Estimation of Salary Given Differences in Draft Picks in the NBA

Sean Hong, Chris Li, Jonathan Lin, Justin Ross, Ben Schwartz

Our Data

60 Draft Picks Per Year

From 1997 to 2013

2nd Contract Salary (1st Non-Rookie Contract)

Minutes Per Game

Winning Percentage

Draft Pick Number

Salary Distribution

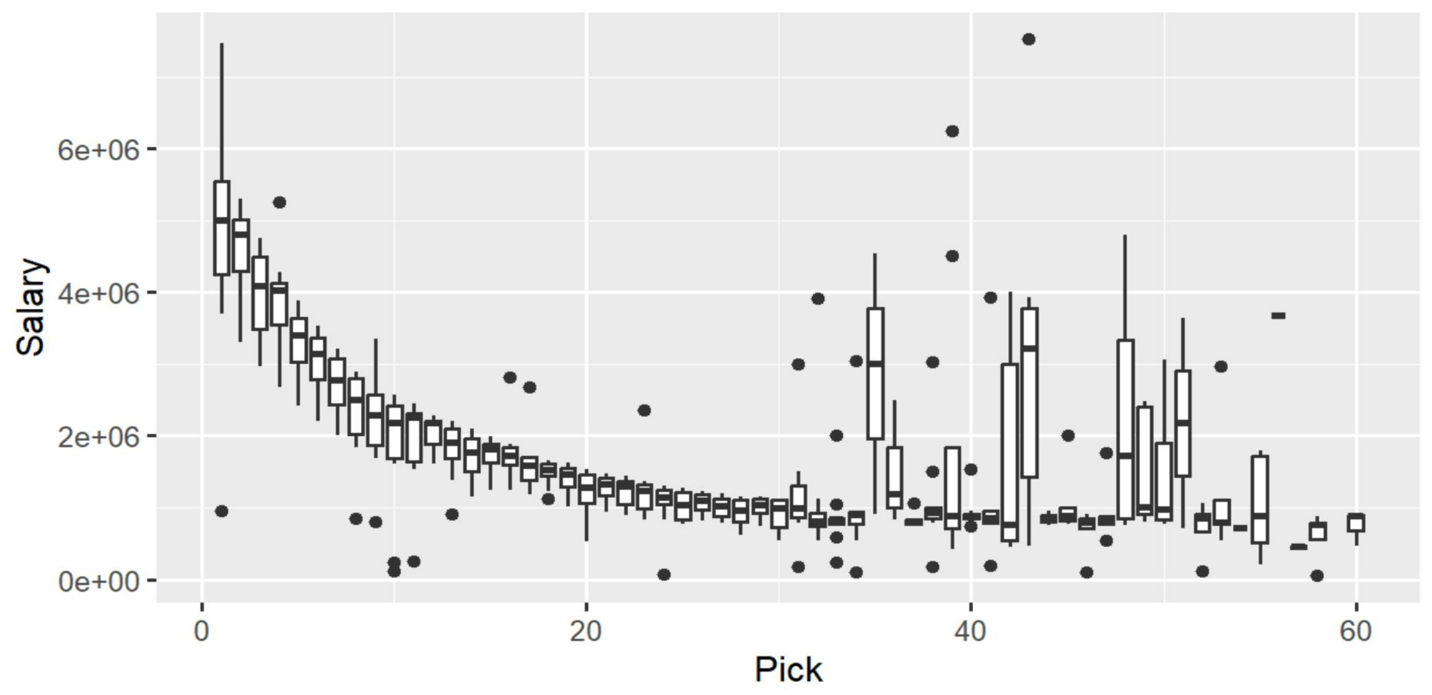


Figure 1. Salary Distribution per Draft Pick

Defining Our Estimation

- **Treatment**
  - **Shifting a player one draft pick down**
    - **IE. A player that would be picked at the 5th draft pick is picked 6th or the player who would be picked 59th is picked 60th**
- **Estimation in terms of potential outcomes**
  - $E(Y_i(Pk_i)) - E(Y_i(Pk_i + 1))$
- **Estimation in terms of what we observed**
  - $u(x_1, x_2, x_3) = E(E(Y_i \mid Pk_i = Pk, MPG, WP) - E(Y_i \mid Pk_i = Pk + 1, MPG, WP))$

Defining Our Estimation Continued...

