# Times Series Analysis (II) – Trend

Jianyong Sun
School of Mathematics and Statistics
Xi'an Jiaotong University

Sep., 2017

## Recap

- Examples of time series
- Concept: time series and stochastic process
- Mean, autocovariance and autocorrelation
- Example stochastic processes (random walk, moving average and white noise
- Stationarity

# Deterministic Trends and Stochastic Trends

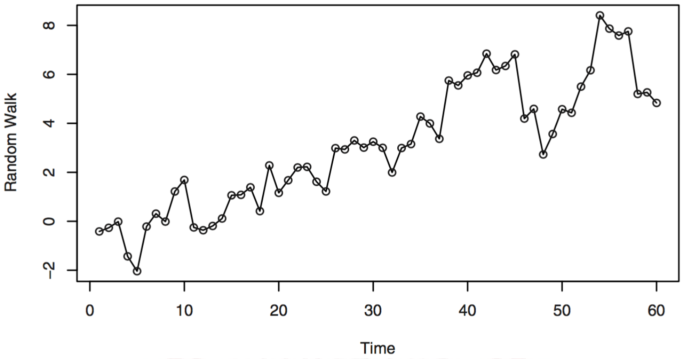



**Figure:** Time Series Plot of a Random Walk

# Deterministic Trends and Stochastic Trends

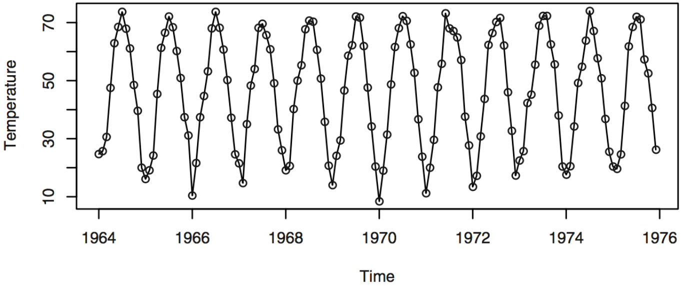



**Figure:** Average Monthly Temperatures, Dubuque, Iowa

## Formula

Let $\{Y_t\}$ be stationary with autocovariance function $\gamma_k$. Let $\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$. We have:

$$Var(\bar{Y}) = \frac{\gamma_0}{n} + \frac{2}{n}\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\gamma_k\right) = \frac{1}{n}\sum_{k=-n+1}^{n-1}\left(1 - \frac{|k|}{n}\right)\gamma_k$$

# Estimation of a Constant Mean

Consider a simple model with constant mean:

$$Y_t = \mu + X_t$$

where $E(X_t) = 0$ for all $t$.

# Estimation of a Constant Mean

To investigate the precision of $\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$ over the observed time series $Y_1, Y_2, \cdots, Y_n$, we assume $\{X_t\}$ is a stationary time series with autocorrelation function (ACF) $\rho_k$, then

$$Var(\bar{Y}) = \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right]$$

- If $\{X_t\}$ is white noise, $\rho_k = 0$ for $k > 0$, then $Var(\bar{Y}) = \frac{\gamma_0}{n}$
- If $\{X_t\}$ is a moving average $e_t - 0.5e_{t-1}$, then $\rho_1 = -0.4$ and $\rho_k = 0$ for $k > 1$, we have $Var(\bar{Y}) = \frac{\gamma_0}{n}\left[1 - 0.8\frac{n-1}{n}\right] \approx 0.2\frac{\gamma_0}{n}$
- If $\rho_k > 0$ for all $k > 1$, $Var(\bar{Y})$ will be larger than $\gamma_0/n$.

## Estimation of a Constant Mean

For stationary processes that $\sum_{k=0}^{\infty} |\rho_k| < \infty$,

$$Var(\bar{Y}) \approx \frac{\gamma_0}{n} \left[ \sum_{k=-\infty}^{\infty} \rho_k \right]$$

for large $n$.

