Chapter 1 - Introduction to time series

AMA4650
Fu Guanxing
guanxfu@polyu.edu.hk



Times series (1/2)

1. **Probability space** $(\Omega, \mathcal{F}, \mathbb{P})$.
   - Sample space Ω: the set of all possible outcomes.
   - Event space $\mathcal{F}$: a σ-algebra. Each element is called an event.
   - Probability measure $\mathbb{P}$: measure the possibility of occurance of each outcome or event.

   Example: the cast of a dice.
   $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\mathbb{P}(\text{outcome} = i) = 1/6$ for $i = 1, \cdots, 6$.

2. **Random variable** (r.v.). A real-valued r.v. is a measurable function/mapping $\xi : \Omega \to \mathbb{R}$.

3. **Stochastic process**: a family of random variables (r.v.s.) indexed by some set $\{X_n, n \in \mathbb{T}\}$[1]. In particular, when the index is time, X is a time series.

4. A **time series** is defined as a family of r.v.s. ordered in time,
   - X_t: r.v.
   - $\mathbb{T}$: index of time, which determines the number and order of observations.

[1] For simplicity, $\{X_n\}$ or X.

Times series (2/2)

- Continuous-time time series and Discrete-time time series.

- Continuous r.v.s. and discrete r.v.s.

Time series data *vs* time series model

Time series data.

- Let $X = \{X_t, t \in \mathbb{T}\}$ be a time series. Time series data $\{x_t, t \in \mathbb{T}\}$ is a *realization* of the time series X, i.e.

$$\text{for each } t \in \mathbb{T},\ x_t = X_t(\omega) \qquad \text{for some } \omega \in \Omega.$$

  There may be many time series data corresponding to one time series.
  There may be dependence between time series observations!
  Example. Stock price in Black-Scholes market. See ... for a realization.

Time series model.

- For a time series X, if each X_t is a discrete r.v., then the time series model for the observed data $\{x_t\}$ is the *joint distributions* of a time series $\{X_t\}$ of which $\{x_t\}$ is a realization

$$\mathbb{P}(X_{t_1} = x_1, \cdots, X_{t_n} = x_n), \quad \text{or} \quad \mathbb{P}(X_{t_1} \le x_1, \cdots, X_{t_n} \le x_n).$$

- If each X_t is a continuous r.v., then the time series model is the *joint density function*

$$f(x_1, \cdots, x_n) \quad \text{or} \quad \mathbb{P}(X_{t_1} \le x_1, \cdots, X_{t_n} \le x_n).$$

  Example (exercise). Show the time series model for the stock price in Black-Scholes market.

- It may be more practical to focus solely on the means and covariances.

For r.v.s. $\mathcal{Y}$ and $\mathcal{Z}$,

- **Variance**: $\mathrm{Var}(\mathcal{Z}) = \mathbb{E}[\mathcal{Z}^2] - (\mathbb{E}[\mathcal{Z}])^2$.
  For any constant c, $\mathrm{Var}(c + \mathcal{Z}) = \mathrm{Var}(\mathcal{Z})$ and $\mathrm{Var}(c\mathcal{Z}) = c^2\mathrm{Var}(\mathcal{Z})$.
- **Covariance**: $\mathrm{Cov}(\mathcal{Y}, \mathcal{Z}) := \mathbb{E}[\mathcal{Y}\mathcal{Z}] - \mathbb{E}[\mathcal{Y}]\mathbb{E}[\mathcal{Z}]$.
  In particular, $\mathrm{Var}(\mathcal{Y}) = \mathrm{Cov}(\mathcal{Y}, \mathcal{Y})$.
  If $\mathcal{Y}$ and $\mathcal{Z}$ are uncorrelated, then $\mathrm{Var}(\mathcal{Z} + \mathcal{Y}) = \mathrm{Var}(\mathcal{Z}) + \mathrm{Var}(\mathcal{Y})$ (Exercise).

Examples of time series (models) (1/3)

Example 1.1 (iid Noise) Let $\{X_t, t = 1, 2, \cdots\}$ be a sequence of independent and identically distributed (iid) r.v.s. with zero mean. It is called iid noise. For any positive integer n and real numbers $x_1, \cdots, x_n$,

$$\mathbb{P}(X_1 \leq x_1, \cdots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \cdots \mathbb{P}(X_n \leq x_n) = F(x_1) \cdots F(x_n),$$

where F is the cumulative distribution function of X_1.

- There is no dependence between observations: for any $h > 0$,

$$\mathbb{P}(X_{n+h} \leq x | X_1 = x_1, \cdots, X_n = x_n) = \mathbb{P}(X_{n+h} \leq x),$$

  which implies that the knowledge of $X_1, \cdots, X_n$ is of no value for predicting the behavior of X_{n+h}.
- Although this means that iid noise is a rather uninteresting process for forecasters, it plays an important role as a building block for more complicated time series models (see the remaining part of this course).

Examples of time series (models) (2/3)

Example 1.2 (A binary process) A very simple example of a stochastic process/time series is the binary process $\{X_t, t = 1, 2, \ldots\}$ of independent r.v.s. with

$$\mathbb{P}(X_t = 1) = \mathbb{P}(X_t = -1) = \frac{1}{2}.$$

The joint distribution/time series model of $(X_1, \cdots, X_n)$ as

$$\mathbb{P}(X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = \frac{1}{2^n},$$

where $i_k = 1$ or -1. [2]

Example 1.3 (Random walk) Let X be an iid noise. A random walk with zero mean is defined as follows

$$S_0 = 0, \quad S_t = X_1 + \cdots + X_t, \quad t = 1, 2, \cdots$$

- $S_t = S_{t-1} + X_t$: random walk.
- If $\{X_t, t = 1, 2, \cdots\}$ is a binary process, then $\{S_t, t = 1, \cdots\}$ is called a simple symmetric random walk.
- Exercise: calculate the variance of S.

