Time Series Analysis

Henrik Madsen

hm@imm.dtu.dk

Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby

Outline of the lecture

Regression based methods, 1st part:

- Introduction (Sec. 3.1)
- The General Linear Model, including OLS-, WLS-, and ML-estimates (Sec. 3.2)
- Prediction in the General Linear Model (Sec. 3.3)
- Examples...

General form of the regression model

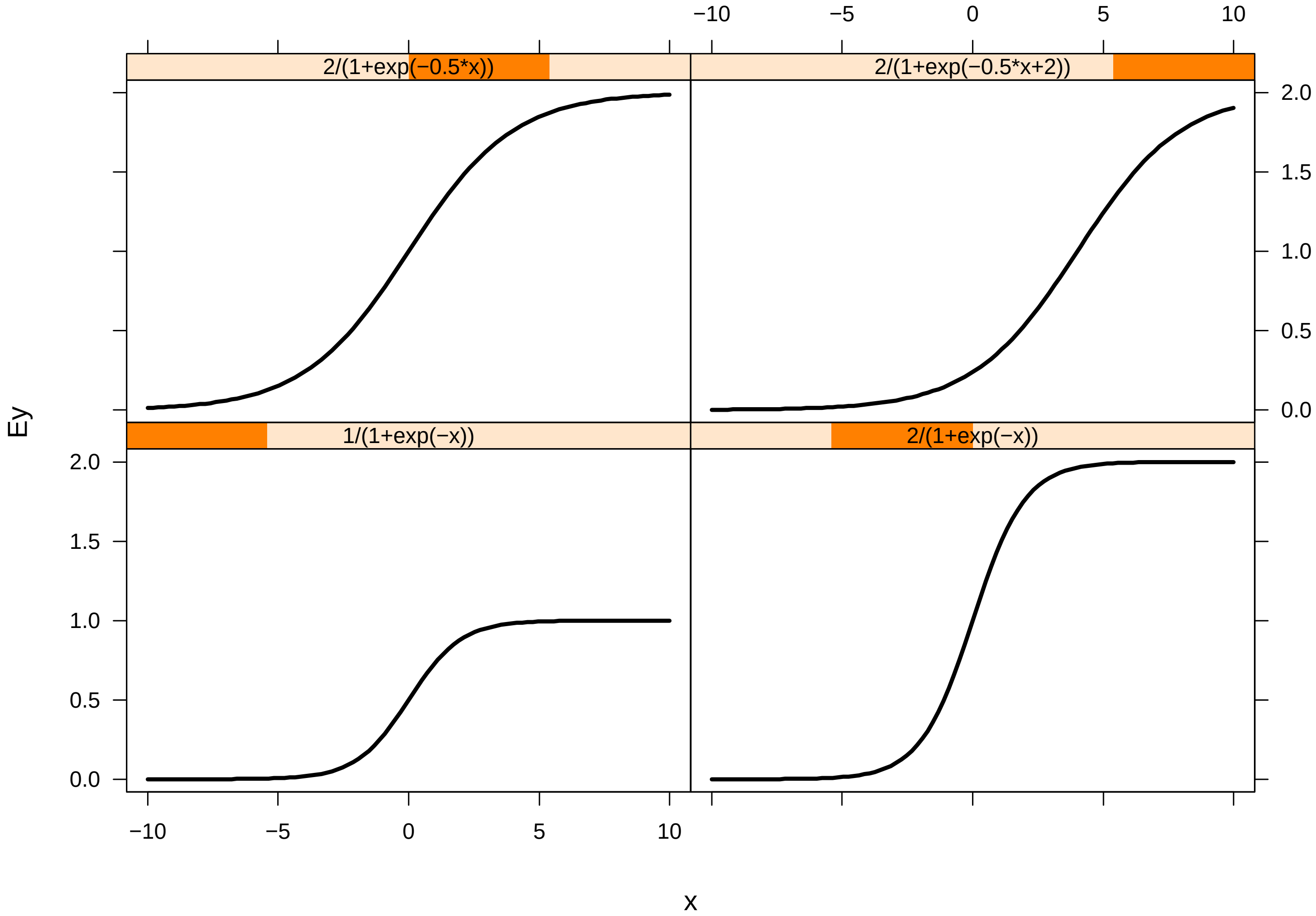
$$Y_t = f(\boldsymbol{X}_t, t; \boldsymbol{\theta}) + \varepsilon_t$$

Where:

