Applied Time Series Analysis
FS 2011 – Week 09

Marcel Dettling

Institute for Data Analysis and Process Design

Zurich University of Applied Sciences

marcel.dettling@zhaw.ch

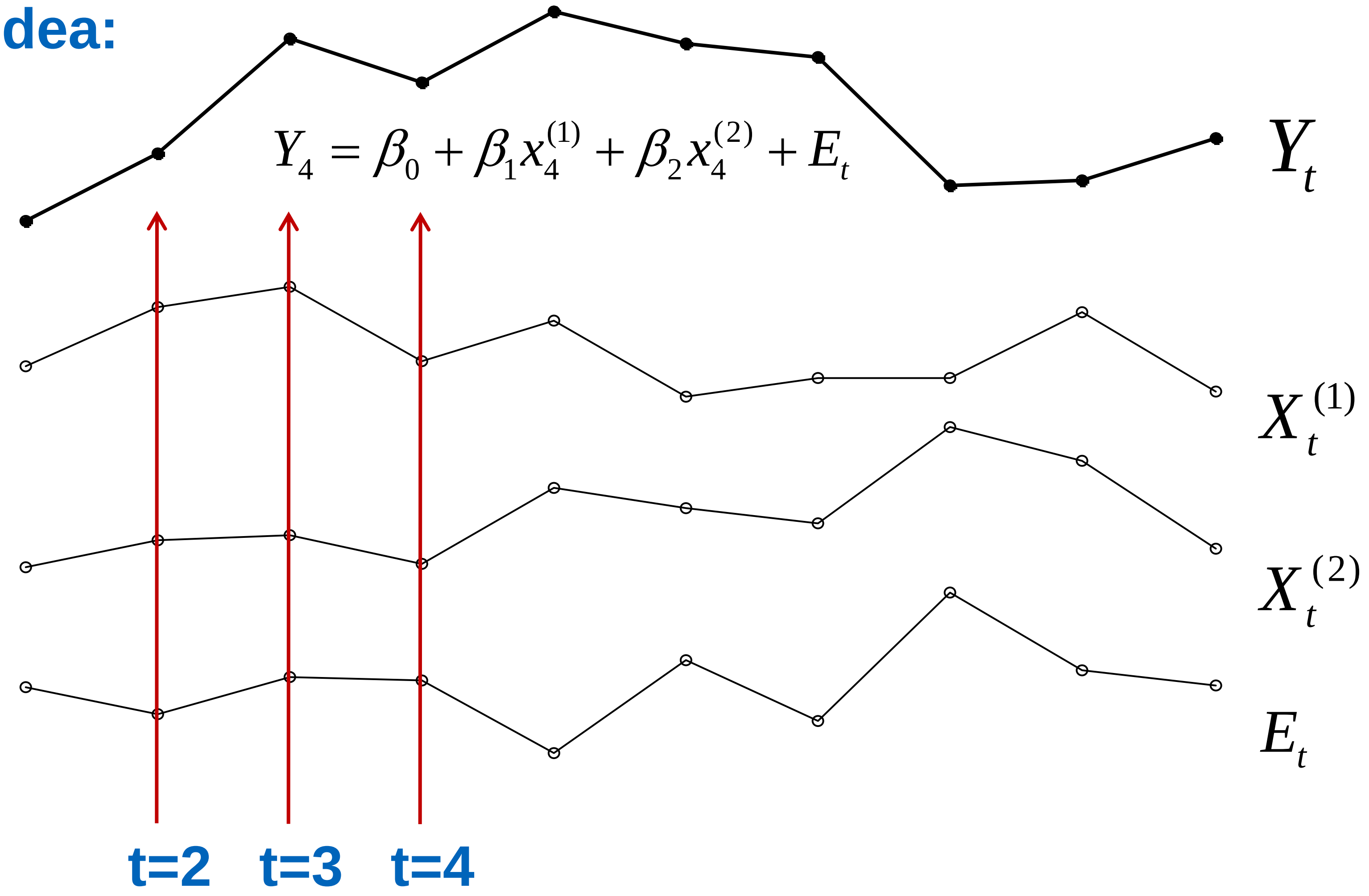
http://stat.ethz.ch/~dettling

ETH Zürich, April 18, 2010

Applied Time Series Analysis
FS 2011 – Week 09

Time Series Regression

Idea:



Applied Time Series Analysis
FS 2011 – Week 09

The Setup

- There is a response time series Y_t
- There is one or several explanatory/descriptive time series $X_t^{(1)},..., X_t^{(k)}$
- The goal is to infer the relation between X and Y, i.e. the β_j
- As long as the error series E_t is i.i.d, the usual regression setup with LS-estimates is perfectly fine

→ **Caution and specific procedures are required if the errors are correlated!**

Dealing with Correlated Errors

- In case of time series regression, the error term E_t is usually correlated and not i.i.d.

- Then, the estimated β_j are still unbiased, but the usual LS-procedure is no longer efficient and the standard errors can be grossly wrong

- There are procedures that correct for correlated errors:
  - **Cochrane-Orcutt-Method**
  - **Generalized Least Squares**

- **They must be applied in case of correlated errors!**

Example 1

A few scenarios, where time series regression is met:

1) „Normal“ Time Series

$(\text{Oxidant})_t = (\text{Temp})_t + (\text{Wind})_t + E_t$

→ The data are from 30 consecutive measurement days at L.A.

→ It‘s plausible that the pollutant levels is influenced by both wind and temperature

→ It’s well conceivable that there is day-to-day “memory” in the pollutant levels, which expresses itself in correlated errors

Applied Time Series Analysis
FS 2011 – Week 09

Example 2

A few scenarios, where time series regression is met:

2) Lagged Time Series

$(\text{Fish Caught})_t = (\text{Young Fish Introduced})_{t-1} + E_t$

→ Data may be available from several years

→ It‘s plausible that the fish caught are influenced by the young fish that were introduced

→ It’s well conceivable that there is year-to-year “memory” in the fish levels, which expresses itself in correlated errors

Applied Time Series Analysis
FS 2011 – Week 09

Example 3

A few scenarios, where time series regression is met:

3) Parametric Input Terms (Time Series)

We are already familiar with:

→ linear and quadratic trends

→ intervention and increasing intervention models

→ intervention with diminishing influence

→ etc.

Time Series Regression Model

$$Y_t = \beta_0 + \beta_1 x_t^{(1)} + \ldots + \beta_q x_t^{(q)} + E_t$$

- $t = 1, \ldots, N$
- no feedback from Y_t onto the predictors (i.e. input series)
- E_t are independent from $x_s^{(j)}$ for all j and all s, t
- E_t (generally) are dependent (e.g. an ARMA(p,q)-process)

Facts When Using Least Squares

In case of correlated errors, the effect on point estimates is:

- the estimated coefficients $\beta_1,...,\beta_q$ are unbiased
- the estimates are no longer optimal: $Var\left(\hat{\beta}_j\right) > \min_* Var\left(\hat{\beta}_j^*\right)$

Important is the effect on the standard errors of the estimates:

- $V\hat{a}r\left(\hat{\beta}_j\right)$ can be grossly wrong!
- often, the standard errors are underestimated
- too small CIs & spuriously significant results

