Regression wrap-up and Granger causality

Nisheeth

Regression Model Building

- Setting: Possibly a large set of predictor variables (including interactions).
- Goal: Fit a parsimonious model that explains variation in Y with a small set of predictors
- Automated procedures and all possible regressions:
  - Backward Elimination (Top down approach)
  - Forward Selection (Bottom up approach)
  - Stepwise Regression (Combines Forward/Backward)

Backward Elimination Traditional Approach

- Select a significance level to stay in the model (e.g. SLS=0.20, generally .05 is too low, causing too many variables to be removed)
- Fit the full model with all possible predictors
- Consider the predictor with lowest *t*-statistic (highest *P*-value).
  - If $P >$ SLS, remove the predictor and fit model without this variable (must re-fit model here because partial regression coefficients change)
  - If $P \leq$ SLS, stop and keep current model
- Continue until all predictors have *P*-values below SLS

Forward Selection – Traditional Approach

- Choose a significance level to enter the model (e.g. SLE=0.20, generally .05 is too low, causing too few variables to be entered)
- Fit all simple regression models.
- Consider the predictor with the highest *t*-statistic (lowest *P*-value)
  - If $P \leq$ SLE, keep this variable and fit all two variable models that include this predictor
  - If $P >$ SLE, stop and keep previous model
- Continue until no new predictors have $P \leq$ SLE

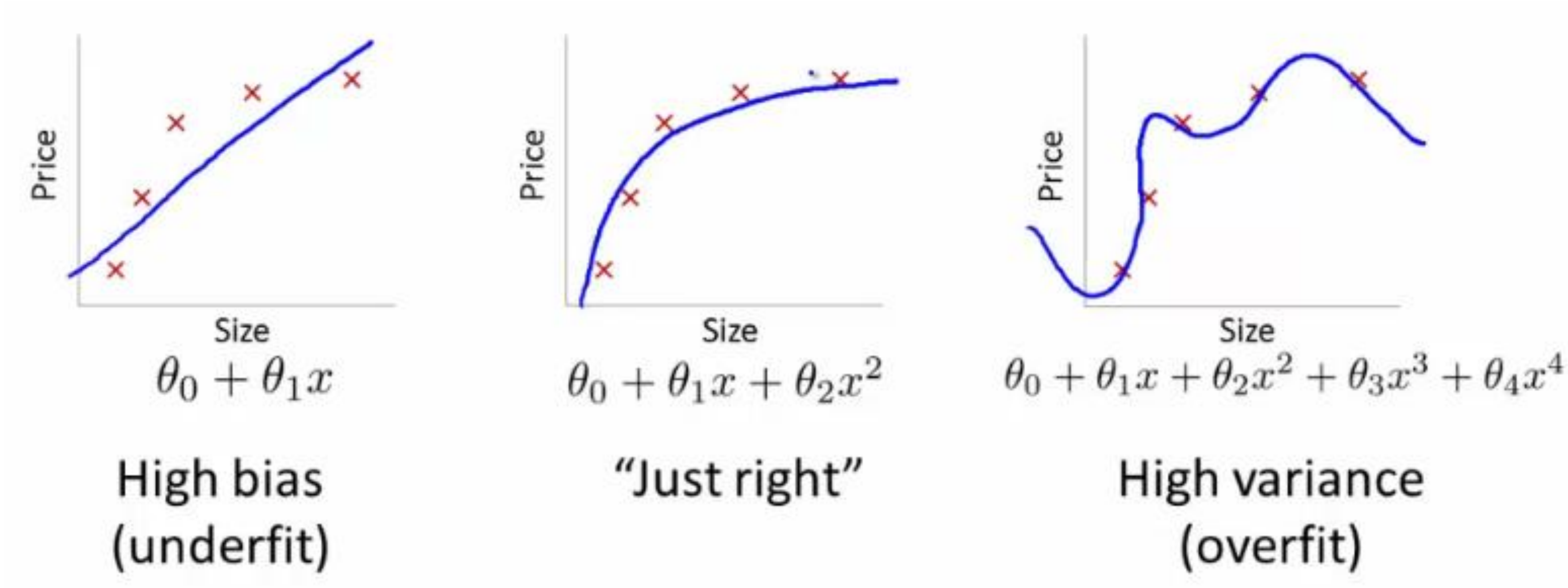
Stepwise Regression – Traditional Approach

- Select SLS and SLE (SLE<SLS)
- Starts like Forward Selection (Bottom up process)
- New variables must have $P \leq$ SLE to enter
- Re-tests all "old variables" that have already been entered, must have $P \leq$ SLS to stay in model
- Continues until no new variables can be entered and no old variables need to be removed

Model-based criteria

- $\frac{dR^2}{dp} < \epsilon$
- R^2_{adj}
- $AIC = 2p - 2\log(\hat{L})$
- $BIC = p\log(n) - 2\log(\hat{L})$
  - Parameter count p
  - Sample count n
  - Model likelihood $\hat{L}$ = p(y|w, x,σ^2)

Overfitting still a concern



Cross validation

- How well does a model fit on one set of data (training sample predict previously unseen data (validation sample).
- Training set should have at least 6-10 times as many observations as potential predictors
- Models should give similar model fits based on regression coefficients and model selection criteria
- Mean Square Prediction Error when training model is applied to validation sample:

$$MSPR = \frac{\sum_{i=1}^{n*}\left(Y_i - \hat{Y}_i\right)^2}{n*} \qquad \hat{Y}_i = b_0^T + b_1^T X_{i1}^V + \ldots + b_{p-1}^T X_{i,p-1}^V$$

Correlation ≠ causation

GRANGER CAUSALITY

- In most regressions, it is very hard to discuss causality. For instance, the significance of the coefficient β in the regression
$$y_i = \beta x_i + \varepsilon_i$$
only tells the 'co-occurrence' of x and y, not that x causes y.
- In other words, usually the regression only tells us there is some 'relationship' between x and y, and does not tell the nature of the relationship, such as whether x causes y or y causes x.

