Time Series Analysis

Lasse Engbo Christiansen

Department of Applied Mathematics and Computer Science
Technical University of Denmark

September 15, 2017

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

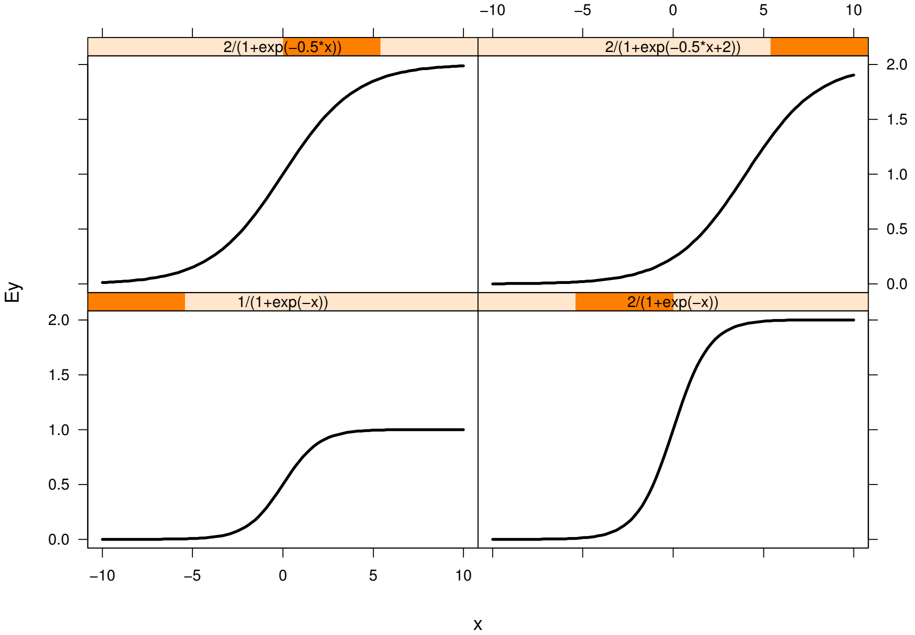
$$Y_t = f(\boldsymbol{X}_t, t; \boldsymbol{\theta}) + \varepsilon_t$$

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^2, and $\mathrm{Cov}[\varepsilon_{t_i}, \varepsilon_{t_j}] = \sigma_t^2 \Sigma_{ij}$

For now we restrict the discussion to the case where $\boldsymbol{X}_t$ is non-random and thus we write $\boldsymbol{x}_t$ instead of $\boldsymbol{X}_t$.

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



Ordinary least squares (OLS) estimates

Observations:

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

Ordinary Least Square (unweighted) estimates are found from

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

where

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

For the unweighted method to result in reliable estimates, the errors must be assumed to all have the same variance and be mutually uncorrelated.

OLS – Variance of error and estimates

If the model errors ε_t are i.i.d.

- The variance of the model errors is estimated as:

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

  where p is the number of estimated parameters.

- The variance-covariance matrix of the estimates is approximately

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

The General Linear Model (GLM)

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

Can this quadratic model in z_t be a GLM?

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

Yes, as it can be written as

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

It is linearity in the parameters that matters!

General Linear Model - sub classes

(Multiple) regression analysis, ex: $Y = \alpha + \beta x + \varepsilon$

Ex: The height of a plant described by age (x_1), concentration of nutrients in soil (x_2), etc.

Analysis of variance, ex: $Y = \alpha_i + \varepsilon$

Ex: The height of plants described by species (i).

Analysis of covariance, ex: $Y = \alpha_i + \beta x + \varepsilon$

Ex: Height of plant described by species *and* age, nutrients...

Comments

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

OLS-estimates

- 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:

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

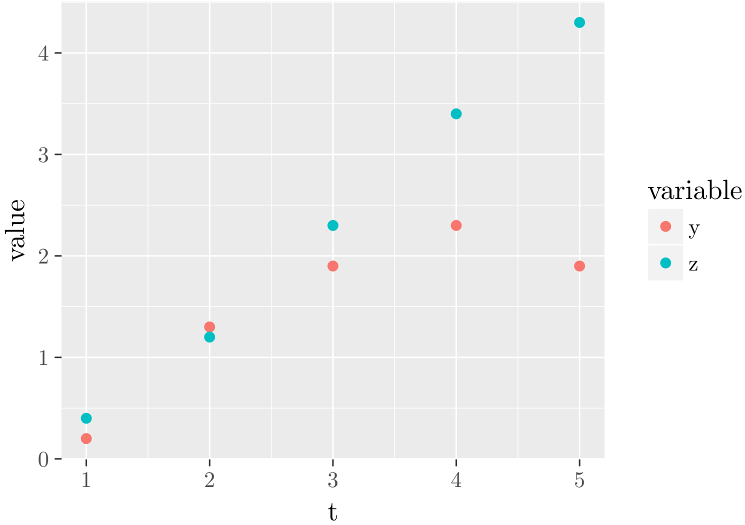
  i.e. we want to minimize $S(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$
- The solution is $\boxed{\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T \boldsymbol{x})^{-1} \boldsymbol{x}^T \boldsymbol{Y}}$ (if $\boldsymbol{x}$ has full rank)
- $\widehat{\sigma}^2 = \boldsymbol{\varepsilon}^T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

