Introduction to Financial Engineering

Week 43: Basic Data Analysis

Nina Lange

Management Science, DTU

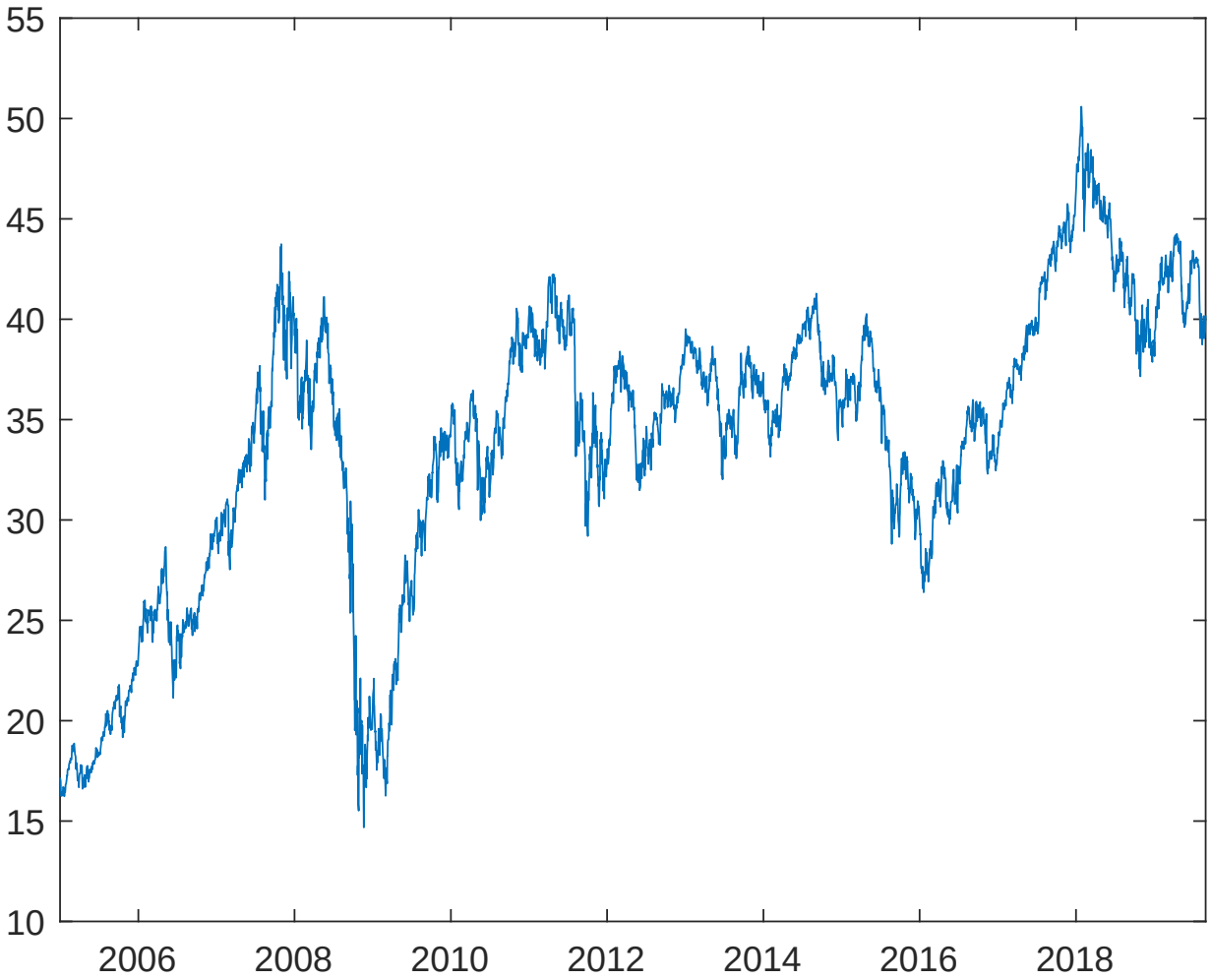
Week 43

Using financial data

- We wish to use financial data to make financial decisions based on given criteria
- The distribution of data provide information about potential risks and possible rewards
- The objective is to say something about the distribution of a financial variable in a day, a week, a month (or every future time point)
- A starting point is to look at historical data and assume that the future is well described by the past
- In later courses, projects or in real life, things needs to be a bit more sophisticated
- But remember; the theory on efficient frontier, CML etc. does not hinge on how estimates were obtained. It just says that if we assume certain mean and covariances, we know how to compose optimal portfolios.

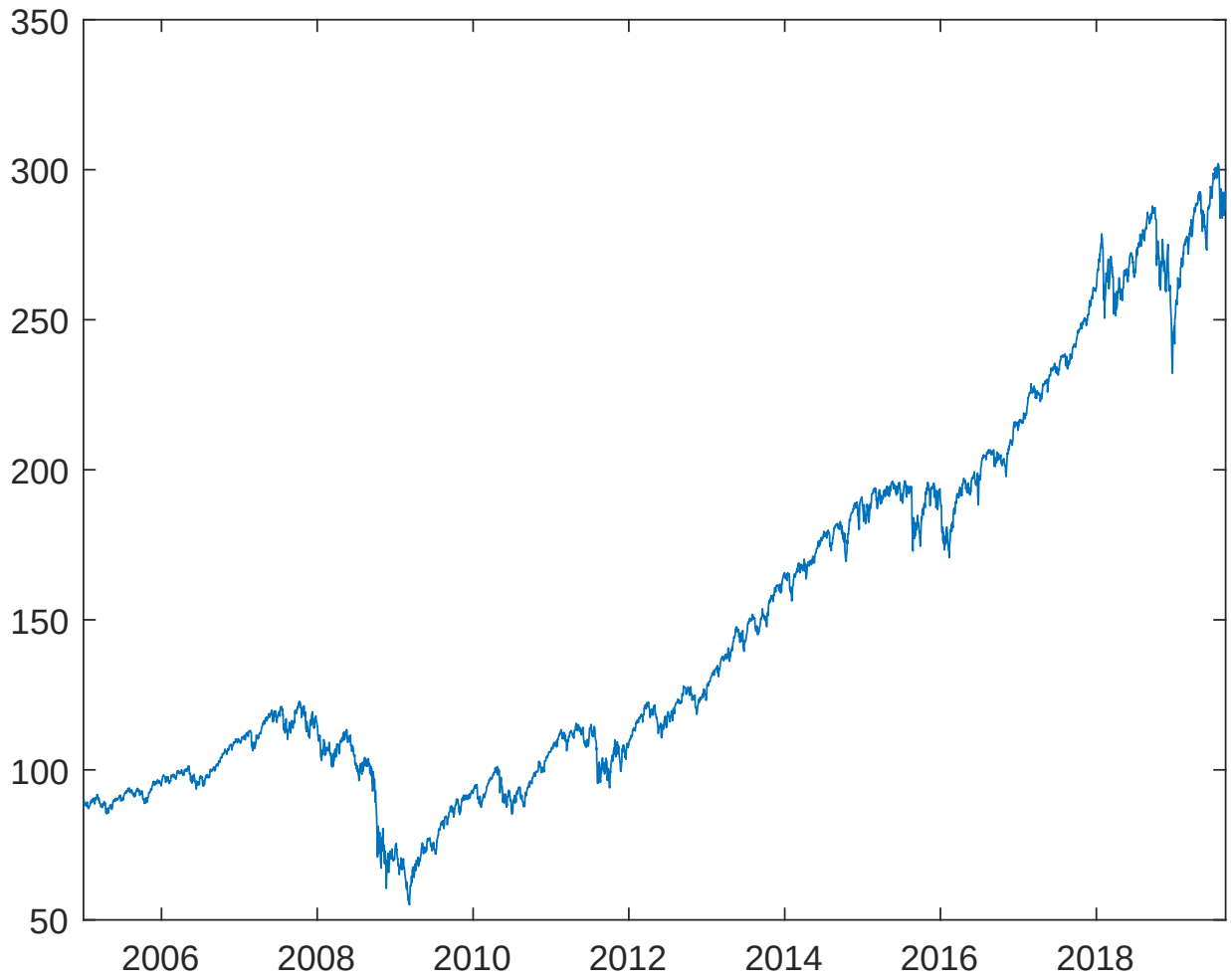
Plot of data

Example (iShares MSCI Emerging Markets ETF (EEM))