**Example**

Suppose that $\rho_k = \phi^{|k|}$ for all $k$, $|\phi| < 1$. Then we have:

$$Var(\bar{Y}) \approx \frac{(1+\phi)\gamma_0}{(1-\phi)n}$$

# Estimation of a Constant Mean

For a nonstationary process but with a constant mean, like White Noise, the precision of the sample mean

$$
\begin{aligned}
Var(\bar{Y}) &= Var\left(\frac{1}{n}\sum_i Y_i\right) \\
&= \frac{1}{n^2} Var\left[\sum_{i=1}^{n}\sum_{j=1}^{i} e_j\right] \\
&= \frac{1}{n^2} Var(e_1 + 2e_2 + \cdots + ne_n) \\
&= \frac{\sigma_e^2}{n^2}\sum_{k=1}^{n} k^2 = \sigma_e^2(2n+1)\frac{n+1}{6n}
\end{aligned}
$$

variance increases alongside the increase of $n$

## Linear and Quadratic Trends in Time

Consider the deterministic time trend

$$\mu_t = \beta_0 + \beta_1 t, Y_t = \mu_t + X_t, X_t \backsim WN(\mu_e, \sigma_e^2)$$

## Estimating $\beta_0$ and $\beta_1$

Classical least squares (or regression) method: to choose an estimates of $\beta_1$ and $\beta_0$ that minimize:

$$Q(\beta_0, \beta_1) = \sum_{t=1}^{n} [Y_t - (\beta_0 + \beta_1 t)]^2$$

Denote the solutions by $\hat{\beta}_0$ and $\hat{\beta}_1$, we find

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{t=1}^{n}(Y_t - \bar{Y})(t - \bar{t})}{\sum_{t=1}^{n}(t - \bar{t})^2} \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{t}\end{aligned}$$

where $\bar{t} = (n+1)/2$ is the average of $1, 2, \cdots, n$.

# Regression Methods: Estimating non-constant mean trend

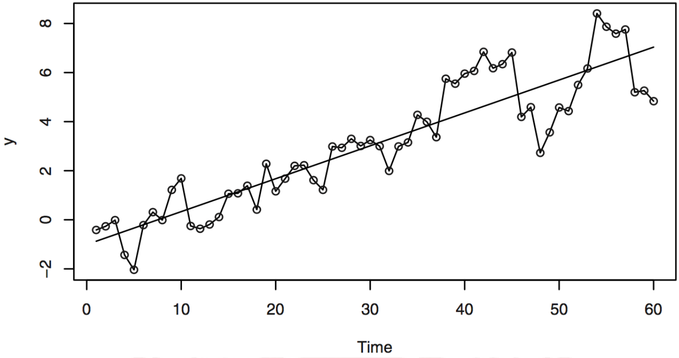



**Figure:** Random Walk with Linear Time Trend

**Question**

Why fitting a line to the data is not appropriate?

## Cyclical or Seasonal Trends

- Again assume $Y_t = \mu_t + X_t$ with $E(X_t) = 0$ for all $t$
- For monthly seasonal data, assume $\mu_t$ consists of 12 constant parameters w.r.t. each month.

$$\mu_t = \begin{cases} \beta_1 & \text{for} \quad t = 1, 13, 25, \cdots, \\ \beta_2 & \text{for} \quad t = 2, 14, 26, \cdots, \\ \vdots & \\ \beta_{12} & \text{for} \quad t = 12, 24, 36, \cdots, \end{cases}$$

$\rightarrow$**seasonal mean model**

| | Estimate | Std. Error |
|---|---|---|
| Intercept | 16.608 | 0.987 |
| February | 4.042 | 1.396 |
| March | 15.867 | 1.396 |
| April | 29.917 | 1.396 |
| May | 41.483 | 1.396 |
| June | 50.892 | 1.396 |
| July | 55.108 | 1.396 |
| August | 52.725 | 1.396 |
| September | 44.417 | 1.396 |
| October | 34.367 | 1.396 |
| November | 20.042 | 1.396 |
| December | 7.033 | 1.396 |

**Figure:** Results for Seasonal Means Model with an Intercept

# Cyclical or Seasonal Trends

Seasonal mean model does not account for the shape of the seasonal trend.

## Cosine Trends

Consider the cosine curve with equation

$$\mu_t = \beta \cos(2\pi f t + \Phi)$$

where $\beta (> 0)$ – amplitude, $f$ – frequency, $\Phi$ – phase.
Note that $\beta \cos(2\pi f t + \Phi) = \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$ where $\beta = \sqrt{\beta_1^2 + \beta_2^2}$, $\Phi = atan(-\beta_2/\beta_1)$. Therefore,

