Time Series For Business: Lecture one

Xiaodong Lin

Outline

- What is time series?
- Examples of time series.
- Objectives of time series analysis.
- Areas of applications.

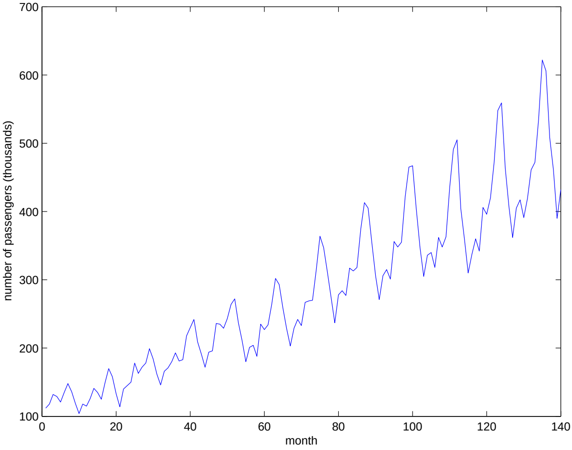
What is time series?

- A time series is a set of observations x_t, each one being recorded at a specific time t.
- A **time series** is a a collection of observations made sequentially through time.
- *"One damn thing after another".* R. A. Fisher.
- There are correlations among successive observations: Autocorrelated.
- Fundamentally different from i.i.d. observations.

Some areas of applications

- Financial investment.
  - Monthly closings of the Dow-Jones industrial index.
  - Annual bond yield in the USA.
  - Daily S&P 500 index of stocks.
  - Johnson and Johnson daily stock closing prices.
- Macro Economics.
  - Monthly percent changes in US wages and salaries.
  - U.S. Annual industrial production.
  - Quarterly U.S. GNP (Billions).
- Micro Economics.
  - Annual Copper prices.
  - Daily morning gold prices.
- Sales.
  - Monthly car sales.
  - Quarterly sales of toys.

Monthly airline passengers



What do you see?

Monthly airline passengers

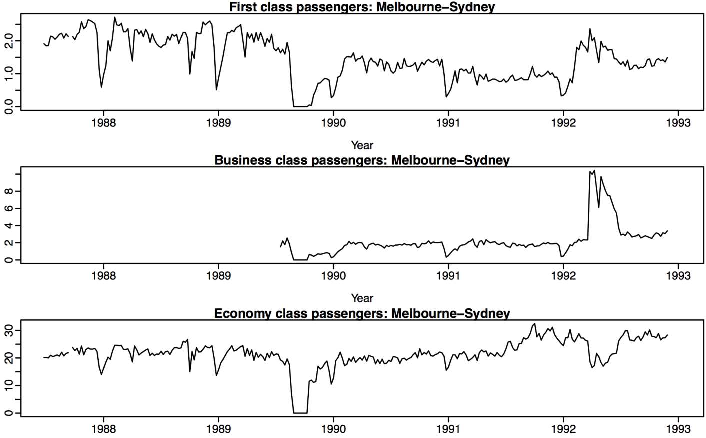
CLASSICAL DECOMPOSITION

$$X_t = T_t + S_t + C_t + E_t$$

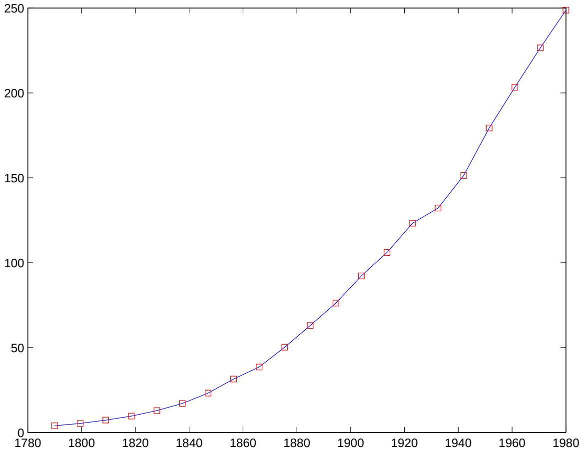
- Trend (T_t) - Long term movement in the mean.
- Seasonal variation (S_t) - Cyclical fluctuations due to calendar.
- Cycles (C_t) - Cyclical fluctuations of larger period. (e.g: Business cycles).
- Residuals (E_t) - random and all other unexplained variations.