[2] Note that a binary processes is an iid noise.

Examples of time series (models) (3/3)

Example 1.4 (Geometric Brownian motion)

$$dS_t = S_t(\mu\, dt + \sigma\, dW_t), \qquad S_0 = s.$$

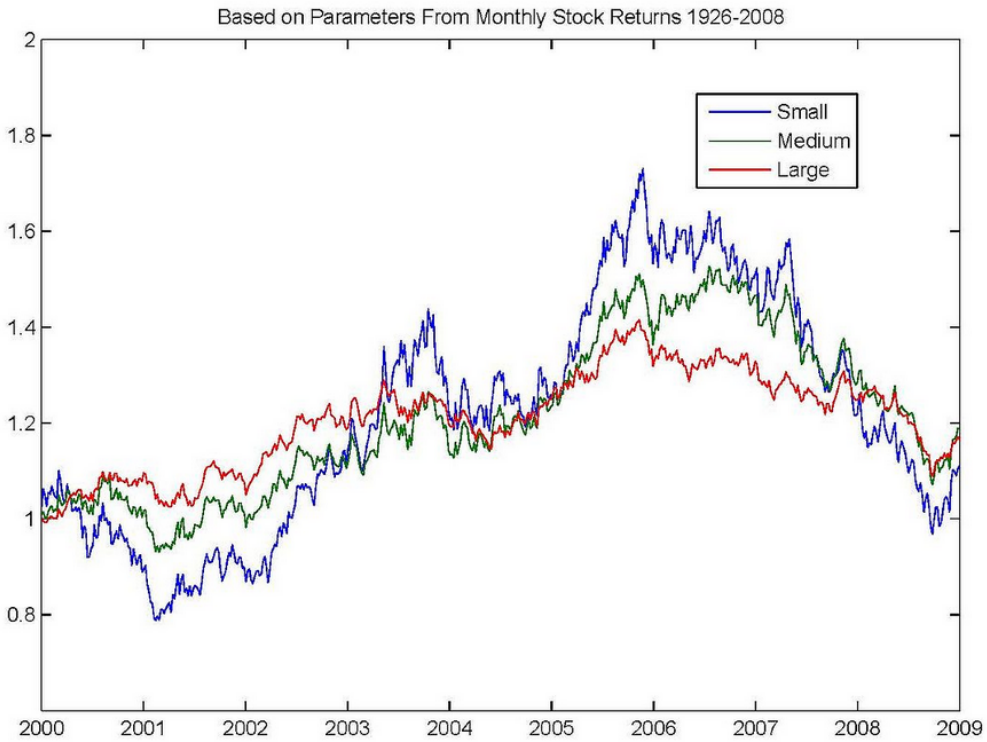


Figure 1.1: Simulation of GBM with parameters from market data.

Examples of time series data (1/3)

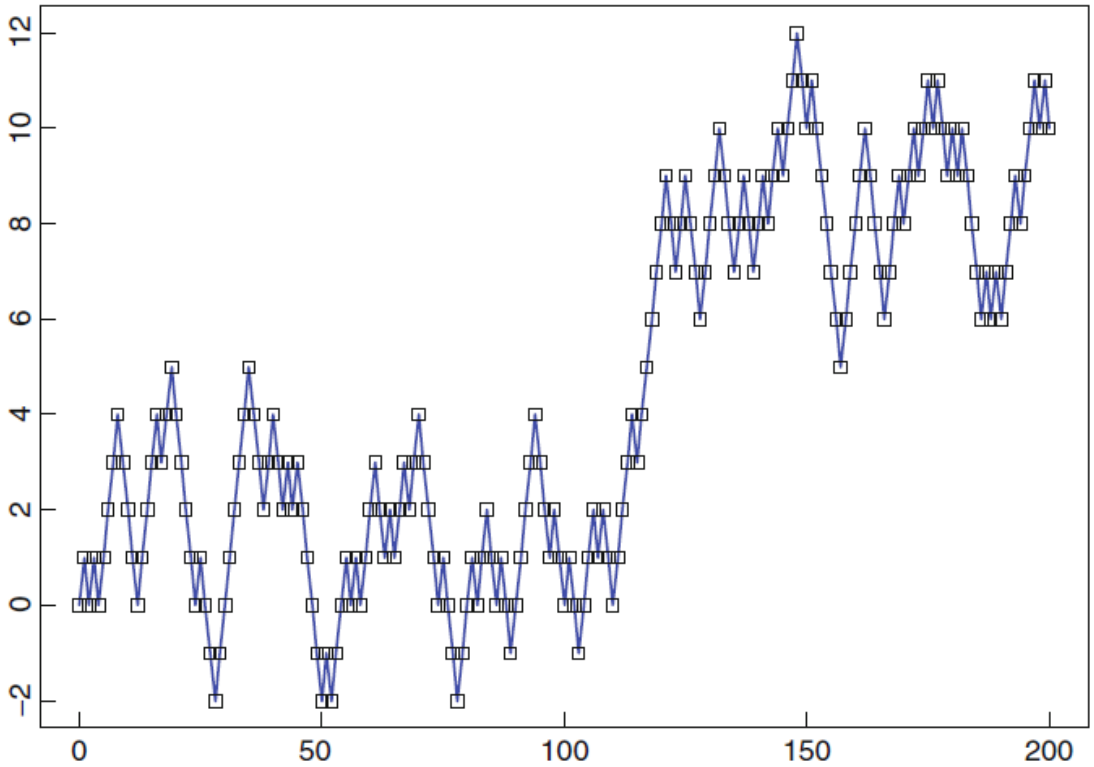


Figure 1.2: One realization of simple random walk.

