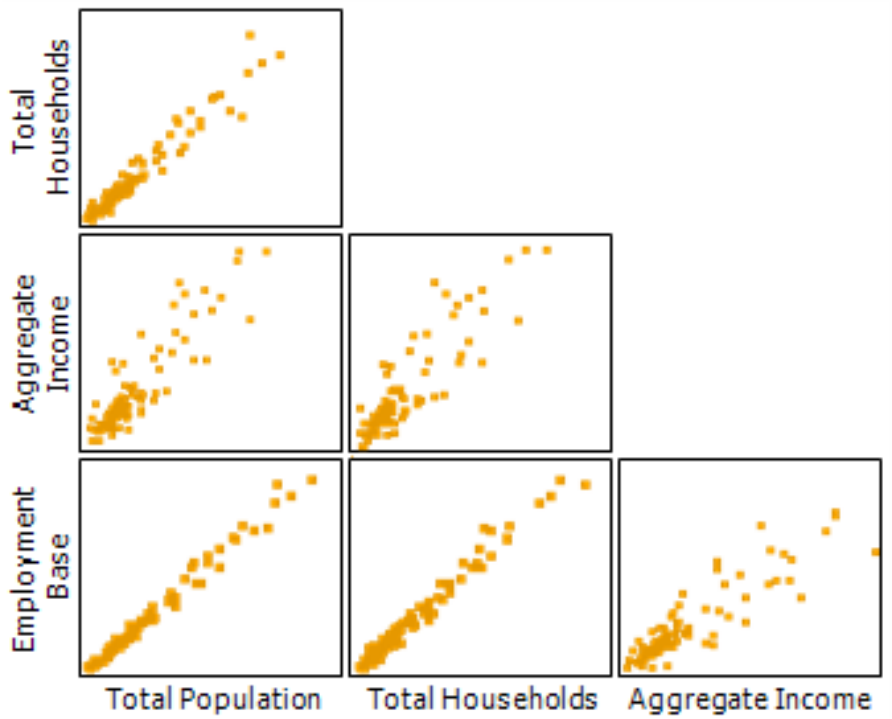
Sports Econometrics

Regression Problems

or “How to make all your results irrelevant”

Problem 1: Multicollinearity

- Multicollinearity occurs when some of the X variables are too highly correlated with each other.
- Multicollinearity leads to impossible coefficient estimates and illegitimate results.



Multicollinearity Example

- Example: Suppose Y is Designated Hitter Salary. There is a high correlation between hits, singles, doubles, home runs etc…
- How to diagnose multicollinearity?
- A. Calculate the correlation between suspected X variables (too close to -1 or 1 is bad).
- B. Remove one of the suspected X variables and rerun the regression; if the coefficient estimate changes dramatically then you have multicollinearity.

Problem 2: Spurious Results

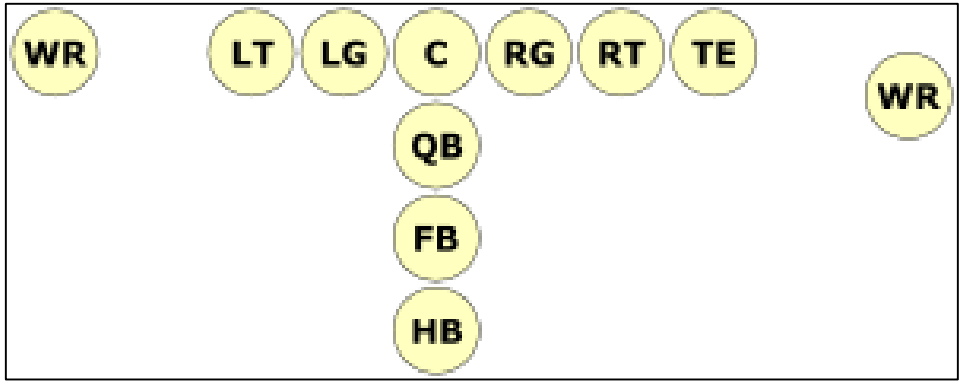
- Spurious results occur when a Y variable and an X variable are correlated BY ACCIDENT.
- Often, statistically significant but not really causation (only correlation)
- http://www.tylervigen.com/spurious-correlations
- How to avoid? Know why your X variable is included in the regression. You're the expert! Don't include an X variable if you can't explain why it may impact your Y

