FE570 Financial Markets and Trading

Lecture 5. Linear Time Series Model and Its Applications

Steve Yang

Stevens Institute of Technology

12/04/2018

Outline

Moving Average Model

- We start, at least in theory, an AR model with infinite order as:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \ldots + a_t. \qquad (1)$$

  However, such an AR model is not realistic because it has infinite many parameters.

- One way to make the model practical is to assume that the coefficients ϕ_i satisfy some constraints so that they are determined by a finite number of parameters.

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_1^2 r_{t-2} + \phi_1^3 r_{t-3} + \ldots + a_t. \qquad (2)$$

  A special case of this idea is: the coefficients depend on a single parameter θ_1 via $\phi_i = -\theta_1^i$ for $i \geq 1$.

  To be stationary, $\theta_1 < 1$ must hold; otherwise, θ_1^i and the series will explode.

Moving Average Model

- The previous representation can be rewritten in a rather compact form:

$$r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + \theta_1^3 r_{t-3} + \ldots = \theta_0 + a_t. \qquad (3)$$

- The model for r_{t-1} is then

$$r_{t-1} + \theta_1^1 r_{t-2} + \theta_1^2 r_{t-3} + \ldots = \theta_0 + a_{t-1}. \qquad (4)$$

- Multiplying the last equation with θ_1 and subtract the result from Eq. (3), we obtain

$$r_t = \phi_0(1 - \theta_1) + a_t - \theta_1 a_{t-1}, \qquad (5)$$

which says that, except for the constant term, r_t is a weighted average of shocks a_t and a_{t-1}. Therefore, the model is called an MA model of order 1 or MA(1) model for short.

General Moving Average Model

- The general form of an MA(1) model is:

$$r_t = c_0 + a_t - \theta_1 a_{t-1}, \text{ or } r_t = c_0 + (1 - \theta_1 B)a_t, \quad (6)$$

where c_0 is a constant and a_t is a white noise series. Similarly, an MA(2) model is in the form

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}, \quad (7)$$

- The general MA(q) model is

$$r_t = c_0 - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q}, \quad (8)$$

$$\text{or } r_t = c_0 - (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)a_t, \text{ where } q > 0. \quad (9)$$

where B is a back-shift operator.

Properties of MA Models

- MA models are always weakly stationary because they are finite linear combinations of a white noise sequence for which the first two moments are time-invariant.
  - For example, consider MA(1) model, we have

$$E(r_t) = c_0$$

    which is time-invariant.
  - Take the variance of MA(1) model Eq. (6), we have

$$Var(r_t) = \sigma_a^2 + \theta_1^2\sigma_a^2 = (1 + \theta_1^2)\sigma_a^2.$$

    and we have $Var(r_t)$ is time-invariant.
- For the general MA(q) model, we have

$$\begin{aligned} E(r_t) &= c_0 \\ Var(r_t) &= (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma_a^2. \end{aligned}$$

Autocorrelation Function of MA(q) Model

- Assume for simplicity that $c_0 = 0$ for an $MA(1)$ model, and we have

$$r_t r_{t-\ell} = r_{t-\ell} a_t - \theta_1 r_{t-\ell} a_{t-1}.$$
$$\gamma_1 = -\theta_1 \sigma_a^2, \text{ and } \gamma_\ell = 0, \text{ for } \ell > 1.$$

- Using the prior result and the fact that $Var(r_t) = (1 + \theta_1^2)\sigma_a^2$, we have

$$\rho_0 = 1, \rho_1 = \frac{-\theta_1}{1 + \theta_1^2}, \rho_\ell = 0, \text{ for } \ell > 1$$

- For MA(2) model, the autocorrelation coefficients are:

$$\rho_1 = \frac{-\theta_1 + \theta_1 \theta 2}{1 + \theta_1^2 + \theta_2^2}, \rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, \rho_\ell = 0, \text{ for } \ell > 2$$

Autocorrelation Function of MA(q) Model

- We can conclude that for MA(1) model, the lag-1 ACF is not zero, but all higher order ACFs are zeros.
- For the MA(2) model, the ACF cuts off at lag-2
- For an MA(q) model, the lag-ℓ ACF is not zero, but $\rho_\ell = 0$ for $\ell > q$. Consequently, an MA(q) series is only linearly related to its first q lagged values - "finite-memory" model.