Example

Data:

t	y	z
1	0.2	0.4
2	1.3	1.2
3	1.9	2.3
4	2.3	3.4
5	1.9	4.3

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



Example

Data:

t	y	z
1	0.2	0.4
2	1.3	1.2
3	1.9	2.3
4	2.3	3.4
5	1.9	4.3

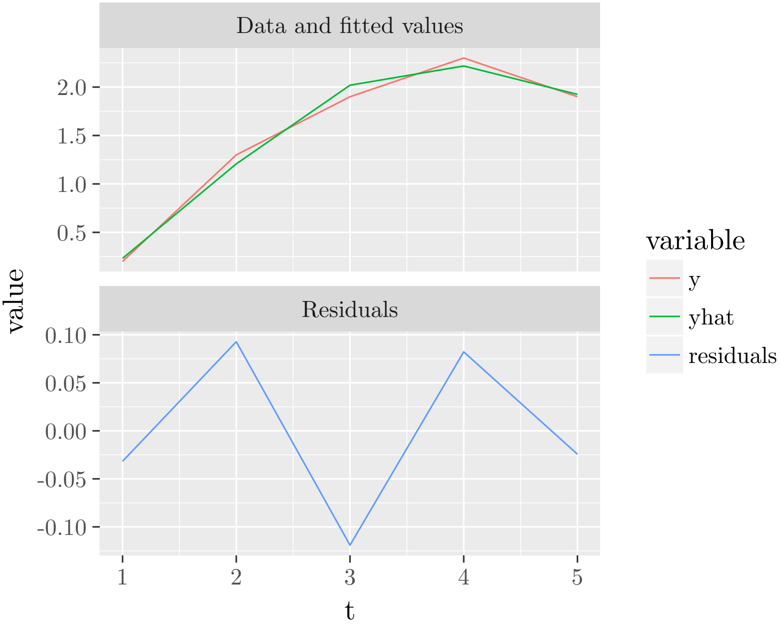
Model: $Y_t = \theta_0 + \theta_1 z_t + \theta_2 z_t^2 + \varepsilon_t$

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

$$\begin{bmatrix} 0.2 \\ 1.3 \\ 1.9 \\ 2.3 \\ 1.9 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 & 0.16 \\ 1 & 1.2 & 1.44 \\ 1 & 2.3 & 5.29 \\ 1 & 3.4 & 11.56 \\ 1 & 4.3 & 18.49 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T \boldsymbol{x})^{-1} \boldsymbol{x}^T \boldsymbol{Y} = \begin{bmatrix} -0.38 \\ 1.62 \\ -0.25 \end{bmatrix}$$

Plot of the model fit



Properties of the OLS-estimator of a GLM

- It is a linear function of the observations $\boldsymbol{Y}$ (and $\widehat{\boldsymbol{Y}}$ is thus 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 (<u>B</u>est <u>L</u>inear <u>U</u>nbiased <u>E</u>stimator), which means that it has the smallest variance among all estimators which are a linear function of the observations.

WLS-estimates

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

Example WLS/OLS: Clear sky radiation

$$
\begin{aligned}
I_N(h(t)) &= a_N(1 - \exp(-b_N h(t))) + \varepsilon_N(t) && (1) \\
\varepsilon_N(t) &\sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}) && (2)
\end{aligned}
$$

Suggested variance structures

- i.i.d

$$\boldsymbol{\Sigma} = \boldsymbol{I}$$

- Only correlation

$$\Sigma_{ij} = \rho^{|t_i - t_j|}$$

- Only variance

$$\Sigma_{ij} = \frac{1}{\sin(h(t_i))\sin(h(t_j))}$$

- Both correlation and variance

$$\Sigma_{ij} = \frac{\rho^{|t_i - t_j|}}{\sin(h(t_i))\sin(h(t_j))}$$

Maximum Likelihood (ML) - estimates

- We now assume that the observations are Gaussian:

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

- $\boldsymbol{\Sigma}$ is assumed known
- The ML-estimator is (here) 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 (minimum variance of unbiased estimators).

Unknown $\boldsymbol{\Sigma}$

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 $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ 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

Theorem 3.8 and 3.9

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

Prediction in the general linear model

- Known parameters:

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

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

- Estimated parameters:

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

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

Prediction in the general linear model – continued

- In practice we have to use an estimate of σ and therefore a $100(1-\alpha)\%$ prediction interval of a future value is calculated as:

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

  where $t_{\alpha/2}(n-p)$ refers to the $\alpha/2$'th quantile of the t-distribution with $n-p$ degrees of freedom.
- For $n-p$ large, percentiles from the normal distribution can be used.

Highlights

- When ε_t are i.i.d then the variance of the model errors is estimated as:

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

- OLS estimator:

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

- WLS estimator (When $V[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2\boldsymbol{\Sigma}$)

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

- ML estimator for $\widehat{\theta}$ is the same as the WLS estimator.
- but not for $\widehat{\sigma}^2$
- The expected mean $E[Y|\boldsymbol{X} = \boldsymbol{x}]$ is the optimal predictor.
- Prediction in GLM

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