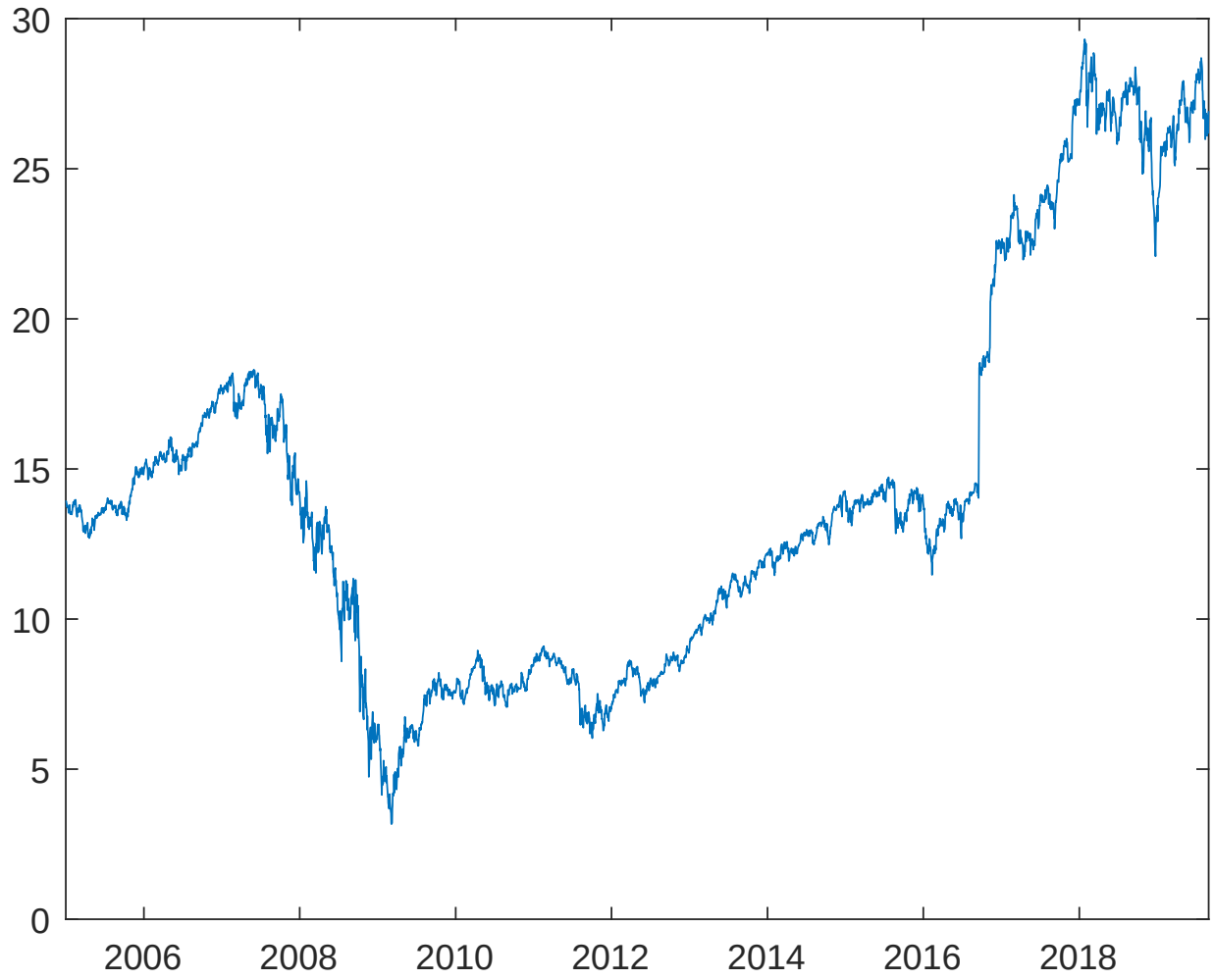
Plot of data

Example (SPDR S&P 500 trust (SPY))



