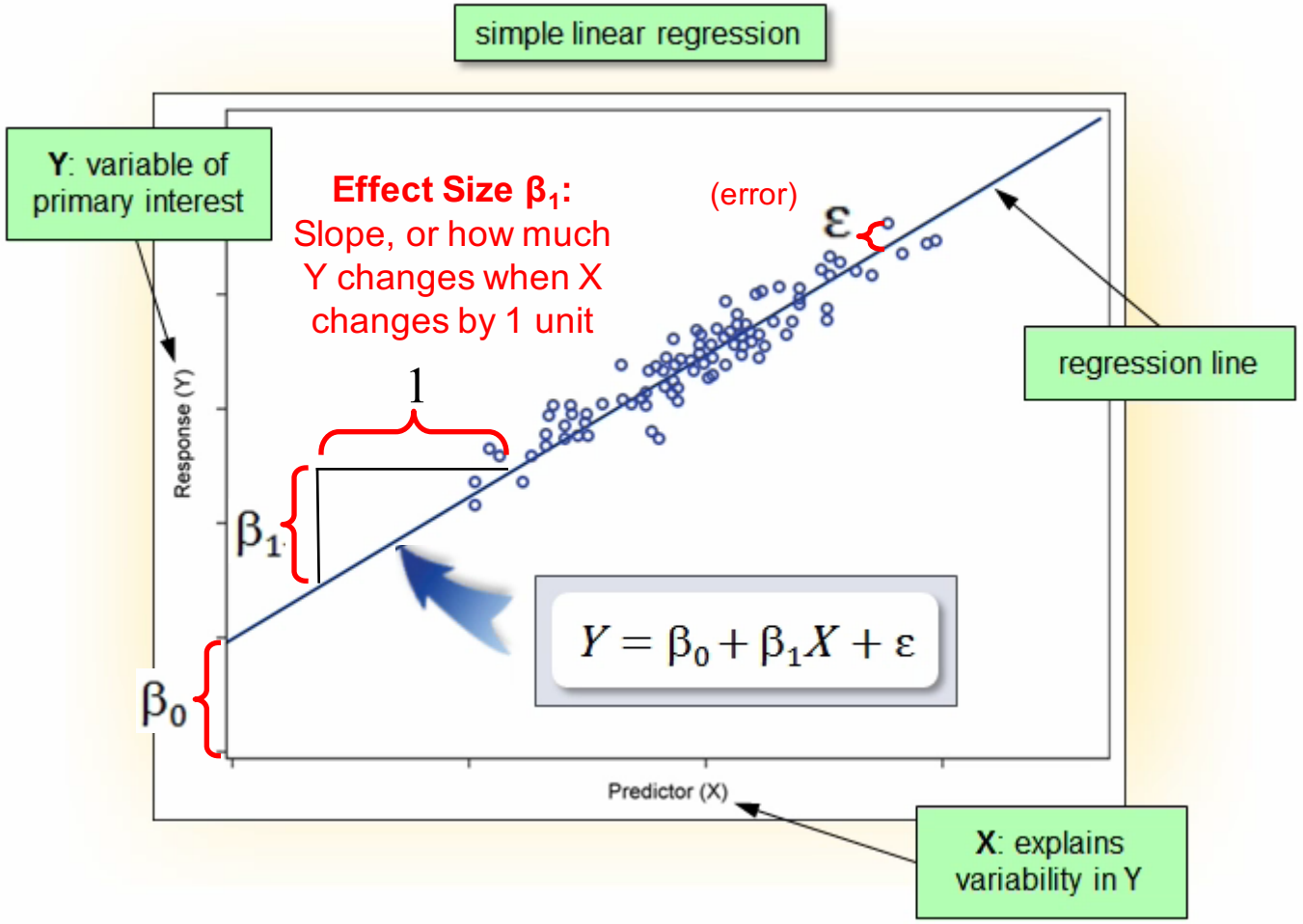
Simple Linear Regression

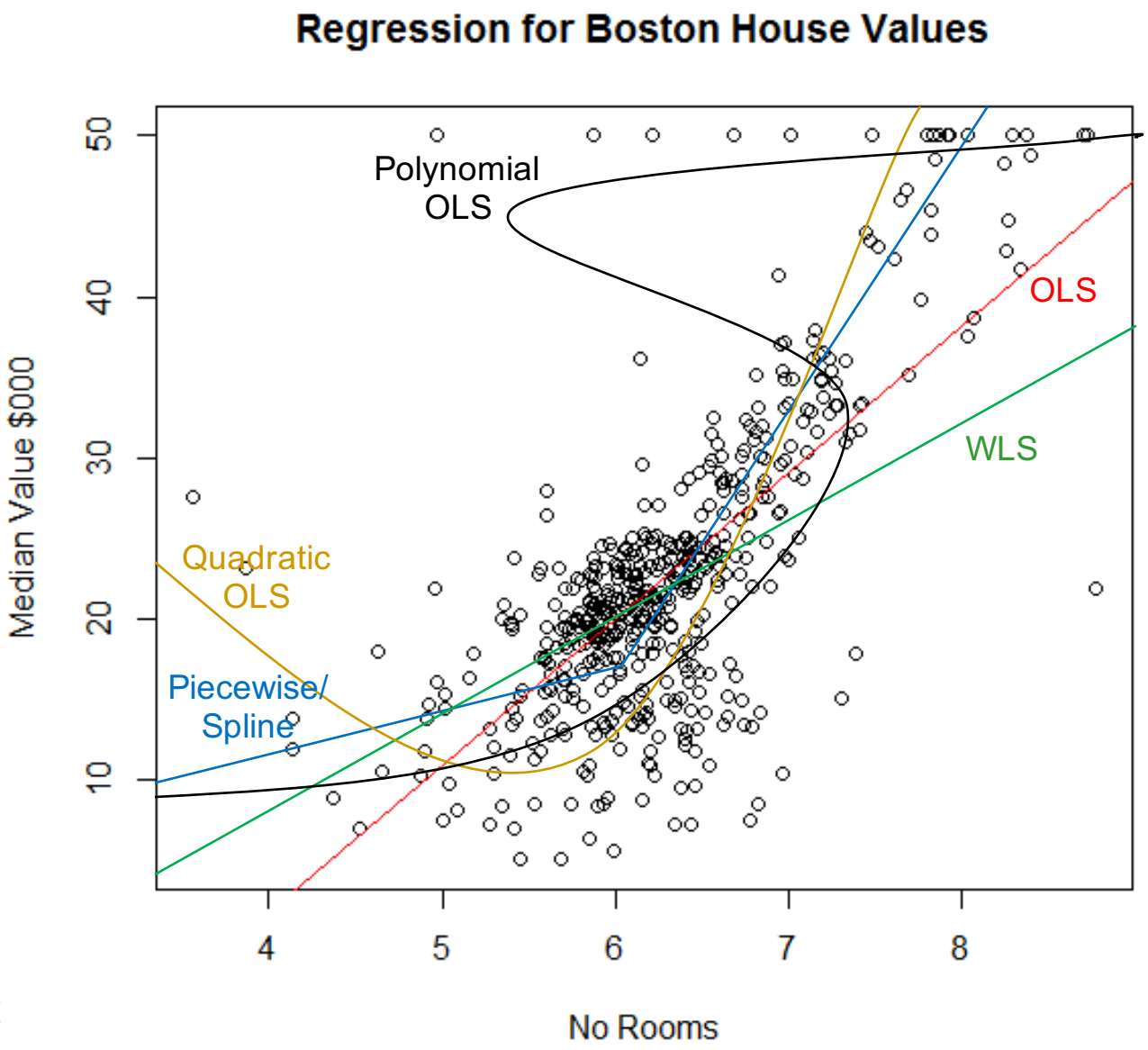
Simple Linear Regression



Questions

- How do you find the regression **line** that best **fits** your **data**?
- How do you **minimize errors** (i.e., deviations from the regression line)?
- In statistical parlance, how do you **"estimate"** the regression line?
- There are many **"estimation methods"**
- The most popular one is called **"Ordinary Least Squares" (OLS)**
- It is the line that **minimizes the sum of errors squared**

Estimation (Boston{MASS})



Questions

- **Estimation** is about the method you use to develop your model
- Is **OLS** the best estimation method?
- **OLS** minimizes the sum of squared errors.
- But weighted least squares (**WLS**) may be more efficient if errors grow or shrink with x
- **Quadratic OLS**, **Piecewise**, or **Spline** regressions may be more accurate predictors
- **Polynomial** regressions may even be more accurate but it suffers from "over-identification".

Selecting your estimation method is key!!

Important Regression Statistics

- **β_0** = Regression line intercept – i.e., Y's predicted value when X is 0
- **β_1** = Slope of the Regression line – i.e., Y's change when X increases by 1 unit or the **"effect size"**
- **Significance** = whether β_1 is significantly **different than 0**
- **p-value** = measures **significance** (same as with correlation) i.e., the probability that **β_0**, **β_1**, etc. are 0 – i.e., the observed effects happened by chance
  → the **smaller** the p-value the **better**
- **Confidence** = 1 - p; e.g., p=0.05 implies we are (0.95 or) 95% confident that the observed effect did not happen by chance

KOGOD SCHOOL
of
BUSINESS