Invertibility

- Rewriting a zero-mean MA(1) model as $a_t = r_t + \theta_1 a_{t-1}$, and

$$a_t = r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + +\theta_1^3 r_{t-3} + ...$$

- Intuitively, θ_1^j should go to zero as j increases because the remote return r_{t-j} should have very little impact on the current shock, if any. We require $|\theta_1| < 1$, so that MA(1) is *invertable*.

Identifying MA(q) Order

- For a time series r_t with ACF ρ_ℓ, if $\rho_q \neq 0$, but $\rho_\ell = 0$ for $\ell > q$, then r_t follows an MA(q) model.

Estimation

- Maximum likelihood estimation is commonly used to estimate MA models. There are two approaches for evaluating the likelihood function of an MA model: *conditional likelihood method* vs. *exact likelihood method*. (We are not going to get into the details, but just note that the latter is preferred over the former).

Forecasting Using MA Models

- Assume that the forecast origin is h and let F_h denote the information available at time h. For the l-step ahead forecast of an MA(1) process, the model says

$$r_{h+1} = c_0 + a_{h+1} - \theta_1 a_h.$$

Forecasting Using MA Models

- Taking the conditional expectation, we have

$$\hat{r}(1) = E(r_{h+1}|F_h) = c_0 - \theta_1 a_h,$$
$$e_h(1) = r_{h+1} - \hat{r}(1) = a_{h+1}$$

- The variance of the 1-step ahead forecast error is $Var[e_h(1)] = \sigma_a^2$.
- In practice, the quantity a_h can be obtained in several ways. For instance, assume that $a_0 = 0$, then $a_1 = r_1 - c_0$, and we can compute a_t for $a \leq t \leq h$ recursively by using $a_t = r_t - c_0 + \theta_1 a_{t-1}$.
- For the 2-step ahead forecast, from the equation

$$r_{h+2} = c_0 + a_{h+2} - \theta_1 a_{h+1}.$$

Forecasting Using MA Models

- Taking the conditional expectation, we have

$$\hat{r}(2) = E(r_{h+2}|F_h) = c_0,$$
$$e_h(2) = r_{h+2} - \hat{r}(2) = a_{h+2} - \theta_1 a_{h+1}$$

- The variance of the 2-step ahead forecast error is $Var[e_h(2)] = (1+\theta_1^2)\sigma_a^2$, which is the variance of the model and is greater than or equal to that of the 1-step ahead forecast error. More generally, $\hat{r}_h(\ell) = c_0$ for $\ell \geq 2$.
- Similarly, for an MA(2) model, we have

$$r_{h+\ell} = c_0 + a_{h+\ell} - \theta_1 a_{h+\ell-1} - \theta_2 a_{h+\ell-2}.$$
$$\hat{r}_h(1) = c_0 - \theta_1 a_h - \theta_2 a_{h-1},$$
$$\hat{r}_h(2) = c_0 - \theta_2 a_h,$$
$$\hat{r}_h(\ell) = c_0, \text{ for } \ell > 2$$

Forecasting Using MA Models

- In general, for an MA(q) model, multi-step ahead forecasts go to the mean after first q steps.

Summary for AR and MA Models

- For MA models, the ACF is useful in specifying the order because the ACF cuts off at lag q for an MA(q) series.
- For AR models, the PACF is useful in order determination because the PACF cuts off at lag p for an AR(p) process.
- An MA series is always stationary, but for an AR series to be stationary, all of its characteristic roots must be less than 1 in modulus
- For stationary series, the multi-step ahead forecasts converge to the mean of the series and the variances of forecast errors converge to the variances of the series.

Building an MA Model

- Specification: Use sample ACF.

  Sample ACFs are all small after lag q for an MA(q) series.
- Constant term? Check the sample mean.
- Estimation: use maximum likelihood method (Exact method is preferred, but it is more computing intensive.)
- Model checking: examine residuals (to be white noise)
- Forecast: use the residuals as $\{a_t\}$, which can be obtained from the data and fitted parameters, to perform forecasts.

- R examples to demonstrate ACF and PACF, and build MA models.
  1. Simulated AR1, AR2, MA1, and MA2
  2. IBM stock return, and E-Mini S&P Simulated time series analysis.

