LUND UNIVERSITY

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

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

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

Least squares

Consider N measurements of y_t, and let

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

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

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

Least squares

Consider N measurements of y_t, and let

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

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 $n_{\boldsymbol{\theta}}$-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

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

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

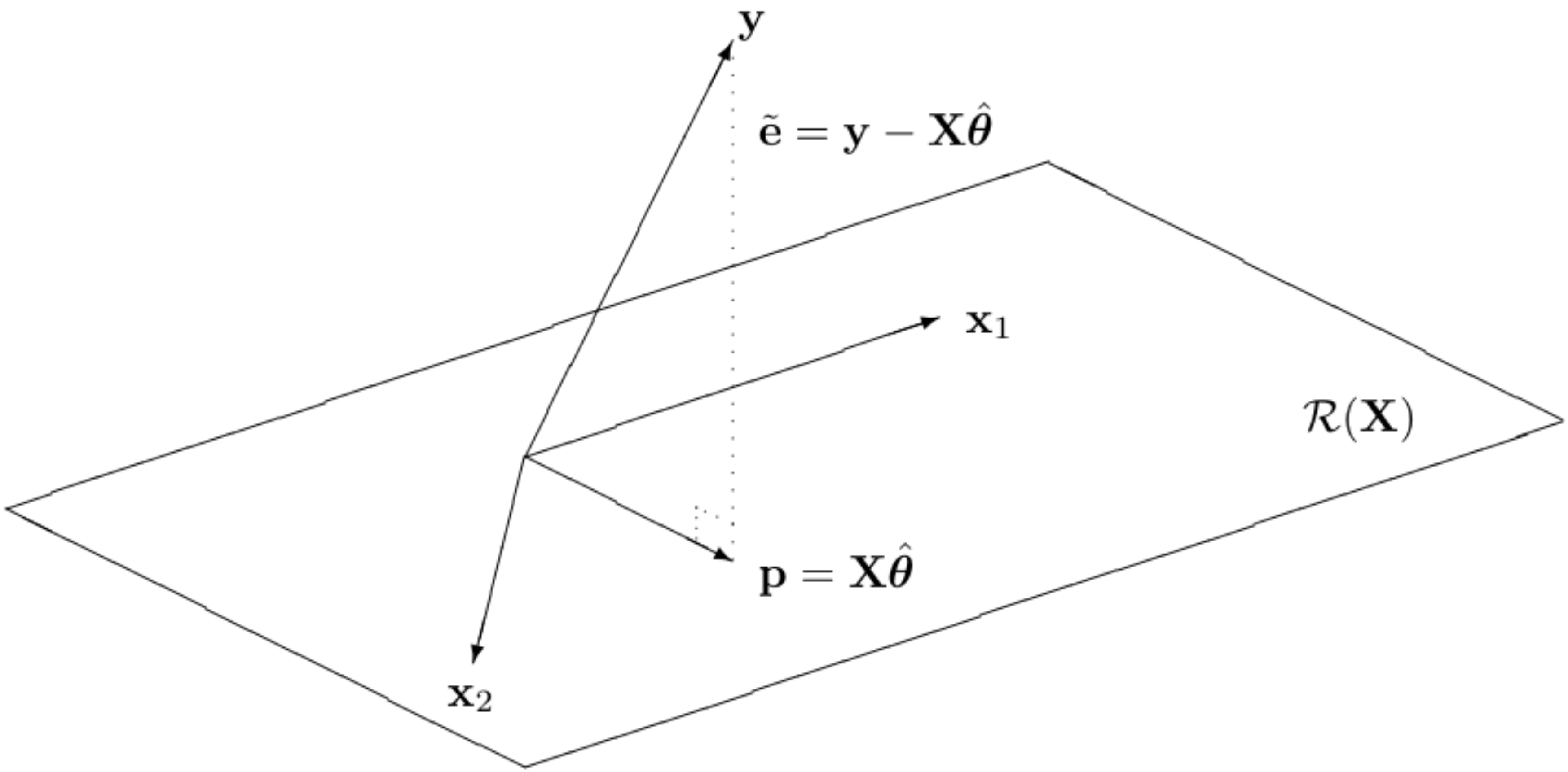
$$\mathbf{X}^\dagger = (\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*$$

and with

$$\boldsymbol{\Pi}_{\mathbf{X}} = \mathbf{X}\mathbf{X}^\dagger = \mathbf{X}\Big(\mathbf{X}^*\mathbf{X}\Big)^{-1}\mathbf{X}^*$$