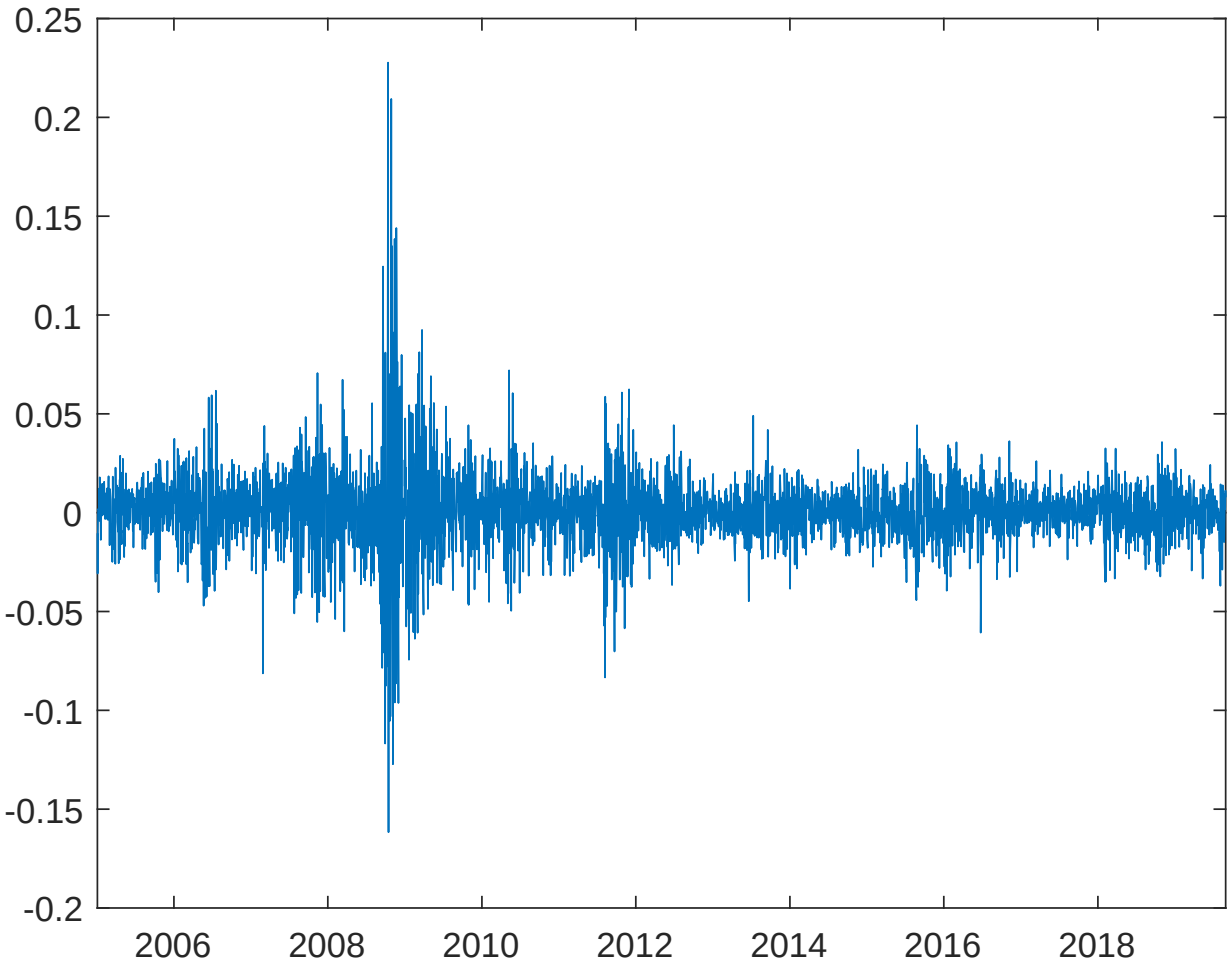
Plot of data

Example (The Financial Select Sector SPDR Fund (XLF))



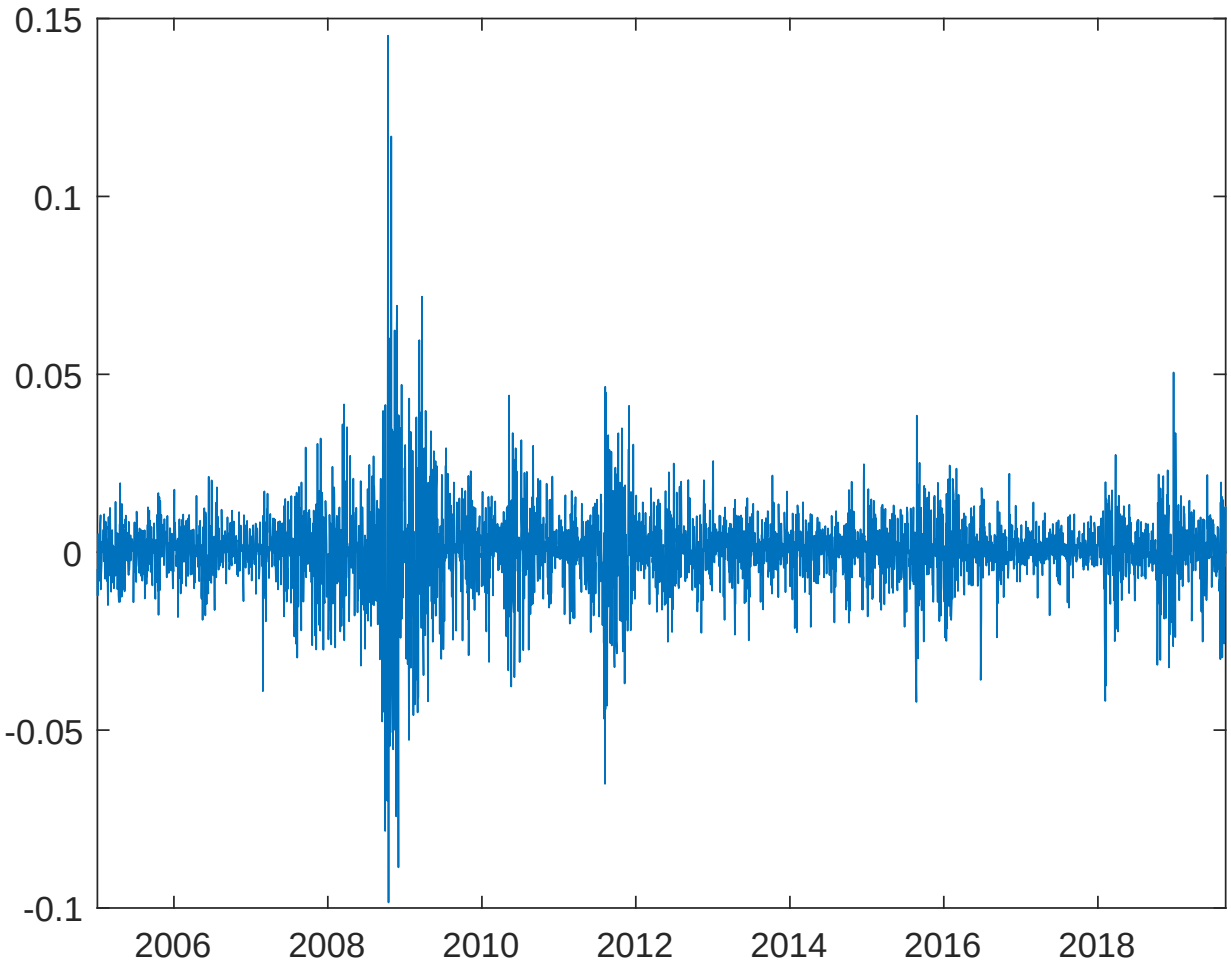
Plot of returns

Example (iShares MSCI Emerging Markets ETF (EEM))