ARMA Models

- In some applications, the AR and MA models discussed in the previous sections become cumbersome because one may need a high-order model with many parameters to adequately describe the dynamic structure of the data.
- To overcome this difficulty, the autoregressive moving-average (ARMA) models are introduced. An ARMA model combines the idea of AR and MA models into a compact form so that the number of parameters used is kept small.

ARMA(1,1) Model

- A time series follows an ARMA(1,1) model if it satisfies:

$$r_t - \phi_1 r_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1} \tag{10}$$

where $\{a_t\}$ is a white noise series.

Properties of ARMA(1,1) Models

- Taking expectation of Eq. 10, we have

$$E(r_t) - \phi_1 E(r_{t-1}) = \phi_0 + E(a_t) - \theta_1 E(a_{t-1}). \tag{11}$$

Because $E(a_i) = 0$ for all i, the mean of r_t is

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1}. \tag{12}$$

provided that the series is weakly stationary.

- Next, we consider the autocovariance function of r_t. We multiply Eq. 10 by a_t and take expectation (simplify $\phi_0 = 0$)

$$E(r_t a_t) = E(a_t^2) - \theta_1 E(a_t a_{t-1}) = E(a_t^2) = \sigma_a^2.$$

$$r_t = \phi_1 r_{t-1} + a_t - \theta_1 a_{t-1}.$$

Properties of ARMA(1,1) Models

- Taking variance of prior equation:

$$Var(r_t) = \phi_1^2 Var(r_{t-1}) + \sigma_a^2 + \theta_1^2\sigma_a^2 - 2\phi_1\theta_1 E(r_{t-1}a_{t-1}).$$

- If the series r_t is weakly stationary, then $Var(r_t) = Var(r_t - 1)$ and we have

$$Var(r_t) = \frac{(1 - 2\phi_1\theta_1 + \theta_1^2)\sigma_a^2}{1 - \phi_1^2}.$$

we need $\phi_1^2 < 1$

- To obtain the autocovariance function of r_t, we assume $\phi_0 = 0$ and multiply the model in 10 by $r_{t-\ell}$ to obtain:

$$r_t r_{t-\ell} - \phi_1 r_{t-1} r_{t-\ell} = a_t r_{t-\ell} - \theta_1 a_{t-1} r_{t-\ell}.$$

Properties of ARMA(1,1) Models

- For $\ell = 1$, taking expectation

$$\gamma_1 - \phi_1\gamma_0 = -\theta_1\sigma_a^2,$$
$$\gamma_\ell - \phi_1\gamma_{\ell-1} = 0, \text{ for } \ell > 1.$$

- In terms of ACF, we have

$$\rho_1 = \phi_1 - \frac{\theta_1\sigma a^2}{\gamma_0}, \rho_\ell = \phi_1\rho_{\ell-1}, \text{ for } \ell > 1.$$

- The ACF of an ARMA(1,1) model does not cut off at any finite lag.
- The stationary condition of an ARMA(1,1) model is the same as that of an AR(1) model, and the ACF of an ARMA(1,1) exhibits a pattern similar to that of an AR(1) model except that the pattern starts at lag 2.

General ARMA Models:

- A general ARMA(p,q) model is in the form

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i a_{t-i}, \tag{13}$$

where $\{a_t\}$ is a white noise series and p and q are no-negative integers. The AR and MA models are special cases of ARMA(p,q) model. Using the back-shift operator, the model can be written as

$$(1 - \phi_1 B - ... - \phi_p B^p) r_t = \phi_0 + (1 - \theta_1 B - ... - \theta_q B^q) a_t.$$

The polynomial $1 - \phi_1 B - ... - \phi_p B^p$ is the AR polynomial of the model. Similarly, $1 - \theta_1 B - ... - \theta_q B^q$ is the MA polynomial. We require that there are no common factors between the AR and MA polynomials.