$$\mu_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$$

# Cyclical or Seasonal Trends

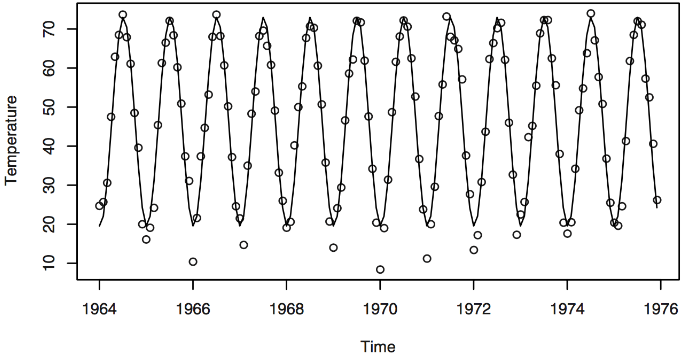



**Figure:** Cosine Trend for the Temperature Series.

# Reliability and Efficiency of Regression Estimates

So far, we assume $Y_t = \mu_t + X_t$, $\mu_t$ is deterministic trend, $\{X_t\}$ zero-mean stationary stochastic process, with autocovariance and autocorrelation $\gamma_k$ and $\rho_k$.
ordinary regression estimation methods – least squares – is used.

# Reliability and Efficiency of Regression Estimates

## Seasonal Means

If we have $N$ years data,

$$\hat{\beta}_j = \frac{1}{N}\sum_{i=0}^{N-1} Y_{j+12i}$$

$$Var(\hat{\beta}_j) = \frac{\gamma_0}{N}\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_{12k}\right]$$

If $X_t$ is white noise, $Var(\hat{\beta}_j) = \frac{\gamma_0}{N}$

## Cosine Trends

For $\mu_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$, if $f = \frac{m}{n}$ where $m$ is an integer satisfying $1 \le m \le n/2$, then we have:

$$\hat{\beta}_1 = \frac{2}{n} \sum_{t=1}^{n} \left[ \cos\left(\frac{2\pi m t}{n}\right) Y_t \right], \hat{\beta}_2 = \frac{2}{n} \sum_{t=1}^{n} \left[ \sin\left(\frac{2\pi m t}{n}\right) Y_t \right]$$

Their variances are:

$$Var(\hat{\beta}_1) = 2\frac{\gamma_0}{n} \left[ 1 + \frac{4}{n} \sum_{s=2}^{n} \sum_{t=1}^{s-1} \cos\left(\frac{2\pi m t}{n}\right) \cos\left(\frac{2\pi m s}{n}\right) \rho_{s-t} \right]$$

Similarly for $Var(\hat{\beta}_1)$ if we replace the cosines by sines.

For this variance,

- If $\{X_t\}$ is WN, we get $2\gamma_0/n$
- If $\rho_1 \neq 0$ and $\rho_k = 0$ for $k > 1$, and $m/n = 1/12$,

$$Var(\hat{\beta}_1) = 2\frac{\gamma_0}{n}\left[1 + \frac{4\rho_1}{n}\sum_{t=1}^{n-1}\cos\left(\frac{\pi t}{6}\right)\cos\left(\frac{\pi t + 1}{6}\right)\right]$$

In case $n = \infty$,

$$\begin{aligned} Var(\hat{\beta}_1) &= \frac{2\gamma_0}{n}\left(1 + 2\rho_1\cos\left(\frac{\pi}{6}\right)\right) \\ &= \frac{2\gamma_0}{n}(1 + 1.732\rho_1) \end{aligned}$$

If $\rho_1 = -0.4$, then $1 + 1.732\rho_1 = 0.307$, the variance reduced about 70% when compared with WN.

## Reliability and Efficiency of Regression Estimates

If the simple cosine model is adequate, how much do we lose if we use the seasonal means model?

# Model comparison

Model parameters are not comparable. To compare the estimates of the trend at comparable time points. Consider the two estimates for the trend in Jan. , i.e. $\mu_1$

- For seasonal means model,

$$Var(\hat{\mu_1}) = \frac{\gamma_0}{N}\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_{12k}\right]$$

- For cosine model,

$$\begin{aligned}
\hat{\mu_1} &= \hat{\beta}_0 + \hat{\beta}_1 \cos\left(\frac{2\pi}{12}\right) + \hat{\beta}_2 \sin\left(\frac{2\pi}{12}\right) \\
Var(\hat{\mu_1}) &= Var(\hat{\beta}_0) + Var(\hat{\beta}_1)\cos\left(\frac{2\pi}{12}\right)^2 + Var(\hat{\beta}_2)\sin\left(\frac{2\pi}{12}\right)^2
\end{aligned}$$

# Model comparison

If $X_t$ is WN,

- for seasonal means model, $Var(\hat{\mu_1}) = \gamma_0/N$
- for cosine model,
  $Var(\hat{\mu_1}) = \gamma_0/n \left\{1 + 2\left[\cos\left(\frac{\pi}{6}\right)\right]^2 + +2\left[\sin\left(\frac{\pi}{6}\right)\right]^2\right\} = 3\frac{\gamma_0}{n}$
- The ratio of the standard deviation is

$$\sqrt{\frac{3\gamma_0/n}{\gamma_0/N}} = \sqrt{\frac{3N}{n}}$$

- For monthly temperature series, $n = 144$, $N = 12$, the ratio is 0.5.

