

Time Series Analysis

Andreas Jakobsson

Content:

- Least squares
- Weighted least squares

The linear model

We will here consider *linear* models, such that

$$y_t = \mathbf{x}_t^* \boldsymbol{\theta} + e_t$$

where e_t is an additive noise process and the regressor \mathbf{x}_t is a *known* vector, whereas $\boldsymbol{\theta}$ is an $n_{\boldsymbol{\theta}}$ -dimensional vector containing the parameters to be estimated.

The linear model

We will here consider *linear* models, such that

$$y_t = \mathbf{x}_t^* \boldsymbol{\theta} + e_t$$

where e_t is an additive noise process and the regressor \mathbf{x}_t is a *known* vector, whereas $\boldsymbol{\theta}$ is an $n_{\boldsymbol{\theta}}$ -dimensional vector containing the parameters to be estimated.

Example: The process

$$\begin{aligned} y_t &= a_0 + a_1 z_t + a_2 z_t^2 + a_3 \sin(z_t) + e_t \\ &= \begin{bmatrix} 1 & z_t & z_t^2 & \sin(z_t) \end{bmatrix} \begin{bmatrix} a_0 & a_1 & a_2 & a_3 \end{bmatrix}^T + e_t \\ &= \mathbf{x}_t^T \boldsymbol{\theta} + e_t \end{aligned}$$

is linear in $\boldsymbol{\theta}$.

Least squares

Consider N measurements of y_t , and let

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} y_1 & \dots & y_N \end{bmatrix}^T \\ \mathbf{X} &= \begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_N \end{bmatrix}^* \\ \mathbf{e} &= \begin{bmatrix} e_1 & \dots & e_N \end{bmatrix}^T \end{aligned}$$

where we assume e_t to be a zero-mean white process with variance σ_e^2 . Thus,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}$$

Least squares

Consider N measurements of y_t , and let

$$\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_N \end{bmatrix}^T$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_N \end{bmatrix}^*$$

$$\mathbf{e} = \begin{bmatrix} e_1 & \dots & e_N \end{bmatrix}^T$$

where we assume e_t to be a zero-mean white process with variance σ_e^2 . Thus,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}$$

The least-squares (LS) estimate of the unknown n_θ -dimensional parameter vector $\boldsymbol{\theta}$ can then be formed as

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_2^2$$

with $\|\mathbf{a}\|_2^2 = \mathbf{a}^* \mathbf{a}$ denoting the 2-norm of the vector \mathbf{a} .

Least squares

Note that

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_2^2 \\ &= \arg \min_{\boldsymbol{\theta}} \{ [\mathbf{y} - \mathbf{X}\boldsymbol{\theta}]^* [\mathbf{y} - \mathbf{X}\boldsymbol{\theta}] \} \\ &= \arg \min_{\boldsymbol{\theta}} \left\{ [\boldsymbol{\theta} - \mathbf{X}^\dagger \mathbf{y}]^* [\mathbf{X}^* \mathbf{X}] [\boldsymbol{\theta} - \mathbf{X}^\dagger \mathbf{y}] + \mathbf{y}^* \mathbf{y} - \mathbf{y}^* \Pi_{\mathbf{X}} \mathbf{y} \right\} \\ &= \mathbf{X}^\dagger \mathbf{y} \end{aligned}$$

where \mathbf{X}^\dagger denotes the (Moore-Penrose) pseudoinverse of \mathbf{X} , defined as

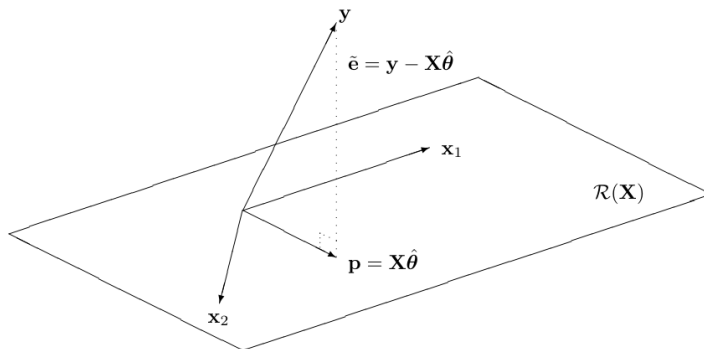
$$\mathbf{X}^\dagger = (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^*$$

and with

$$\Pi_{\mathbf{X}} = \mathbf{X} \mathbf{X}^\dagger = \mathbf{X} (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^*$$

is the *projection* matrix onto the range space of \mathbf{X} . Here, we have also assumed that \mathbf{X} has full (column) rank.

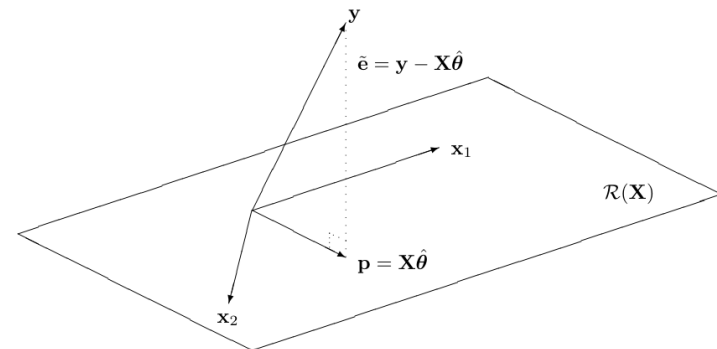
Least squares



$$\mathbf{p} = \mathbf{X} \mathbf{X}^\dagger \mathbf{y} = \mathbf{X} (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \mathbf{y} = \Pi_{\mathbf{X}} \mathbf{y}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\theta}} = (\mathbf{I} - \Pi_{\mathbf{X}}) \mathbf{y} = \Pi_{\mathbf{X}}^\perp \mathbf{y}$$

