Thurs, 4-18-19

Multiple Linear Regression (cont.)

1. **Automated selection of predictors**
2. **Demo11a: interactive tool "stepwise"**
3. **Sample runs of geosa11**

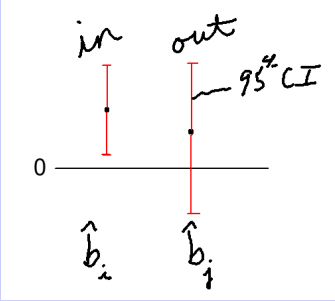
Assignment a11: due next Tuesday

Regression: Automated selection of predictors

$x_1, x_2, \ldots x_m$ Potential predictors

DEMO011A (STEPWISE TOOL)

GEOSA11 (CALIBRATION MODE)



-enter or remove at each step

-criterion: significant "t" statistic

$$t = \frac{\hat{b}_i}{\mathrm{SE}\left(\hat{b}_i\right)}$$

$p_{enter} = 0.05$

$p_{remove} = 0.10$

-in order of max reduction of $\mathrm{var}\left(\hat{e}^2\right)$

-entry only; no removal

-no cutoff criteria (all may enter)

Stepwise Alternatives

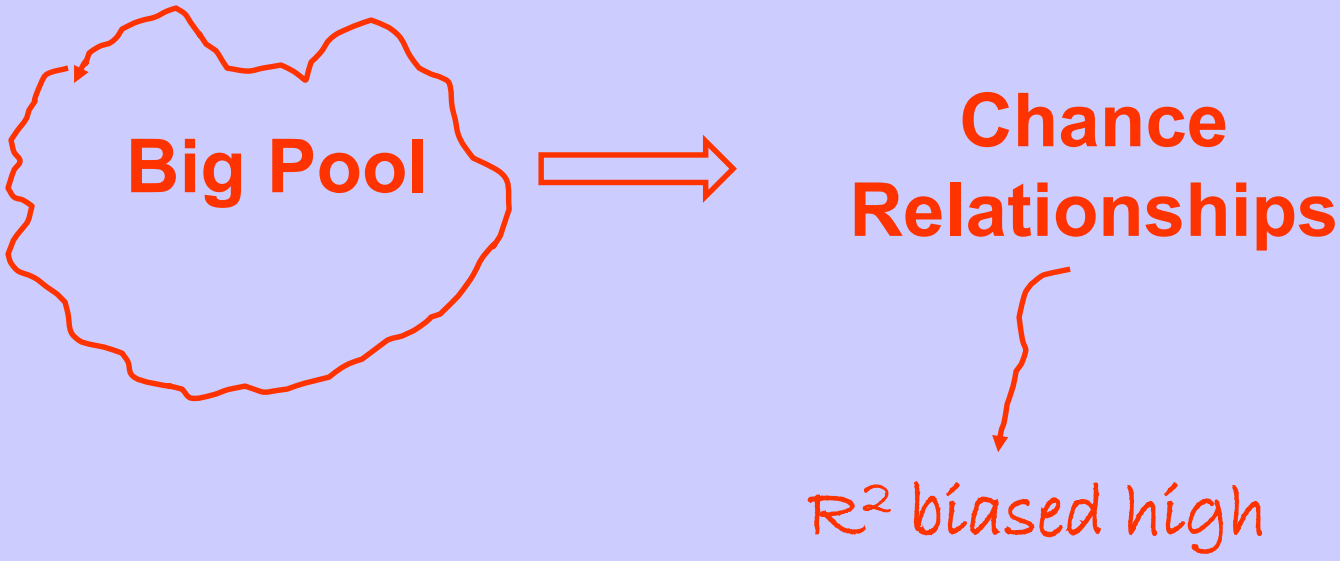
- Forward
- Backward
- Bidirectional (used by demo11a)
- Best subsets

Stepwise

Pro: Natural guard against multicolinearity

Con: Data “dredging”

Caveat to Automated Selection



See Rencher and Pun (1980) reference

Demo11a

1. **Matlab's s "stepwise" tool**
2. **vif function to compute variance inflation factor**

- Stepwise is a graphical interactive tool to explore stepwise regression
- Assumes
  - Input matrix of predictors, X, with no "ones" column
  - Input vector of predictand, y

Demo11a – sample problem

1) y_t Annual discharge of North Fork American River,

2) **X** Matrix of 9 Apr 1 SWE at 9 snow course in Sierra Nevada for same period, PLUS and additional fake predictor:

$$x_{10} = x_6 + \nu,$$

X_6 is the Castle Creek snow course, whick I know has strongest bivariate relationship with the river discharge)

$$\text{where} \quad \nu \sim \mathrm{N}\left(0, \frac{s_{x_6}^2}{400}\right)$$

Random noise with mean 0 and variance 1/400 the variance of predictor x_1 (ratio of standard deviations is 1/20)

Function vif for Variance inflation factor

$$\mathrm{VIF}_i = \frac{1}{1 - R_i^2}$$

Recall this equation (last lecture), which give VIF for the i^{th} predictor as a function of how much variance of that precictor is explained in a regression on the other predictors.

Structure with various fields, including
- VIF for set of selected predictors
- VIF_i for individual predictors
- Variance-explained quantities

Predictor matrix

An options setting: use k=1

$$\mathrm{R} = \mathrm{vif}\left(\mathrm{X}, \mathrm{k}\right)$$

User-written Matlab function

Run demo11a

Sample run of geosa11 (Calibration mode)…