ISMT S-136 Time Series Analysis with Python

Harvard Summer School

Dmitry Kurochkin

Summer 2021
Lecture 2

Contents

Contents

Definition of White Noise

Def.

White Noise is defined as
uncorrelated random variables w_t
with mean 0 and variance σ_w^2. We denote the process as:

$$w_t \sim \mathsf{wn}(0, \sigma_w^2).$$

Contents

White Independent Noise

Def.

White Independent Noise is defined as independent identically distributed (iid) random variables w_t with mean 0 and variance σ_w^2. We denote the process as:

$$w_t \sim \text{iid}(0, \sigma_w^2).$$

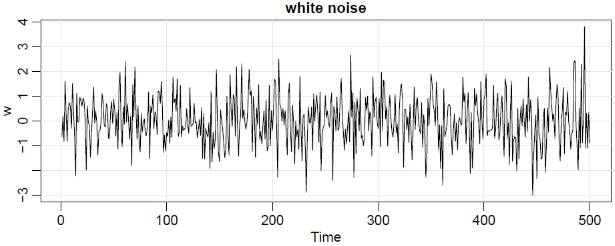
Contents

Gaussian White Noise

Def.

Gaussian White Noise is defined as independent identically distributed (iid) random variables w_t drawn from $\mathcal{N}(0, \sigma_w^2)$. We denote the process as:

$$w_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2)$$



Source: *Time Series Analysis and Its Applications: With R Examples* by R. Shumway and D. Stoffer

Contents

Problem 1.6

Consider the time series $x_t = \beta_1 + \beta_2 t + w_t$ where β_1 and β_2 are known constants and w_t is a white noise process with variance σ_w^2.

(a) Determine whether x_t is stationary.
(b) Show that the process $y_t = x_t - x_{t-1}$ is stationary.
(c) Show that the mean of the moving average

$$v_t = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t-j}$$

is $\beta_1 + \beta_2 t$, and give a simplified expression for the autocovariance function.

Contents

Problem 1.15

Let w_t , for $t = 0, \pm 1, \pm 2, \ldots$ be a normal white noise process, and consider the series

$$x_t = w_t w_{t-1}$$

Determine the mean and autocovariance function of x_t, and state whether it is stationary.

Contents

Problem 1.16

Consider the series

$$x_t = \sin(2\pi U t),$$

$t = 1, 2, \ldots$, where U has a uniform distribution on the interval $[0, 1]$.
(a) Prove x_t is weakly stationary.
(b) Prove x_t is not strictly stationary.

Contents

Linear Regression Model

Assume dependent random variable x_t (can be time series!), $t = 1, 2, \ldots n$, is generated via

$$x_t = \beta_0 + \beta_1 z_{t1} + \beta_2 z_{t2} + \ldots + \beta_q z_{tq} + w_t,$$

$z_{t1}, z_{t2}, \ldots, z_{tq}$ are some inputs we can measure and $w_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2)$. This model is called *linear regression*.

Contents

Estimating a Linear Trend via Regression

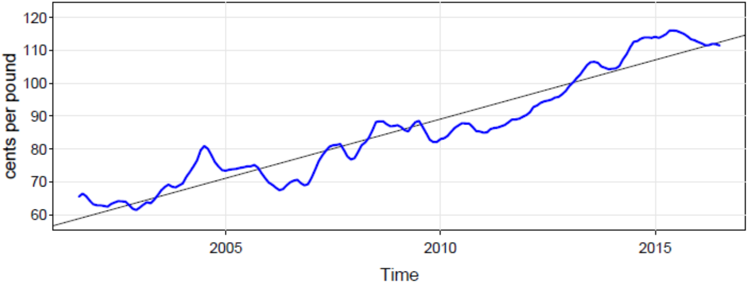
We can take $q = 1$ and $z_{t1} = t$, then the linear regression model becomes:

$$x_t = \beta_0 + \beta_1 t + w_t$$

where $w_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2)$.

Estimating a Linear Trend via Regression: Example

Ex.: The price of chicken: monthly whole bird spot price, Georgia docks, US cents per pound, August 2001 to July 2016, with fitted linear trend line.



Source: *Time Series Analysis and Its Applications: With R Examples* by R. Shumway and D. Stoffer

Estimating a Linear Trend via Regression: Example

We can take $z_t = 2001\frac{7}{12}, 2001\frac{8}{12}, 2001\frac{9}{12}, \ldots, 2016\frac{6}{12}$
that correspond to August 2001 to July 2016
and model the process as follows:

$$x_t = \beta_0 + \beta_1 z_t + w_t, \quad \text{where} \quad w_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2).$$

Estimating a Linear Trend via Regression: Example

We can take $z_t = 2001\frac{7}{12}, 2001\frac{8}{12}, 2001\frac{9}{12}, \ldots, 2016\frac{6}{12}$
that correspond to August 2001 to July 2016
and model the process as follows:

$$x_t = \beta_0 + \beta_1 z_t + w_t, \quad \text{where} \quad w_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2).$$

Given β_0 and β_1, define $\hat{x}_t = \beta_0 + \beta_1 z_t$. The ordinary least squares (OLS) is then

$$\mathcal{Q}(\beta_0, \beta_1) = \sum_{t=1}^{n} (x_t - \hat{x}_t)^2 = \sum_{t=1}^{n} (x_t - \underbrace{[\beta_0 + \beta_1 z_t]}_{\hat{x}_t})^2.$$

The minimization of $\mathcal{Q}$ with respect to its arguments β_0 and β_1 results in

