Tues, 4-16-19

11. Multiple Linear Regression

* Lightning talk

•Feedback on A10

1.Model and assumptions

2.Analysis of residuals

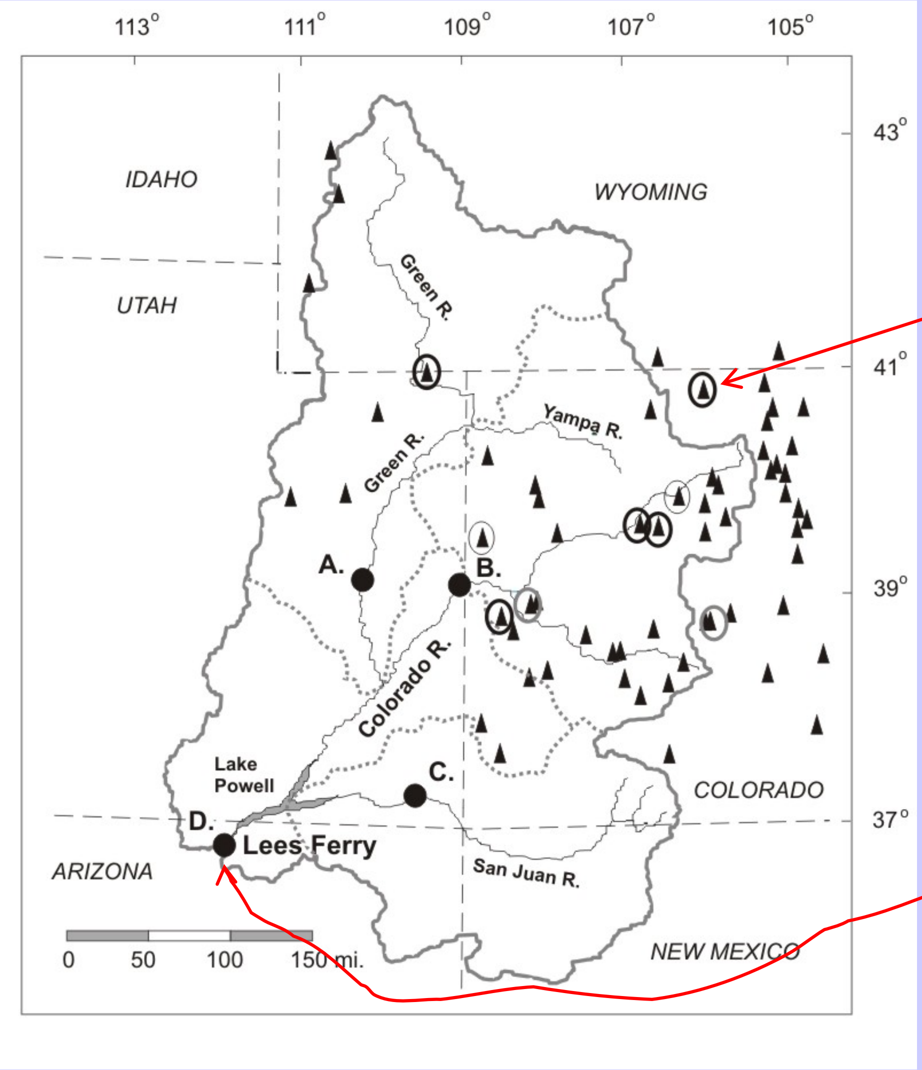
3.Multicolinearity

Read notes_11.pdf

A10 Feedback

1. Download A10x.pdf from D2L
2. Automatic points, for running assignment and having uploaded by due time, is already marked in parentheses at top of first page
3. Each assignment has maximum possible 10 points; if you make no deductions, score is 10/10
4. A10x is color coded for points; purple=1; yellow=0.5; blue=0.5
5. Open your copy of the same assignment pdf you uploaded
6. In Acrobat Reader, using “Add text box,” mark in right margin for deductions only, with deduction and segment reference : (eg., -0.5 A); round to tenths in deductions (e.g., no -0.25)
7. At top of your pdf, mark grade like this : 9.5/10
8. If necessary, put any comments at top near the grade
9. Upload your self-graded pdf to folder A10_graded in D2L