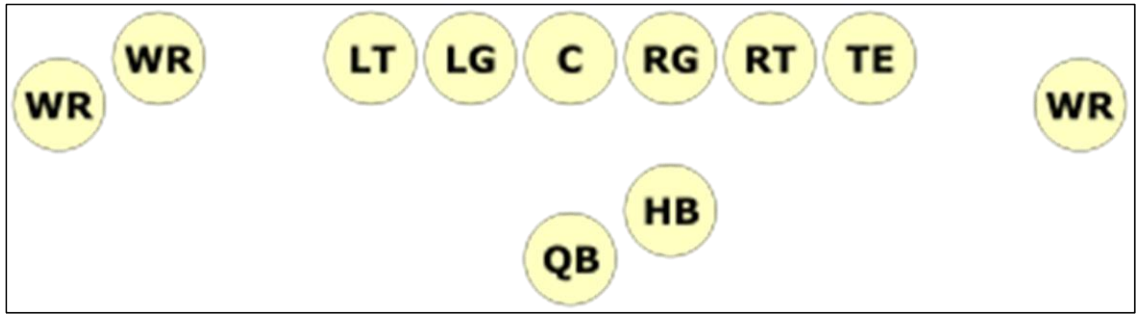
Problem 3: Endogeneity

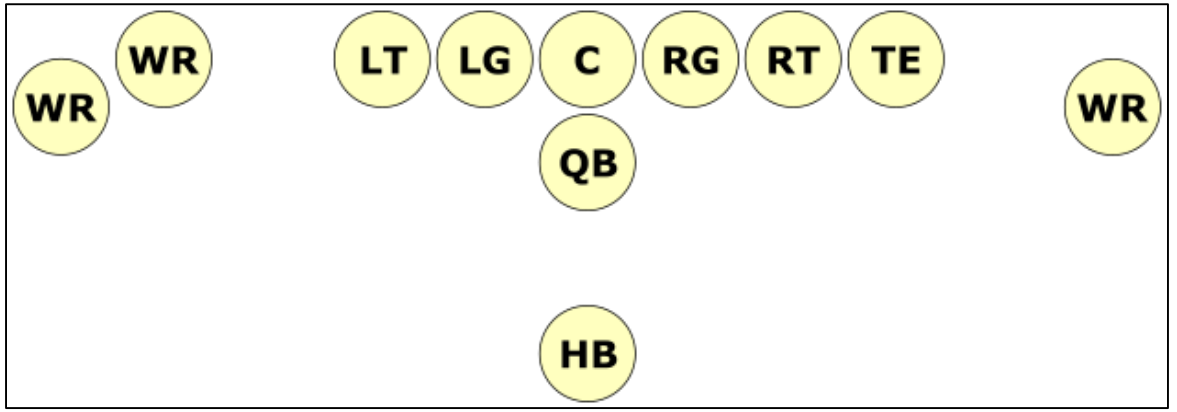
- Endogenous = within the system, Exogenous = outside the system
- Endogenous variables are X variables that are decided at the same time as your Y variable.
- Endogeneity means that your coefficient estimate and significance of the X variable in question is WRONG.

Endogenous Example

- Example: Y variable is number of Wide Receivers for a given play. One of the X variable is number of Running Backs for the play.
  - Why is it endogenous?
  - Because the personnel decision on any given play occurs at the same time.







Avoiding Endogeneity?

- Don’t be dumb.
- Again, you’re the expert. If X and Y are decided at the same time then you need to exclude X.
- OR Include the lagged X variable instead
- OR ask for my help
  - Two Stage Least Squares (proves causation instead of correlation)
  - Seemingly Unrelated Regression (estimates both X and Y at the same time)
  - Econometrics here at the school can help

Omitted Variable Bias

- Ideally, all of your X variables should control for all theoretical factors.
- If you omit an important factor from your X variables then you have an omitted variable bias.
- Omitted variable bias can lead to all your other X variables having incorrect coefficients and statistical significance.
- Omitted variable bias can lead to all your forecasts to be inaccurate.

Diagnosing an omitted variable bias

- Ask somebody.
- Is your R-squared “too low”?
- What data are you missing?

Fixing an omitted variable bias

- 1. Don’t omit the variable.
- 2. CONTROL variables.
- 3. Dummy variables for fixed effects.
- 4. Lag your Y variable and use the lagged variable as an X variable
  - if it is important to Y than it was important to your past period’s Y variable