Examples of time series data (2/3)

The next figure shows the time series plot of global Apple iPhone sales from 3rd quarter 2007 to 3rd quarter 2017 (in million units).

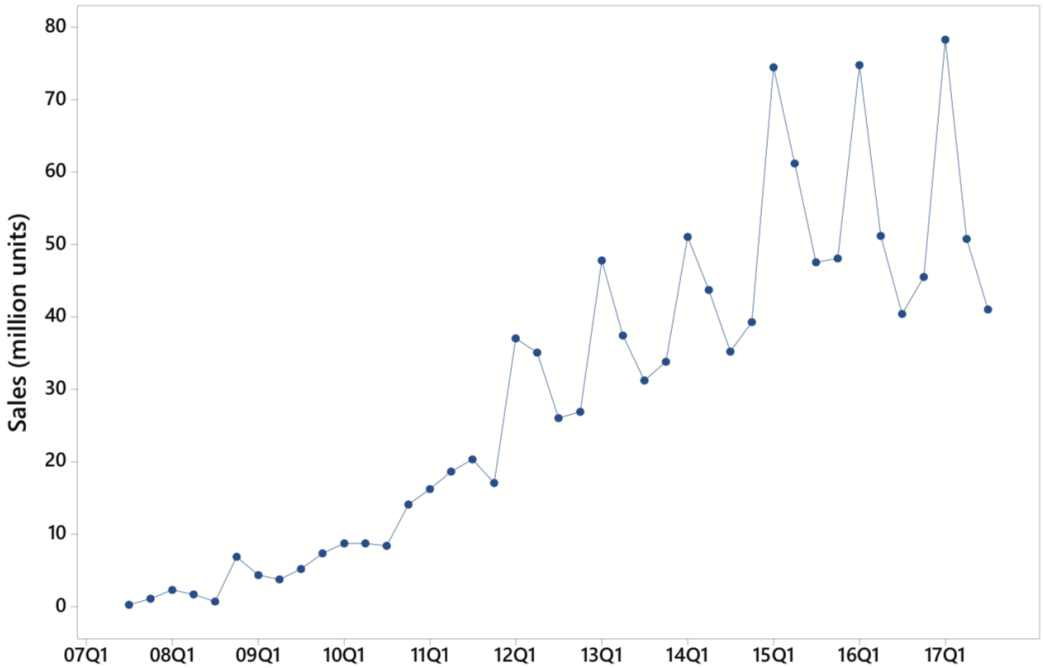


Figure 1.3: Global Apple iPhone sales.

(Source: https://www.statista.com)

Examples of time series data (3/3)

Figure 1.4 shows the time series plot of Apple Inc. stock price from 2007 to 2017.