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

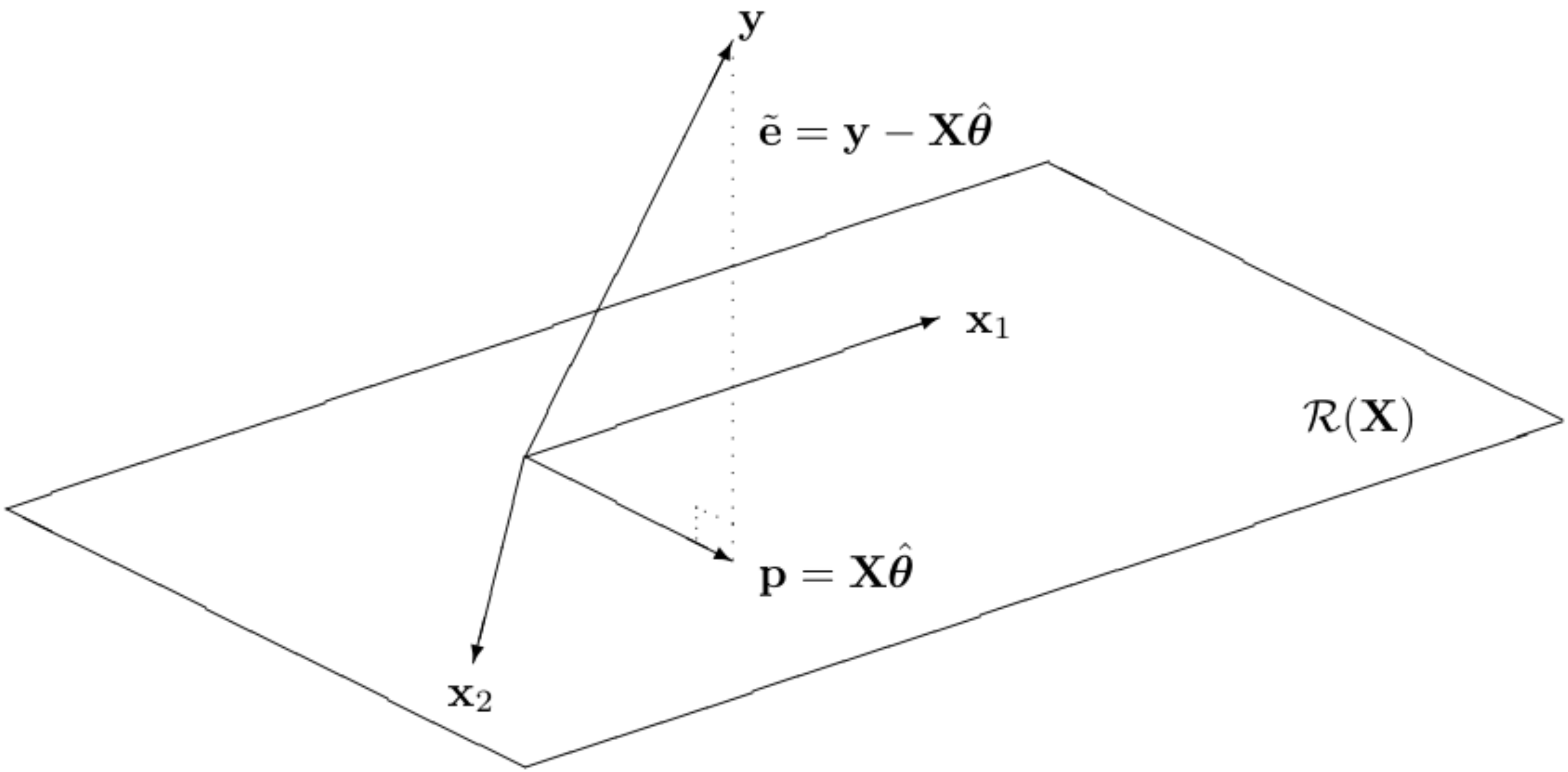
Least squares



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

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

Least squares



$$\mathbf{p} = \mathbf{X}\mathbf{X}^{\dagger}\mathbf{y} = \mathbf{X}\left(\mathbf{X}^{*}\mathbf{X}\right)^{-1}\mathbf{X}^{*}\mathbf{y} = \mathbf{\Pi}_{\mathbf{X}}\mathbf{y}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}} = (\mathbf{I} - \mathbf{\Pi}_{\mathbf{X}})\mathbf{y} = \mathbf{\Pi}_{\mathbf{X}}^{\perp}\mathbf{y} \qquad \Rightarrow \qquad \hat{\sigma}_e^2 = \frac{1}{N - n_{\boldsymbol{\theta}}}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}}\right)^{*}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}}\right) = \frac{\mathbf{y}^{*}\mathbf{\Pi}_{\mathbf{X}}^{\perp}\mathbf{y}}{N - n_{\boldsymbol{\theta}}}$$

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

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

implying that the variance

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

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

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

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

$$e_t = y_t + \left[\begin{array}{ccc} y_{t-1} & \ldots & y_{t-p} \end{array} \right] \left[\begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_p \end{array} \right] = y_t + \mathbf{x}_t^T \boldsymbol{\theta}$$

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

$$\begin{aligned} \mathbf{y} &= \left[\begin{array}{ccc} y_{p+1} & \ldots & y_N \end{array} \right]^T \\ \mathbf{X} &= \left[\begin{array}{ccc} \mathbf{x}_{p+1} & \ldots & \mathbf{x}_N \end{array} \right]^* \\ \mathbf{e} &= \left[\begin{array}{ccc} e_{p+1} & \ldots & e_N \end{array} \right]^T \end{aligned}$$

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

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

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

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

Weighted least squares

Then,

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

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

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

which implies that

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

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