- Y_t is the output we aim to model
- $\boldsymbol{X}_t$ indicates the p independent variables $\boldsymbol{X}_t = (X_{1t}, \cdots, X_{pt})^T$
- t is the time index
- $\boldsymbol{\theta}$ indicates m unknown parameters $(\theta_1, \cdots, \theta_m)^T$
- ε_t is a sequence of random variables with mean zero, variance σ_t, and $Cov[\varepsilon_{t_i}, \varepsilon_{t_j}] = \sigma\Sigma_{ij}$

We restrict the discussion to the case where $\boldsymbol{X}_t$ is non-random and we write $\boldsymbol{x}_t$

$$Y_t = \theta_1/(1 + \exp(-\theta_2 x_t + \theta_3)) + \varepsilon_t$$



Least squares estimates

Observations:

$$(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \cdots, (y_n, \boldsymbol{x}_n)$$

Ordinary Least Square (unweighted) estimates is found from

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} S(\boldsymbol{\theta})$$

where

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{n} [y_t - f(\boldsymbol{x}_t; \boldsymbol{\theta})]^2 = \sum_{t=1}^{n} \varepsilon_t^2(\boldsymbol{\theta})$$

The unweighted method assumes that the errors all have the same variance and are mutually uncorrelated.

Variance of error and estimates

If the model errors ε_t are i.i.d.

- The variance of the model errors is estimated as:

$$\widehat{\sigma}^2 = \frac{S(\widehat{\boldsymbol{\theta}})}{n-p}$$

- The variance-covariance matrix of the estimates is

$$V[\widehat{\boldsymbol{\theta}}] = 2\widehat{\sigma}^2 \left[\frac{\partial^2}{\partial^2 \boldsymbol{\theta}} S(\boldsymbol{\theta}) \right]^{-1} \Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

The General Linear Model

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t$$

The General Linear Model

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t$$

Note that the quadratic model

$$Y_t = \theta_0 + \theta_1 z_t + \theta_2 z_t^2 + \varepsilon_t$$

can be written

$$y_t = \begin{pmatrix} 1 & z_t & z_t^2 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix} + \varepsilon_t$$

and hence it is a general linear model.

General Linear Models

- Some examples in the book
- (Multiple) regression analysis, ex: $Y = \alpha + \beta x + \varepsilon$
- Analysis of variance, ex: $Y = \alpha_i + \varepsilon$ (i indexes the treatment)
- Analysis of covariance, ex: $Y = \alpha_i + \beta x + \varepsilon$

For ANOVA and ANCOVA the treatments must be coded into a number of x-variables.

OLS – solution

- Non-linear regression: Numerical optimization is required; see the book for a simple example (Newton-Raphson)
- For the general linear model a closed-form solution exists. For all observations the model equations are written as:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad or \quad \boldsymbol{Y} = \boldsymbol{x}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

  i.e. we want to minimize $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$

- The solution is $\boxed{\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{Y}}$ (if $\boldsymbol{x}$ has full rank)
- $\widehat{\sigma}^2 = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}/(n-p)$ and $V[\widehat{\boldsymbol{\theta}}] = \widehat{\sigma}^2(\boldsymbol{x}^T\boldsymbol{x})^{-1}$

Example

Data:

t	y	x
1	0.2	0.4
2	1.2	1.2
3	1.9	2.3
4	2.3	3.4
5	1.9	4.3

Model:

$$Y_t = \theta_0 + \theta_1 x_t + \theta_2 x_t^2 + \varepsilon_t$$

Example

Data:

t	y	x
1	0.2	0.4
2	1.2	1.2
3	1.9	2.3
4	2.3	3.4
5	1.9	4.3

Model:

$$Y_t = \theta_0 + \theta_1 x_t + \theta_2 x_t^2 + \varepsilon_t$$

$$\boldsymbol{Y} = \boldsymbol{x}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

Example

Data:

t	y	x
1	0.2	0.4
2	1.2	1.2
3	1.9	2.3
4	2.3	3.4
5	1.9	4.3

Model:

$$Y_t = \theta_0 + \theta_1 x_t + \theta_2 x_t^2 + \varepsilon_t$$

$$\boldsymbol{Y} = \boldsymbol{x}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} 0.2 \\ 1.2 \\ 1.9 \\ 2.3 \\ 1.9 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 & 0.16 \\ 1 & 1.2 & 1.44 \\ 1 & 2.3 & 5.29 \\ 1 & 3.4 & 11.56 \\ 1 & 4.3 & 18.49 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

Example

Data:

t	y	x
1	0.2	0.4
2	1.2	1.2
3	1.9	2.3
4	2.3	3.4
5	1.9	4.3

Model:

$$Y_t = \theta_0 + \theta_1 x_t + \theta_2 x_t^2 + \varepsilon_t$$

$$\boldsymbol{Y} = \boldsymbol{x}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} 0.2 \\ 1.2 \\ 1.9 \\ 2.3 \\ 1.9 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 & 0.16 \\ 1 & 1.2 & 1.44 \\ 1 & 2.3 & 5.29 \\ 1 & 3.4 & 11.56 \\ 1 & 4.3 & 18.49 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{Y} = \begin{bmatrix} -0.40 \\ 1.61 \\ -0.25 \end{bmatrix}$$