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} (x_t - \bar{x})(z_t - \bar{z})}{\sum_{t=1}^{n} (z_t - \bar{z})^2} \quad \text{and } \hat{\beta}_0 = \bar{x} - \hat{\beta}_1 \bar{z}.$$

Contents

Removing Trend via Regression

Given time series x_t, assume that

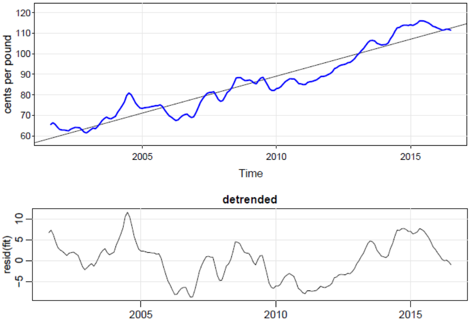
$$x_t = \mu_t + y_t,$$

where μ_t denotes the trend and y_t is the remainder.[1]

[1]If $E(y_t) \neq 0$, then μ_t and y_t can be replaced with $\mu_t + E[y_t]$ and $y_t - E[y_t]$, respectively.

Removing Trend via Regression: Example

Ex.: The price of chicken: monthly whole bird spot price, Georgia docks, US cents per pound, August 2001 to July 2016, with fitted linear trend line.



Source: *Time Series Analysis and Its Applications: With R Examples* by R. Shumway and D. Stoffer

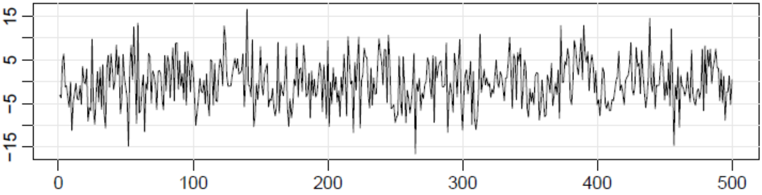
Removing Trend via Regression: Example

Ex.: $n = 500$ observations are generated as follows:

$$x_t = A\cos(2\pi\omega t + \phi) + w_t,$$

where $\omega = 1/50$, $A = 2$, $\phi = 0.6\pi$, and

$$w_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2) \quad \text{with} \quad \sigma = 5.$$



Source: *Time Series Analysis and Its Applications: With R Examples* by R. Shumway and D. Stoffer

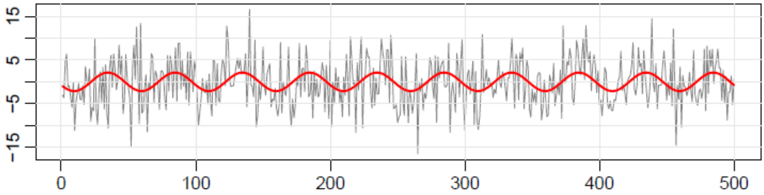
Removing Trend via Regression: Example

$x_t = A\cos(2\pi\omega t + \phi) + w_t$, where we notice that

$$\cos(2\pi\omega t + \phi) = \underbrace{A\cos\phi}_{\beta_1}\cos(2\pi\omega t) + \underbrace{(-A\sin\phi)}_{\beta_2}\sin(2\pi\omega t).$$

We then can model the trend μ_t as follows:

$$\mu_t = \beta_1 \cos\left(\frac{2\pi t}{50}\right) + \beta_2 \sin\left(\frac{2\pi t}{50}\right).$$



Source: *Time Series Analysis and Its Applications: With R Examples* by R. Shumway and D. Stoffer

Contents

Removing Trend via Differencing

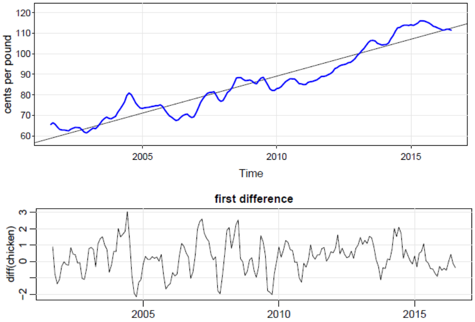
Def.

Given time series x_t, its *first difference* is defined as follows:

$$\nabla x_t = x_t - x_{t-1}.$$

Removing Trend via Regression: Example

Ex.: The price of chicken: monthly whole bird spot price, Georgia docks, US cents per pound, August 2001 to July 2016, with fitted linear trend line.



Source: *Time Series Analysis and Its Applications: With R Examples* by R. Shumway and D. Stoffer

Removing Trend via Differencing

Def.
Given time series x_t, its *second difference* is defined as follows:

$$\nabla^2 x_t = \nabla(\nabla x_t)$$

Removing Trend via Differencing

Def.
Given time series x_t, its *second difference* is defined as follows:

$$\nabla^2 x_t = \nabla(\nabla x_t)$$

Note:

$$\begin{aligned}
\nabla(\nabla x_t) &= \nabla(x_t - x_{t-1}) \\
&= \nabla x_t - \nabla x_{t-1} \\
&= (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) \\
&= x_t - 2x_{t-1} + x_{t-2}
\end{aligned}$$

Removing Trend via Differencing

Def.

Define the *backshift operator* as

$$Bx_t = x_{t-1}.$$

Note:

The first and second differences can be written as

$$\nabla x_t = (1 - B)x_t \text{ and } \nabla^2 x_t = (1 - B)^2 x_t,$$

respectively.

Removing Trend via Differencing

Def.

The *difference of order* d is defined as

$$\nabla^d = (1 - B)^d,$$

where $d = 1, 2, \ldots$ (if $d = 1$, the difference is denoted by ∇).