Consistency of Bivariate OLS

Organizing Remark

- Most textbooks teach the small-sample properties, followed by the large-sample properties.
  - Wooldridge, Chapter 2: 4 assumptions to prove that bivariate OLS coefficients are unbiased.
  - I'm going to start with a look at a large sample property: consistency.
- There are several reasons I want to do this:
  - Today's data scientists work most often with large samples.
  - More time to talk about exogeneity – a central assumption in statistics that take time to internalize.
  - Language to understand omitted variables

Consistency

- Suppose you run a bivariate regression and you have a large sample.
  - Rule of thumb: n>30 is considered large
- One of the first thing you want to know: are your coefficients consistent?
  - E.g. $plim_{n\rightarrow\infty}\hat{\beta}_j = \beta_j$
  - If we don't have consistency, our coefficients are approaching some other values, not the true parameters.
- It takes 4 assumptions to establish consistency.
  - I'll state them for bivariate regression.
  - Later, you'll see how to extend these to multiple regression.

Bivariate OLS Assumptions

- Assumption SLR.1 (Linear in parameters)

$$y = \beta_0 + \beta_1 x + u$$

- Assumption SLR.2 (Random sampling)
- $\{(x_i, y_i) : \ i = 1, \ldots n\}$ is a random sample from the population model.
- All datapoints are i.i.d.
  - No clustering
  - No autocorrelation

Bivariate OLS Assumptions

- Assumption SLR.3 (Sample variation in explanatory variable)

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$$

- Assumption SLR.4�much (Exogeneity)

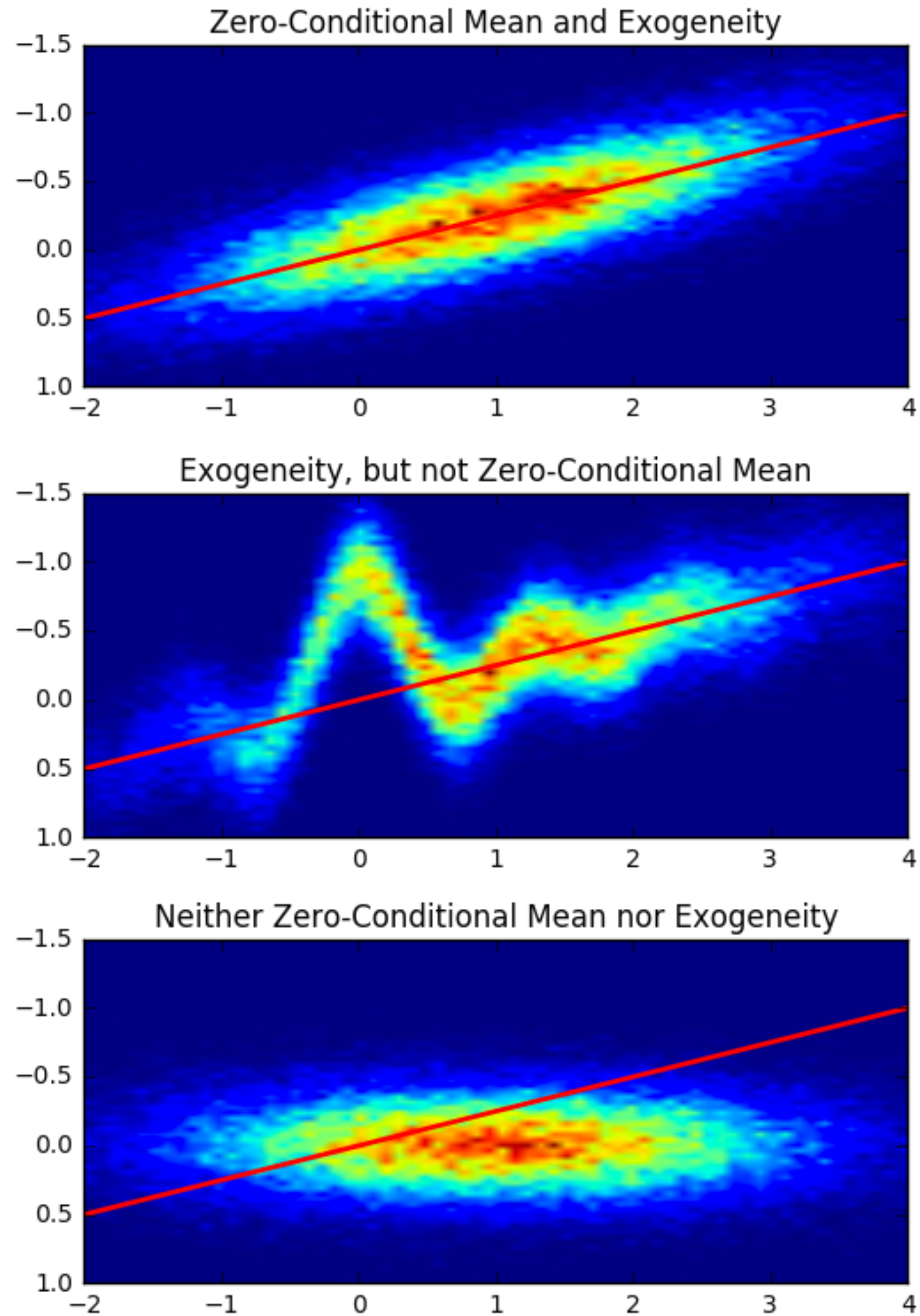
  $E(u) = 0$ and $Cov(x_i, u) = 0$

Zero-conditional mean:

- Conditional error has no relationship of any kind with x

Exogeneity:

- No overall linear relationship between u and x



- **<u>Theorem 2.1 (Consistency of OLS)</u>**

SLR.1-SLR3, SLR4' $\Longrightarrow \; plim_{n\to\infty}\hat{\beta}_0 = \beta_0,$

$$plim_{n\to\infty}\hat{\beta}_1 = \beta_1$$

- Remember that this is a probabilistic statement about the limiting behavior of our coefficients.
- For any one sample, our estimates could be very far off.
- We'll need to quantify the uncertainty in our coefficients in order to give meaning to our estimates.
- To do that, we'll need more assumptions.