Example Setting—Streamflow Reconstruction



Predictors: Indices of tree-ring width at 31 sites – 7 used

$x_{t,i}$

Predictand: annual natural flow of the Colorado River at Lee Ferry, Arizona

y_t

Woodhouse et al. 2006. Water Resources Research, 42, W05415, doi:10.1029/2005WR004455

Model

$$y_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \ldots + b_K x_{i,K} + e_i$$

$x_{i,j}$ = value of j^{th} predictor in year i

b_0 = regression constant

b_j = coefficient on the j^{th} predictor

K = total number of predictors

y_i = predictand in year i

e_i = error term

Predictions

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{i,1} + \hat{b}_2 x_{i,2} + \ldots + \hat{b}_K x_{i,K}$$

$x_{i,j}$ = value of j^{th} predictor in year i, $\quad j \le K$

$\hat{b}_0, \hat{b}_1, \cdots \hat{b}_K$ = estimated regression constant and coefficients

$\hat{y}_i$ = predicted value for year i

Residuals

$$\hat{e}_i = y_i - \hat{y}_i$$

$$y_i = \text{observed value of predictand in year } i$$

$$\hat{y}_i = \text{predicted value of predictand in year } i$$

Assumptions

1. Relationships linear
2. Residuals uncorrelated with predictors
3. Residuals have constant variance
4. Residuals not autocorrelated
5. Residuals normally distributed

