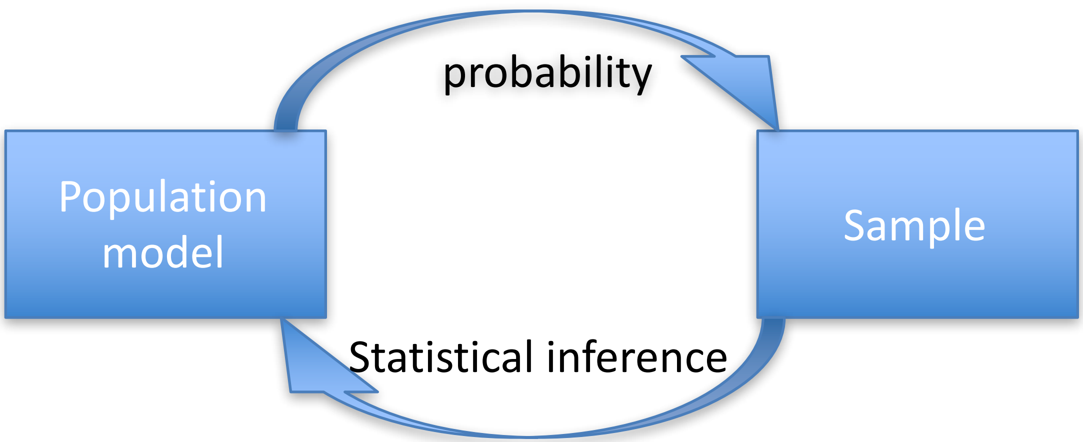
Introducing the Multiple OLS Population Model

The Multiple OLS Population Model

- Just like with simple regression, we can use multiple regression in a strictly descriptive way.
- More commonly, we are interested in more than just the one sample of data, so we need a population model.
- The multiple OLS population model is a direct extension of the bivariate OLS population model.
- We will study it extensively later, for now, I will introduce some key assumptions.



Multiple OLS Assumptions

Assumption MLR1: Linear in Parameters

- The population model can be written

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$$

- y is linear in the betas.
  - We can't multiply betas together, or raise an x to the power of a beta, etc.
- At this point, we haven't constrained u, so this describes any joint distribution.

Multiple OLS Assumptions

Assumption MLR2: Random Sampling

- $\{(x_{i1}, x_{i2}, ..., x_{ik}, y_i)\}$ is a random sample from the population model.
- Data points are i.i.d.
  – No clustering
  – No autocorrelation
- Typically, we'll need background knowledge to assess this assumption.

Multiple OLS Assumptions

Assumption MLR3: No Perfect Multicollinearity

- None of the x's is constant and there is no perfectly linear relationship among the x's.
- Ex: we accidentally include both number of eggs and number of dozens of eggs in our regression.
  - There would be no unique variation in either variable, since they vary together.
  - Usually, this isn't a big problem because you just need to drop one variable.
- Sometimes, multicollinearity emerges among more than two variables.
  - For example, campaign spending by the winning candidate, spending by the losing candidate, and total spending
  - A simple correlation table would not reveal this type of mutlicollinearity.
- The assumption only prohibits *perfect* multicollinearity.
  - Imperfect collinearity will lower precision and increase our standard errors, but it doesn't violate any assumptions.

Multiple OLS Assumptions

- At this point, Wooldridge introduces MLR4, zero-conditional mean.
  - This is used to prove that OLS coefficients are unbiased.
- I'll focus on exogeneity, which lets us prove consistency.