Identifying ARMA Models:

- The ACF and PACF are not informative in determining the order of an ARMA model. There are two ways to determine the order of ARMA models:
  - *The Extended Autocorrelation Function (EACF)* - If we can obtain a consistent estimate of the AR component of an ARMA model, then we can derive the MA component. From the derived MA series, we can use the ACF to identify the order of the MA component.
  - *The Information Criteria (AIC or BIC)* - Typically, for some prespecified positive integers P and Q, one computes AIC (or BIC) for ARMA(p,q) models, where $0 \leq p \leq P$ and $0 \leq q \leq Q$, and selects the model that gives the minimum AIC (or BIC).
- Once an ARMA(p,q) model is specified, its parameters can be estimated by either the conditional or exact likelihood method.
- The Ljung-Box statistics of the residuals can be used to check the adequacy of a fitted model.

Forecasting Using ARMA Models:

- Denote the forecast origin by h and the available information by F_h. The 1-step ahead forecast of r_{h+1} becomes:

$$\hat{r}_h(1) = E(r_{h+1}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i} - \sum_{i=1}^{q} \theta_i a_{h+1-i}.$$

  and the forecast error is $e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$. The variance of 1-step ahead forecast error is $Var[e_h(1)] = \sigma_a^2$.

- For the ℓ-step ahead forecast, we have:

$$\hat{r}_h(\ell) = E(r_{h+\ell}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(\ell - i) - \sum_{i=1}^{q} \theta_i a_h(\ell - i).$$

  where it is understood that $\hat{r}_h(\ell - i) = r_{h+\ell-i}$ if $\ell - i \leq 0$ and $a_h(\ell - i) = 0$ if $\ell - i > 0$ and $a_h(\ell - i) = a_{h+\ell-i}$ if $\ell - i \leq 0$.

Forecasting Using ARMA Models:

- Thus, the multi-step ahead forecasts of an ARMA model can be computed recursively. The associated forecast error is

$$e_h(\ell) = r_{h+\ell} - \hat{r}_h(\ell - i).$$

Three Model Representations for an ARMA Model

- The three representations of a stationary ARMA(p,q) models serve different purposes. Knowing these representations can lead to a better understanding of the model.
- The first representation is the ARMA(p,q) model in Eq. 14. This representation is compact and useful in parameter estimation. It is also useful in computing recursively multi-step ahead forecasts of r_t.
- For the other two representations, we use long division of two polynomials.

Three Model Representations for an ARMA Model

- Given the following two polynomials:

$$\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$$

$$\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$$

we can obtain, by long divisions, that

$$\frac{\theta(B)}{\phi(B)} = 1 + \psi_1 B + \psi_2 B^2 + ... \equiv \psi(B) \tag{14}$$

$$\frac{\phi(B)}{\theta(B)} = 1 + \pi_1 B + \pi_2 B^2 + ... \equiv \pi(B) \tag{15}$$

from the definition $\psi(B)\pi(B) = 1$.

Three Model Representations for an ARMA Model

- For instance, if $\theta(B) = 1 - \theta_1 B$ and $\phi(B) = 1 - \phi_1 B$, then

$$\frac{\phi_0}{\theta_1} = \frac{\phi_0}{1 - \theta_1 - ... - \theta_q} \text{ and } \frac{\phi_0}{\phi_1} = \frac{\phi_0}{1 - \phi_1 - ... - \phi_p}$$

- **AR Representation** (shows the dependence of the current return r_t on the past return r_{t-i} where $i > 0$)

$$r_t = \frac{\phi_0}{1 - \theta_1 - ... - \theta_q} + \pi_1 r_{t-1} + \pi_2 r_{t-2} + \pi_3 r_{t-3} + ... \quad (16)$$

- **MA Representation** (shows explicitly the impact of the past shock $a_{t-i}(i > 0)$ on the current return r_t.)

$$r_t = \frac{\phi_0}{1 - \phi_1 - ... - \phi_p} + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + ... \quad (17)$$

Unit-Root Nonstationary Time Series

- Consider an ARMA model, if one extends the model by allowing the AR polynomial to have 1 as a characteristic root, then the model becomes the well-known autoregressive integrated moving-average (ARIMA) model. A conventional approach for handling unit-root nonstationarity is to use *differencing*.
- A time series is said to be an ARMA(p, 1, q) process if the change series

$$c_t = y_t - y_{t-1} = (1 - B)y_t,$$

follows a stationary and invertable ARMA(p, q) model. In finance, price series are commonly believed to be nonstationary, but the log return series is stationary.

$$r_t = ln(p_t) - ln(p_{t-1})$$

Unit-Root Nonstationary Time Series

- In some scientific field, a time series y_t may contain multiple unit roots and need to be differenced multiple times to become stationary. More complex model ARIMA(p, d, q) can be applied.
- *Unit-Root Test*

  whether the log price p_t of an asset follows a random walk or a random walk with drift.