Analysis of residuals

$$\hat{e}_t = y_t - \hat{y}_t, \quad t = 1, N$$

Observed – Predicted

N = number of observations in calibration period

Diagnostic Plots for Residuals Analysis

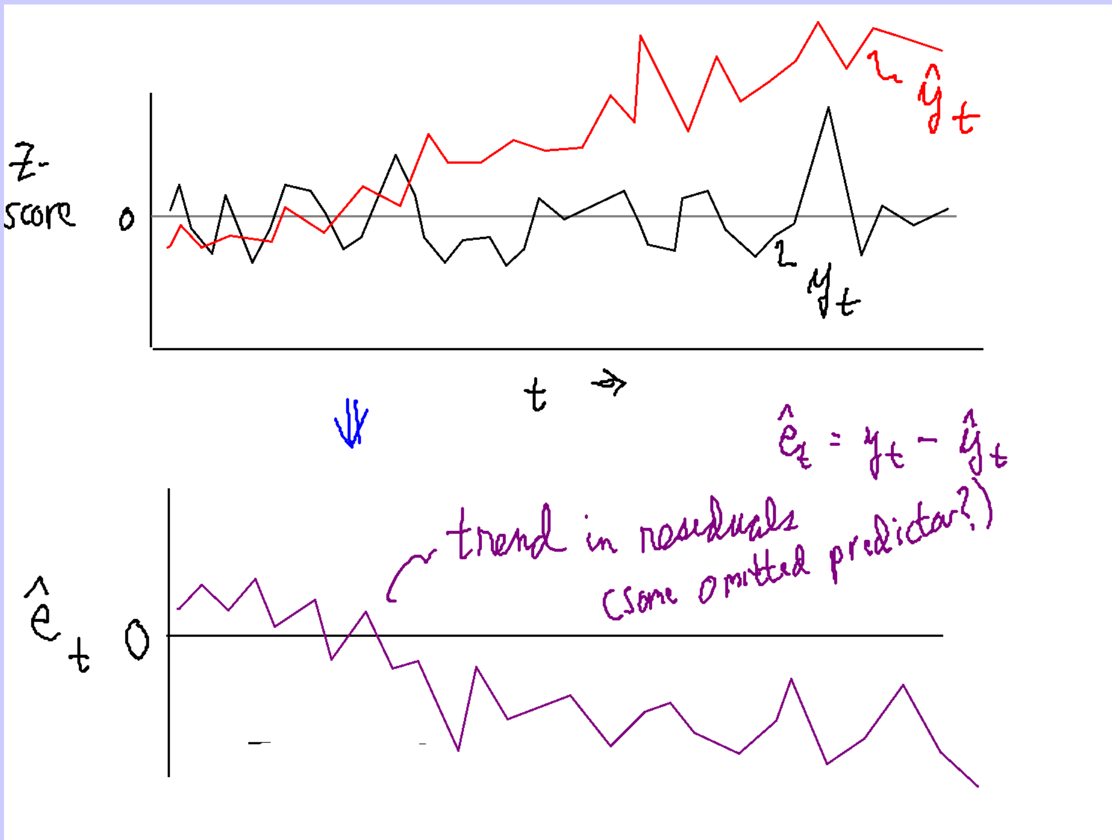
1 Histogram of $\hat{e}_t$

2* Time plot of $\hat{e}_t$

3* Acf of $\hat{e}_t$

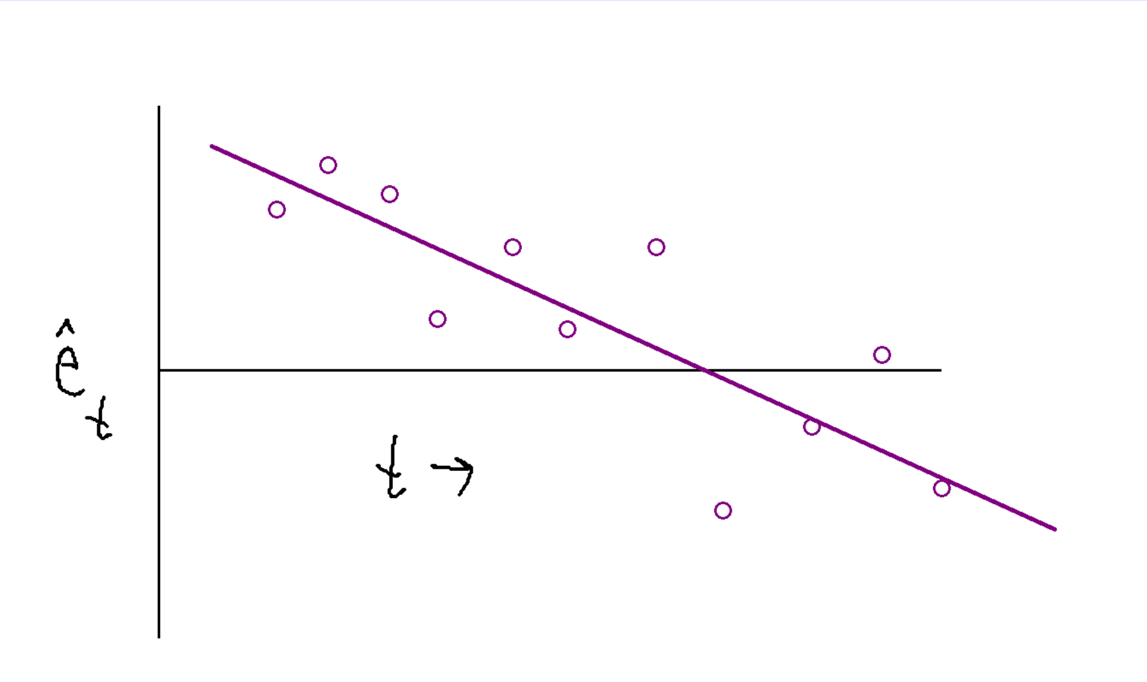
4* Scatter of $\hat{e}_t$ on $\hat{y}_t$

Time plot of residuals: trend?



- Omitted predictor?
- e.g., CO2 fertilization distorting a tree-ring reconstruction of precipitation