Assumption MLR4': Exogeneity

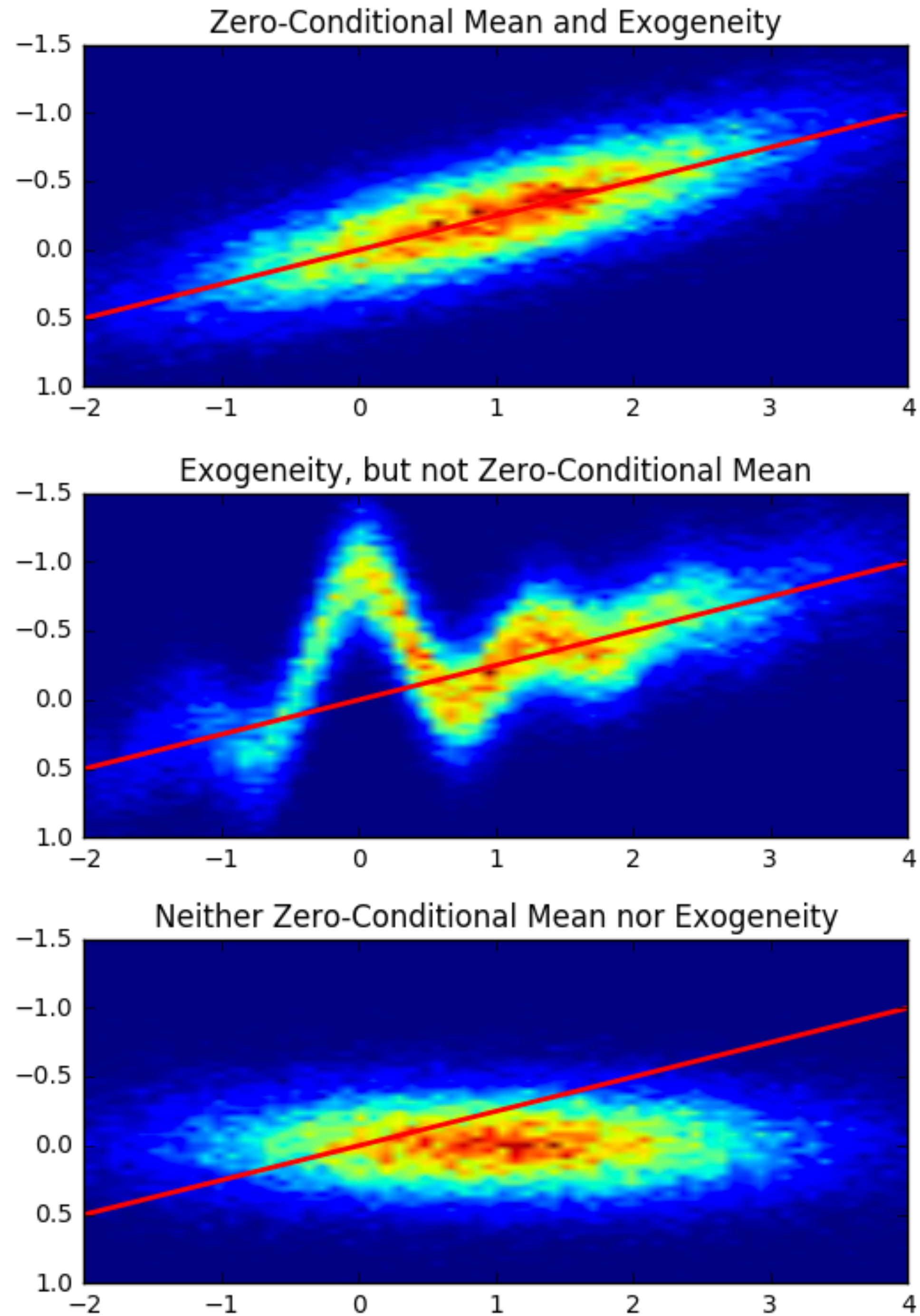
- $E(u) = 0$ and $cov(x_i, u) = 0$ for all x_i.
- There's no correlation between each x and the error.
- In other words, there's no omitted variable that's correlated with x

Zero-conditional mean:

- Conditional error has no relationship of any kind with x

Exogeneity:

- No overall linear relationship between u and x



Consistency of OLS

- Under MLR1-3 and MLR4'

$$plim_{n\rightarrow\infty}\hat{\beta}_j = \beta_j$$

- Our coefficients approach the true parameter values in probability.
  - Each sample has randomness, so our betas will never equal the true parameters exactly.
  - There could still be considerable bias for small n.