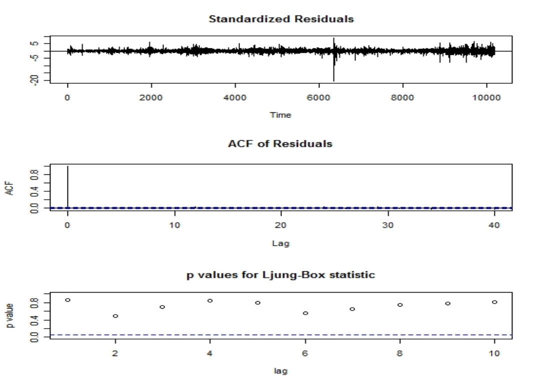
$$p_t = \phi_1 p_t + e_t,$$
$$p_t = \phi_0 + \phi_1 p_{t-1} + e_t,$$

  the null hypotheses $H_0 : \phi_1 = 1$ versus the alternative hypothesis $H_a : \phi_1 < 1$. This is well-known *Dickey-Fuller* test.
- ARIMA also removes the trend and provides a trend-stationary time series.

Example

Use R ARIMA module to fit an ARMA model

- Identify Model
- Estimate Parameters
- Model Checking
- Forecast Using Fitted Model



Overview of Back-Test of Trading Strategies

Overview of Back-Test of Trading Strategies *"It is hard to make predictions, especially about the future" - Mark Twain*.

- Forecasting has been widely used in economics, finance, and natural science (e.g., weather forecasting, climate change, global warming, outbreaks war, etc.) - arguably in every field where time series analysis is involved.
- The term *back-testing* implies that the forecasting models are fitted and tested using past empirical data. Usually, the entire available data set is split into two parts, one of which (earlier data) is used for *in-sample* calibration of the predicting model while the other is reserved for *out-of-sample* testing of the calibrated model. In simple models, it is usually assumed that the testing sample is variance-stationary; that is, the sample volatility is constant.
- In general, it is not always true that "more is better", as the long time series may be non-stationary.

考点

Overview of Back-Test of Trading Strategies

- Very often we find that the optimal strategy parameters may evolve in time. In this case, moving-window sampling may be appropriate. For example, if a 10-year data sample is available, the first five-year data are used for in-sample calibration and the sixth-year data are used for out-of-sample testing. Then, the data from the second to the sixth year data are used for in-sample calibration and the seventh year is tested as out-of-sample data so on and so forth.

- Time series may also experiences *regime shifts* caused by macro-economic events or changes in regulatory policies (e.g., introduction of the Euro in 1999, changes in uptick rule, etc.). Markov-switching models are sometimes used for a unified description of data samples with regime shifts.

- Lastly, there is always a danger of model over-fitting (too good to be true with regard to in-sample accuracy).

Overview of Back-Test of Trading Strategies

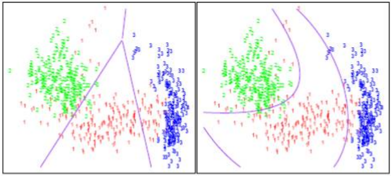


Figure 4.1: *The left plot shows some data from three classes, with linear decision boundaries found by linear discriminant analysis. The right plot shows quadratic decision boundaries. These were obtained by finding linear boundaries in the five-dimensional space* $X_1, X_2, X_{12}, X_1^2, X_2^2$. *Linear inequalities in this space are quadratic inequalities in the original space.*

Figure: In this case, the model is fitted to the in-sample noise rather than to the deterministic relationship. The maximum likelihood-based criteria that have been used to evaluated the fitness of the model, such as, AIC, BIC, ABIC, DIC (Deviance information criterion), etc.

Overview of Back-Test of Trading Strategies

考点

- *Data Snooping Bias* - this bias can appear during the testing of different strategies with the same data set.

