns

- There are distributions that can handle skewness and positive excess kurtosis
- Using moment of methods or maximum likelihood we are able to fit such distributions
- Below is one example of a distribution more ”flexible” than the normal

# NIG distribution

- The Normal Inverse Gaussian distribution may be used for modeling returns and allows for excess kurtosis and skewness

- Density is

$$f(x;\alpha,\beta,\delta,\mu)$$
$$= \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\pi\sqrt{\delta^2 + (x-\mu)^2}} e^{\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)}$$

# NIG distribution

- $K_1$ is the modified Bessel function of the second kind (of order 1)

- The parameters $\alpha, \beta, \delta, \mu$ are ”tail heaviness”, asymmetry, scale and location

- May be estimated in matlab using the ”NIG package” by Ralf Werner available at online at matlabcentral.com

# Parameter relations

- Letting $\gamma = \sqrt{\alpha^2 - \beta^2}$ and $X \sim NIG(\alpha, \beta, \delta, \mu)$, we have

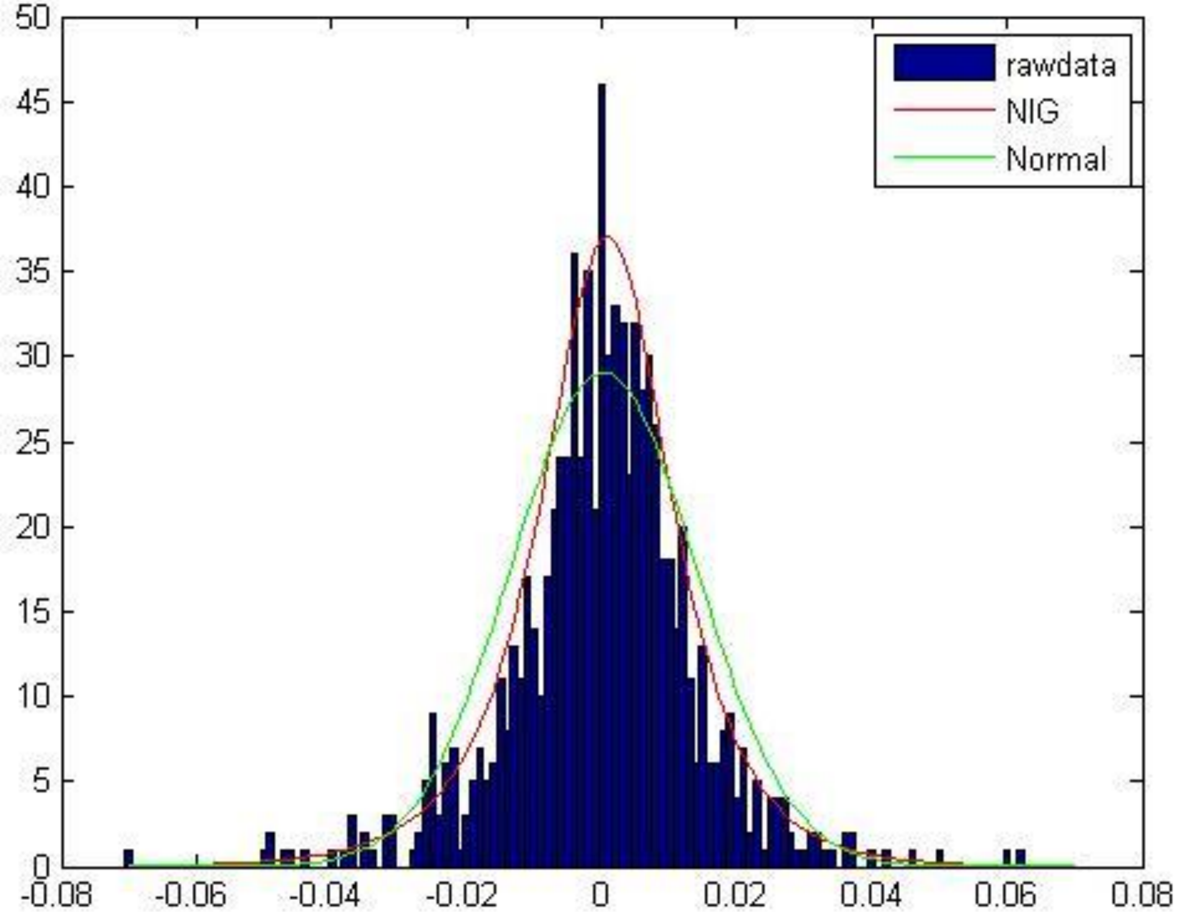
- $E[X] = \mu + \delta\beta/\gamma$
- $Var(X) = \delta\alpha^2/\gamma^3$
- $S(X) = 3\beta/\left(\alpha\sqrt{\delta\gamma}\right)$
- $K(X) = 3 + 3(1 + 4\beta^2/\alpha^2)/(\delta\gamma)$