Simulation Study: Model

We want to study the effect of correlated errors on the quality of estimates when using the least squares approach:

$$x_t = t / 50$$

$$y_t = x_t + 2x_t^2 + E_t$$

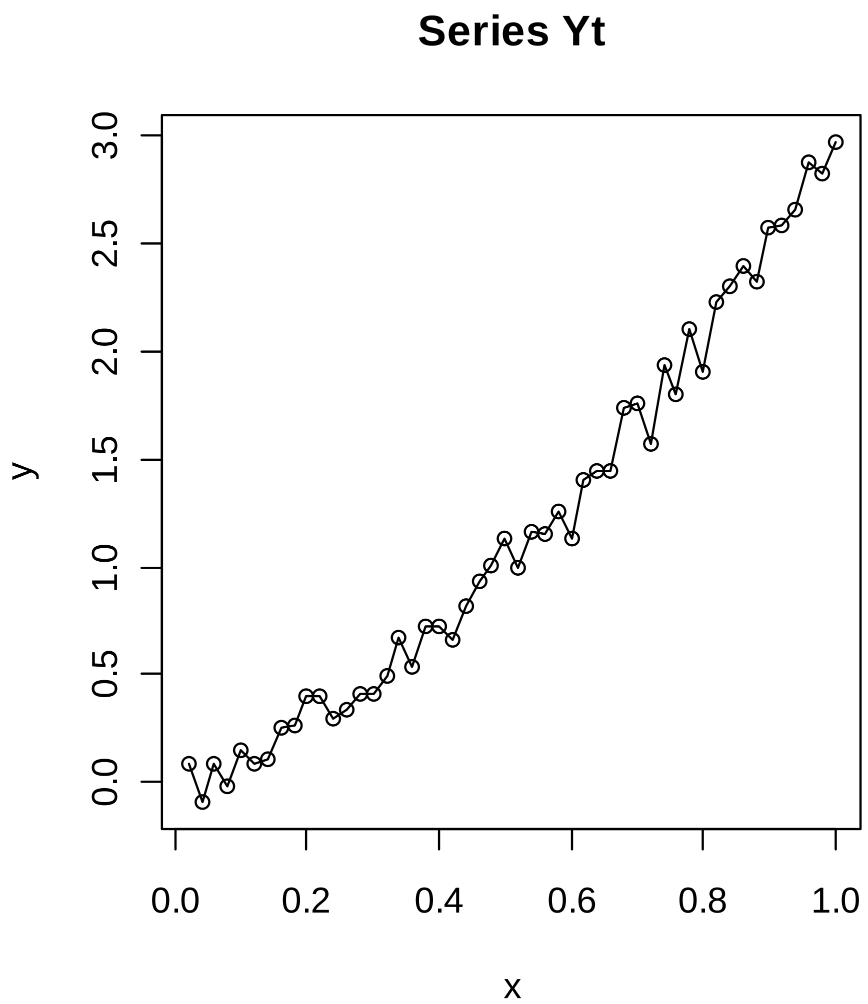
where E_t is from an AR(1)-process with $\alpha = -0.65$ and $\sigma = 0.1$.

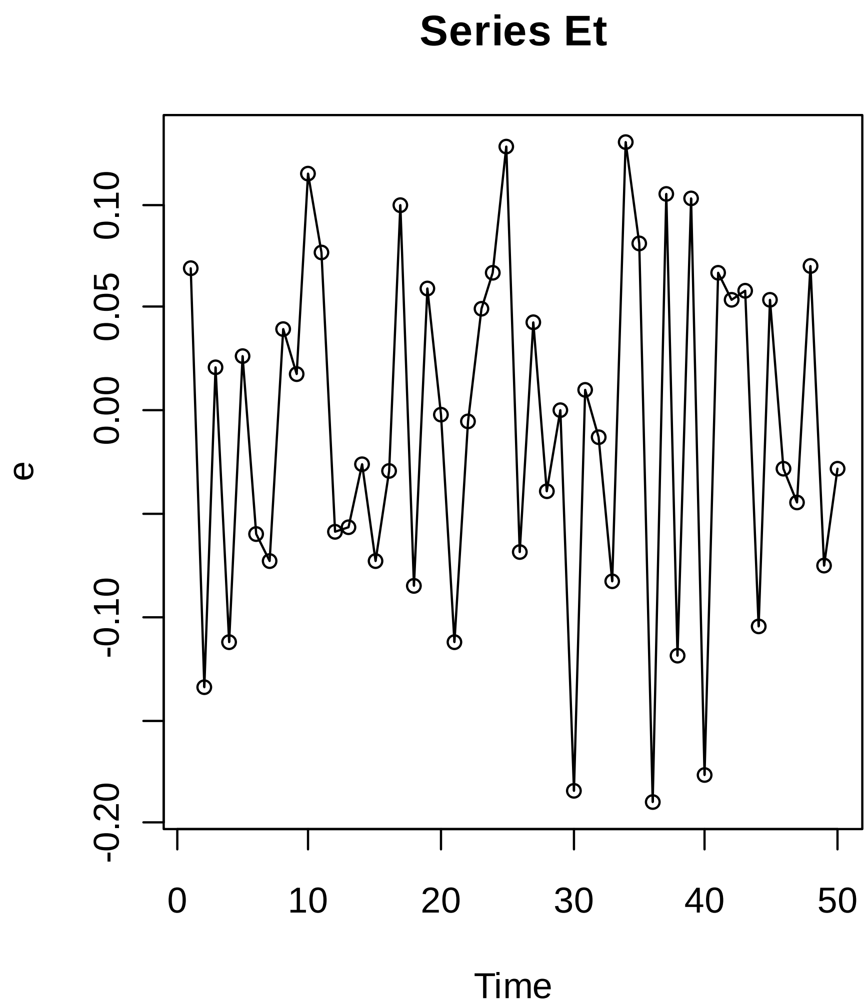
We generate 100 realizations from this model and estimate the regression coefficient and its standard error by:

1) LS
2) GLS

Applied Time Series Analysis
FS 2011 – Week 09

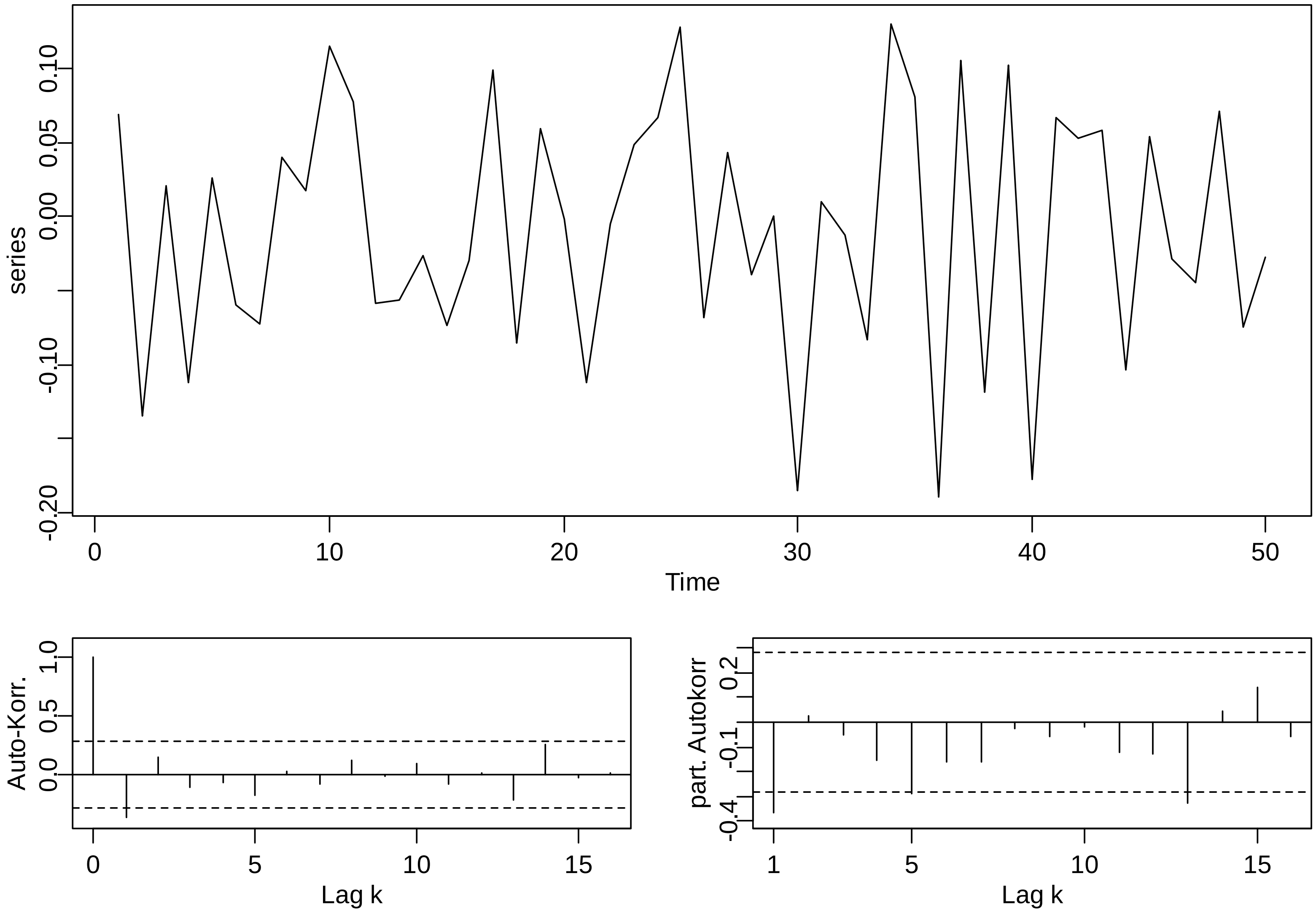
Simulation Study: Series





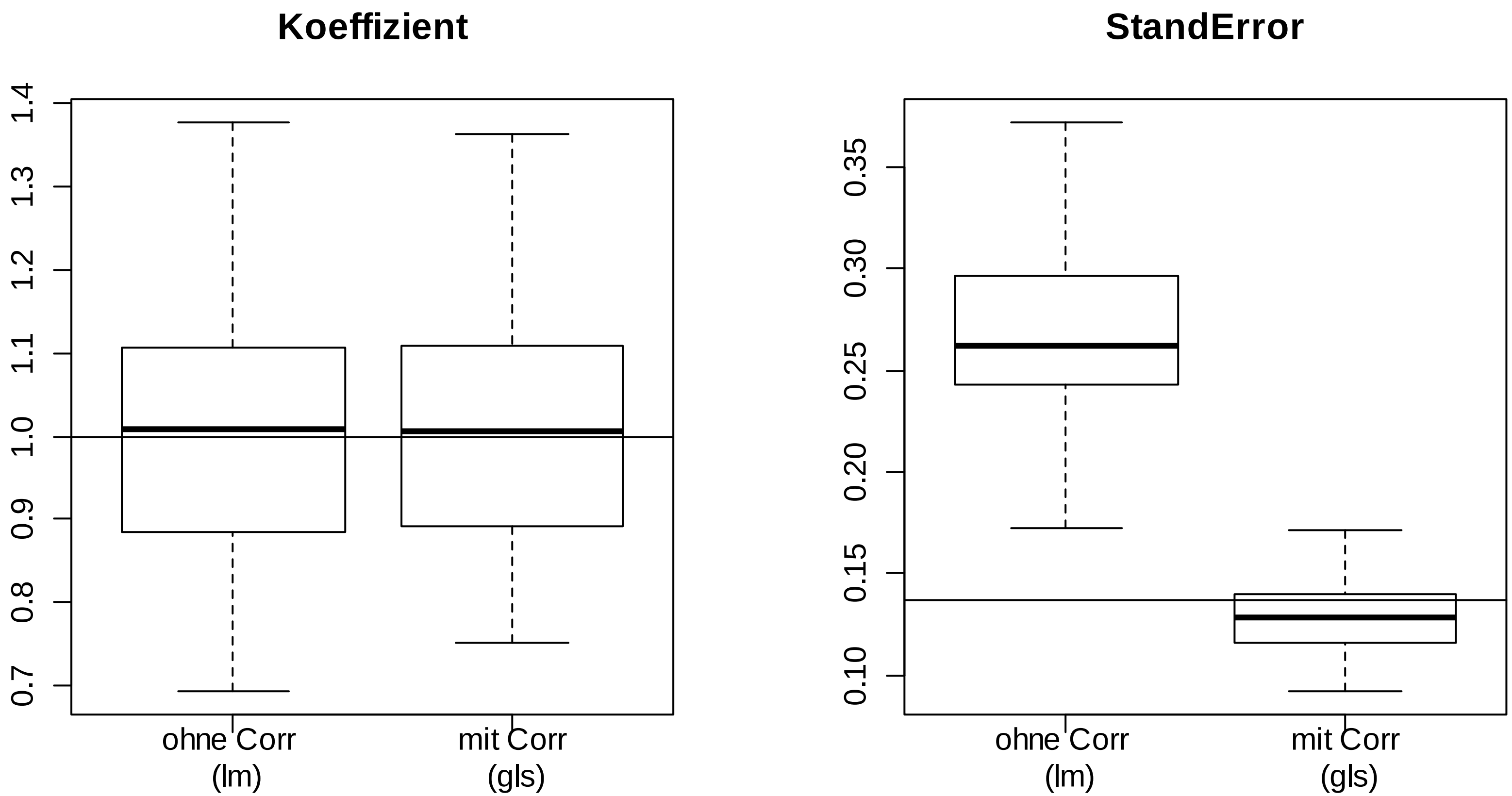
Applied Time Series Analysis
FS 2011 – Week 09

Simulation Study: ACF of the Error Term



Applied Time Series Analysis
FS 2011 – Week 09

Simulation Study: Results



Pollutant Example

A few scenarios, where time series regression is met:

1) „Normal" Time Series

$$(\text{Oxidant})_t = (\text{Temp})_t + (\text{Wind})_t + E_t$$

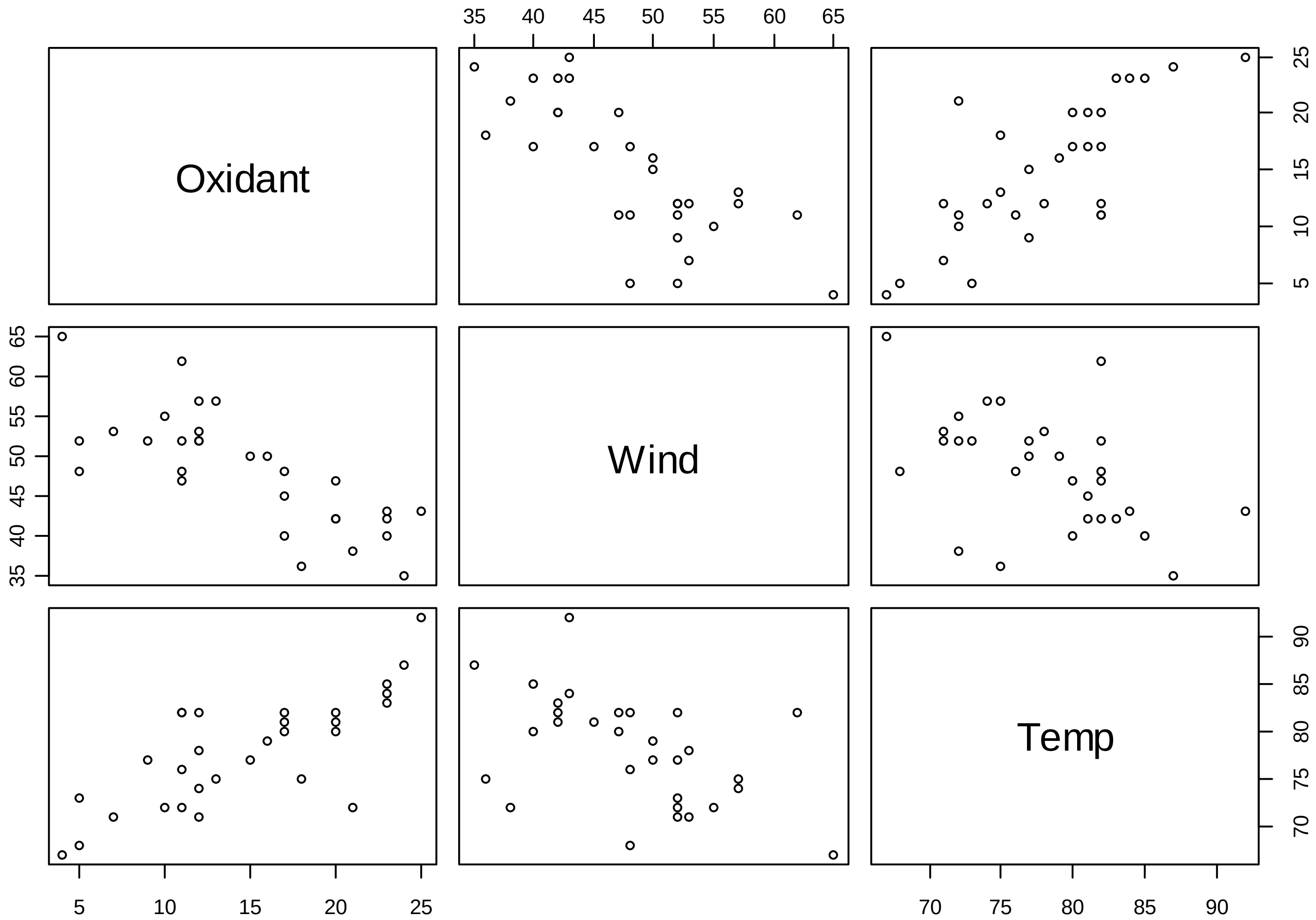
→ The data are from 30 consecutive measurement days at L.A.

→ It's plausible that the pollutant levels is influenced by both wind and temperature

→ It's well conceivable that there is day-to-day "memory" in the pollutant levels, which expresses itself in correlated errors

Applied Time Series Analysis
FS 2011 – Week 09

Pollutant Example



Applied Time Series Analysis
FS 2011 – Week 09

Pollutant Example

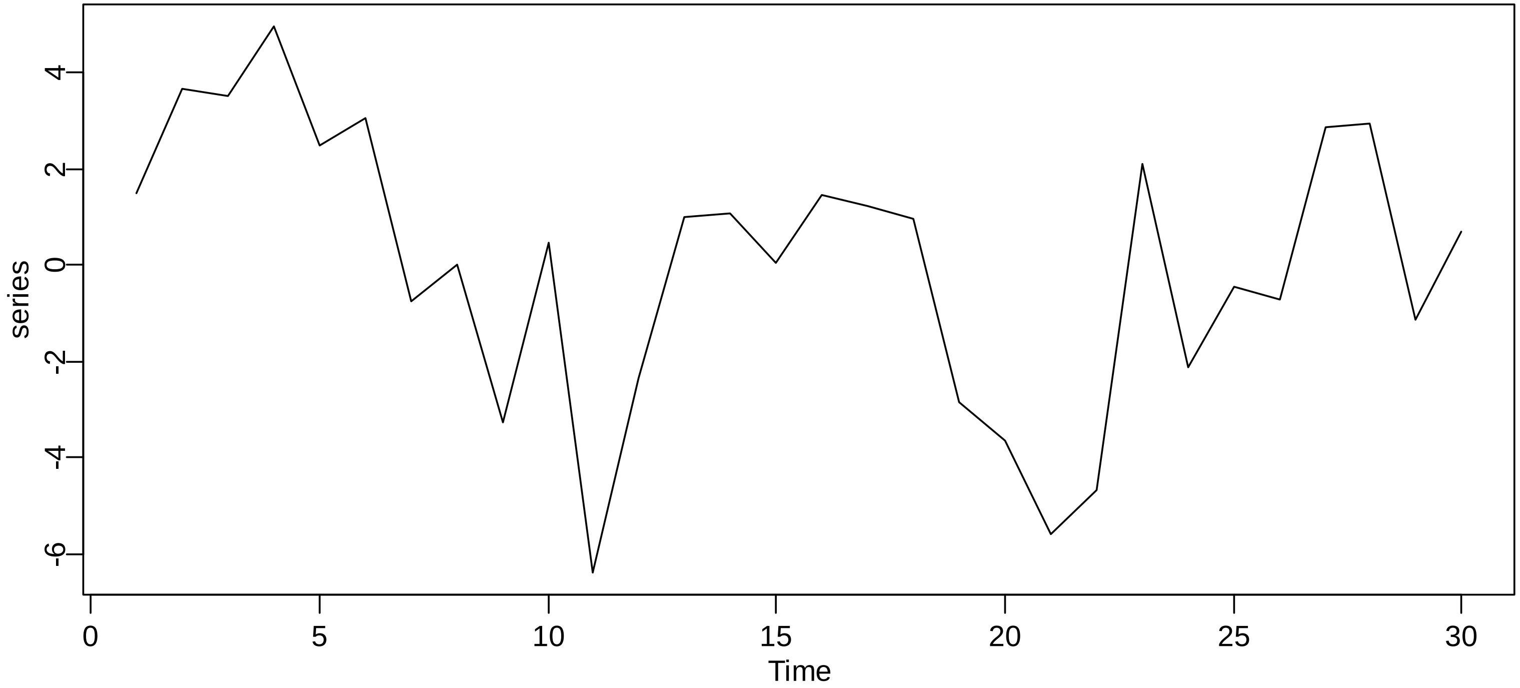
```
> summary(erg.poll,corr=F)
Call: lm(formula = Oxidant ~ Wind + Temp, data = pollute)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)

(Intercept) -5.20334   11.11810  -0.468    0.644
Wind        -0.42706    0.08645  -4.940 3.58e-05 ***
Temp         0.52035    0.10813   4.812 5.05e-05 ***

Residual standard error: 2.95 on 27 degrees of freedom
Multiple R-squared: 0.7773, Adjusted R-squared: 0.7608
F-statistic: 47.12 on 2 and 27 DF,  p-value: 1.563e-09
```

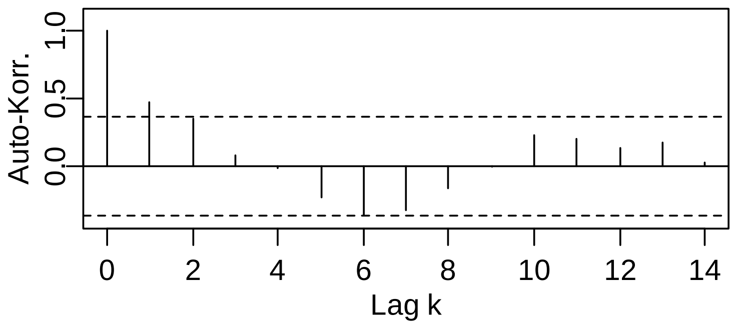
Pollutant Example

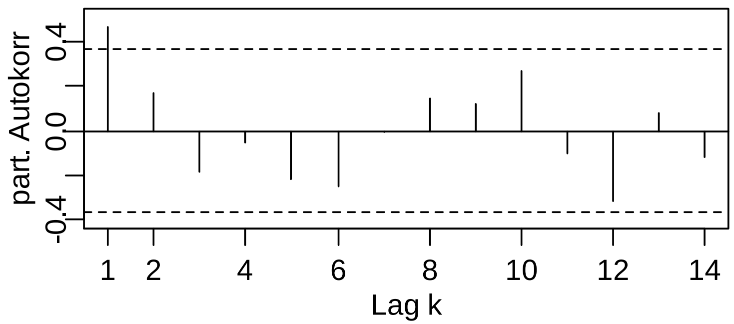
```
> summary(erg.poll,corr=F)
Call: lm(formula = Oxidant ~ Wind + Temp, data = pollute)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.20334   11.11810  -0.468    0.644
Wind        -0.42706    0.08645  -4.940 3.58e-05 ***
Temp         0.52035    0.10813   4.812 5.05e-05 ***

Residual standard error: 2.95 on 27 degrees of freedom
Multiple R-squared: 0.7773, Adjusted R-squared: 0.7608
F-statistic: 47.12 on 2 and 27 DF,  p-value: 1.563e-09
```