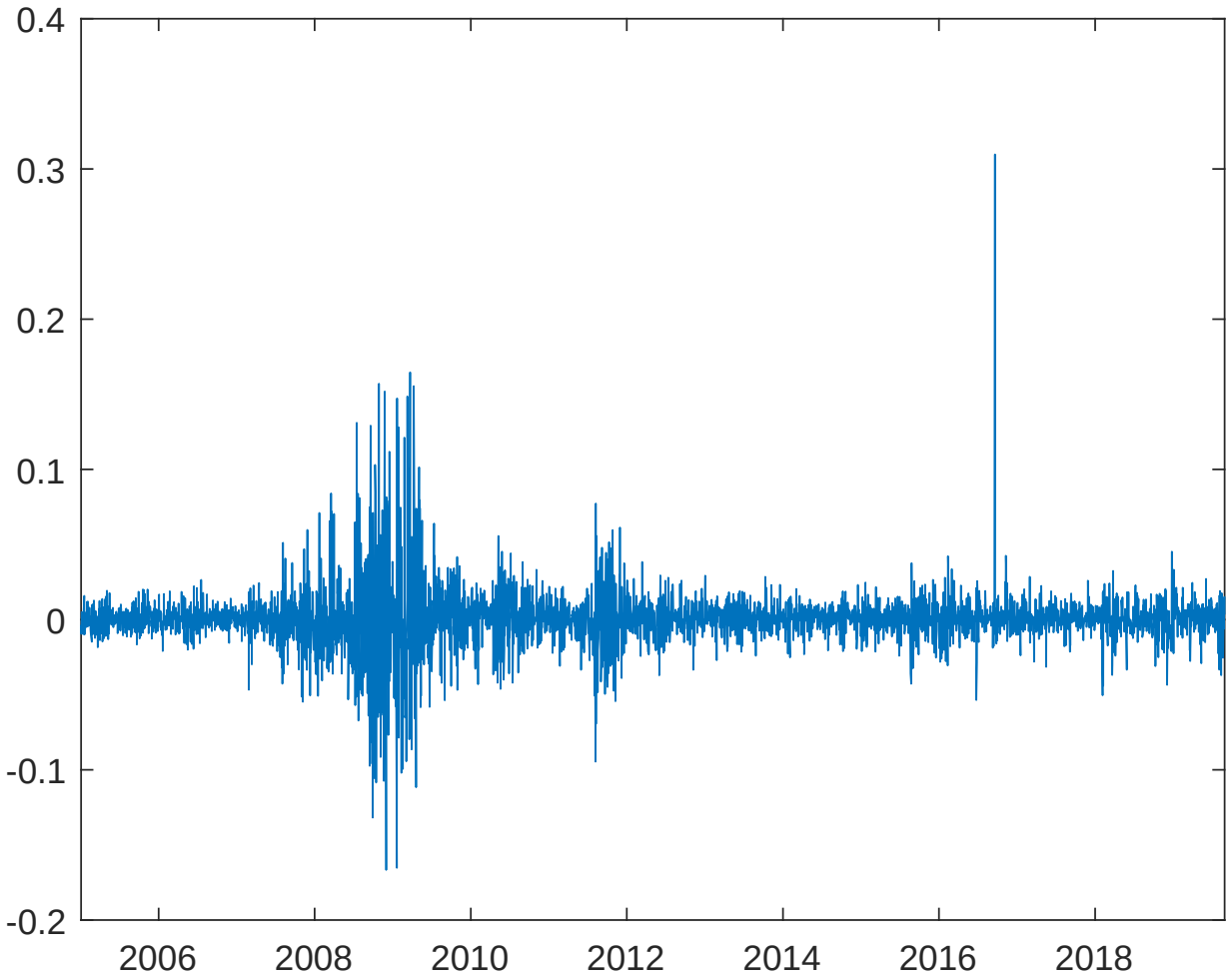
Plot of returns

Example (SPDR S&P 500 trust (SPY))



Plot of returns

Example (The Financial Select Sector SPDR Fund (XLF))



Histogram

Example (The Financial Select Sector SPDR Fund (XLF))

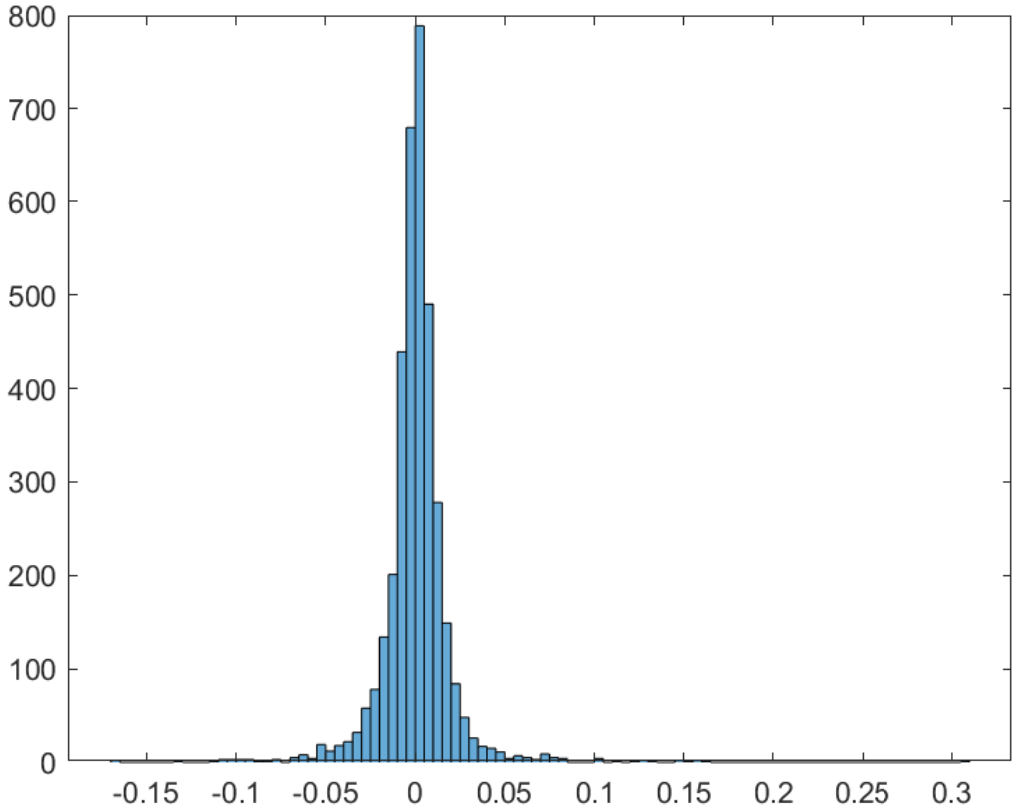
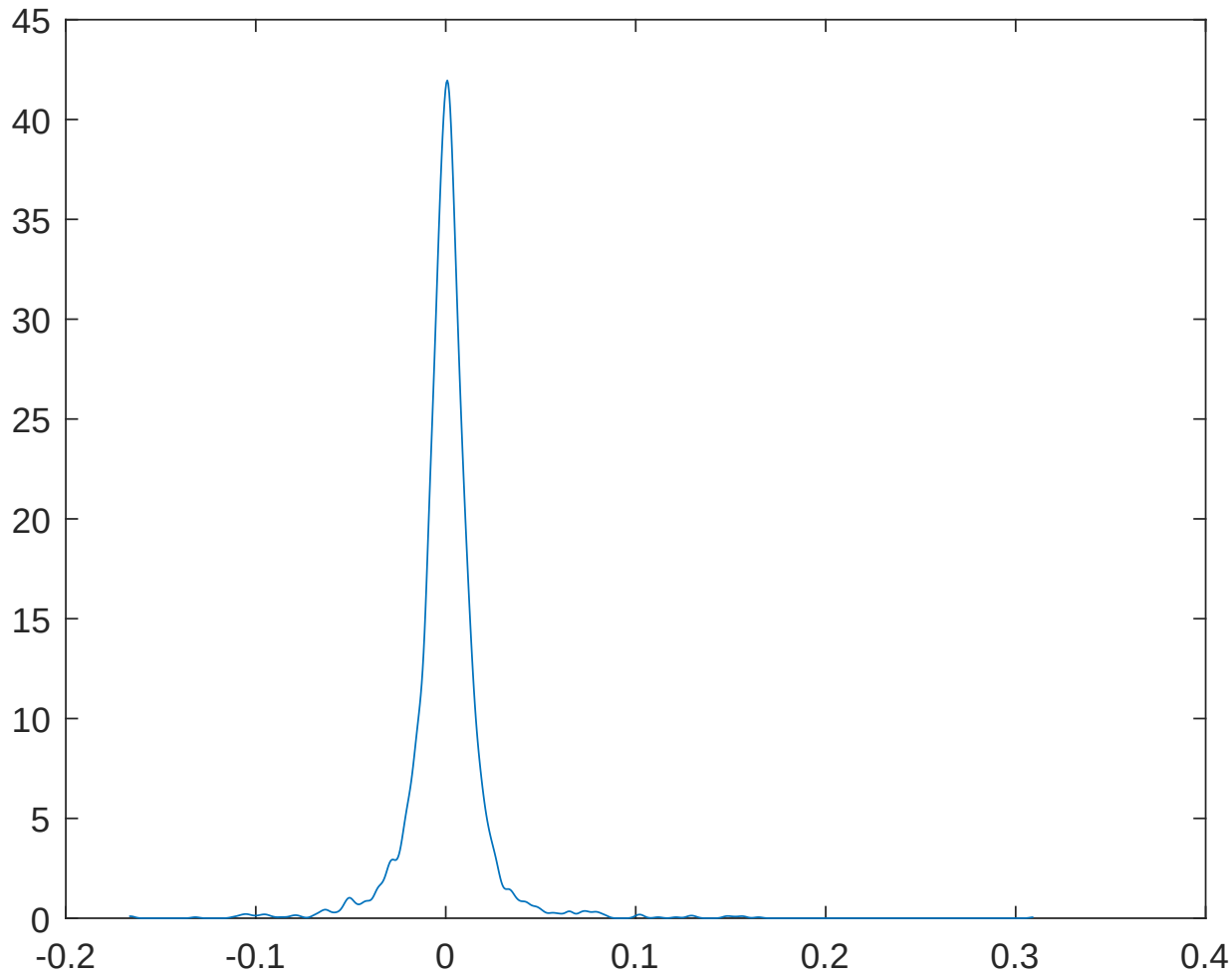