# Example, OMXS30 log return density

- Using the matlab commands mean, var, skewness and kurtosis, we may find the NIG parameters using the nigpar command (Werner)

- For OMXS30 log returns (100104-130414), we find that $\alpha = 77.47, \beta = -5.307, \delta = 0.0145, \mu = 0.0012$

# NIG fit for OMXS30 log returns



# Your typical time series model for returns

- We modeling returns in the time series or econometrics frame work we often assume that

$$r_t = \mu + \sigma_t \varepsilon_t$$
$$\sigma_t = f(r_{t-1}, \sigma_{t-1})$$

- Here $\{\varepsilon_t\}$ is an i.i.d series and the dependence structure and in some cases skewnees and excess kurtosis is captured in the volatility series $\{\sigma_t\}$

# Basic time series models

- The most simple time series models are moving averages (MA) and auto regressive (AR) series

MA: $r_t = \mu + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_n \varepsilon_{t-n}$

AR: $r_t = \beta_0 + \beta_1 r_{t-1} + \cdots + \beta_n r_{t-n} + \varepsilon_t$

# Basic time series

- In most cases we refer to a times series as a collection of random variables $\{r_t\}_{t \in \mathbb{N}}$, where the index set is the natural numbers
- We define a mean function $\mu_t = E(r_t)$
- And an autocovariance function $\gamma_{s,t} = Cov(r_s, r_t)$

# Basic time series

- We say that $\{r_t\}_{t\in\mathbb{N}}$ is (*weakly) stationary* if

$$E\left({r_t}^2\right) \text{ is finite}$$

$$\mu_t \text{ is constant}$$

and

$$\gamma_{t+h,t} \text{ does not depend on } t \text{ for any } h.$$

# Example

- Let $X_t = Acos(\omega t + \Theta)$, where $A \sim Exp(1)$ and $\Theta \sim Uni(0,2\pi)$ and independent

- Exercise: show (weak) stationarity for $\{X_t\}_{t \in \mathbb{N}}$

# Autocorrelation function (ACF)

- For a (weakly) stationary series $\{r_t\}_{t \in T}$ we may define

$$\rho_h = \frac{\gamma_h}{\gamma_0}$$

where $\gamma_h = \gamma_{t+h,t}$

and $\gamma_0 = Var(r_t)$ (by definition)

# Sample autocorrelation

- We define the sample autocorrelation at lag $h$ for (the observed sample) $\{r_1, \dots, r_T\}$ as

