

# Sports Econometrics

Getting more from your data:  
Interaction Variables & Data Mining

# Interaction Variables

- By multiplying two X variables together we create an interaction variable.



## Effect of Interaction

- Given:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$
- Without interaction term, effect of  $X_1$  on  $Y$  is measured by  $\beta_1$
- With interaction term, effect of  $X_1$  on  $Y$  is measured by  $\beta_1 + \beta_3 X_2$
- Effect changes as  $X_2$  changes

# Example: Predicting Pass Play



Figure 7. 2013 Iron Bowl: Forecasts overlay of Next Play, Drive Points, Win Probability and Final Score

# Example: Predicting Pass Play

- What matters?
- Down AND Distance
- Interact DOWN variable with DISTANCE to 1<sup>st</sup> Down/End Zone
  - 1<sup>st</sup> & 10 = 50/50 Run v Pass
  - 3<sup>rd</sup> & 10 = 20/80 Run v Pass
  - 3<sup>rd</sup> & 1 = 60/40 Run v Pass

# Example

- Sample Data Set – CF\_Line\_Outcome
- Interact Home Team's Offense with Away Team's Defense
- Restrict two coefficients to equal values
  - Add the two variables together
  - Or Divide them....

# Data Mining

- Data Mining (running multiple different regression models) can be problematic
- Running too many regressions to hunt for a statistical relationship is dangerous. (aka p-hacking)
  - You may find a statistically significant result that isn't valid
- BUT, Data Mining to avoid multicollinearity OR to reduce the number of X variables is virtuous.
- Regressions v Decision Trees v Black Boxes
- Why mine your data? To have more concise and more accurate model