GRANGER CAUSALITY

- In principle, the concept is as follows:

- If X causes Y, then, changes of X happened first then followed by changes of Y.

GRANGER CAUSALITY

- If X causes Y, there are two conditions to be satisfied:

1. X can help in predicting Y. Regression of X on Y has a big R^2
2. Y can not help in predicting X.

GRANGER CAUSALITY

- If we restrict ourselves to linear functions, y fails to Granger-cause x if

$$MSE\left[\hat{E}\left(x_{t+s} \middle| x_t, x_{t-1}, \cdots\right)\right] = MSE\left[\hat{E}\left(x_{t+s} \middle|_{t+s} \middle| x_t, x_{t-1}, \cdots, y_t, y_{t-1}, \cdots\right)\right]$$

- Equivalently, we can say that x is exogenous in the time series sense with respect to y, or y is not linearly informative about future x.

Possible outcomes

1. X → Y
2. X ← Y
3. X ←→ Y
4. X Y

TESTING GRANGER CAUSALITY

- The simplest test is to estimate the regression which is based on

$$x_t = c_1 + \sum_{i=0}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{p} \beta_j y_{t-j} + u_t$$

using OLS and then conduct a F-test of the null hypothesis

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0.$$

TESTING GRANGER CAUSALITY

1. Run the following regression, and calculate RSS (full model)

$$x_t = c_1 + \sum_{i=0}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{p} \beta_j y_{t-j} + u_t$$

2. Run the following limited regression, and calculate RSS (Restricted model).

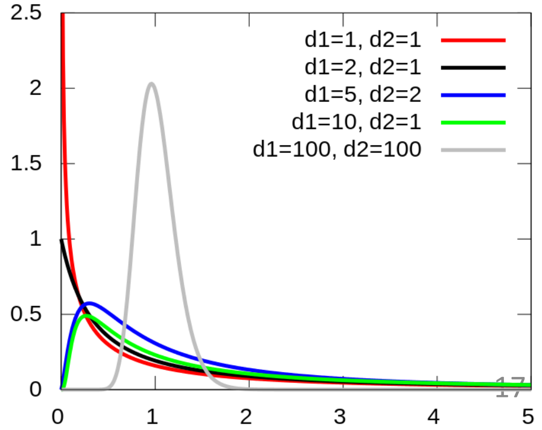
$$\boldsymbol{x_t} = \boldsymbol{c}_1 + \sum_{\boldsymbol{i}=0}^{\boldsymbol{p}} \boldsymbol{\alpha_i} \boldsymbol{x_{t-i}} + \boldsymbol{u_t}$$

TESTING GRANGER CAUSALITY

3. Do the following F-test using SSR obtained from stages 1 and 2:

$$F = \frac{SSR_{small} - SSR_{big}}{SSR_{big}} \times \frac{\#samples - \#params\_big}{\#params\_small}$$

F(p_big-p_small, n-p_big) will give p-value



TESTING GRANGER CAUSALITY

5. If H_0 rejected, then X causes Y.

- This technique can be used in investigating whether or not Y causes X.

Example of the Usage of Granger Test

World Oil Price and Growth of US Economy (Hamilton, 1996)

- Does the increase of world oil price influence the growth of US economy or does the growth of US economy effects the world oil price?
- Model:

$$Z_t = a_0 + a_1 Z_{t-1} + \ldots + a_m Z_{t-m} + b_1 X_{t-1} + \ldots b_m X_{t-m} + \varepsilon_t$$

$Z_t = \Delta P_t$; changes of world price of oil

$X_t = log\ (GNP_t / GNP_{t-1})$

World Oil Price and Growth of US Economy

- There are two causalities that need to be observed:

(i) H_0: Growth of US Economy does not influence world oil price

Full:

$$Z_t = a_0 + a_1 Z_{t-1} + \ldots + a_m Z_{t-m} + b_1 X_{t-1} + \ldots + b_m X_{t-m} + \varepsilon_t$$

Restricted:

$$Z_t = a_0 + a_1 Z_{t-1} + \ldots + a_m Z_{t-m} + \varepsilon_t$$

World Oil Price and Growth of US Economy

(ii) H_0 : World oil price does not influence growth of US Economy

- Full :

$$X_t = a_0 + a_1 X_{t-1} + \ldots + a_m X_{t-m} + b_1 Z_{t-1} + \ldots + b_m Z_{t-m} + \varepsilon_t$$

- Restricted:

$$X_t = a_0 + a_1 X_{t-1} + \ldots + a_m X_{t-m} + \varepsilon_t$$

World Oil Price and Growth of US Economy

- F Tests Results:

1. Hypothesis that world oil price does not influence US economy is rejected. It means that the world oil price does influence US economy .
2. Hypothesis that US economy does not affect world oil price is not rejected. It means that the US economy does not have effect on world oil price.

World Oil Price and Growth of US Economy

- Summary of Results

Null Hypothesis (H_0)	(I)F(4,86)	(II)F(8,74)
I. Economic growth ≠→World Oil Price	0.58	0.71
II. World Oil Price≠→Economic growth	5.55	3.28

World Oil Price and Growth of US Economy

- Remark: The first experiment used the data 1949-1972 (95 observations) and *lag=4*; while the second experiment used data 1950-1972 (91 observations) and *lag=8*.
- How to decide what lag to use
  - – Model selection. See demo for a working example.

Granger causality ≠ causality

- Even if x_1 does not cause x_2, it may still help to predict x_2, and thus Granger-causes x_2 if changes in x_1 precedes that of x_2