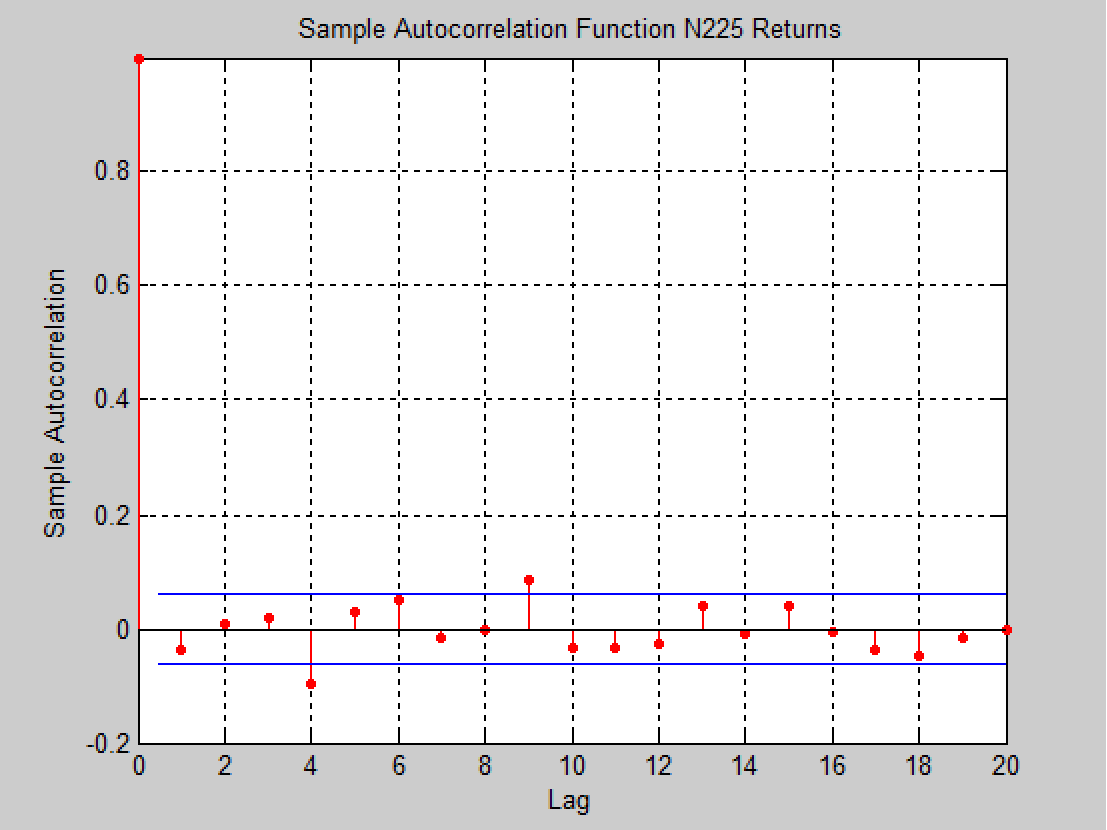
$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0},$$

where $\hat{\gamma}_h = \frac{1}{T}\sum_{j=1}^{T-h}(r_{j+h} - \bar{r})(r_j - \bar{r})$

# Sample autocorrelation

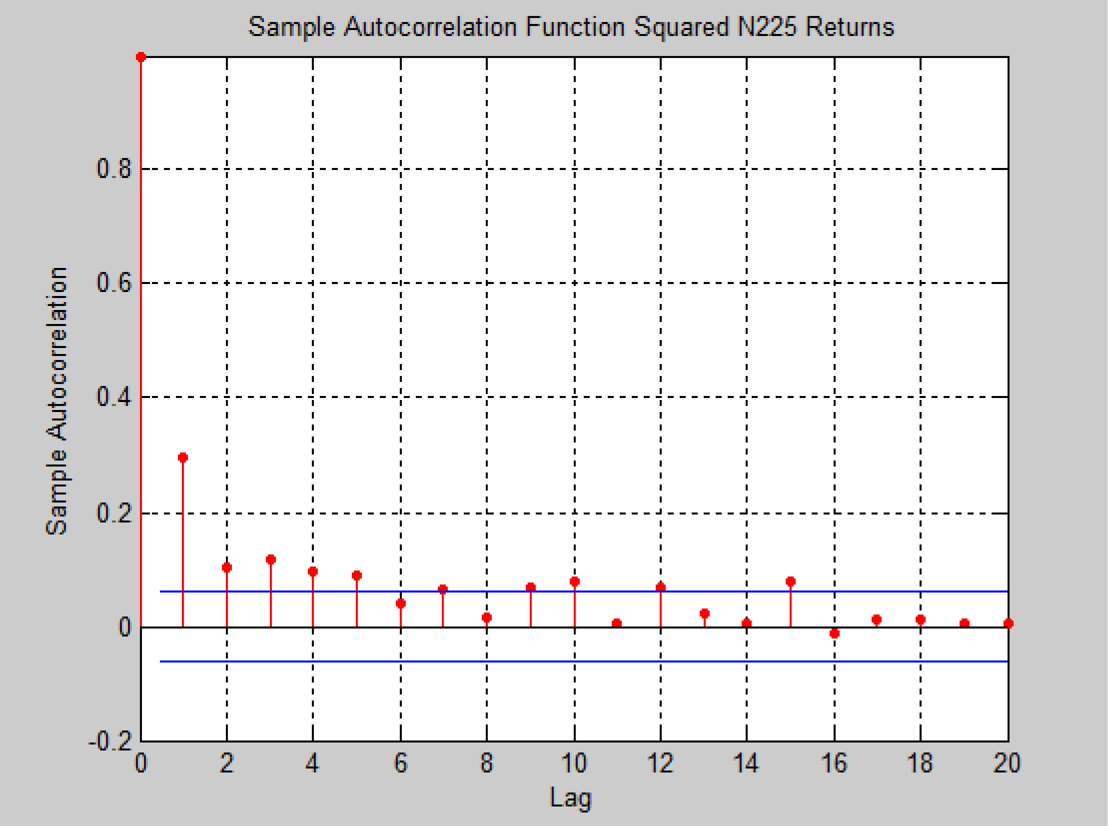
- Why bother?
- The sample autocorrelations function at different lags will give hints on how to model our time series
- In particular for AR and MA models
- Below we used autocorrelation to once again establish the well known fact that though a series financial returns may not exhibit autocorrelation the returns are (in most cases) not independent