Figure:

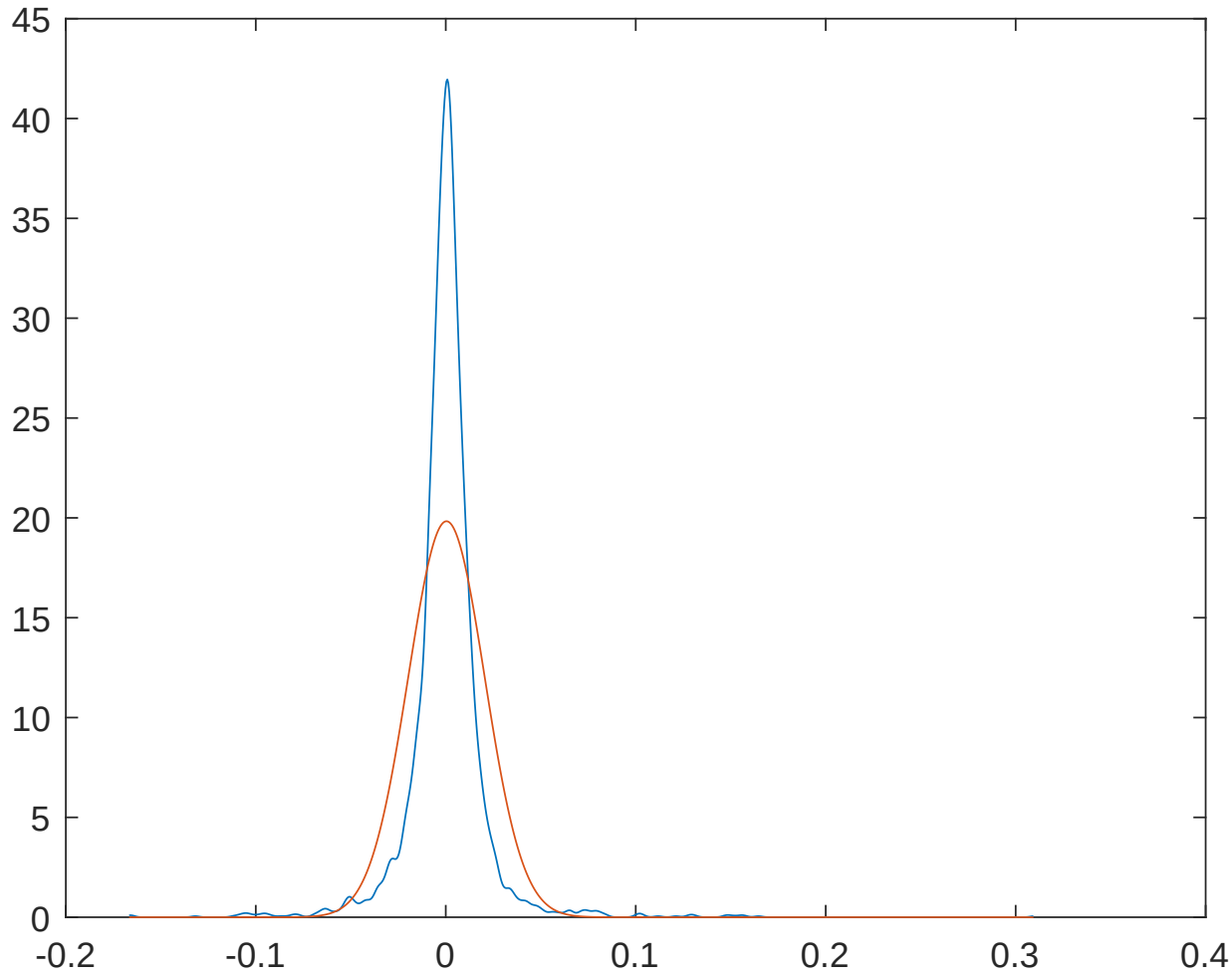
Empirical density

Example (The Financial Select Sector SPDR Fund (XLF))



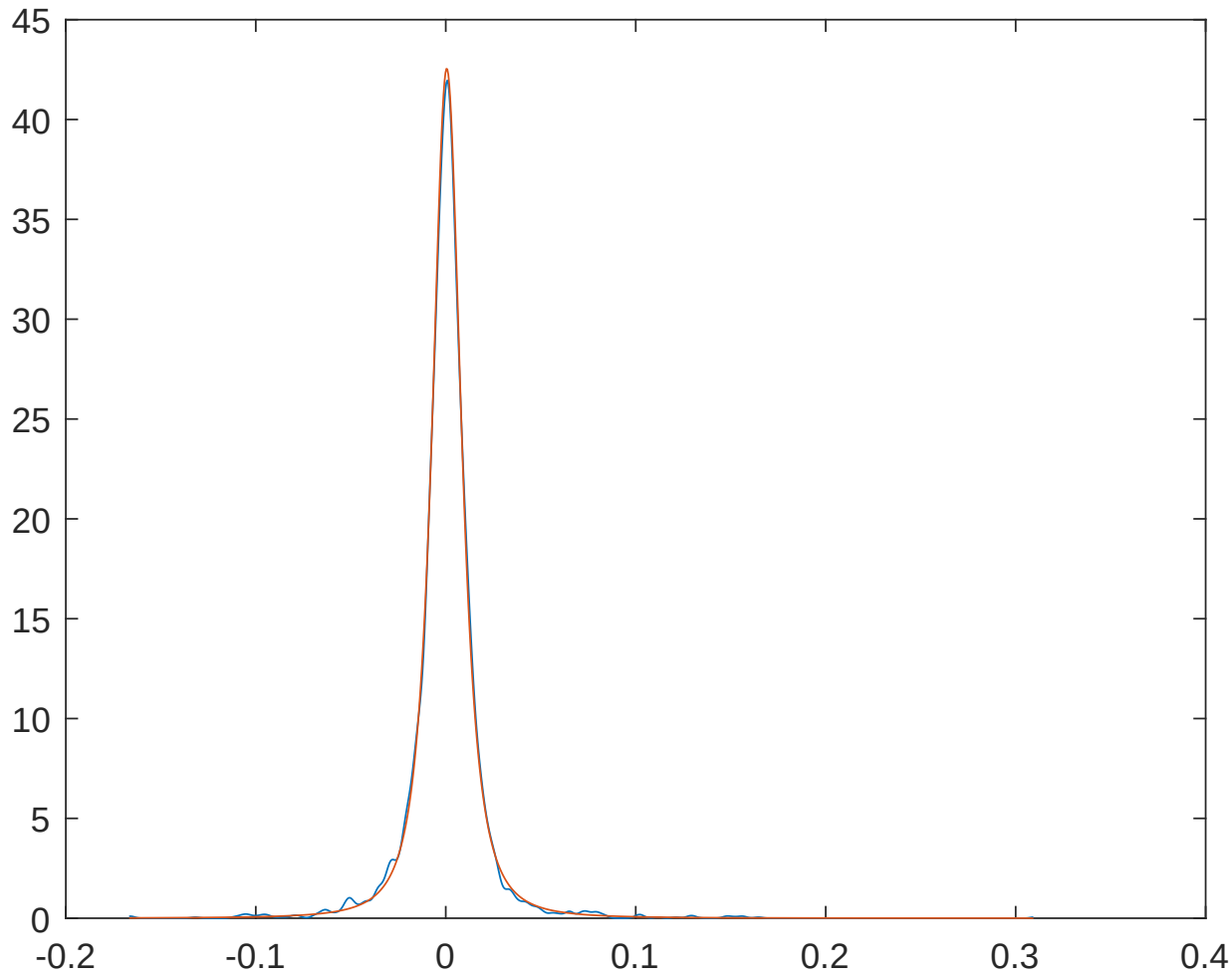
Empirical density vs Normal distribution