Applied Time Series Analysis
FS 2011 – Week 09

Pollutant Example

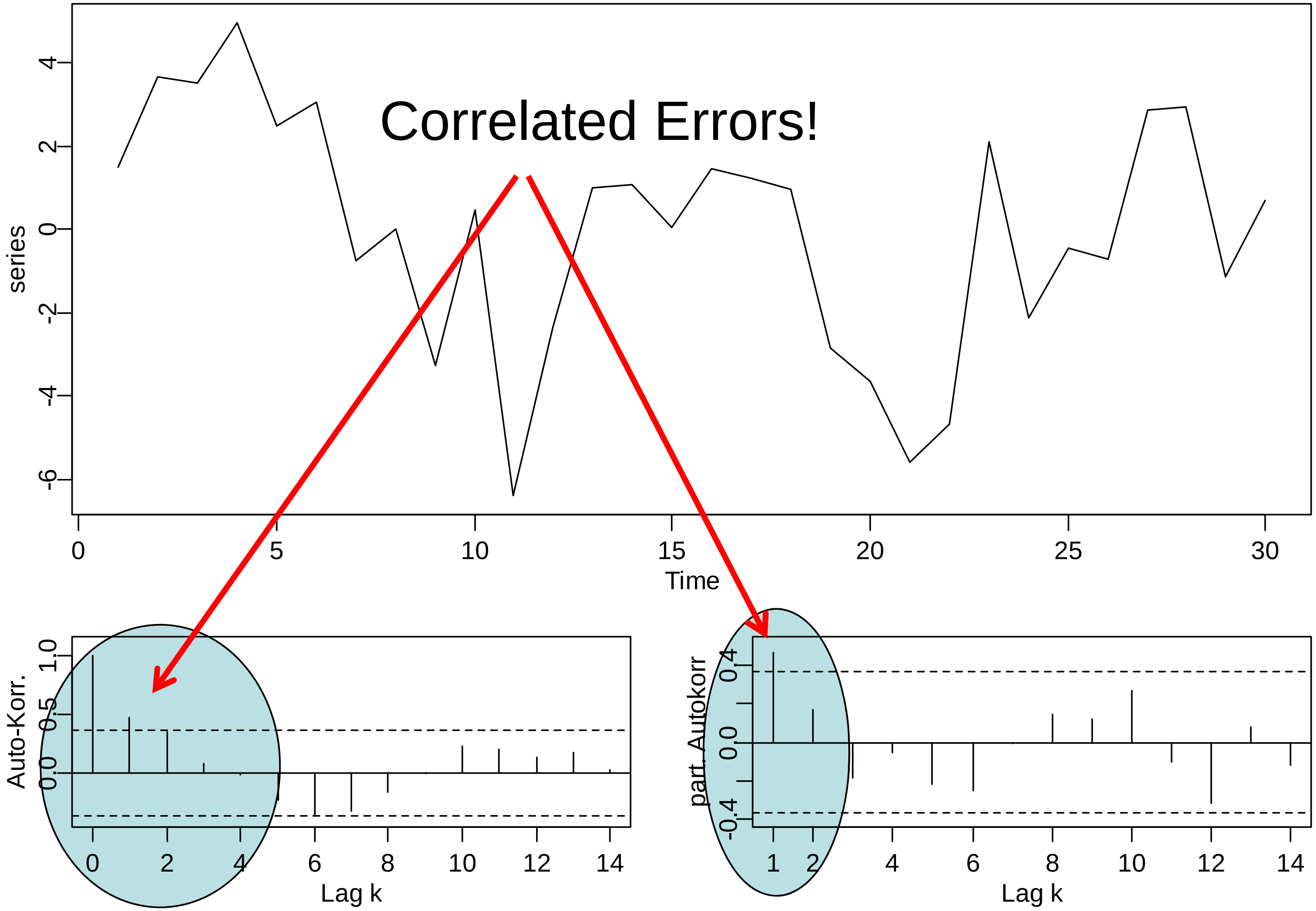






Applied Time Series Analysis
FS 2011 – Week 09

Pollutant Example



Durbin-Watson Test

also see the blackboard...

- The Durbin-Watson approach is a dull test for (auto)-correlated errors in regression modeling
- Many statistics software packages automagically yield a decision or p-value for this test
- A rejection of its null hypothesis should always be taken as a serious hint for correlated errors
- **A non-rejection doesn‘t mean much!**

→ **Better to check ACF/PACF of residuals!**

Pollutant Example

We observe clearly correlated errors/residuals in the pollutant example. They need to be taken into account.

The two major options are:

1) **Cochrane-Orcutt (for AR(p) correlation structure only)**
stepwise approach: i) β, ii) α, iii) β

2) **GLS (Generalized Least Squares, for ARMA(p,q))**
simultaneous estimation of β and α

Cochrane-Orcutt

Stepwise approach: i) β, ii) α, iii) β

→ **see blackboard…**

Generalized Least Squares

simultaneous estimation of β and α

→ **see blackboard…**

Missing Input Variables

- (Auto-)correlated errors are often caused by the non-presence of crucial input variables.