# Sample autocorrelation




Using ”autocorr” in matlab we see that N225 returns do not exhibit significant autocorrelations at any lags between 1 and 20. 95% conifidence limits are computed under the assumption of Gaussian white noise, so limits are $\pm 1.96/\sqrt{T}$

# Sample autocorrelation




But the squared returns have significant positive autocorrelations at lags 1,2,3,4 and 5…

# Tests for autocorrelation

- We may be interested in testing if a return series $\{r_t\}$ exhibits significant autocorrelation at som lag $h$
- The hypotheses are

$$H_0: \rho_h = 0$$
$$H_a: \rho_h \neq 0$$

- And the test statistic

$$\frac{\hat{\rho}_h}{\sqrt{\left(1 + 2\sum_{i=1}^{h-1} \hat{\rho}_i{}^2\right)/T}}$$

Is asymptotically normal if $\{r_t\}$ is stationary Gaussian with $\rho_j = 0$ for $j > h$

# Tests for autocorrelation: Portmanteau

- We may be interested in testing

$$H_0: \rho_1 = \cdots = \rho_m = 0$$
$$H_a: \rho_i \neq 0, i \in \{1, \dots, m\}$$

- The test statistic

$$T \sum_{l=1}^{m} \hat{\rho}_l^{\,2}$$

is asymptotically $\chi^2{}_m$ if $\{r_t\}$ is i.i.d.

# Improved tests for autocorrelation: Ljung-Box

- Modificatio of Portmanteau for higher power
- The test statistic

$$T(T+2)\sum_{l=1}^{m}\frac{\hat{\rho}_l^{\,2}}{T-l}$$

is asymptotically $\chi^2{}_m$ if $\{r_t\}$ is i.i.d.