Trend in residuals -- identifying



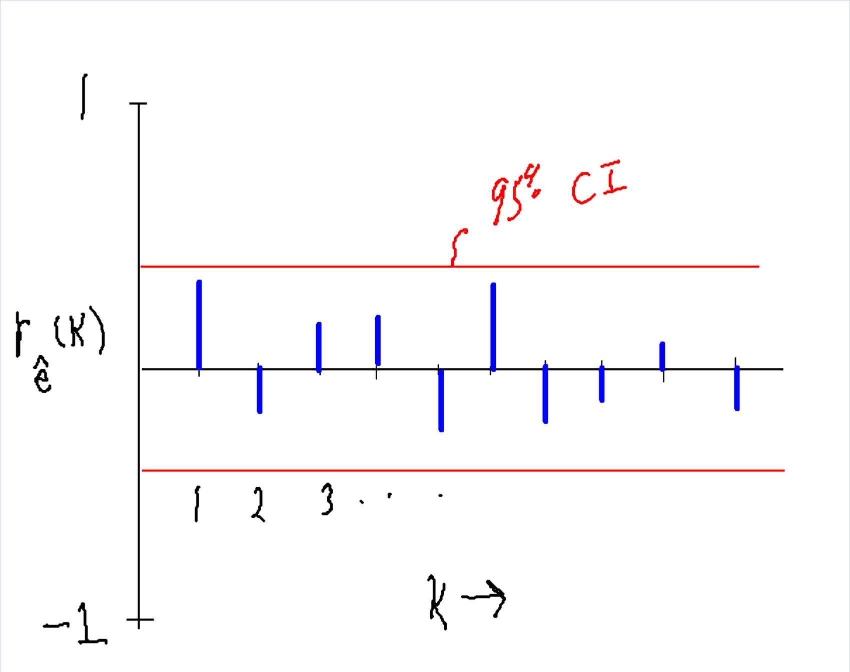
$$\hat{e}_t = y_t - \hat{y}_t$$

Regress $\hat{e}_t$ on t

$\downarrow$

Slope significantly different from 0?

ACF of residuals



a) Individual $r_e(k)$ small?

b) Sequence of $r_e(k)$ small?

Portmanteau statistic

$$Q = N\left(\sum_{k=1}^{K} r_e^2(k)\right)$$

Sample length N
(eqn 18 in notes)

Caveat on low-lag autocorrelations of regression residuals

- Confidence interval computed as if series were an observed time series rather than regression residuals not strictly applicable
- Actual CI at low lags may be narrower than indicated

Durbin-Watson (DW) statistic

DW test -- hypothesis

The UNKNOWN true errors, as opposed to the regression residuals

Assume e_t generated by an AR(1) process with coefficient p

$$e_t = pe_{t-1} + n_t, \text{ where } n_t \text{ is normally distributed random noise}$$

H0: $p = 0 \quad \rightarrow \quad$ errors not autocorrelated

H1: $p > 0$ or $p < 0$

DW test -- statistic

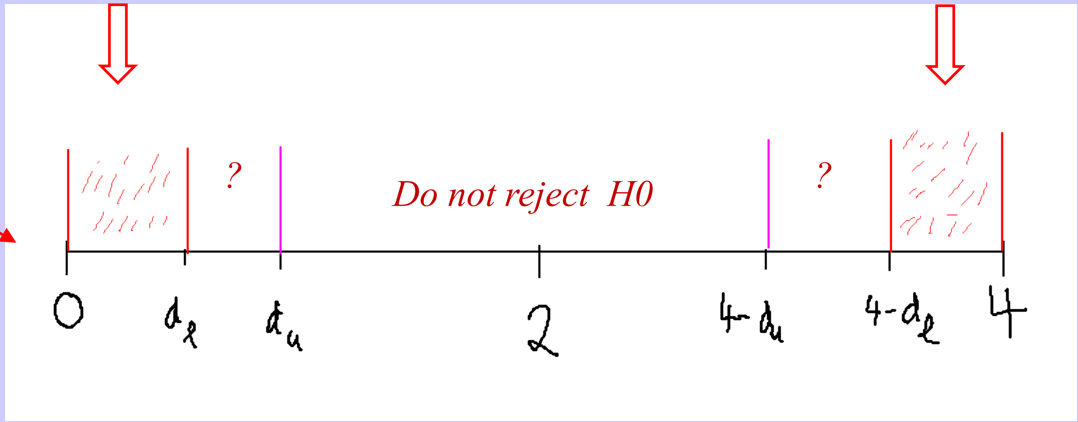
$$d = \frac{\sum_{i=1}^{N} \left(\hat{e}_i - \hat{e}_{i-1}\right)^2}{\sum_{i=1}^{N} \hat{e}_i^2}, \quad \text{where } \hat{e}_i, i = 1, N \text{ are the regression residuals}$$

