Econ 174, Section 101/103 Week 14

Joshua Blumenstock

Today

- General questions?
- Standard Errors
- Stata tips

- Next week?
  - Send me suggestions!

Standard Errors

- What are they?
  - A measure of the statistical accuracy of an estimate

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- We estimate the residual as:

$$\hat{u}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

- OLS choose the values of beta that minimize the sum of squared residuals

OLS Assumptions

- OLS assumes:
  - Dep. var. is a linear function of independent vars.
  - u_i has mean zero
  - Independent vars. can be treated as fixed (IV)
  - u_i have constant variance
  - uncorrelated across observations $(E(u_i u_j) = 0, i \neq j).$

- What if these assumptions are violated?
  - Bias problems (beta estimates inaccurate)
  - Inference problems (standard errors inaccurate)

Non-spherical disturbances

- Estimates are unbiased (this is good)
- But standard errors are off (this is bad)
  - Who cares if standard errors are off?

- Always remember: *funny-looking errors should at first be interpreted as signaling a specification error, not a nonspherical error* (Kennedy 2010)

Heteroskedasticity

- Assumption: u_i have constant variance (homoskedastic)
  - Off-diagonal elements of variance matrix are zero, but diagonal elements not necessarily equal
- Common violation: error varies with indep variable
  - E.g. consumption is a function of income.
  - At higher levels of income, consumers are more whimsical
  - Errors associated with measuring income are higher
- So what?
  - Those super-noisy observations contribute more than their share to the estimates of the error
- "Solution": robust standard errors, aka sandwich/huber/white
  - Heteroscedasticity-consistent standard errors
  - $(X'X)^{-1}X'GX(X'X)^{-1}$: G is diagonal matrix with squared OLS residuals

Autocorrelation

- u_i uncorrelated across observations $\quad (E(u_i u_j) = 0, i \neq j).$
- When could this happen?
  - Husband/wives responses related
  - 20 People rating 10 photographs for hotness
  - Individual responses over time
  - Persistence of a shock over time, across space
  - Treatment assigned to existing groups of people
  - Data was duplicated
- What this means: each "independent observation" is not independent, and contains less information than if it were truly independent
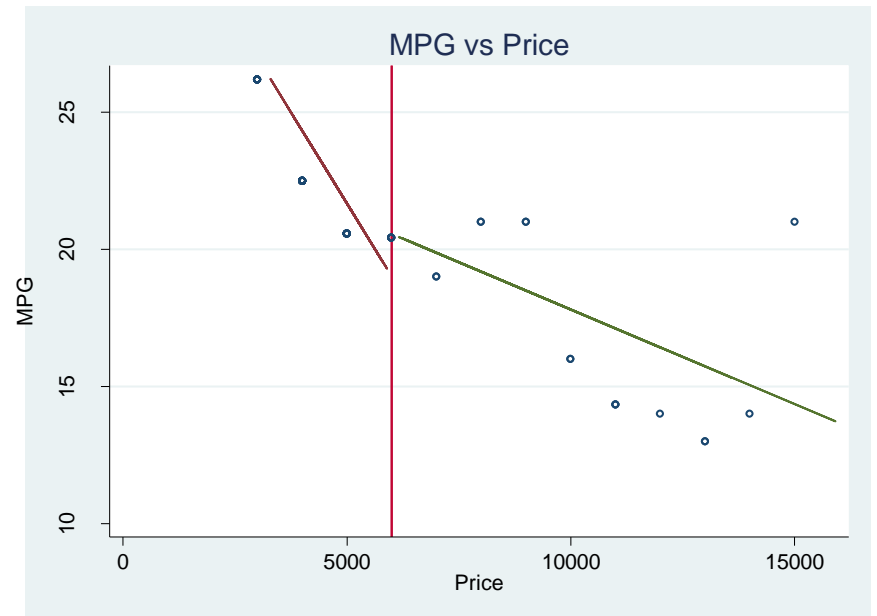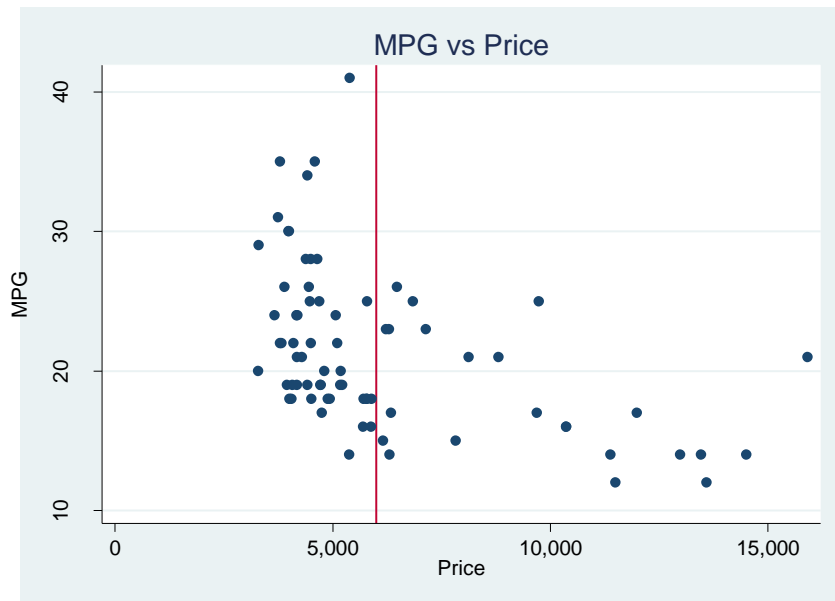
Autocorrelation

- How to deal with it: cluster standard errors!

- Accounts for arbitrary within-group variance

- $V_{OLS} = s^2 * (X'X)^{-1}$
  - Where $s^2 = (1/(N - k)) \sum e_i^2$

- $V_{cluster} = (X'X)^{-1} * \sum^{nc} u_j'*u_j * (X'X)^{-1}$
  - Where $u_j = \sum e_i*x_i$ and n_c is the total number of clusters.

Some notes for pset 4

- Goal: plot average mpg for each price decile

Stata tutorial

- Creating bins
  - `gen bin = int(price/1000) * 1000`
  - `egen bin_avg = mean(mpg), by(bin)`

- Spline regressions
  - `reg ... if left==1`
  - `predict leftside`
  - `reg ... if left==0`
  - `predict rightside`