DO WE NEED SOME TRANSFORMATIONS OF THE TIME SERIES?

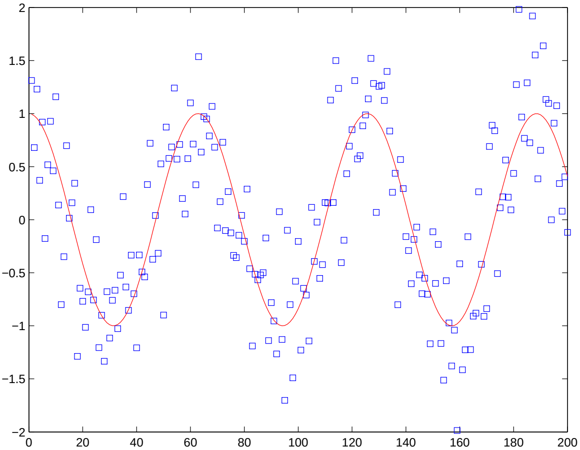
Forecast passenger traffic on major routes



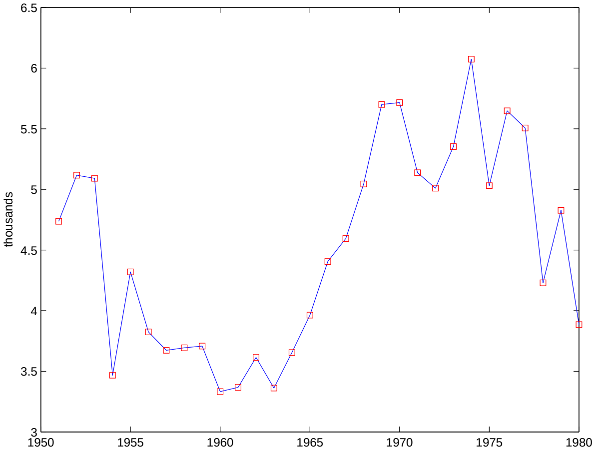
Example 3: US Population at 10-year intervals



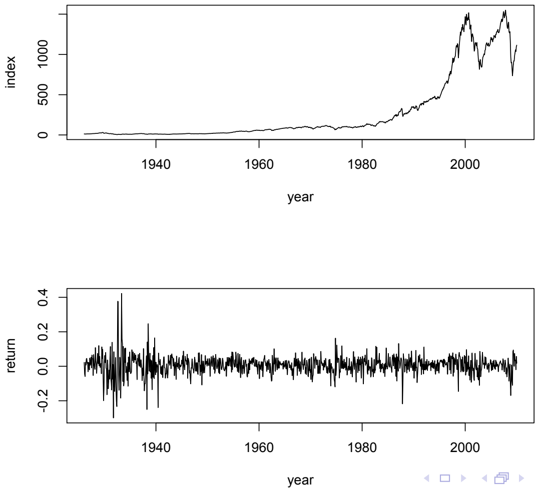
Example 4: The series from cosine



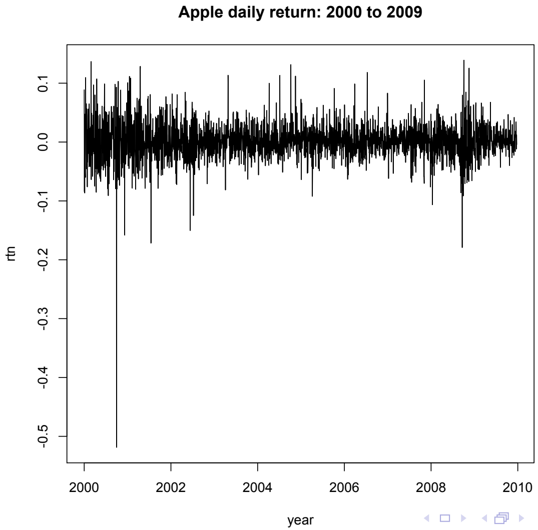
Example 5: Strikes in the USA, 1951-1980