# Model comparison

Suppose the stochastic component $\{X_t\}$ is such that $\rho_1 \neq 0$ and $\rho_k = 0$ for $k > 1$, the variance of the seasonal means model is the same, while for the cosine model, it becomes

$$Var(\hat{\mu}_1) = \frac{\gamma_0}{n}\left\{3 + 2\rho_1\left[1 + 2\cos\left(\frac{\pi}{6}\right)\right]\right\}$$

If $\rho_1 = -0.4$, then we have $0.814\gamma_0/n$. The ratio of the standard deviation is $\sqrt{\frac{0.814N}{n}}$.
Take $n = 144$ and $N = 12$, the ratio is 0.26.

# Residue Analysis

The unobserved stochastic process $\{X_t\}$ can be estimated or predicted by the **residual**:

$$\hat{X}_t = Y_t - \hat{\mu}_t$$

and **residual standard deviation**

$$s = \sqrt{\frac{1}{n-p}\sum_{t=1}^{N}(Y_t - \hat{\mu}_t)^2}$$

where $p$ is the number of parameters estimated, $n - p$ — *degrees of freedom*

- If the model is reasonably correct, the residues should behave roughly like the true stochastic component
- various assumptions about $X_t$ can be assessed by looking at the residues.

# Residue Analysis

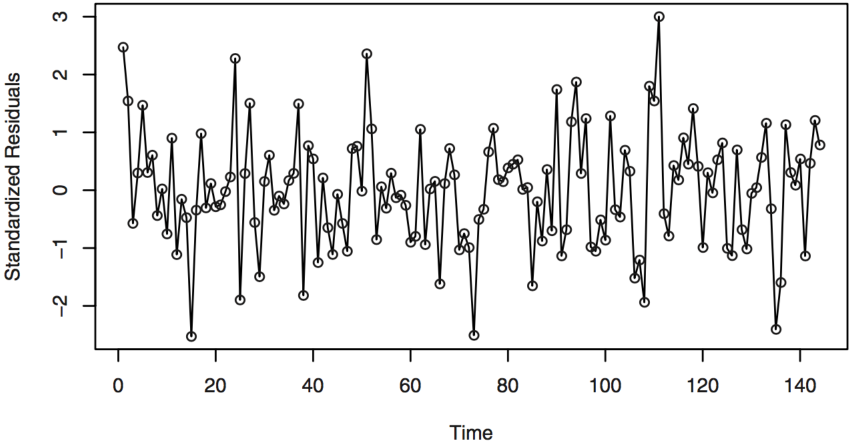



**Figure:** Residuals versus Time for Temperature Seasonal Means.

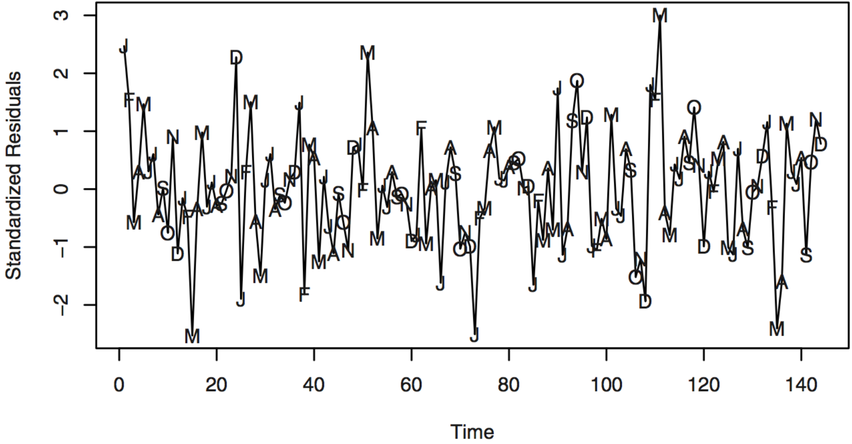



**Figure:** Residuals versus Time with Seasonal Plotting Symbols.

# Residue Analysis

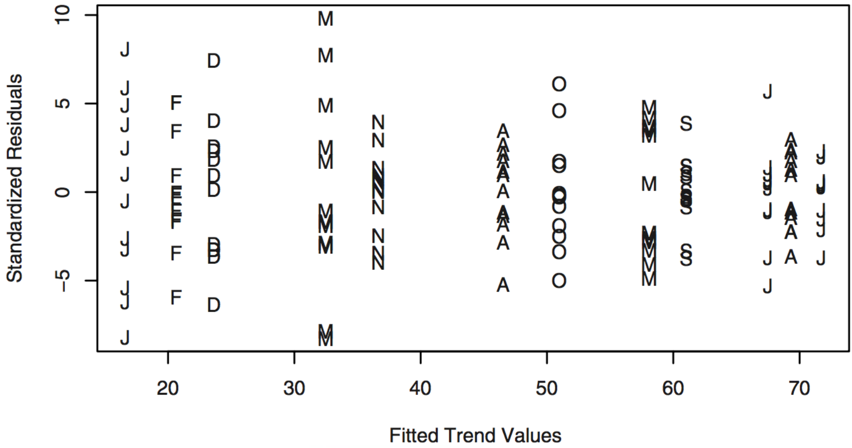
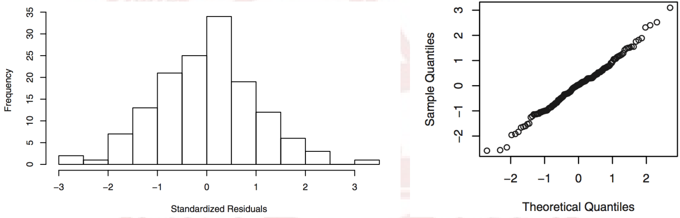



**Figure:** Standardized Residuals versus Fitted Values for the Temperature Seasonal Means Model

Normality Test: Histogram, QQ-plot, and Shapiro-Wilk Test (calculate correlation between residuals and normal quantitles)




