Econometrics Modeling

Key Aspects

1. Select correct explanatory variables
2. Correctly specify regression equation

How to choose Model?

1. Collect possible explanatory vars
2. Analize potential type of dependency: linear/non-linear (x^2, $\ln$, $x_1 \cdot x_2$)
   Also think about factors correlation.
3. Include all the factors and exclude one-by-one (the most insignificant, repunning every time) untill all are significant

Achtung!

4. Add transformations, when required.

1. Non-normality (of residuals)
2. Heteroscedasticity (var. of residuals)
3. Autocorrelation (for time-series)
4. Endogeneity (dependence in factors)
5. Multicollinearity (factors are correlated)
6. Simultaneity [$x_i = f(y)$]

Transformations

1) Log-linear model

$$\ln y = \ln b_0 + b_1 \ln x_1 + \dots + b_k \ln x_k + e \iff$$

$$\iff y = b_0 \cdot x_1^{b_1} \cdot \dots \cdot x_k^{b_k} \cdot \exp^{(e)} \iff$$

$$\iff \frac{\partial y}{\partial x_i} = b_i \cdot \frac{y}{x_i} \Rightarrow b_i = \frac{\partial y}{y} \Big/ \frac{\partial x_i}{x_i}$$

$\frac{\partial y}{y}$ - is change of y (in %)

$\frac{\partial x_i}{x_i}$ - change of x_i (in %)

Why? in 1 dimension:

$$\frac{dy}{\Delta x} = tg\alpha \rightarrow dy = \frac{\partial y}{1} = \Delta y$$

($\Delta x = \partial x_i$)

→ number which describes how will change "Δy" if Δx changes by 1 unit: $\Delta y = \Delta x \cdot tg\alpha$

$$b_i = \frac{\Delta y}{y} \Big/ \frac{\Delta x}{x} \Rightarrow \frac{\Delta y}{y} = b_i \cdot \left(\frac{\Delta x}{x}\right)$$ change in x.

Ex

$y_1 = 4 + 1.2 \log(x)$

$y_2 = 4 + 1.2 \log(x + \Delta x)$

$y_2 - y_1 = 1.2 \left(\log(x+\Delta x) - \log(x)\right)$

$$\Delta y \approx 1.2 \log \frac{\Delta x + x}{x} = 1.2 \log\left(1 + \frac{\Delta x}{x}\right) \approx 1.2 \cdot \frac{\Delta x}{x}$$ (Maclaren row)

So 100% change in "x" leads to 1.2 change in "y". If "y" with log ⇒ % change in "y"

2) Semi-log model:

$\ln(y) = b_0 + b_1 x_1 + \dots + b_k x_k + e$

Change in % of y is given by $b_i \cdot 100$ if x_i changes by 1 unit

3) Dummy variables:

if you want to use qualitative variable (like "industry") you should create k dummies for each possible variant. dummy = 0 or 1 (except one - reference category ⇒ (n-1) otherwise linear dependency)

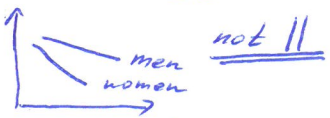
Coefficient δ_i shows change in $\hat{y}$ if shifting from ref. category to category i

4) Interactions:

$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2$

usually dummy variable is used to capture difference in behaviour (ex. between men and women)

Ex



not ||

this picture means that, say, drug influence women more than men. We need to include interaction to improve the model

Note: don't try to interpret importance of main factors in