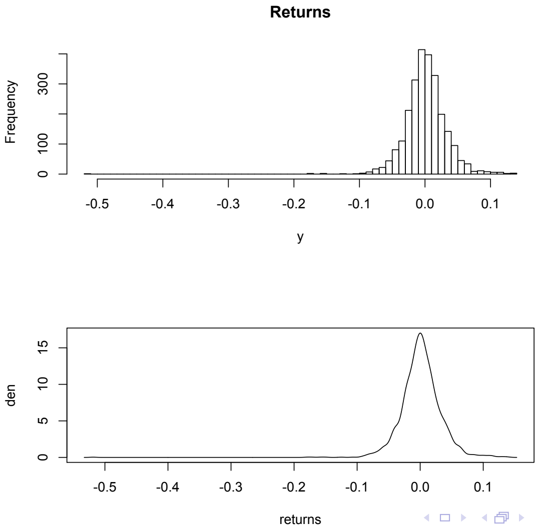
Example 6: S&P index and return



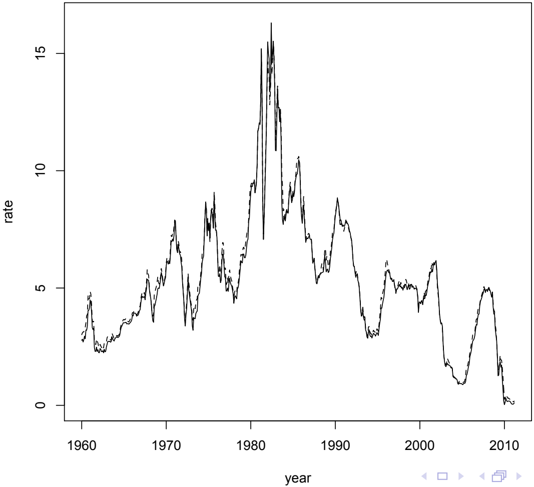
Example 7: Apple daily return



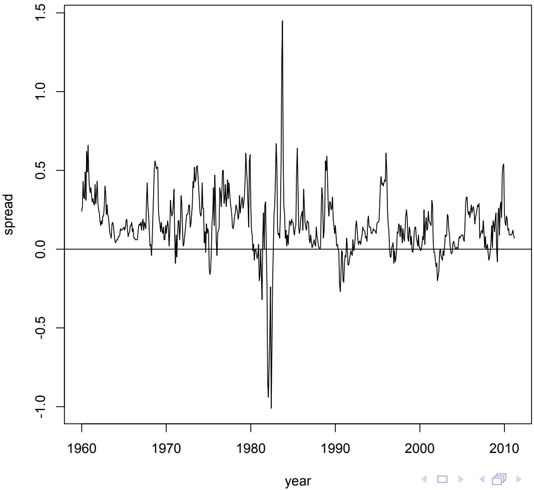
Example 8: Apple daily return



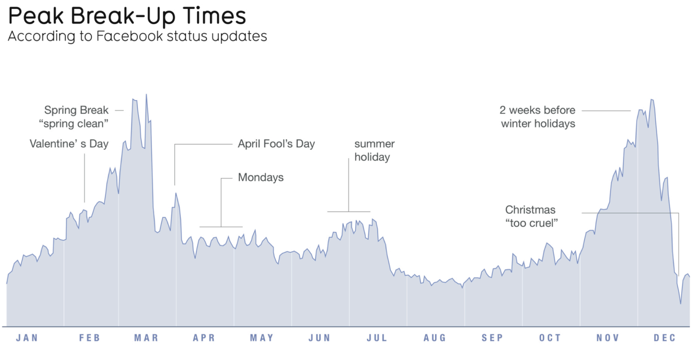
Example 9: Monthly T-bill rate, 3 and 6 month