Example (The Financial Select Sector SPDR Fund (XLF))



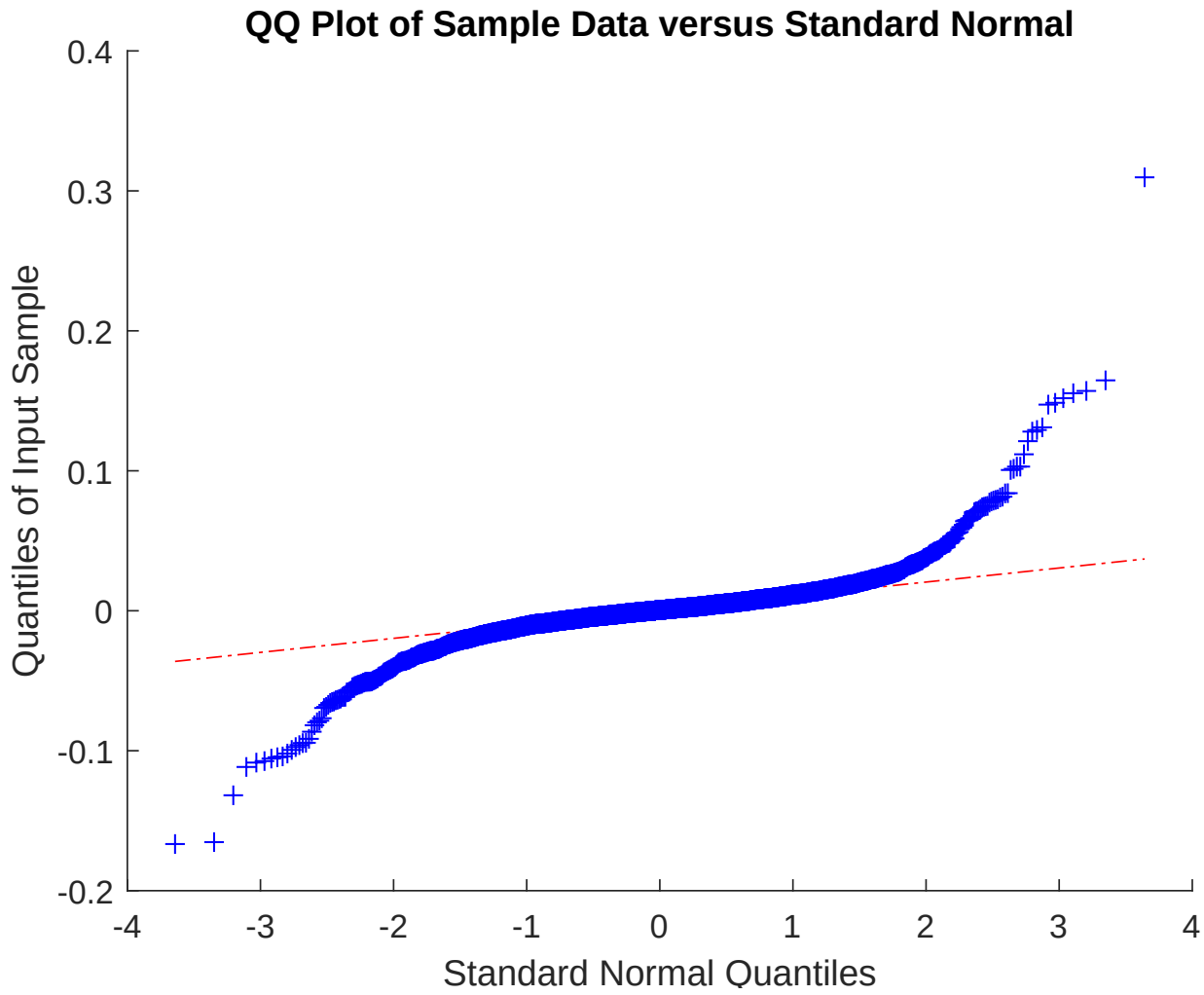
Empirical density vs t-distribution

Example (The Financial Select Sector SPDR Fund (XLF))



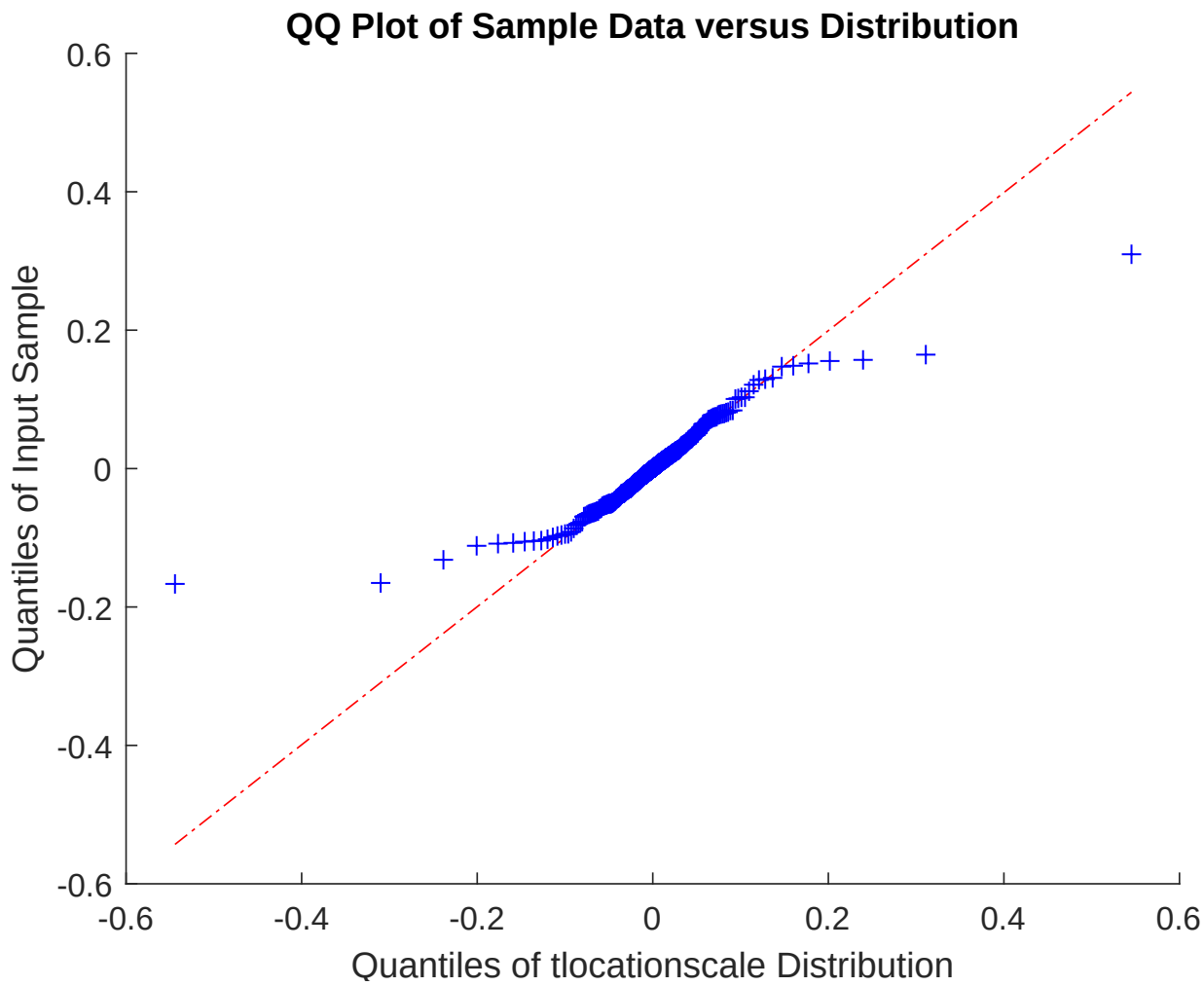
Normal probability plot – testing normality