(a) (b)

**Figure:** (a). Histogram of Standardized Residuals from Seasonal Means Model. (b). QQ-plot: Standardized Residuals of Seasonal Means Model.

**Runs test** to examine against independence. runs: the number of residuals above or below their **median**

### Runs test

Under the null hypothesis, the number of runs in a sequence of $N$ elements is a random variable whose conditional distribution given the observation of $N_+$ positive values and $N_-$ negative values ($N = N_+ + N_-$) is approximately normal.

# Sample Autocorrelation Function

To examine **dependence**, another important tool: **sample autocorrelation function** defined as follows:

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}\left(Y_t - \bar{Y}\right)^2}$$

A plot of $r_k$ versus lag $k$ is called a **correlogram**.

## Sample Autocorrelation Function

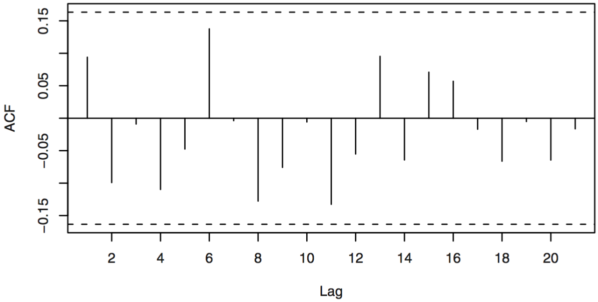



**Figure:** Sample ACF of Seasonal Means Model.

for $k = 1, 2, ..., 21$, none of the hypotheses $\rho_k = 0$ can be rejected at the usual significance levels, and it is reasonable to infer that the stochastic component of the series is white noise.

# Sample Autocorrelation Function

Example: standardized residuals from fitting a straight line to the random walk time series

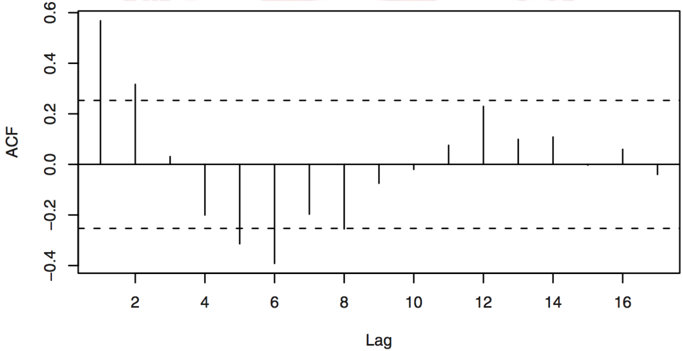



**Figure:** Sample ACF of Residuals from Straight Line Model.

Questions?