Properties

- It is a linear function of the observations $\boldsymbol{Y}$ (and $\widehat{\boldsymbol{Y}}$ is a linear function of the observations)
- It is unbiased, i.e. $E[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$
- $V[\widehat{\boldsymbol{\theta}}] = E\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right] = \sigma^2(\boldsymbol{x}^T\boldsymbol{x})^{-1}$
- $\hat{\theta}$ is BLUE (Best Linear Unbiased Estimator), which means that it has the smallest variance among all estimators which are a linear function of the observations.

WLS-estimates

- Equations for all observations: $\boldsymbol{Y} = \boldsymbol{x}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$
- $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ and $V[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is known
- We want to minimize $(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{\theta})$ (why?)
- The solution is

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$$

(if $\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}$ is invertible)

- An unbiased estimate of σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{n-p}(\boldsymbol{Y} - \boldsymbol{x}\widehat{\boldsymbol{\theta}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\widehat{\boldsymbol{\theta}})$$

Example WLS/OLS

- H. Madsen & P. Thyregod (1988). *Modelling the Time Correlation in Hourly Observations of Direct Radiation in Clear Skies.* Energy and Buildings, **11**, 201–211.
- See the examples in the book.

ML-estimates

- We now assume that the observations are Gaussian:

$$\boldsymbol{Y} \sim \mathsf{N}_n(\boldsymbol{x\theta}, \sigma^2\boldsymbol{\Sigma})$$

- $\boldsymbol{\Sigma}$ is assumed known
- The ML-estimator is the same as the WLS-estimator:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$$

- The ML-estimator for σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{x}\widehat{\boldsymbol{\theta}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\widehat{\boldsymbol{\theta}})$$

Properties of the ML-estimator

- It is a linear function of the observations which now implies that it is normally distributed
- It is unbiased, i.e. $E[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$ and
- The variance $V[\widehat{\boldsymbol{\theta}}] = E[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] = (\boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x})^{-1} \sigma^2$
- It is an efficient estimator

Unknown Σ

Relaxation algorithm:

a) Select a value for $\boldsymbol{\Sigma}$ (e.g. $\boldsymbol{\Sigma} = \boldsymbol{I}$).

b) Find the estimates for this value of $\boldsymbol{\Sigma}$ e.g. by solving the normal equations.

c) Consider the residuals $\{\widehat{\varepsilon}_t\}$ and calculate the correlation and variance structure of the residuals. Then select a new value for $\boldsymbol{\Sigma}$ which reflects that correlation and variance structure.

d) Stop if convergence - otherwise go to b).

See (Goodwin and Payne, 1977) for details.

Prediction

- If the expected value of the squared prediction error is to be minimized, then
- we must use the expected mean $E[Y|\boldsymbol{X} = \boldsymbol{x}]$ as the prediction.

Prediction in the general linear model

- Known parameters:

$$\widehat{Y}_{t+\ell} = E[Y_{t+\ell} | \boldsymbol{X}_{t+\ell} = \boldsymbol{x}_{t+\ell}] = \boldsymbol{x}_{t+\ell}^T \boldsymbol{\theta}$$

$$V[Y_{t+\ell} - \widehat{Y}_{t+\ell}] = V[\varepsilon_{t+\ell}] = \sigma^2$$

- Estimated parameters:

$$\widehat{Y}_{t+\ell} = E[Y_{t+\ell} | \boldsymbol{X}_{t+\ell} = \boldsymbol{x}_{t+\ell}] = \boldsymbol{x}_{t+\ell}^T \widehat{\boldsymbol{\theta}}$$

$$V[Y_{t+\ell} - \widehat{Y}_{t+\ell}] = V[\varepsilon_{t+\ell}] = \sigma^2 [1 + \boldsymbol{x}_{t+\ell}^T (\boldsymbol{x}^T \boldsymbol{x})^{-1} \boldsymbol{x}_{t+\ell}]$$

Prediction in the general linear model – continued

- We must use an estimate of σ and therefore a $100(1-\alpha)\%$ prediction interval of a future value is calculated as:

$$\widehat{Y}_{t+\ell} \pm t_{\alpha/2}(n-p)\hat{\sigma}\sqrt{1+\boldsymbol{x}_{t+\ell}^T(\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}_{t+\ell}}$$