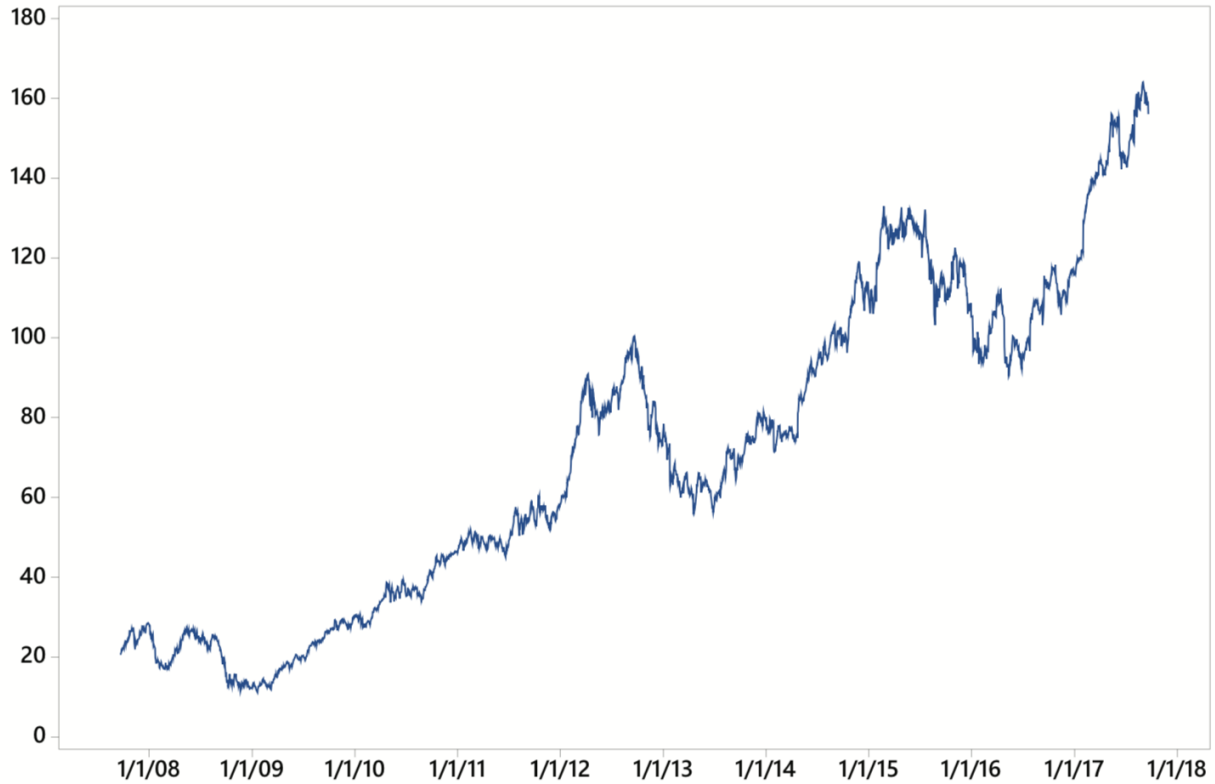
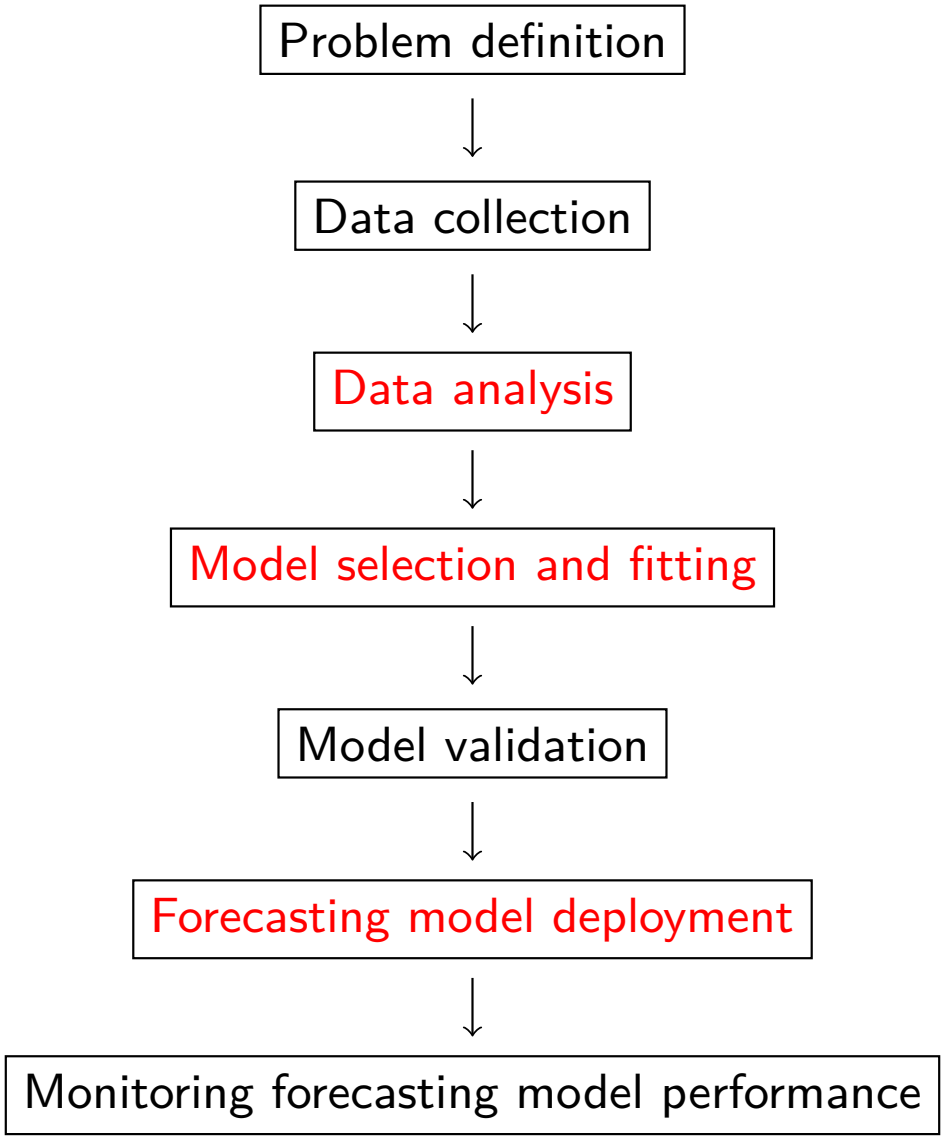


Figure 1.4: Apple Inc. stock price 2007–2017.

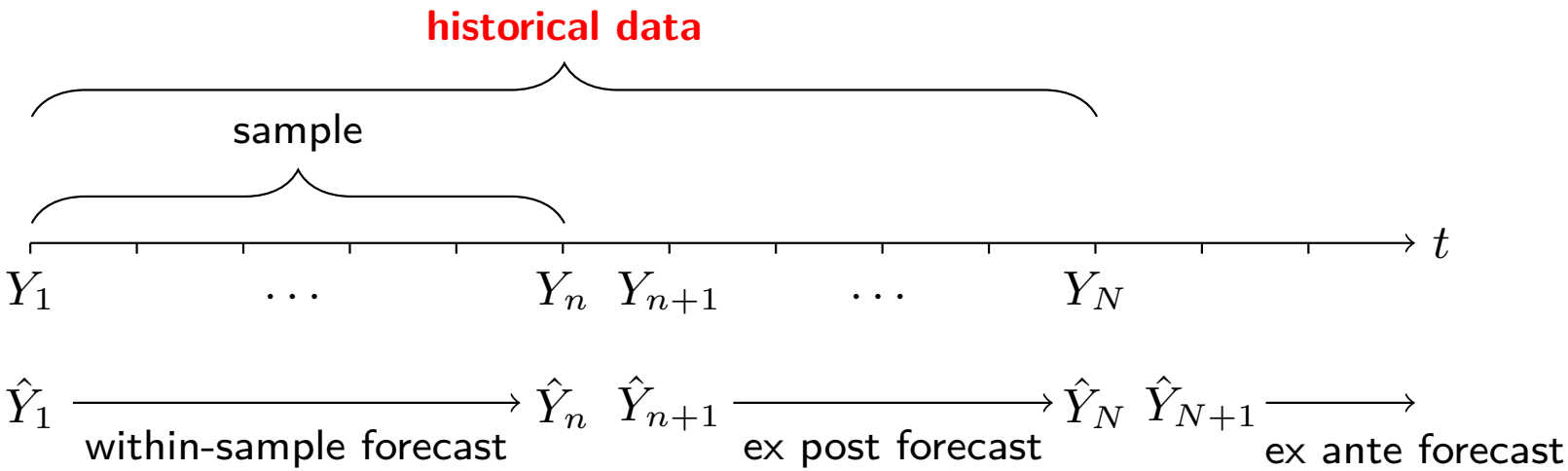
(Source: http://www.nasdaq.com)