Example (The Financial Select Sector SPDR Fund (XLF))



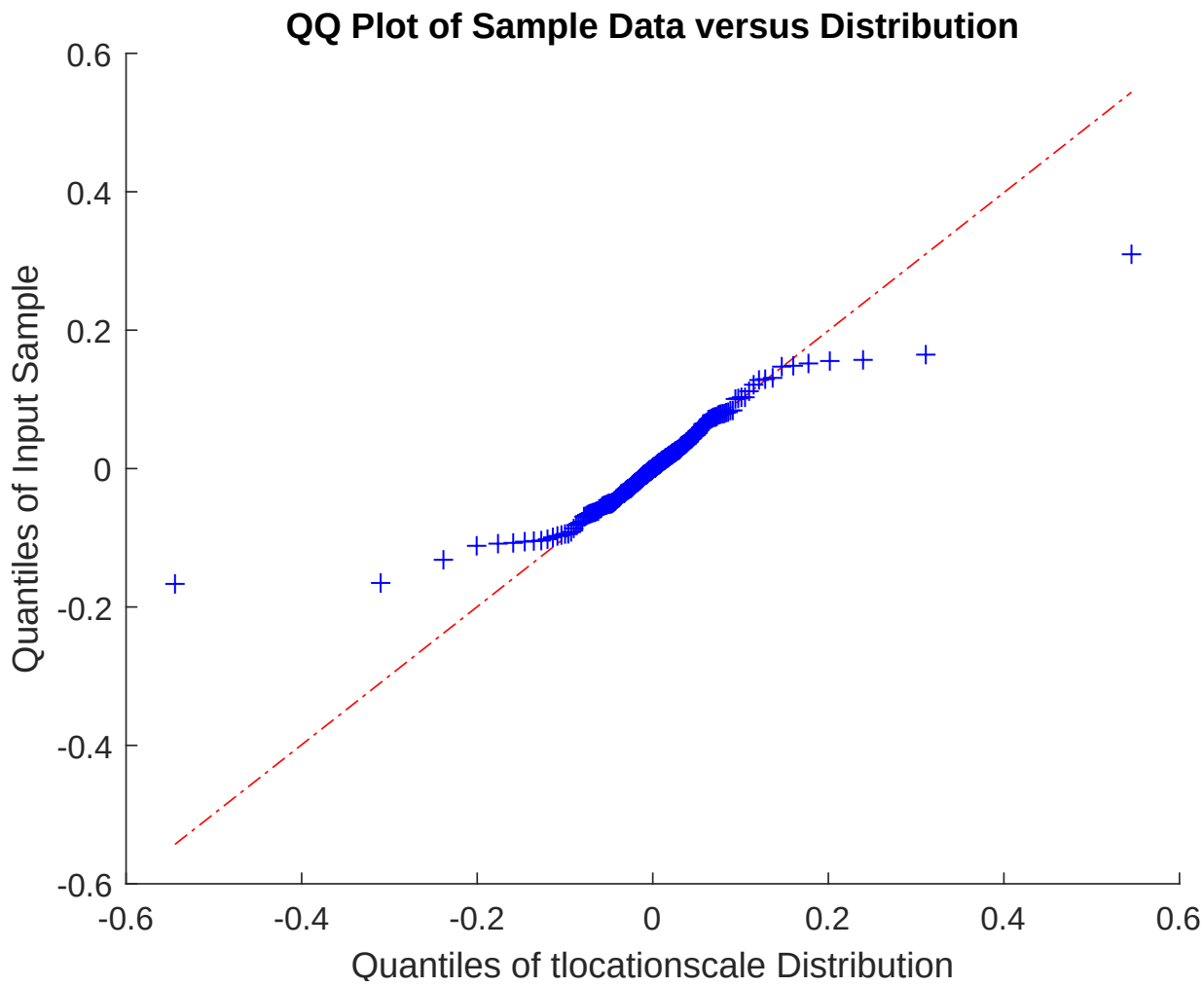
Normal probability plot – testing t-distribution

Example (The Financial Select Sector SPDR Fund (XLF))



