1 Regression between time series

1.1 Introduction

Classical regression model

$$Y_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \ldots + \beta_m x_i^{(m)} + E_i$$

assumes independent E_is.
When obs. form time series ($i \to t$), this is often voilated.

Consequence: Improper statistical inference
(often too significant p-values, too short confidence intervals)
$\longrightarrow$ Methods to treat this situation correctly are needed!

1.2 Notions and models for time series

a **Stationary time series:**

- Random variables Y_t, $t = 1, 2, ..., T$. $Y_t \sim \mathcal{N}\langle \mu, \sigma^2 \rangle$.
  Expectation and standard deviation indep. of t.
  Correlated! Consecutive observations are more similar than observations at farther time lags.

- Autocorrelation $\rho\langle h \rangle = corr\langle Y_t, Y_{t-h} \rangle$
  (assumed indep. of t)

1.2

b Simple Model: Autoregression (of order 1)

$$(Y_t - \mu) = \alpha(Y_{t-1} - \mu) + D_t \qquad (\alpha < 1)$$

$D_t \sim \mathcal{N}\langle 0, \tau^2\rangle$ independent (from Ys, too)

(Y_t depends only on the preceding time point!

„Markov property")

Autocorrelation: $\rho\langle 1\rangle = \alpha$

$\rho\langle h\rangle = \alpha^h$. Autocorrelation function is exponential.

1.2

c Autoregression of order p:

$(Y_t - \mu) = \alpha_1(Y_{t-1} - \mu) + ... + \alpha_p(Y_{t-p} - \mu) + D_t$

Autocorrelations are functions of parameters $\alpha_1, ..., \alpha_p$.

d Estimation of autocorrelations from the observed series:

$\widehat{\rho}\langle h\rangle = \frac{1}{T}\sum_{t=2}^{T}(Y_t - \widehat{\mu})(Y_{t-1} - \widehat{\mu})$

1.2

e For an AR(1), the corr. for lag 2, $corr\langle Y_t, Y_{t-2}\rangle = \rho\langle 2\rangle = \alpha^2$, is only indirect, via dependence of Y_t on Y_{t-1}, which in turn depends on Y_{t-2}.

The partial autocorrelation (for lag 2) is the corr. in the conditional joint distribution of Y_t and Y_{t-2}, given Y_{t-1}.
It is 0 for AR(1).
The part. autocorr. for lag $p+1$ is 0 for AR(p).
$\longrightarrow$ Model selection ($=$ selection of p)

1.2

f Outlook: Time series models can include

- trend and
- seasonal cycles

... and many, many other features ...

1.3 Regression with AR errors

a Model

$$Y_t = \sum_j \beta_j X_t^{(j)} + E_t \,, \qquad E_t \sim \mathrm{AR}\langle 1 \rangle$$

$\longrightarrow$ covariance matrix of errors

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \alpha & \alpha^2 & \ldots & \alpha^{T-1} \\ \alpha & 1 & \alpha & \ldots & \alpha^{T-2} \\ \vdots & & & \vdots & \\ \alpha^{T-1} & \alpha^{T-2} & \ldots & \alpha & 1 \end{bmatrix}$$

1.3

b Use theory of Generalized Least Squares!
Theory: Matrix formulation of regression model!

R-function `gls` in package `nlme`

c The AR parameter must be estimated from data = errors E_i!

- Estimate errors from fitting ordinary LS (= residuals)
  Estimate α from residual series and use for `gls`
  „Cochran-Orcutt".

- Joint estimation of α and regression coefficients.
  `gls` does this, too.

Messages

- Models for stationary time series include autoregression (of order p)
- These may be used to model correlations of random errors in “ordinary” regression.
- If the corr. structure of the errors is determined Generalized Least Squares estimates regr. coefficients with correct inference (p-values and conf. int.)
- If inference is based on ordinary least squares it may be completely useless.