Least squares



$$\mathbf{p} = \mathbf{X} \mathbf{X}^\dagger \mathbf{y} = \mathbf{X} (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \mathbf{y} = \Pi_{\mathbf{X}} \mathbf{y}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\theta}} = (\mathbf{I} - \Pi_{\mathbf{X}}) \mathbf{y} = \Pi_{\mathbf{X}}^\perp \mathbf{y} \quad \Rightarrow \quad \hat{\sigma}_e^2 = \frac{1}{N - n_\theta} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\theta}})^* (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\theta}}) = \frac{\mathbf{y}^* \Pi_{\mathbf{X}}^\perp \mathbf{y}}{N - n_\theta}$$

Least squares

If $\mathbf{e} \in \mathbb{N}(\mathbf{0}, \sigma_e^2 \mathbf{I})$, then $\hat{\boldsymbol{\theta}}$ is Normal distributed with mean $E\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta}$ and

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \mathbf{y} - \boldsymbol{\theta} = (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \mathbf{e}$$

implying that the variance

$$\begin{aligned} V\{\hat{\boldsymbol{\theta}}\} &= V\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\} = E\left\{(\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \mathbf{e} \mathbf{e}^* \mathbf{X} (\mathbf{X}^* \mathbf{X})^{-1}\right\} \\ &= (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* E\{\mathbf{e} \mathbf{e}^*\} \mathbf{X} (\mathbf{X}^* \mathbf{X})^{-1} \\ &= \sigma_e^2 (\mathbf{X}^* \mathbf{X})^{-1} \end{aligned}$$

Furthermore, $\hat{\boldsymbol{\theta}}$ is the *best linear unbiased estimate* (BLUE) of $\boldsymbol{\theta}$, i.e., this is the estimate that has the lowest variance among all possible linear unbiased estimators.

Least squares

Example: Consider N samples of an AR(p) process, such that

$$y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = e_t$$

for $t = p+1, \dots, N$, where e_t is a zero-mean white noise process with variance σ_e^2 . Expressed differently,

$$e_t = y_t + \begin{bmatrix} y_{t-1} & \dots & y_{t-p} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = y_t + \mathbf{x}_t^T \boldsymbol{\theta}$$

To form the LS estimate of $\boldsymbol{\theta}$, we let

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} y_{p+1} & \dots & y_N \end{bmatrix}^T \\ \mathbf{X} &= \begin{bmatrix} \mathbf{x}_{p+1} & \dots & \mathbf{x}_N \end{bmatrix}^* \\ \mathbf{e} &= \begin{bmatrix} e_{p+1} & \dots & e_N \end{bmatrix}^T \end{aligned}$$

yielding $\mathbf{e} = \mathbf{y} + \mathbf{X}\boldsymbol{\theta}$. Thus, $\hat{\boldsymbol{\theta}} = -(\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \mathbf{y}$.

Weighted least squares

We have so far considered the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}$$

with \mathbf{e} being a white noise. How should we handle the case when the noise is colored?

Let $\mathbf{R}_e = E\{\mathbf{e} \mathbf{e}^*\}$ denote the covariance matrix of the additive noise and let \mathbf{W} denote a matrix square root of \mathbf{R}_e^{-1} , i.e., so that

$$\mathbf{R}_e^{-1} = \mathbf{W}^* \mathbf{W}$$

We then form the weighted least squares (WLS) estimate as

$$\hat{\boldsymbol{\theta}}_{WLS} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_{\mathbf{R}_e}^2$$

with $\|\mathbf{a}\|_{\mathbf{R}}^2 = \mathbf{a}^* \mathbf{R}^{-1} \mathbf{a}$ denoting the \mathbf{R} -norm of the vector \mathbf{a} .

Weighted least squares

Then,

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{WLS} &= \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_{\mathbf{R}_e}^2 \\ &= \arg \min_{\boldsymbol{\theta}} \{[\mathbf{y} - \mathbf{X}\boldsymbol{\theta}]^* \mathbf{W}^* \mathbf{W} [\mathbf{y} - \mathbf{X}\boldsymbol{\theta}]\} \end{aligned}$$

Let $\tilde{\mathbf{y}} = \mathbf{W}\mathbf{y}$, $\tilde{\mathbf{X}} = \mathbf{W}\mathbf{X}$, and $\tilde{\mathbf{e}} = \mathbf{W}\mathbf{e}$. Then,

$$\hat{\boldsymbol{\theta}}_{WLS} = \arg \min_{\boldsymbol{\theta}} [\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\theta}]^* [\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\theta}]$$

which implies that

$$\hat{\boldsymbol{\theta}}_{WLS} = \tilde{\mathbf{X}}^\dagger \tilde{\mathbf{y}} = (\mathbf{X}^* \mathbf{R}_e^{-1} \mathbf{X})^{-1} \mathbf{X}^* \mathbf{R}_e^{-1} \mathbf{y}$$

and that $V\{\hat{\boldsymbol{\theta}}_{WLS}\} = (\mathbf{X}^* \mathbf{R}_e^{-1} \mathbf{X})^{-1}$.