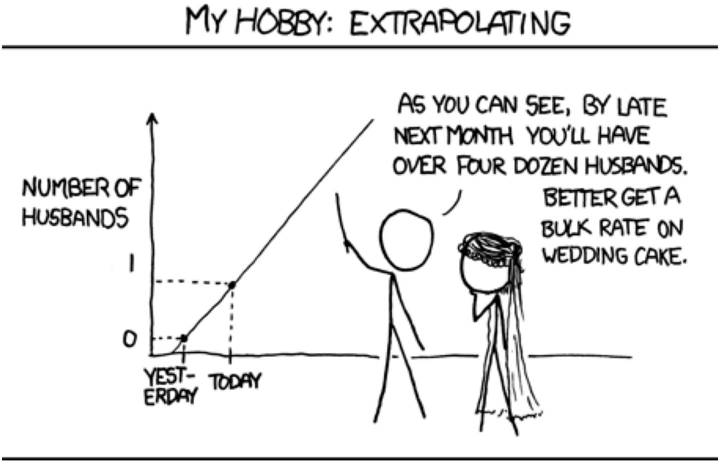
Example 10: T-bill spread



"Knowledge Discovery" from time series data



Are you sure?


MY HOBBY: EXTRAPOLATING
NUMBER OF HUSBANDS
1
0
YEST-ERDAY
TODAY
AS YOU CAN SEE, BY LATE NEXT MONTH YOU'LL HAVE OVER FOUR DOZEN HUSBANDS.
BETTER GET A BULK RATE ON WEDDING CAKE.

Tasks and objectives of Time Series Analysis

- Preliminary analysis
  - Graph the data. Are there consistent patterns, trends, seasonality, evidence of business cycles? Are there outliers or other variables that might affect/improve our analysis?
- Modeling or Model building
  - Set up a hypothetical model (or family of models) to represent the data generating mechanism.
- Estimation
  - Estimate the parameters of the postulated model.
- Model checking
  - Check the goodness of fit of the model to the data.
- Forecasting
  - Predict the future price of a share of a given stock.

Descriptive statistics

- Measures of central tendency: mean, median, mode ...

Month	Sales	Month	Sales	Month	Sales
1	3	10	8	19	5
2	4	11	1	20	7
3	5	12	13	21	4
4	1	13	4	22	5
5	5	14	4	23	2
6	3	15	7	24	6
7	6	16	3	25	4
8	2	17	4		
9	7	18	2		

- Median=mode=4. Mean=4.6.

Dispersion, variance and risk

- You have two choices, which one would you choose?
  1. Get 1000$ with probability 1.
  2. Get 1 million with prob. 1/1000, and 0 otherwise.
- Stock return.
  1. Stock A: 3% return with prob. 0.7; -1% return with prob. 0.3: **Mean=1.8%, std=1.83**
  2. Stock B: 6% return with prob. 0.7; -8% return with prob. 0.3: **Mean=1.8%, std=6.42**
  3. Stock C: 45.42% return with prob. 0.7; -100% return with prob. 0.3: **Mean=1.8%, std=66.64**

Basic probability

- Joint, marginal, and conditional

$$P(X, Y), \quad P(X), \quad P(X|Y)$$

- Bayes theorem

$$P(X|Y) = ?$$

- Independence, dependence, mutual exclusive, correlation
  Correlation in time domain.
- Dependence, causality, prediction.
  Even without a causal relationship, the association can still be used for prediction

Linear combination of r.v.