  A well-publicized example of data snooping is offered by Sullivan et al. (1999), who quoted Leinweber's finding that the best predictor in the United Nations database for the S& P 500 stock index is the production of butter in Bangladesh.

- A general solution to the snooping bias is data **Resampling**. Bootstrap is arguably the most popular resampling procedure. Other resampling technique *jackknife*, and *permutation* tests.

- The **Markov Chain Monte Carlo** (MCMC) simulation is another resampling approach, where the original sample is used to estimate probabilities of new returns conditioned on current and past returns.

- Finally, the **Random Entry Protocol** is a simple resampling procedure that addresses the problem of correlation in two coupled time series.

Performance Measures

Performance Measures for trading strategies usually relate to a rather long time period (at least a quarter or a year), during which a given strategy is used multiple times. Due to market factors, such as, finite liquidity, bid/ask spread, and transaction cost, practical performance measures should be evaluated after fees.

- **Total Return:**
  - Total return is the ultimate performance benchmark. It is calculated over some trading period after all long and short positions in the trading portfolio are closed. In other words, it is realized return that matters.

Note that the compounded return usually listed in asset management statements is not realized and may include dividends and reinvestment. We are interested in pure trading strategy performance and therefore use the same notional amount in every round-trip trade.

Performance Measures

- **Kelly's Criterion:**
  - Percent of winning trades, p, is another important performance measure.

  If this percentage and the ratio of the average winning amount to average losing amount, r, is assumed to be stable, one can use the Kelly's criterion for estimating the optimal fraction of trading capital, f, to be used in each trade:

$$f = (p * r - 1 + p)/r$$

It can be shown that the Kelly's criterion is equivalent to choosing the trading size that maximizes the geometric mean of outcomes. Note that the Kelly's formula yields an estimate that is valid only asymptotically.

Therefore, risk-averse practitioners are advised to use a value of f lower than the Kelly's criterion suggests. The total number of trades for a given period is also important. Frequent trading may lead to more volatile outcomes.

Performance Measures

- **Average Return and Variance:**
  - Multiple trades with a given strategy generate a probability distribution that can be used for hypothesis testing, in particular for comparing different strategies. Hence, the average return, μ, and its variance, σ, are very important performance measures.
- **The Student t Test:**
  - Note that a positive return being accompanied with a high variance does not guarantee the strategy's quality. Provided that the return distribution is normal, one can use the t-statistic for testing the hypothesis that the distribution mean is zero. Namely, for a given number of round-trip trades, N, one calculates the t-value:

$$t = \frac{\mu}{(\sigma^2/N)^{1/2}}$$

    then, the t-value can be used for finding *statistical significance* (also called *p-value*) from the Student's distribution.
  - Usually, the null hypothesis in an analysis of trading strategies is that the strategy return is zero.

Performance Measures

- **ANOVA Test:**

  - If two strategies have the same average return, the one with lower variance is more attractive. Provided that the return distributions are normal, two trading strategies A and B can be compared using the t-statistic:

$$t = \frac{\mu_A - \mu_B}{(\sigma_A^2/N_A + \sigma_B^2/N_B)^{1/2}}$$

  In this case, the degree of freedom equal $N_A + N_B - 2$. The t-statistic can also be used for analysis of profitability of a single strategy if indexes A and B refer to buy and sell signals.

- **Information Ratio:**

  - The Sharpe ratio is often used in performance analysis. Sometimes, the *Sortino ratio* is chosen instead. In the latter ratio, only negative returns are included in calculating the standard deviation σ. If a trading strategy performance is compared with the performance of an index (or buy-and-hold strategy), the *information ratio* can be used.

Performance Measures

- *Information ratio* is calculated as:

$$IR = (E[r_i] - E[r_0])/\sigma_{i0}$$

where r_i is given return, r_0 is return of an index, and σ_{i0} is the tracking error (standard deviation between returns of the strategy and returns of the index).

- **Maximum Drawdown:**

- The *maximum drawdown* (MD) is another important risk measure, particularly for leveraged trades. For a process $X(t)$ on the interval $[0, T]$,

In other words, MD is the largest drop of price after its peak.

$$MD = max[max(X[s] - X[t])], t \in [0, T], s \in [0, t]$$

However, if drift can be neglected, expectation of MD has a simple analytic form:

$$E[MD] = 1.2533\sigma\sqrt{T}$$

Bootstrap:

- The t-statistic can be misleading when it is applied to non-normal distributions. In the simple case, the boostrap protocol is based on picking up at random an element of a given sample size N, copying it into the new sample, and putting it back (replacement). This random selection continues until the new sample has the same number of elements as the original one.

