ECON-C4100 - Capstone: Econometrics I

Lectures 12: Recap

Otto Toivanen

Topics

- Bias - sources and how to detect?
- Interpretation of regression results.
- IV.
- ESS / TSS / R^2; relevance for nonlinear models.
- Can the exogenous variable be correlated with the error term without causing omitted variable bias?
- Strict exogeneity / homoskedasticity.
- Example of a real-life example where exogenous variable not i.i.d.
- Exam.

Bias

- Definition:

$$\mathbb{E}[\hat{\beta}] \neq \beta$$

- Key condition for unbiased estimates (for linear models, which is what we deal with in this course):

$$\mathbb{E}[u|\boldsymbol{X}] = 0$$

- In practice, need an omitted variable problem or something else that causes the condition not to hold ("correlation between explanatory variables and error term").

Bias

- Q: How to recognize bias?

Bias

- Q: How to recognize bias?
- A: This is difficult in practice.
- Compare coefficients to prior results.
- Check robustness to changes in vector of control variables.
- Utilize theory to think through what mechanisms could cause correlation between explanatory variable(s) and the unobservables (=error term).

Interpretation of regression results

- Economic significance of (key) coefficients.
- Statistical significance of (key) coefficients.
- Statistical significance of (vectors of) control variables.
- Statistical performance of the regression as a whole (R^2, F-test, choice of standard errors, ...).

IV

- Root cause for need of IV: (suspected) breakdown of $\mathbb{E}[u|\boldsymbol{X}] = 0$.
- The key properties of an instrumental variable.
- The algebra of IV.
- Think back to what an experiment allows the researcher to do.

ESS / TSS / R^2; relevance for nonlinear model

Explained sum of squares (ESS): $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$.

Total sum of squares (TSS): $\sum_{i=1}^{n}(Y_i - \bar{Y})^2$.

Residual sum of squares (RSS): $\sum_{i=1}^{n}(u_i)^2$.

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \in [0,1] \tag{1}$$

ESS / TSS / R^2; relevance for nonlinear model

$$TSS = ESS + RSS$$

- See App. 4.3 of SW.