- $$E(a_1X_1 + a_2X_2) = a_1E(X_1) + a_2E(X_2).$$
- $$VAR(a_1X_1 + a_2X_2) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\sigma_1\sigma_2\rho$$
- $$E\left(\sum_{i=1}^{n} a_iX_i\right) = \left(\sum_{i=1}^{n} a_iE(X_i)\right)$$
- $$VAR\left(\sum_{i=1}^{n} a_iX_i\right) = \sum_{i=1}^{n} a_i^2\sigma_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_ia_j\sigma_i\sigma_j\rho$$

Confidence interval

- Sample statistics is a point estimate for the population parameter.
- It is very likely to be wrong, so we use interval estimate.
- Consider a large bank that wants to estimate the average amount owed by delinquent debtors μ. A random sample of size 100 is selected and found the sample mean is \$230. Suppose it's known that the standard deviation of the amount owed for all delinquent accounts is $\sigma = 90\$$. Get a 95% confidence interval for μ.

$$\bar{x} \pm 1.96\sigma_{\bar{x}} = 230 \pm 17.64.$$

Confidence interval

- Assume the sample size is 10, and the sampled population is approximately normal. If we only know the sample std, and the value is 80. We have student t with n-1 degree of freedom.

$$\bar{x} \pm t_{10-1,0.025}\sigma_{\bar{x}} = 230 \pm 2.262 \times 8 = 230 \pm 18.096.$$

- Note that t-distribution comes with a heavier tail.
- Get a 95% confidence interval for the true mean of the SP500 index returns.

  We take a sample of 162 days, mean is 0.00311, std is 0.04823, rough CI is (-.000432, 0.01054).

Hypothesis testing

- Concepts of hypothesis testing, p-value, χ^2 test and F-test. type I, type II errors.
- Suppose a portfolio manager is interested in whether the variance of the daily return of a specific index is greater than 0.00015. Thus we are testing:

$$H_0 : 0 < \sigma^2 \leq 0.00015 \;\; v.s. \;\; \sigma^2 > 0.00015.$$

The test statistics is

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Suppose in the past 50 days, the sample variance of the index is 0.0001, thus the test statistics is 32.67. Do not reject the null hypothesis.

The simple linear regression

The regression model:

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

where ϵ is usually assumed to have mean 0, standard deviation σ. Given a random sample (X_i, Y_i), we have

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

with ϵ_i i.i.d. with mean 0 and std σ.

- Y_i follows a distribution with mean $\beta_0 + \beta_1 X_i$ and variance σ^2.
- The mean of Y lies on a straight line of X
- Variance of Y is constant across X
- The slope β_1 is the amount of increase in the mean of Y when X increased by one unit

The simple linear regression

- The parameters β_0 and β_1 can be estimated based on the least square criterion given data (x_i, y_i)

$$(b_0, b_1) = \arg\min \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

- The estimators (b_0, b_1) are random variables with certain sampling distribution. Based on these we can perform hypothesis testing and compute confidence intervals of the parameters.
- Sampling distribution, test statistics, confidence intervals involved in simple regression.

Galton height example using R

```
> xx=matrix(scan(file="Galton_parent_child_height.txt"),
      ncol=2,byrow=T)
Read 1856 items
> dim(xx)
[1] 928   2
>  y=xx[,1] \# child
>  x=xx[,2] \# mid-parent
>  out=lm(y~x)
>   summary.lm(out)
```

Galton height example using R

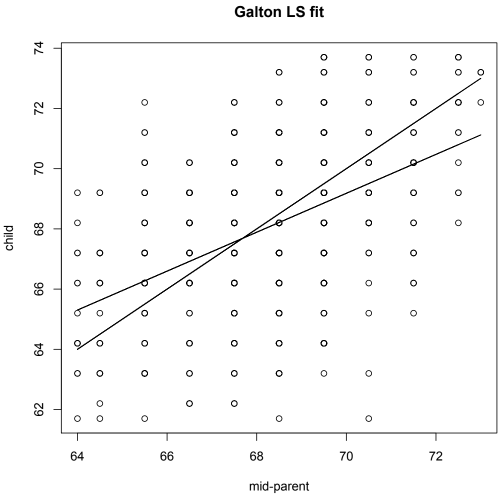
```
lm(formula = y ~ x)
Residuals:
    Min      1Q  Median      3Q     Max
-7.8050 -1.3661  0.0487  1.6339  5.9264
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 23.94153    2.81088   8.517   <2e-16 ***
x            0.64629    0.04114  15.711   <2e-16 ***
Multiple R-squared: 0.2105, Adjusted R-squared: 0.2096
F-statistic: 246.8 on 1 and 926 DF,  p-value: < 2.2e-16
> anova.lm(out)
Response: y
           Df Sum Sq Mean Sq F value    Pr(>F)
x           1 1236.9 1236.93  246.84 < 2.2e-16 ***
Residuals 926 4640.3    5.01
```