A forecasting procedure follows:



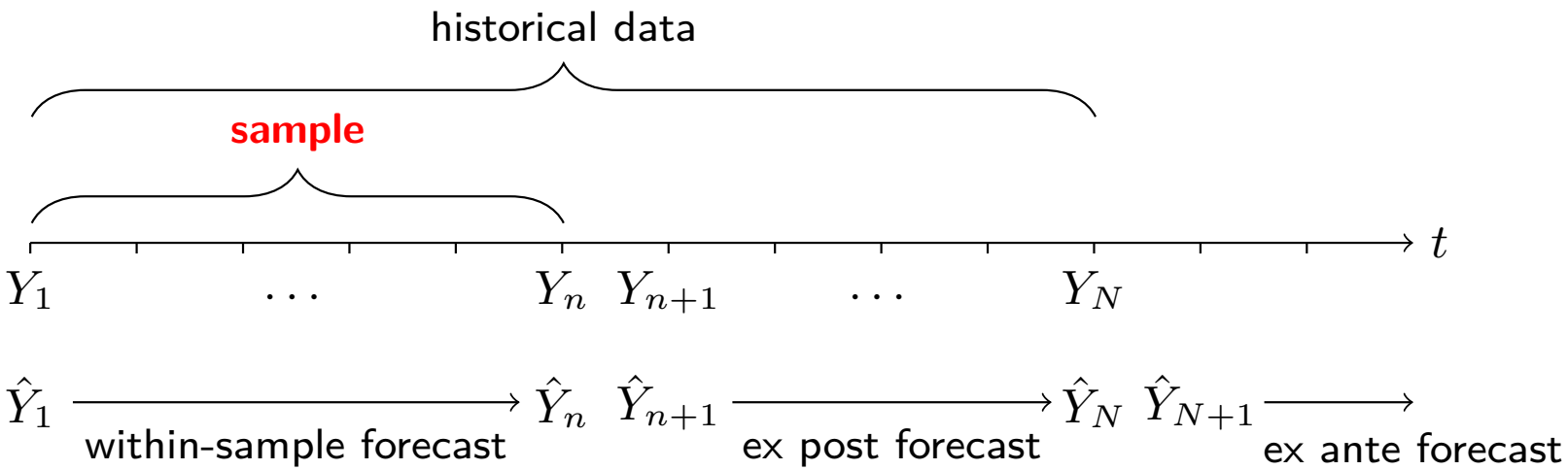
Data splitting (1/4)

Consider a time series $\{Y_t\}$. To build a forecast model for Y_t, a researcher usually collects a sequence of observations from the available data. The data are usually partitioned in the following fashion:



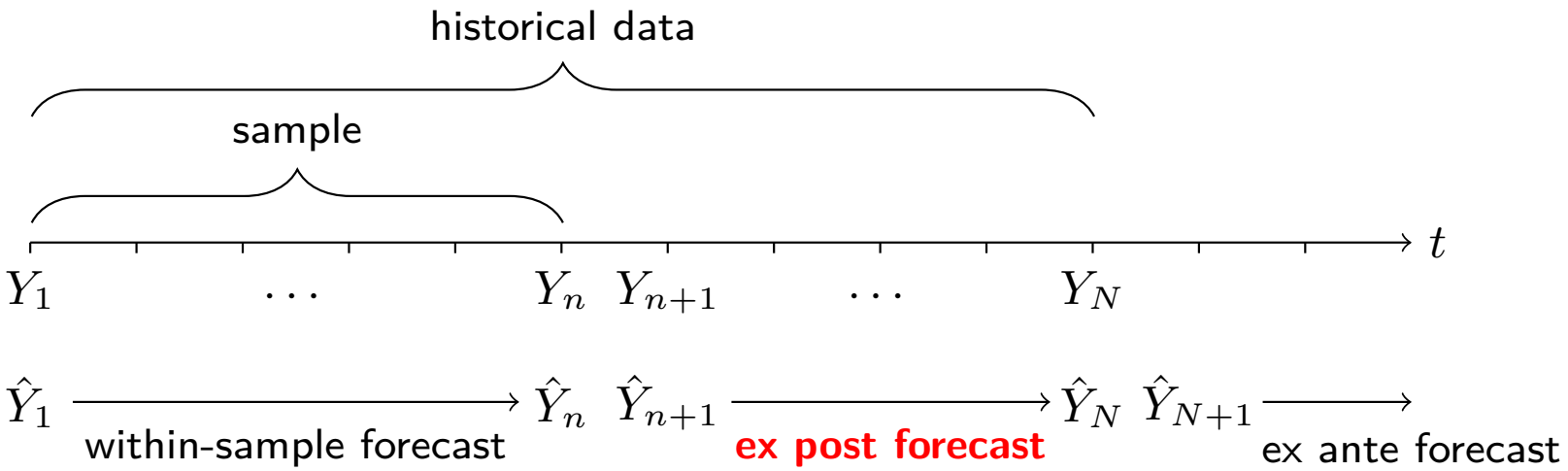
- *Historical data* are provided from time period $Y_{BEG} = Y_1$ to $Y_{END} = Y_N$. We define Y_{BEG} as that time period in which data collection on the variable began and Y_{END} as the most recent observation on the variable. For our purpose Y_{END} might be considered today's observation.