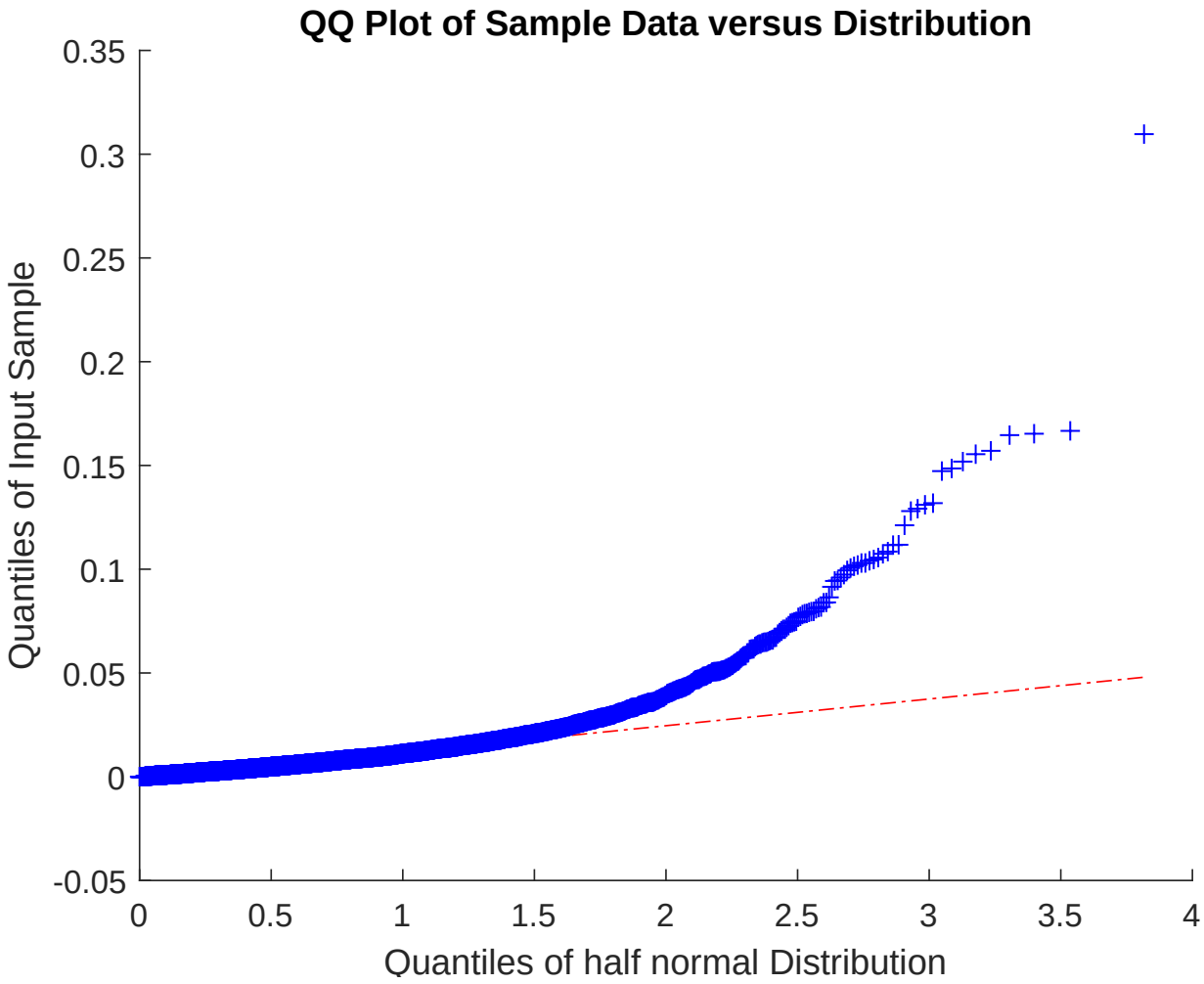
Normal probability plot – testing t-distribution

Example (The Financial Select Sector SPDR Fund (XLF))



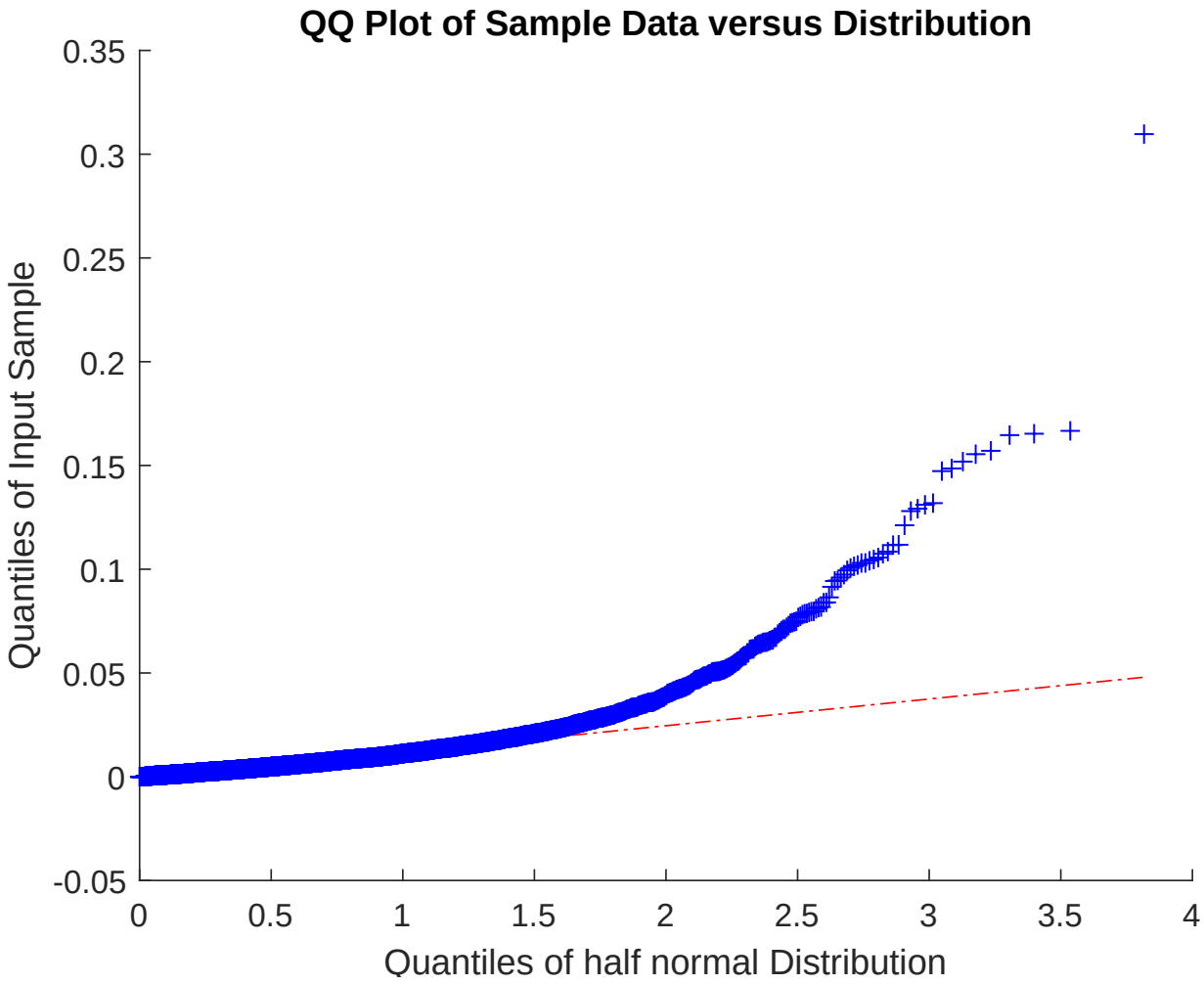
Half-normal plot – outliers

Example (The Financial Select Sector SPDR Fund (XLF))



Half-normal plot – outliers

Example (The Financial Select Sector SPDR Fund (XLF))



Overview of data

Besides graphical inspection and presentation, it's often relevant to display a summary of data. For instance:

- The historical average rate of return (annualised)
- The historical variance or standard deviation of returns (again, annualised)
- The correlation of returns
- Higher moments, such as (excess) kurtosis and skewness
- Statistical tests for normality, for instance Jarque-Bera or Shapiro-Wilk
- Auto-correlation (and autocorrelatioin of squared returns)

Jarque-Bera test

- The Jarque-Bera test is (one of many) test(s) for normality
- Normally distributed data has skewness S of 0 and excess kurtosis of 0 (kurtosis K equal to 3)
- The JB-test tests if data can be assumed to have zero skewness and excess kurtosis
- $JB = n\left(S^2/6 + (K-3)^2/24\right)$
- The test-statistic converges to a $\chi^2(2)$-distribution.
- The test (and many other) are available in R and Matlab

More advanced modelling

- The analysis done so far has dealt with *unconditional* modelling of the relevant variable
- Sometimes this is also referred to as a marginal distribution, i.e., the distribution of a future value given no knowledge about the past
- It would be relevant to include information about past returns in order to capture auto-correlation or volatility ($\sim$ standard deviation) clustering
- If you want to explore this further, Financial Products, Financial Risk Management, Time Series Analysis and others are relevant courses