Galton height example using R

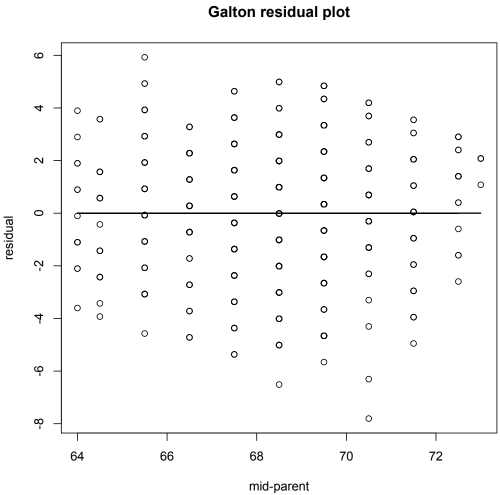
Plot of least square fit and residuals.

```
plot(x,y,ylab='child',xlab='mid-parent')
 lines(x,x)
 lines(x,out$fit)
 title('Galton LS fit')
 plot(x,out$res, ylab='residual',xlab='mid-parent')
 lines(x,x*0)
 title('Galton residual plot')
```

Galton height example using R



Residual plot



Capital Asset Pricing Model(CAPM)

- Assume that the market is efficient
  $E(R_i - R_f) = \beta_1 E(R_M - R_f)$. Thus given β_1 and R_M, one can determine the efficient price of asset i.
- If we consider $R_M - R_f$ and $R_i - R_f$ as having a joint distribution, the previous model is equivalent to

$$R_i - R_f = \beta_1(R_M - R_f) + \epsilon$$

  where ϵ has mean 0 and is independent of the R_M.
- Usually we also consider the intercept in the model

$$R_i - R_f = \beta_0 + \beta_1(R_M - R_f) + \epsilon$$

- In practice, the β_1 is estimated through simple linear regression. Usually S&P 500 return is used as market return.

CAPM example on Apple data

AAPL example. We use 3 month T-bill.

```
> xx=read.table(file="SP500_and_3mTCM.txt",header=TRUE)
> yy=read.table(file="m_logret_10stocks.txt",header=TRUE)
> marketrtn=xx[,"sp500"]-0.01*xx[,"X3mTCM"]
> aaplrtn=yy[,"AAPL"]-0.01*xx[,"X3mTCM"]
> plot(marketrtn,aaplrtn,xlab='SP500',ylab='AAPL',
ylim=c(-0.45,0.22))
> title('excess return: AAPL vs SP500')
> out1=lm(aaplrtn~marketrtn)
> lines(marketrtn,out1$fit)
> summary.lm(out1)
```

R outputs

```
lm(formula = aaplrtn ~ marketrtn)
Residuals:
     Min       1Q   Median       3Q      Max
-0.33016 -0.03384  0.00160  0.04178  0.14422
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.021279   0.009149   2.326   0.0213 *
marketrtn   1.491264   0.215342   6.925 1.11e-10 ***

Residual standard error: 0.06245 on 154 degrees of freedom
Multiple R-squared: 0.2375, Adjusted R-squared: 0.2325
F-statistic: 47.96 on 1 and 154 DF,  p-value: 1.11e-10
```

Regression plot