- Sometimes, blocks of several sequential elements of a given sample are picked at once (block boostrap).

  Usually, the block boostrap is implemented with replacement and blocks are not overlapping.

  Such an approach may preserve short-range autocorrelations present in the original sample. While a simple estimate of an optimal block size $L \sim N^{-1/3}$ can be used, choice of L in the general case is not trivial.

  This method ensures the stationarity of samples bootstrapped from the stationary data - *stationary boostrap*.

Bootstrap

The block size L in *stationary boostrap* is randomly drawn from the geometric distribution:

$$Pr(L = k) = (1 - p)^{k-1} p$$

- The average block size for the stationary bootstrap case equals $1/p$, which can serve as a bridge between the size of the simple block bootstrap and the stationary bootstrap parameter p.

- A more sophisticated approach implies estimating a mathematical model that fits the given sample and bootstrapping the model's residuals. Typical models used for stock prices are the random walk with drift, the AR models and the GARCH models.

- The number of bootstrapped samples needed for good accuracy may reach from several hundreds to several thousands.

Markov Chain Monte Carlo

- Markov process is a generic stochastic process determined with relationships between its future, present, and past values.

- By definition, the Markov chain of the kth order is such a sequence of random variables $X_1, X_2, \ldots$, which satisfies the following equation:

$$\begin{aligned} &Pr(X_n = x | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1) \\ &= Pr(X_n = x | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_{n-k} = x_{n-k}) \end{aligned}$$

  In other words, only k past values (sometimes called initial conditions) determine the present value.

  In particular, for $k = 1$, only one initial condition is needed:

$$\begin{aligned} Pr(X_n = x | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1) \\ = Pr(X_n = x | X_{n-1} = x_{n-1}) \end{aligned}$$

Markov Chain Monte Carlo

- A Markov chain is stationary (or time-homogeneous) if probability in the left-hand side of the general form does not depend on index n.

- Generally, Markov variables can assume only a finite number of values (states). Stationary Markov chains of the 1st order with N states are determined with N^2 probabilities $Pr(X_n = x_k | X_{n-1} = x_i) = p_{ik}, i, k = 1, 2, ..., N$. These probabilities are called the *transition kernel*, and the complete set of values p_{ik} is called the *transition matrix*.

  Note that for each k

  $$\sum_{i=1}^{N} p_{ik} = 1$$

  Similarly, the high order Markov chains of kth order are determined by N^k probabilities.

Markov Chain Monte Carlo

- In MCMC-based resampling, the transition matrix is assumed stationary and is calculated using the original sample. Then, drawings from the uniform distribution are mapped onto transition probabilities for generating new samples.

- For example, consider a two-state Markov chain with $p_{11} = p, p_{22} = q$ (which implies that $p_{12} = 1 - p$ and $p_{21} = 1 - q$). Say the current state is 1. If a drawing from the uniform distribution is less than or equal to p, then the next state is 1; otherwise, it is 2. If the current state is 2 and a drawing from the uniform distribution is less than or equal to q, then the next state is 2; otherwise, it is 1.

- Since the transition matrix size grows with the Markov chain's order as the power law, the use of higher orders for multi-state models can become a computational challenge. Normally, the financial returns do not have long memory, and low-order Markov chains should suffice for their resampling.

Random Entry Protocol

- In the general case, trading strategies can be determined not only by price dynamics but also by some liquidity measure(s), such as the bid/ask spread and the asset amount available at the best price, which also varies with time. Depending on the problem addressed with resampling, one may either want to preserve or destroy correlations between two time series.
- If correlations between two samples are weak, MCMC can be implemented for both coupled samples independently, and bootstrap for coupled samples can be reduced to picking up pairs of variables at the same time.
- The random entry protocol is choosing a random time within the given sample for each newly submitted order. It helps to avoid the bias due to autocorrelations in returns but preserves autocorrelations in order block size. It is similar to the stationary boostrap approach.