- ”lbqtest” in matlab

# Strict stationarity

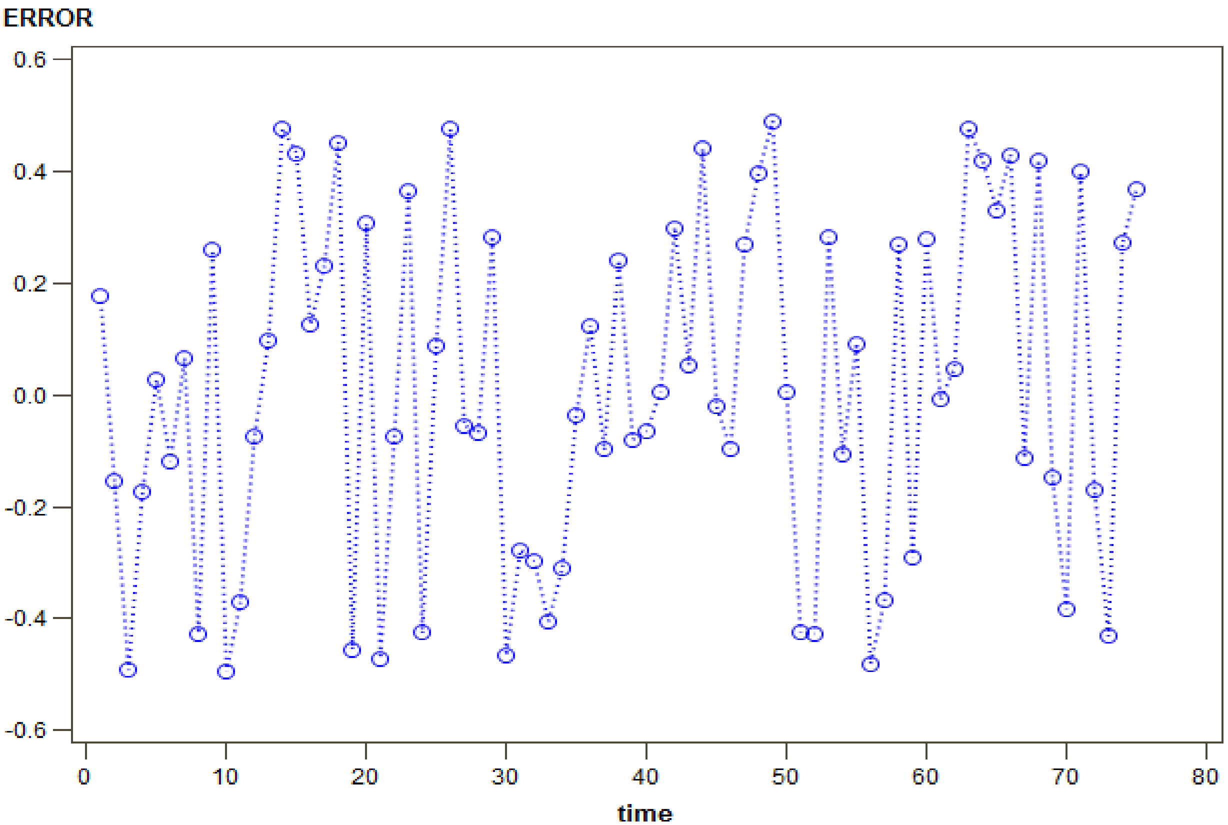
- By (strict) stationarity we mean a series $\{r_t\}_{t \in T}$ for which the finite dimensional distributions are the same at all lags, i.e. for any $i_1, \dots, i_n$ and $h$

$$P\left(r_{i_1} \leq x_{i_1}, \dots, r_{i_n} \leq x_{i_n}\right)$$

$$= P\left(r_{i_1+h} \leq x_{i_1}, \dots, r_{i_n+h} \leq x_{i_n}\right)$$

# Example

- Let $r_t$ be i.i.d. $\mathrm{Uni}(-0.5,0.5)$



# Relation

- We say that a time series is *Gaussian* if the finite dimensional distributions are multivariate normal

- A (weakly) stationary Gaussian time series is strictly stationary

- Why?

# Prediction theory for stationary processes

- One of our main tasks is to be able predict $r_{n+h}$ given observations of $r_1, \dots, r_n$

- To do this, we want to find the linear combination $\hat{r}_{n+h} = \alpha_0 + \alpha_1 r_1 + \cdots + \alpha_n r_n$ that minimizes

$$E[(\hat{r}_{n+h} - r_{n+h})^2]$$

# Prediction theory for stationary processes

- We may write
$$S(\alpha_0, \alpha_1, \dots, \alpha_n) = E[(\hat{r}_{n+h} - r_{n+h})^2]$$
and treat this as an ”ordinary minimization problem”

- So, we take partial derivatives w.r.t. $\alpha_0, \dots, \alpha_n$, set these derivatives to zero and solve the equations

