Financial data

Sample statistics

Liutauras Petrucionis

Department of Economics
University of Essex

November 2011

Outline

Problems in financial data sets

- Financial data tends to be non-stationary and serially correlated.
- OLS regression requires iid data.

Arithmetic returns

Definition

$r_i = \frac{P_t + D_{t-1}}{P_{t-1}}$ r_i- returns at i, D_i - dividends at i, P_i - price at i

- Arithmetic returns are used for the analysis involving cross-sectional methodology, such as portfolio formation.

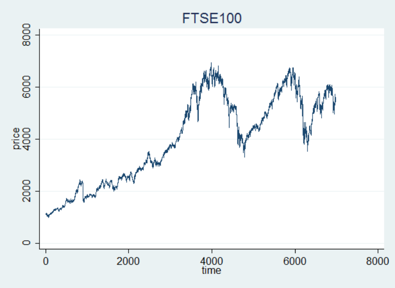
Logarithmic returns

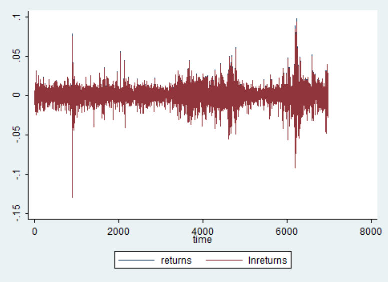
Definition

$r_i = ln(\frac{P_t + D_{t-1}}{P_{t-1}}) = ln(P_t + D_{t-1}) - ln(P_{t-1})$ r_i- returns at i, D_i - dividends at i, P_i - price at i

- Logarithmic returns are used for the analysis involving time dimension, such as returns over several periods of time.

Visual comparison



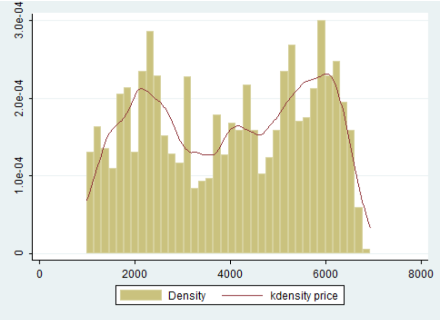


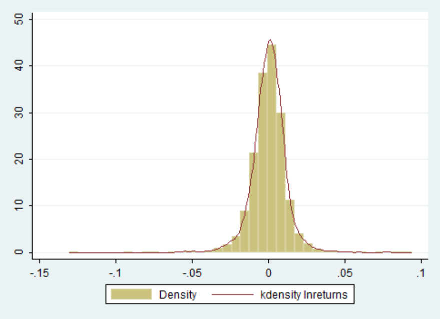
- Price data is non-stationary, while both returns transformations are almost identical and stationary.

Implications

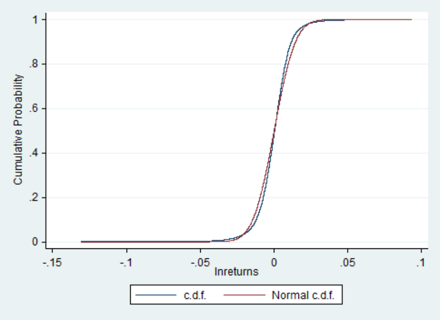
- If returns are normally distributed then prices are log-normally distributed.
- Normally distributed returns do not violate limited liability assumption.
- Data is stationary distributed around the mean, with varying variance.

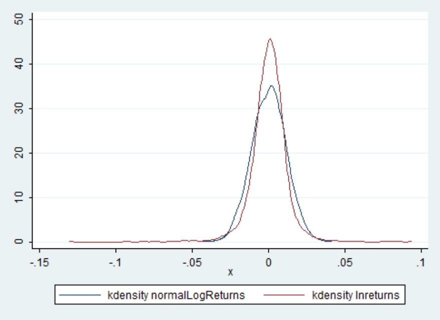
Sample distribution





Comparing with Gaussian distribution





- Skewness-Kurtosis test for Normality can be used for numerical results (sktest command). Test is looking for evidence to reject the null hypothesis of normality.

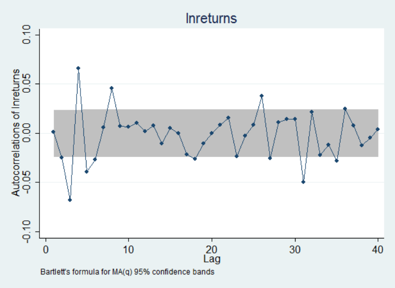
Simple sample statistics

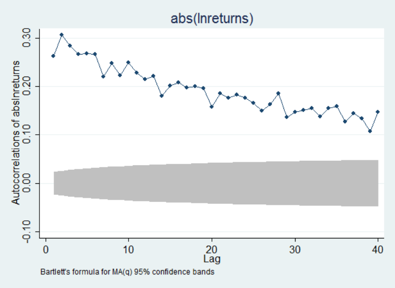
- Data description section should include a table summarizing simple sample statistics.

Example

Log returns statistics	
Mean	.00023
Variance	.00013
Skewness	-0.38
Kurtosis	11.38
Min	-0.130
Max	.094
Obs	6976

Auto-correlation function





Stylized facts of daily returns

- The distribution of returns exhibits fat-tails.
- The distribution of returns is asymmetrically distributed.
- In general returns do not exhibit linear correlation.
- Volatility (absolute returns) exhibit long-range dependency (slow decay rate of the auto-correlation function).

Remarks

- OLS methodology assumes that data is independently and identically distributed, but in previous slides it was displayed that data is heteroskedastic, serially correlated and not independently distributed.
- Normal distribution assumption does not hold as was displayed in kernel estimation graphs.