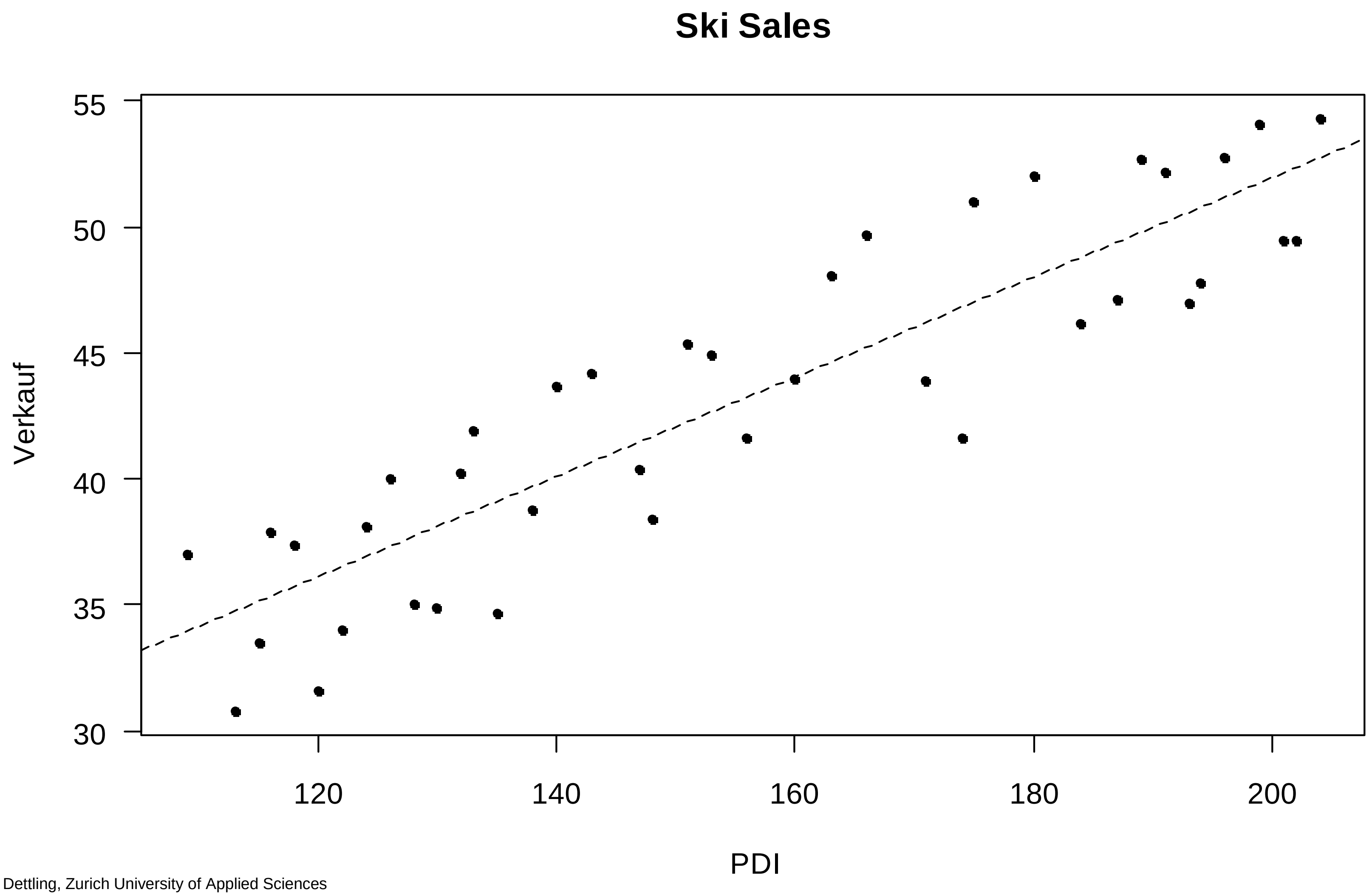
- In this case, it is much better to identify the not-yet-present -variables and include them in the analysis.

- However, this isn‘t always possible.

→ **regression with correlated errors can be seen as a sort of emergency kit for the case where the non-present variables cannot be added.**

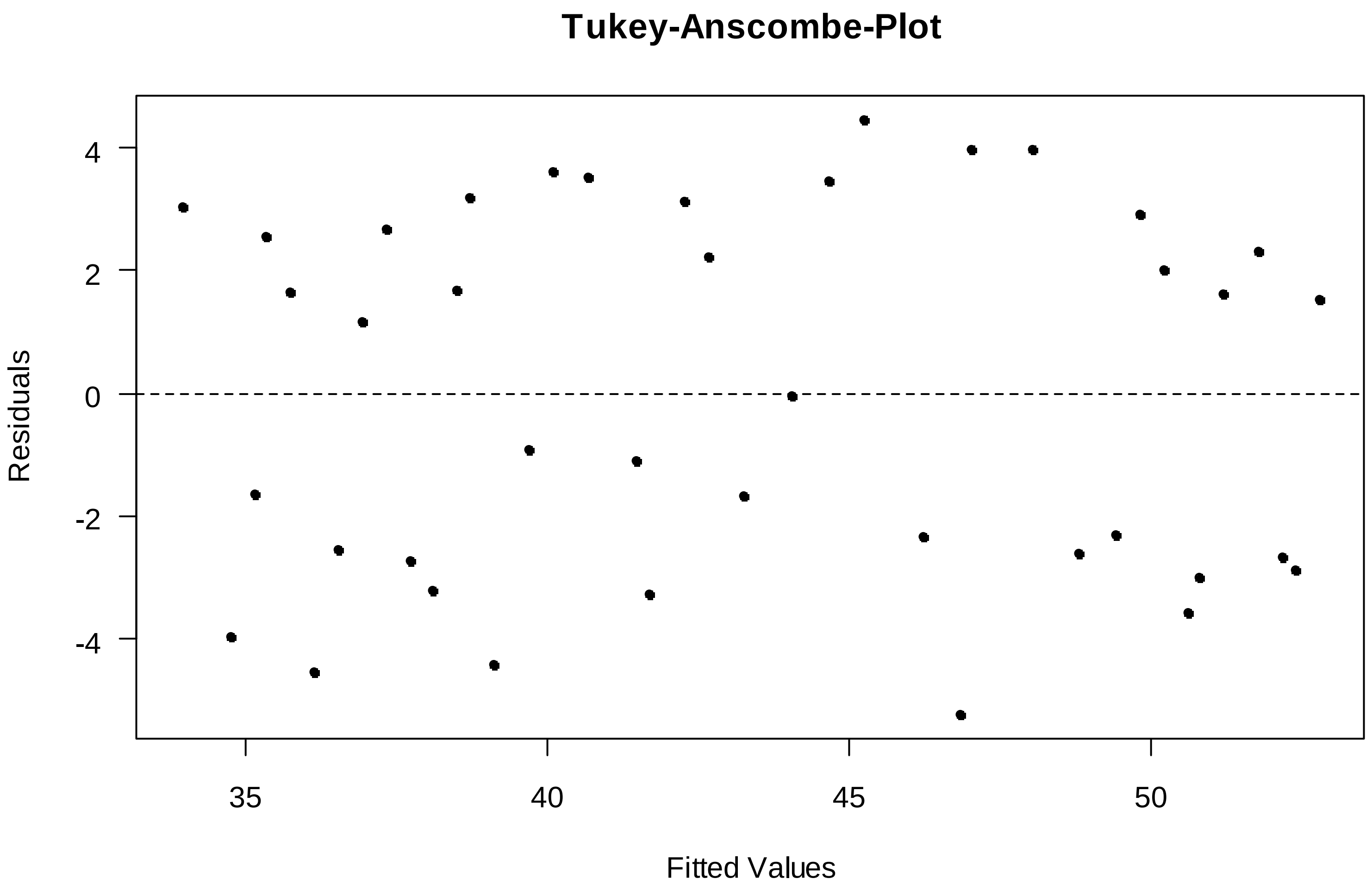
Applied Time Series Analysis
FS 2011 – Week 09