Data splitting (2/4)



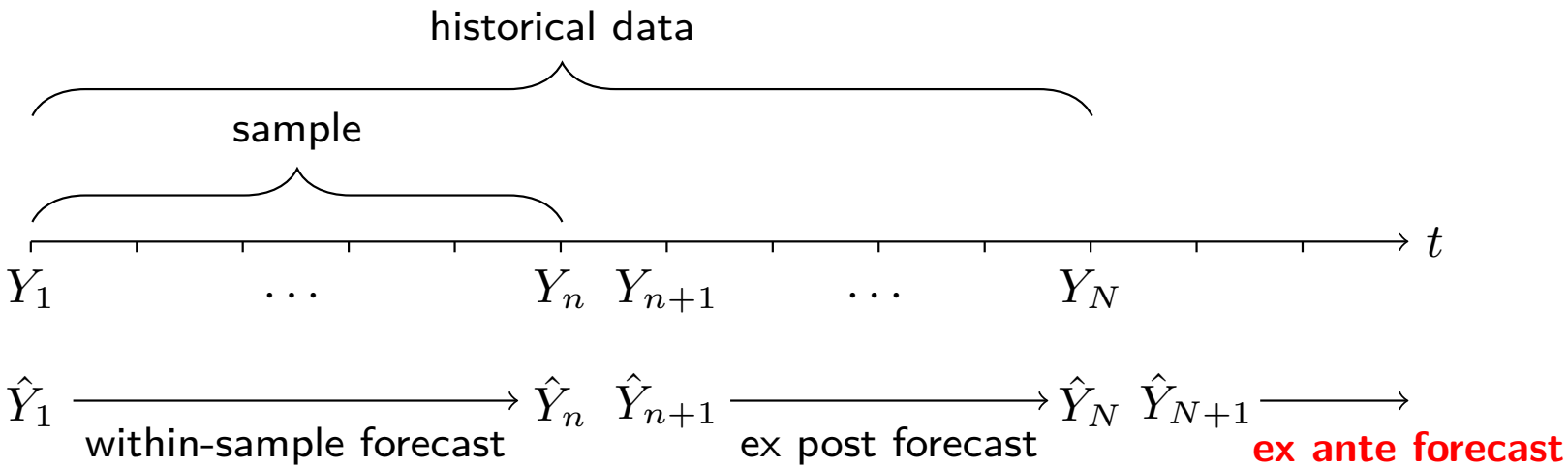
- *The sample period for analysis*, $Y_1, \ldots, Y_n$, is the period over which we will be building or estimating the forecasting model. We use $\{Y_1, \ldots, Y_n\}$ to estimate a model and use the estimated model to forecast $\{Y_{n+1}, \ldots, Y_N\}$. Here we denote by $\hat{Y}_t$ the forecast of Y_t.

Data splitting (3/4)



- *The ex post period* is defined as the time period from the first observation after the end of the sample period for analysis (Y_{n+1}) to the most recent observation (Y_N).
- The most important characteristic of this time period is that the researcher has available both observations and estimations. Hence we can use the observed $\{Y_{n+1}, \ldots, Y_N\}$ to compare against the forecast $\{\hat{Y}_{n+1}, \ldots, \hat{Y}_N\}$.

Data splitting (4/4)



- *The ex ante forecast period* is defined as that time period in which no observations on the time series variable exist.

Measuring the forecast error (1/)

Definition

We define the difference between the actual and the forecast values as the *forecast error*,

$$e_t = Y_t - \hat{Y}_t.$$

If the expected forecast error $\mathbb{E}[e_t] = 0$, we say that the forecast $\hat{Y}_t$ is *unbiased* for Y_t.

We provide four criteria/indice to measure the forecast error.

- The mean absolute error

$$\text{MAE} = \frac{\sum_{t=1}^{n} |e_t|}{n}.$$

- The mean squared error

$$\text{MSE} = \frac{\sum_{t=1}^{n} e_t^2}{n}.$$

  - Unit of MSE is different from the observations.

Measuring the forecast error (2/)

- The root mean squared error

$$\text{RMSE} = \sqrt{\text{MSE}}$$

  - MAE, MSE and RMSE are based on "absolute errors". In particular, they are sensitive to outliers, i.e., they are less robust.

- The mean of the absolute percentage error

$$\text{MAPE} = \frac{\sum_{t=1}^{n} \left(|e_t| / |Y_t| \right)}{n},$$

  where Y_t is the actual value in time period t; n is the number of forecast observations in the estimation period.

  - It cannot be use when some $Y_t = 0$.

Remark.

- If all error criteria of a model is smaller, then this model is better.
- Otherwise, it depends $\curvearrowright$.

Example 1.5

An owner of a small business has been using two models to forecast annual sales during the past four years and now is interested in determining which of the two forecast models is better. The data for annual revenues and the forecasts generated by each model (Model 1 and Model 2) are given in Table 1.1.

Model 1	Actual	Predicted	Error (e_t)	$\lvert e_t \rvert$	e_t^2
	15	15.5	-0.5	0.5	0.25
	20	20.0	0.0	0.0	0.00
	19	18.5	0.5	0.5	0.25
	23	27.0	-4	4.0	16
				MAE $= 1.25$	MSE $= 4.125$

Model 2	Actual	Predicted	Error (e_t)	$\lvert e_t \rvert$	e_t^2
	15	14	1	1	1
	20	18	2	2	4
	19	21	-2	2	4
	23	24	-1	1	1
				MAE $= 1.5$	MSE $= 2.5$

Table 1.1: Comparison of MAE and MSE for two models.