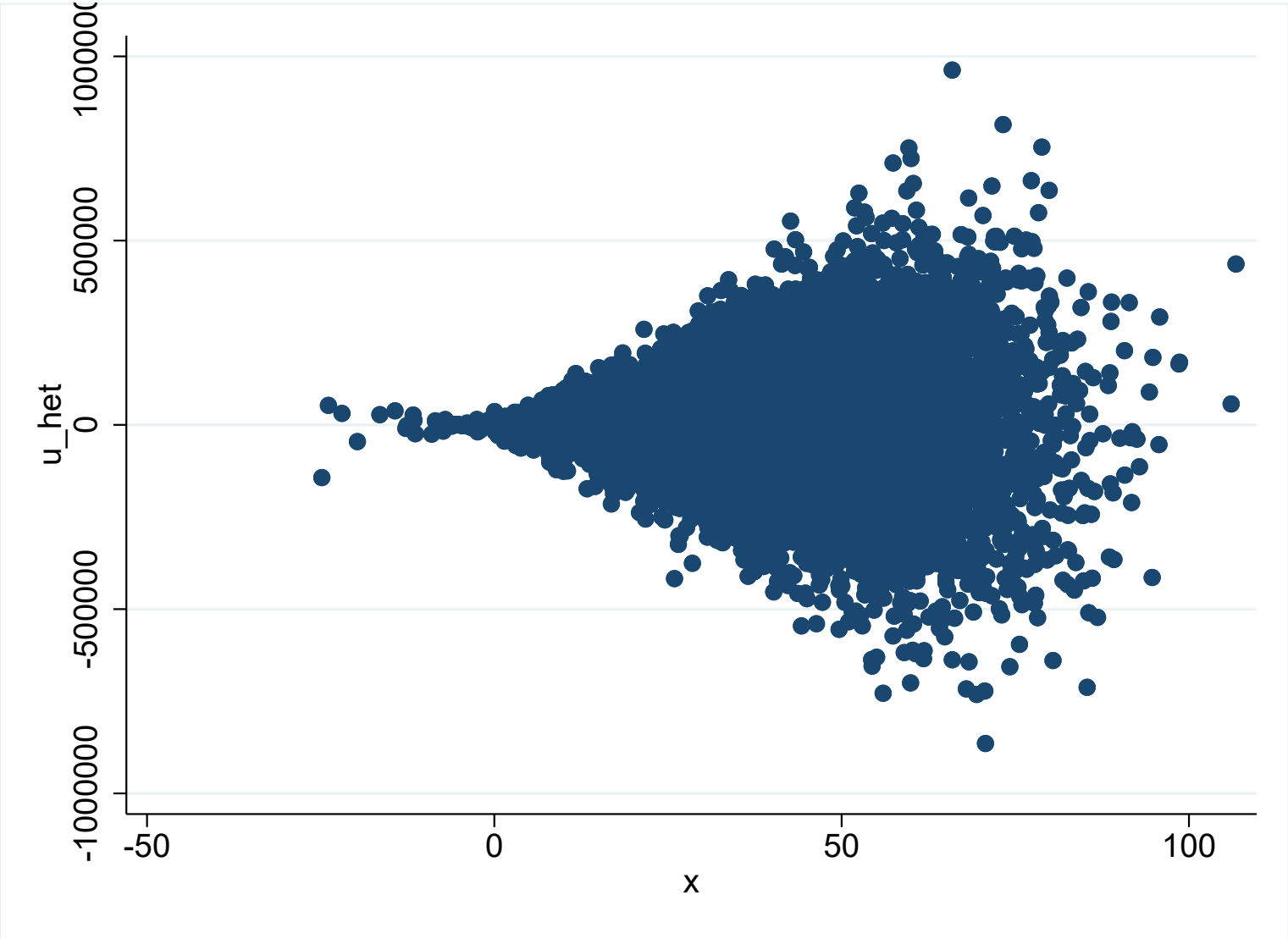
ESS / TSS / R^2; relevance for nonlinear model

- What we call a nonlinear model means that Y is a **nonlinear function of (some) explanatory variables $\boldsymbol{X}$**, such as