$$d = 2\left(1 - p\right)$$

Reject H0

Reject H0

d approaches 0 for perfect autocorrelation

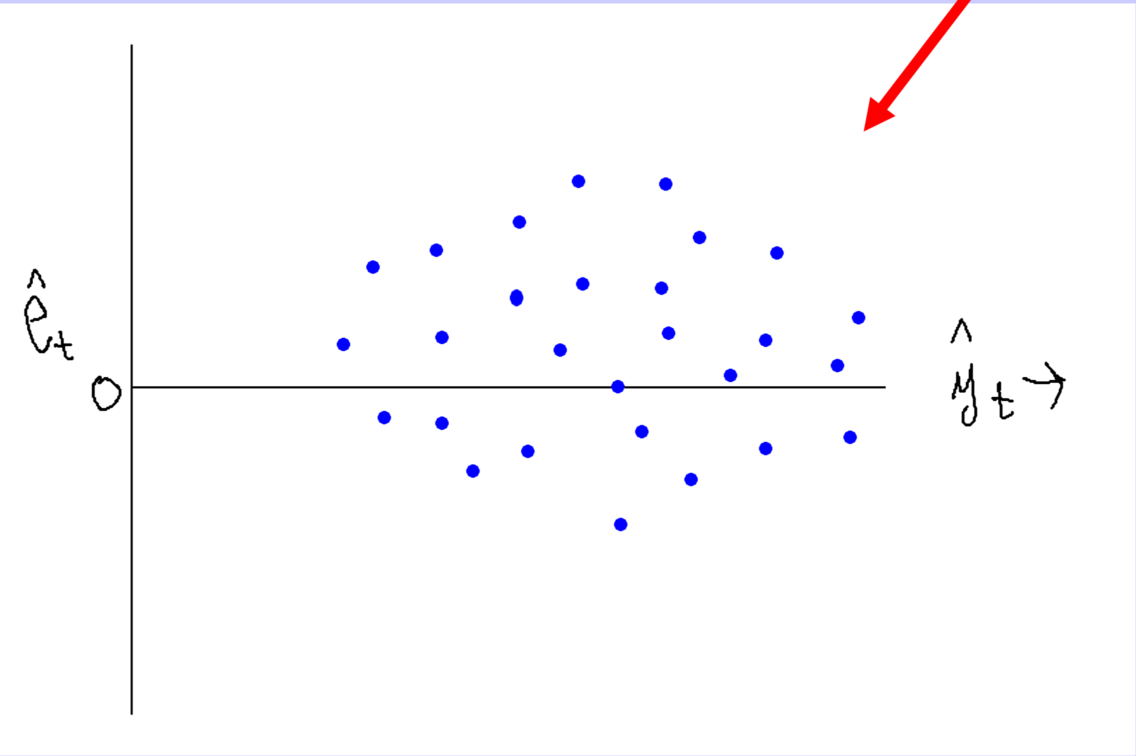


+ autocorrelation

- autocorrelation

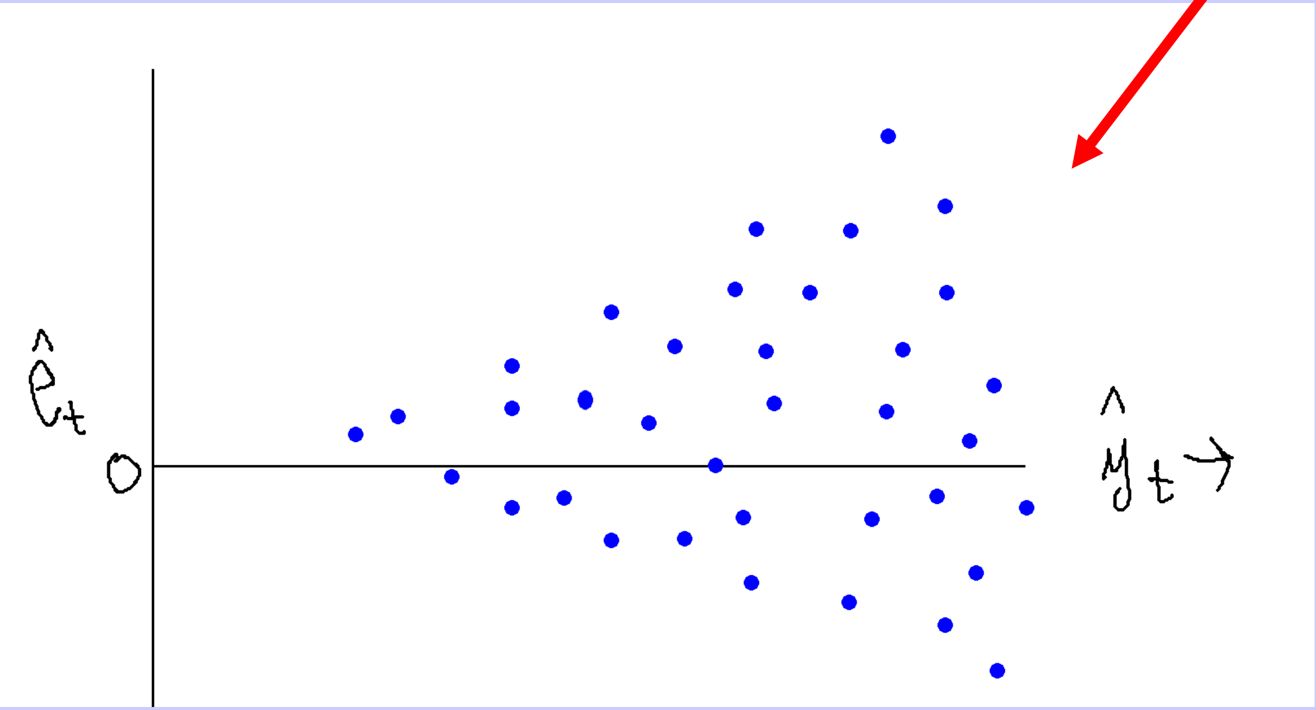
Residuals vs predictions

Ideal pattern



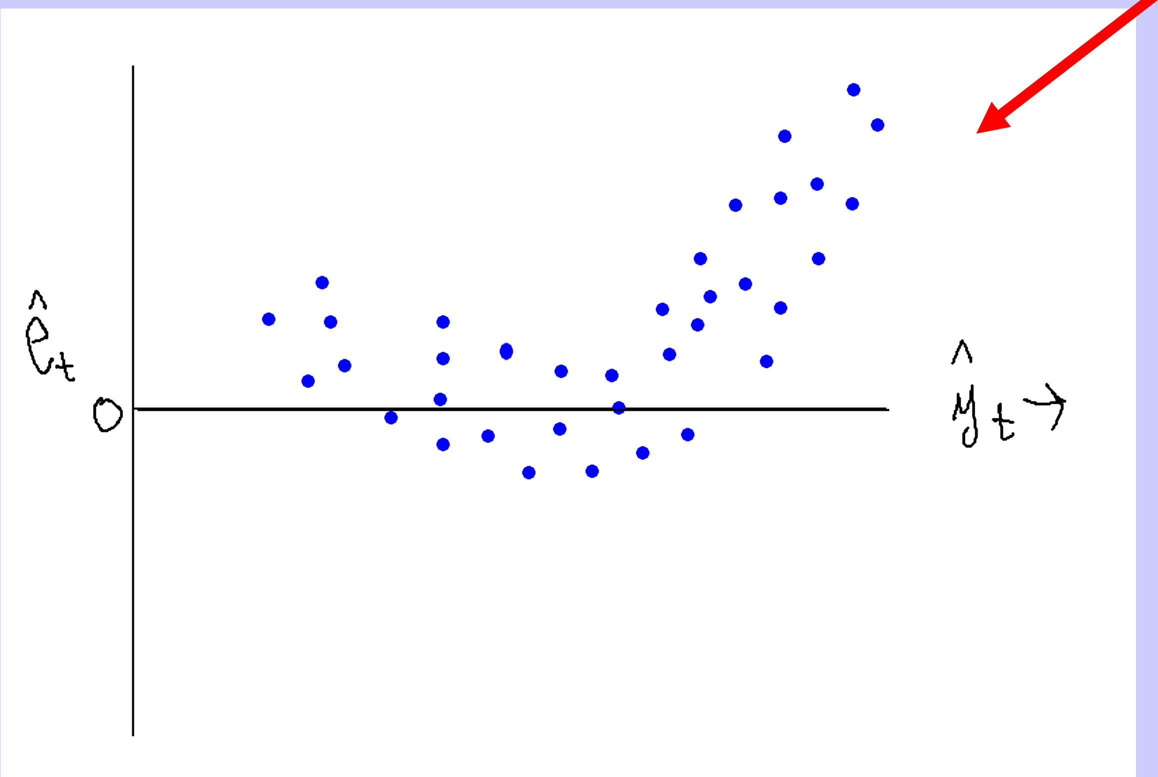
Residuals vs predictions

“heteroskedastic”: try transform of predictand



Residuals vs predictions

Curvature:
try transforms of predictors and predictand



Multicolinearity

$$\hat{y}_t = \hat{b}_0 + \hat{b}_1 x_{t,1} + \hat{b}_2 x_{t,2} + \cdots + \hat{b}_k x_{t,k}$$

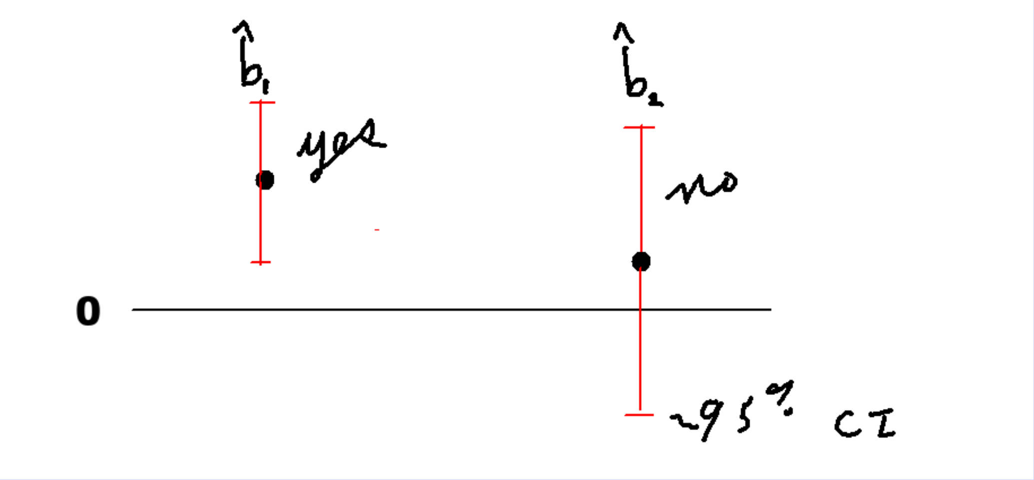
