

Empirical Methodology

Caleb Woessner

My project aims to look at NFL teams offensive line investment on the 2020-2023 seasons to analyze its impact on the team's offensive efficiency. The offensive line investment consists of two variables, offensive line capital percentage and offensive line draft percentage. Both aim to determine the exact percentage of the team's total spending allocated to the offensive line. The three independent variables I will be looking at are the number of sacks suffered, rushing EPA, and receiving EPA to measure a team's effectiveness.

I will be using multiple linear regression models to quantify these relationships. Each model takes the form:

$$Y = \beta_0 + \beta_1 \cdot \text{olCapPCT} + \beta_2 \cdot \text{olDraftPCT} + \varepsilon$$

where Y represents different offensive efficiency outcomes:

- **Receiving EPA** (effectiveness of passing game)
- **Rushing EPA** (effectiveness of run game)
- **Sacks Suffered** (offensive line pass protection)

Using multiple linear regression will allow me to estimate the impact of offensive line investment on offensive efficiency metrics to investigate whether teams may be undervaluing the position.

I will also utilize hypothesis testing to check the significance of the coefficients. I will test each coefficient against all three different efficiency metrics, and each will have a general form of:

Null Hypothesis (H_0): $\beta_i = 0$ (Offensive line investment has no impact on the given efficiency metric)

Alternative Hypothesis (H_A): $\beta_i \neq 0$ (Offensive line investment has a statistically significant effect)

Where "i" either 1 or 2. I will then check the p-values of this to determine the significance at a level of .05.

There are a few potential problems with this estimation strategy, as I have not included any control variables. These could include quarterback quality, offensive scheme, and total offensive spending. These control variables may help account for external influences, meaning the variables might not be as isolated as possible. Another issue with this regression model is the impact of unforeseen circumstances, such as injuries. The model considers the percentage of capital spent on players but does not account for their availability. A player who gets injured early in the season could negatively impact offensive performance. Lastly, the sacks suffered metric has been widely debated as a quarterback statistic versus an offensive line statistic. If this is the case, we may not gain any new information on offensive line spending.