Example: Ski Sales



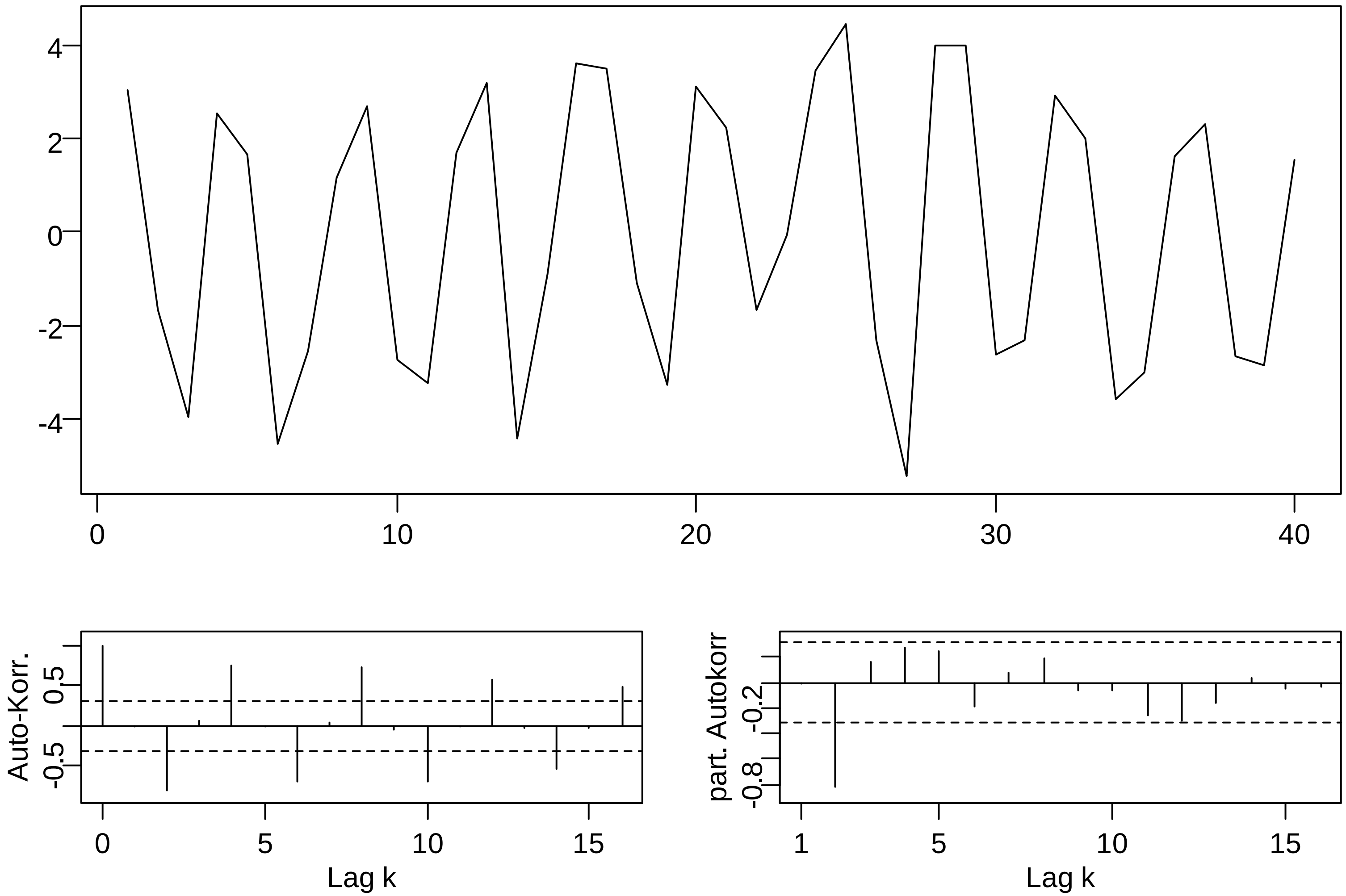
Applied Time Series Analysis
FS 2011 – Week 09

Example: Ski Sales



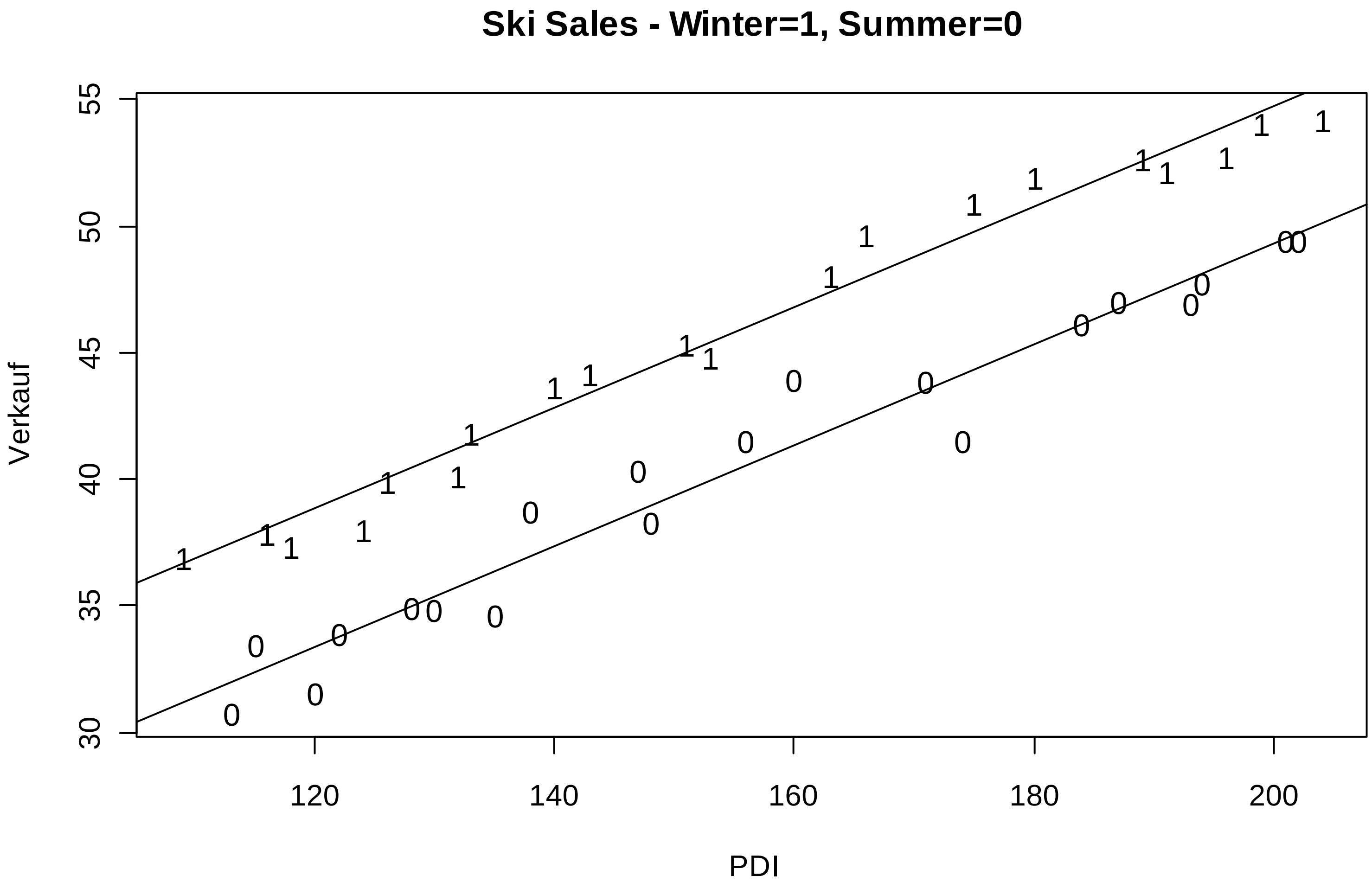
Applied Time Series Analysis
FS 2011 – Week 09