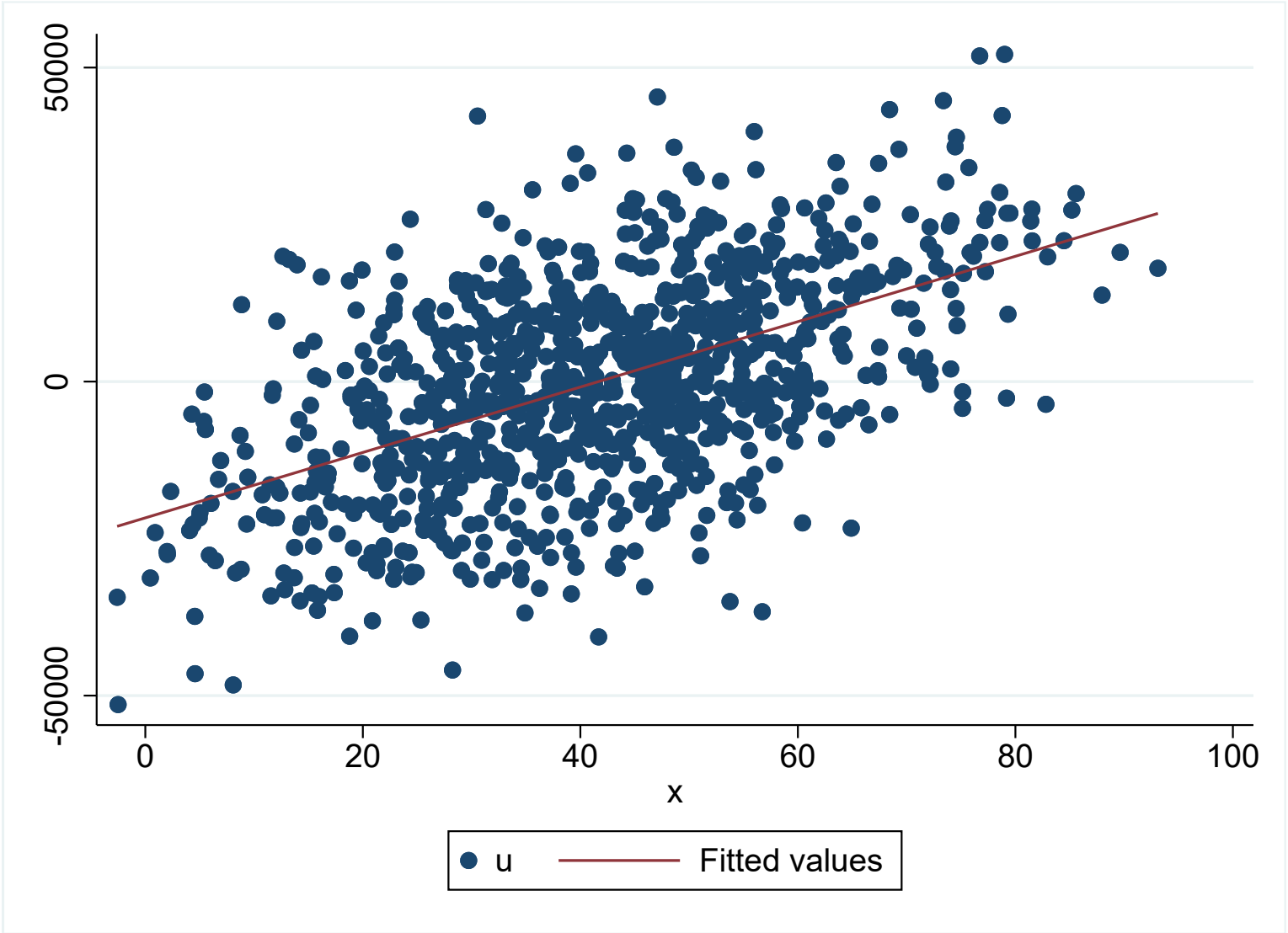
$$Y = \beta_0 + \beta_1 X + \beta_2 x^2 + u$$

- ... but this is still a linear model ("**linear in parameters**").
- A **nonlinear model (in parameters)** is one in which Y cannot be expressed as a linear function of parameters (the βs).
- R^2 is relevant for the nonlinear (in variables) models we deal with in this course. For nonlinear (in parameters) models this is not necessarily the case.

Heteroskedasticity vs endogeneity



Heteroskedasticity vs endogeneity: $\rho_{XU} = 0.5$



Heteroskedasticity vs endogeneity

- Heteroskedasticity: $\mathbb{E}[u|\boldsymbol{X}] = 0$.
- ... but $\sigma(u|X = x) \neq \sigma(u|X = \tilde{x})$ if $x \neq \tilde{x}$.
- Endogeneity: $\mathbb{E}[u|\boldsymbol{X}] \neq 0$
- ...irrespective of whether $\sigma(u|X = x) \neq \sigma(u|X = \tilde{x})$ or $\sigma(u|X = x) = \sigma(u|X = \tilde{x})$.

Example of non-i.i.d. explanatory variable

- Example #1: agricultural experiment where each plot gets same treatment over and over again.
- ... compare to randomization the treatment each year for each plot.
- Example #2: time-series, such as #employees in a given firm month by month.
- Example #3: An explanatory variable that is partly determined by the unobservables ($\mathbb{E}[u|\boldsymbol{X}] \neq 0$).

Exam

- Emphasis not on rote learning.
- If Stata-related questions, then not about commands etc., but rather asking to implement something using a data-set made available.
- Interpretation, understanding.
- If calculations, they are either Stata-based, or can be done with the help of a calculator (e.g. excel); will require documentation of the calculations.
- Will make some old exam questions available.