Example 1.5, *cont'd*

- The owner selects two measures, MAE and MSE, to evaluate the accuracy of the forecasts.
- Model 1 is the better model in the sense that the forecast values are close to the actual values in all but one time period.
- However, the one large error produces a MSE larger than that of Model 2.
- From MAE point of view: Model 1 is better. From MSE point of view: Model 2 is better.
- Conclusion:
  - If the owner is conservative (does not like extreme values), he/she will penalize extreme values, thus he/she trusts MSE and choose Model 2.
  - If the owner is aggressive and thinks one extreme value is acceptable, he/she will trust MAE and choose Model 1.

Difference between MAE and MSE

- Minimization of MSE leads to mean while minimization of MAE leads to median!

- Estimators by median are more robust to ourliers.

y_t	-1	1	-1	1	0

Calculate the mean/median, and the mean/median by changing the first -1 to -10.

Some simple forecast methods (1/5)

(1) **The naïve forecast**

$$\hat{Y}_t = Y_{t-1}$$

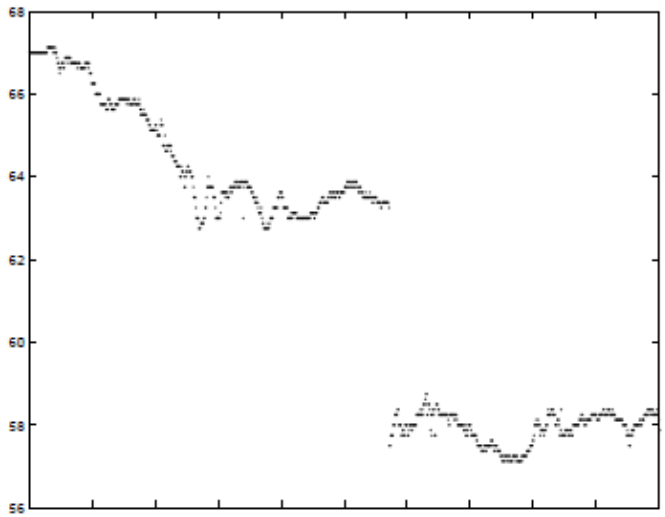


Figure 1.5: Stock price of SLM Corporation (NYSE), February 1993.

- From the name of the forecast, the prediction is based on "guess", with the minimal data.
- It is useful when data has inertia.

Some simple forecast methods (2/5)

(2) **Simple moving average**

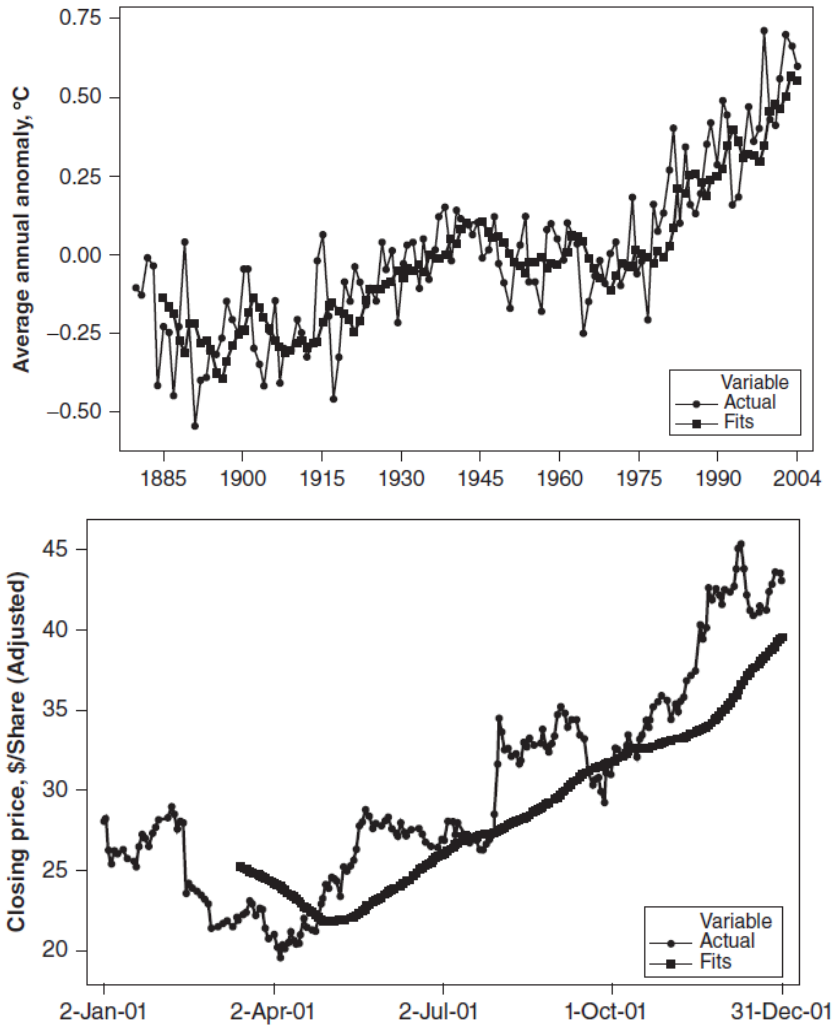
$$k\text{-point moving average forecast}^{3} = \hat{Y}_t = \frac{Y_{t-1} + \cdots + Y_{t-k}}{k}$$

- To use k-period moving average to forecast Y_t, the most recent k observations are assigned a common weight $1/k$, and other observations are assigned zero weight.
- The moving average has less variability than the original observations. For example, if the variance of each Y_t is σ^2, and $\{Y_t, t = 1, \cdots\}$ is uncorrelated, then $\text{Var}(\hat{Y}_t) < \sigma^2$.