Example: Ski Sales



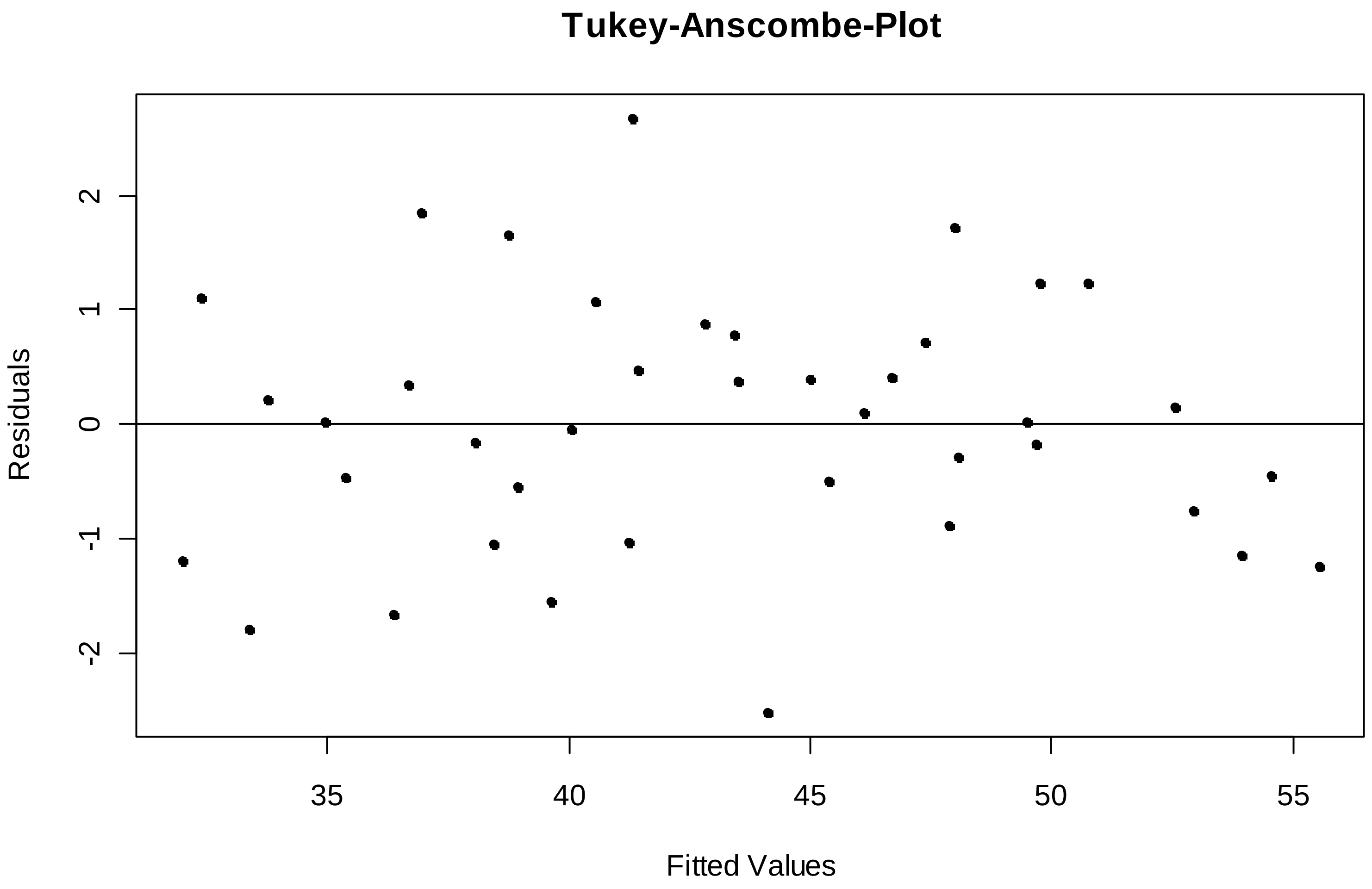
Applied Time Series Analysis
FS 2011 – Week 09

Example: Ski Sales



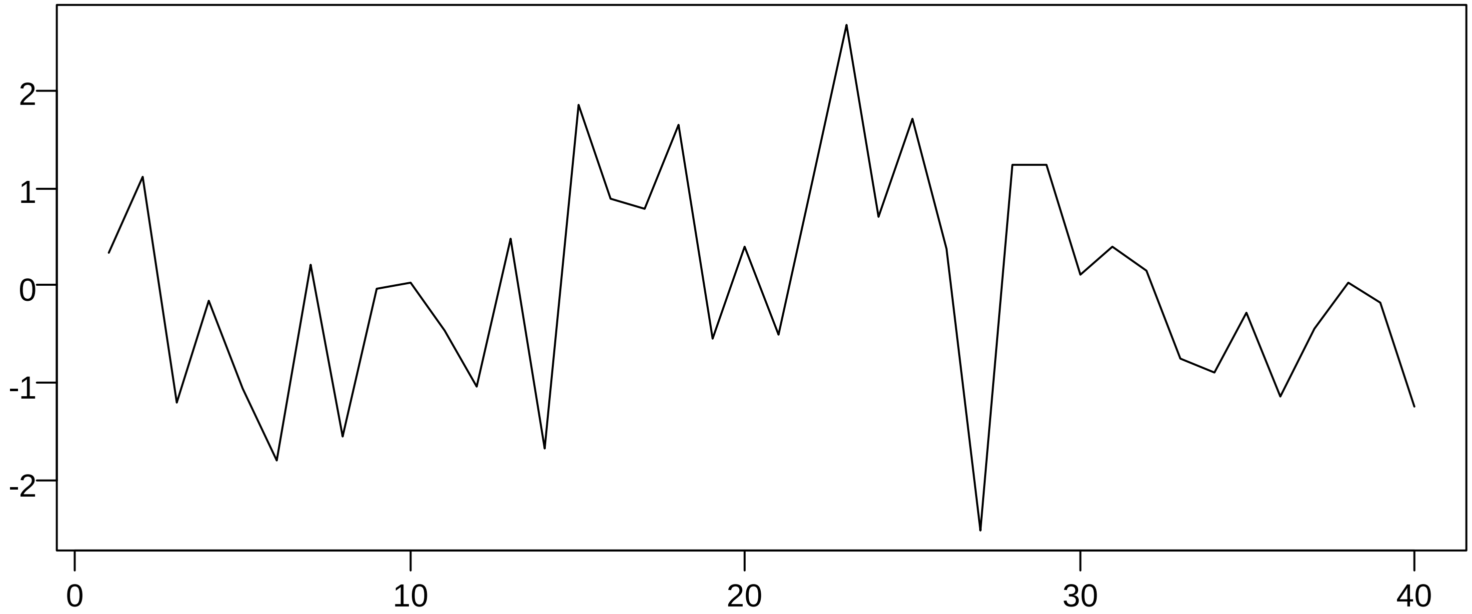
Applied Time Series Analysis
FS 2011 – Week 09

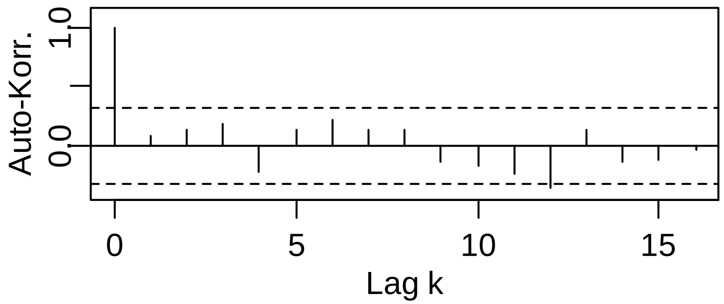
Example: Ski Sales

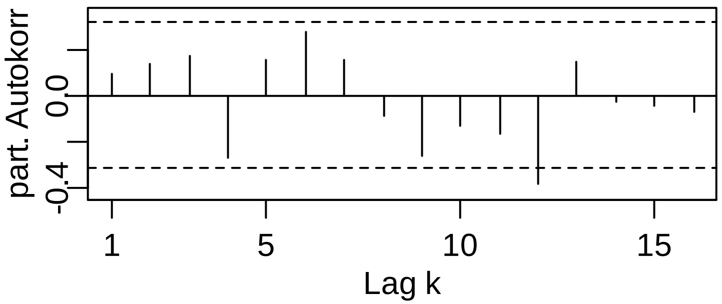


Applied Time Series Analysis
FS 2011 – Week 09

Example: Ski Sales







Ski Sales: Summary

- the first model (sales vs. PDI) showed correlated errors
- the Durbin-Watson test failed to indicate this correlation
- this apparent correlation is caused by ommitting the season
- adding the season removes all error correlation!

→ ***the emergency kit „time series regression" is, after careful modeling, not even necessary in this example. This is quite often the case!***