# Prediction theory for stationary processes

- If $E[r_i] = \mu$, one will find the solutions

$$\alpha_0 = \mu\left(1 - \sum_{i=1}^{n} \alpha_i\right)$$

and

$$[\alpha_1 \cdots \alpha_n]^T = \Gamma^{-1}\gamma$$

where $\Gamma_{ij} = \gamma_{i-j}$ for $1 \leq i, j \leq n$ and

$$\gamma = [\gamma_h \cdots \gamma_{h+n-1}]^T$$

# Prediction theory for stationary processes

- This gives that

$$\hat{r}_{n+h} = \mu + \sum_{i=1}^{n} \alpha_i \, (r_i - \mu)$$

- The prediction error is

$$E[(\hat{r}_{n+h} - r_{n+h})^2]$$

$$= \gamma_0 - (\alpha_1 \gamma_h + \cdots + \alpha_n \gamma_{h-n+1})$$

# More general

- Assume that $Y$ and $W_1, \ldots, W_n$ are any r.v.'s with finite second moment with means $\mu_Y$ and $\mu_i$

- Assume also that the covariances

$$\gamma^T = [Cov(Y, W_1) \cdots Cov(Y, W_n)]$$

and $\Gamma$ where $\Gamma_{ij} = Cov(W_i, W_j)$, for $1 \leq i, j \leq n$ are known

# Prediction operator

- We may define the prediction operator $P(\cdot \mid \cdot)$ as

$$P(Y|W_1, \dots, W_n) = \mu_Y + \sum_{i=1}^{n} \alpha_i (W_i - \mu_i)$$

where

$$[\alpha_1 \cdots \alpha_n]^T = \Gamma^{-1}\gamma$$

# Properties

- MSE:

$$E\left[\left(Y - P(Y|W_1, \dots, W_n)\right)^2\right] = Var(Y) - \sum_{i=1}^{n} \alpha_i \gamma_i$$

- Linearity:

$$P(a_1 Y_1 + a_2 Y_2 | W_1, \dots, W_n) =$$

$$a_1 P(Y_1 | W_1, \dots, W_n) + a_2 P(Y_2 | W_1, \dots, W_n)$$

# Properties

- Orthogonality:

$$\sum_{i=1}^{n} E\left[\left(Y - P(Y|W_1, \dots, W_n)\right)W_i\right] = 0$$

- Unbiasedness:

$$E\left[\left(Y - P(Y|W_1, \dots, W_n)\right)\right] = 0$$

# Example, missing value

- For an AR(1) process defined by

$$r_t = \varphi r_{t-1} + \varepsilon_t$$

and have observed $r_1$ and $r_3$ but $r_2$ is missing

- To estimate $r_2$, we want to determine $P(r_2|r_1, r_3)$

# Example

- We know that

$$\gamma^T = [Cov(r_2, r_1) \;\; Cov(r_2, r_3)] = \frac{\sigma^2}{1-\varphi^2}[\varphi \; \varphi]$$

and

$$\Gamma = \frac{\sigma^2}{1-\varphi^2}\begin{bmatrix} 1 & \varphi^2 \\ \varphi^2 & 1 \end{bmatrix}$$

# Example

- This means that

$$[\alpha_1 \ \ \alpha_3]^T = \Gamma^{-1}\gamma = \frac{1}{1-\varphi^4}\begin{bmatrix} 1 & -\varphi^2 \\ -\varphi^2 & 1 \end{bmatrix}\begin{bmatrix} \varphi \\ \varphi \end{bmatrix} = \frac{1}{1+\varphi^2}\begin{bmatrix} \varphi \\ \varphi \end{bmatrix}$$

- So that

$$P(r_2|r_1, r_3) = \frac{\varphi}{1+\varphi^2}(r_1 + r_3)$$

# Example

- The MSE is

$$Var(r_2) - \alpha_1 Cov(r_2, r_1) - \alpha_3 Cov(r_2, r_3)$$

$$= \frac{\sigma^2}{1-\varphi^2} - 2\frac{\varphi}{1+\varphi^2}\frac{\sigma^2\varphi}{1-\varphi^2} = \frac{\sigma^2}{1+\varphi^2}$$