[3] Also called k-period moving average.

Some simple forecast methods (3/5)

A remark on simple moving average: data smoothing to make identification of the patterns more obvious.



Some simple forecast methods (4/5)

(3) **Linear trend forecast**

If we assume a linear regression over time model,

$$Y_t = \alpha + \beta t + \epsilon_t,$$

then the forecast is given by

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta} t,$$

where $\hat{\alpha}$ and $\hat{\beta}$ are estimators of α and β, respectively (see Chapter 2).

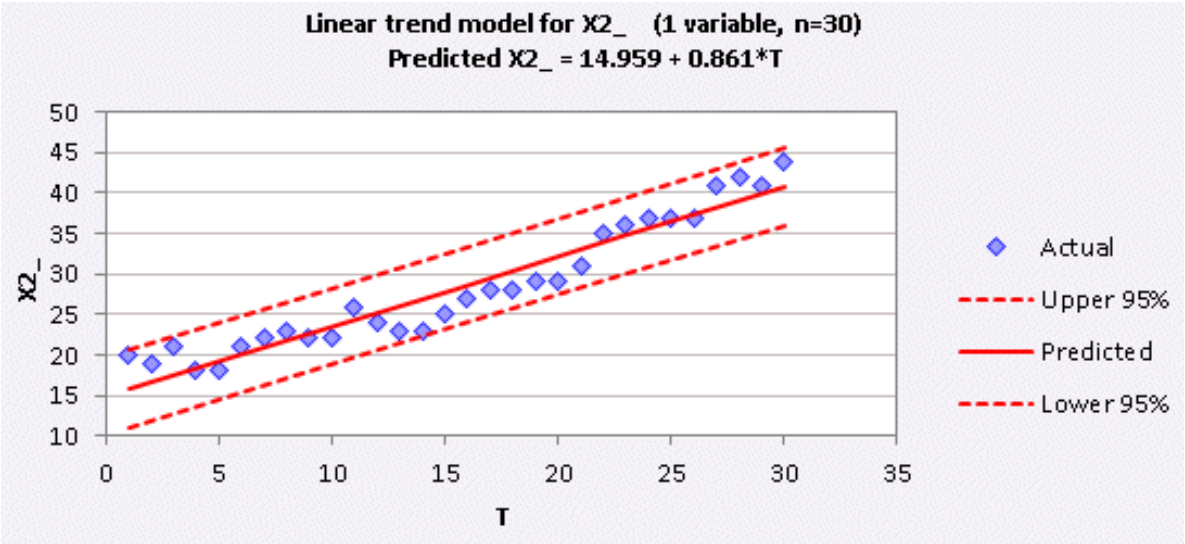


Figure 1.6: Linear trend.

Some simple forecast methods (5/5)

(4) **Causal regression model**

$$Y_t = \alpha + \beta X_t + \epsilon_t,$$

where Y_t = actual values for the time series variable; explained variable;
X_t = explanatory variable;
ϵ_t = random error.

Then

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t,$$

where $\hat{\alpha}$ and $\hat{\beta}$ are estimators of α and β, respectively (see Chapter 2).

Remark.

- In the causal regression model, we want to use the observed data of X to forecast/explain Y. In order to do so, estimate α and β.
- Causal regression is also called linear regression.
- When $X_t = t$, causal regression model reduces to linear trend model.

Example 1.5

The data below refers to some annual revenues of a company.

Year	70	71	72	73	74	75	76
Rev.	375.2	414.2	458.5	511.9	542	588.1	656.4
Year	77	78	79	80	81	82	83
Rev.	722.5	804.2	896.8	957.4	1038.7	1069.3	1167.9

Table 1.2 gives the actual revenue, naïve, 3-period moving average, and linear trend forecasts for 1976–1983. The linear trend equation for 70 to 75 is

$$\widehat{R}_t = -2628.19 + 42.89429t.$$

Year	Actual	Naïve	3-period MA	Trend
76	656.4	588.1	547.3	631.8
77	722.5	656.4	595.5	674.7
78	804.2	722.5	655.7	717.6
79	896.8	804.2	727.7	760.5
80	957.4	896.8	807.8	803.4
81	1038.7	957.4	886.1	846.3
82	1069.3	1038.7	964.3	889.1
83	1167.9	1069.3	1021.8	932.0

Table 1.2: Revenues and forecasts.

Theil's U

Definition

Theil's U is defined as

$$\text{Theil's U} = \frac{\text{RMSE(proposed model)}}{\text{RMSE(naïve model)}}.$$

Remark. Theil's U statistic is a relative accuracy measure that compares the forecasted results with the results of forecasting with minimal historical data (naïve model).

1. If $U < 1$,
2. If $U = 1$,
3. If $U > 1$,

Time series data and cross-sectional data

Cross-sectional data: n observations at a given time point.

- Linear regression: $Y_i = \alpha + \beta X_i + \epsilon_i$.
- Gauss-Markov assumption.
- $\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$ and $\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x}$.

Time series data: one observation at each time point; n time points in total.

- Linear regression: $Y_t = \alpha + \beta X_t + \epsilon_t$.
- Gauss-Markov assumption?
- $\widehat{\beta} = \frac{\sum_{t=1}^{n}(x_t - \overline{x})(y_t - \overline{y})}{\sum_{t=1}^{n}(x_t - \overline{x})^2}$ and $\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x}$.