Predictors strongly intercorrelated

Amplified variance of estimated coefficients

Multicolinearity

$$\hat{y}_t = \hat{b}_0 + \hat{b}_1 x_{t,1} + \hat{b}_2 x_{t,2} + \cdots + \hat{b}_k x_{t,k}$$

Regression coefficient significant?



CI depends on :

1. MSE, or strength of regression
2. **Correlation matrix of predictors**

Possible Consequences of Multicolinearity

1. Physically important x_i has "insignificant" coefficient
2. Magnitude & sign of coefficients illogical
3. Slight change to problem (e.g., delete an observation) gives large change in estimated $\widehat{b}_i$
4. Estimated $\widehat{b}_i$ not amenable to physical interpretation

BUT

PREDICTIONS MAY STILL BE VALID

Testing for Multicolinearity

Variance inflation factor (VIF)

Regression model with K predictors $x_1, x_2, \ldots, x_K$

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{i,1} + \hat{b}_2 x_{i,2} + \ldots + \hat{b}_K x_{i,K}$$

1) Regress each predictor on all other predictors in MLR

2) Compute R_i^2 for regression of x_i on $x_j, \quad j \neq i$

3) $\mathrm{VIF}_i = \dfrac{1}{\left(1 - R_i^2\right)}$

Rules of Thumb for Multicolinearity

$$R_i^2 > 0.80 \qquad\qquad R_i^2 > 0.90$$

1) $\text{VIF}_i > 5, \quad \text{or} \quad \text{VIF}_i > 10$

2) $\text{VIF} \gg 1$