- **Implications of Estimation**
  - Based on the estimated difference in salaries that was found between each draft pick, teams can make informed decisions on the cost-benefit of keeping or trading 2nd contract players
  - Risk is involved in deciding whether to trade or keep the player given their draft pick (negotiate salaries)
  - From the player's perspective, they can gain insight into a sort of expected salary decrease/increase based on their pick (pre-draft motivation)

```
      (Intercept)                      Pk                 MPG
       705832.737               19432.653           80920.621
Winning.Percentage Pk:Winning.Percentage              Pk:MPG
       414297.286              -40408.833           -1735.515
```

Semi-Interactive Model

$$\text{salary} = \text{Pk} + \text{MPG} + \text{WP} + \text{MPG:Pk} + \text{WP:Pk}$$

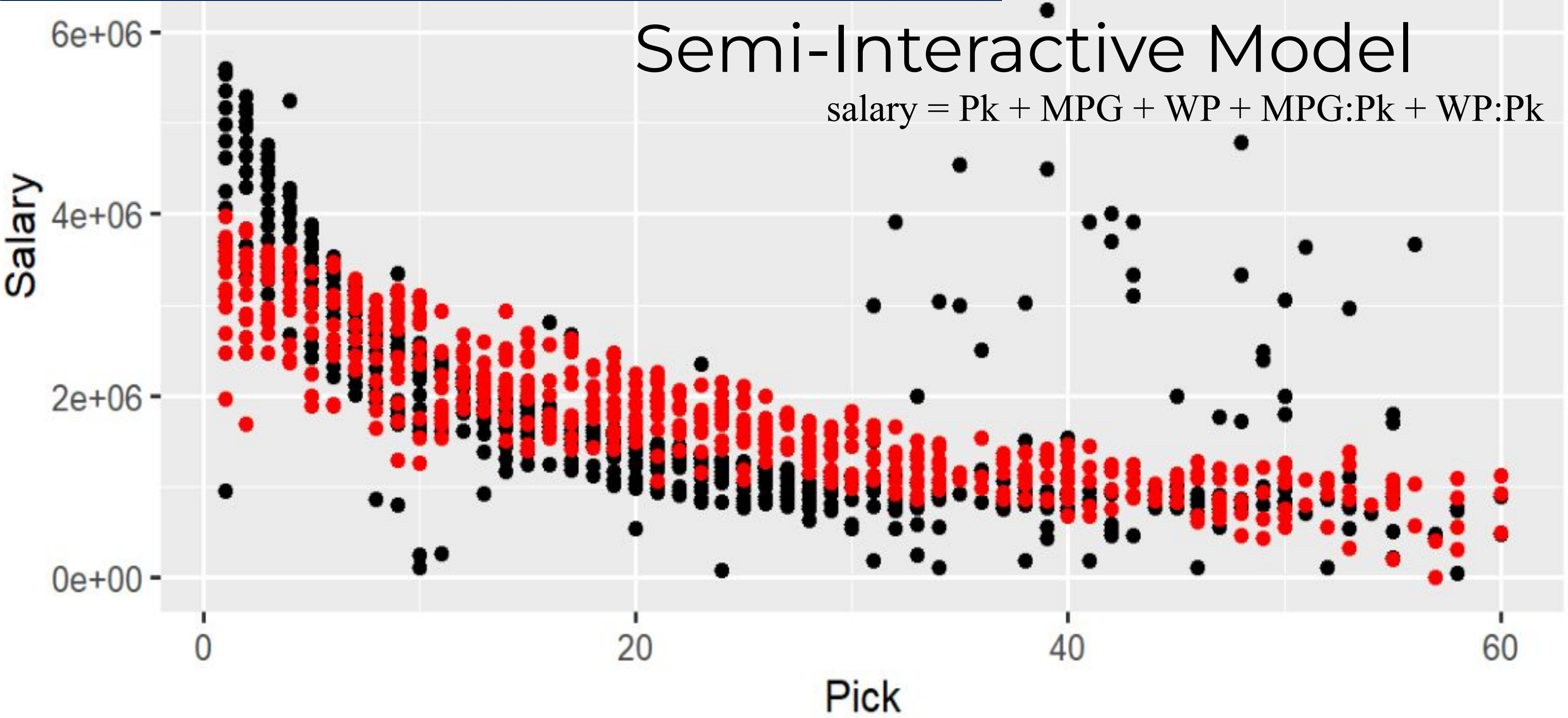


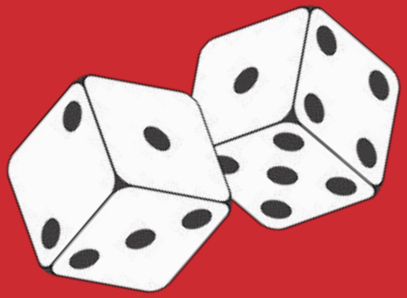
Figure 2. Predictions from our regression in red. True values in black.

Generated Data?

Generated data is much faster and efficient to generate than finding real data points and creating data sets from scratch. Since a large number of data points is necessary in reaching valid conclusions, by creating fake data and comparing its regression with the real data, the regression on the real data can be validated and used to make conclusions.

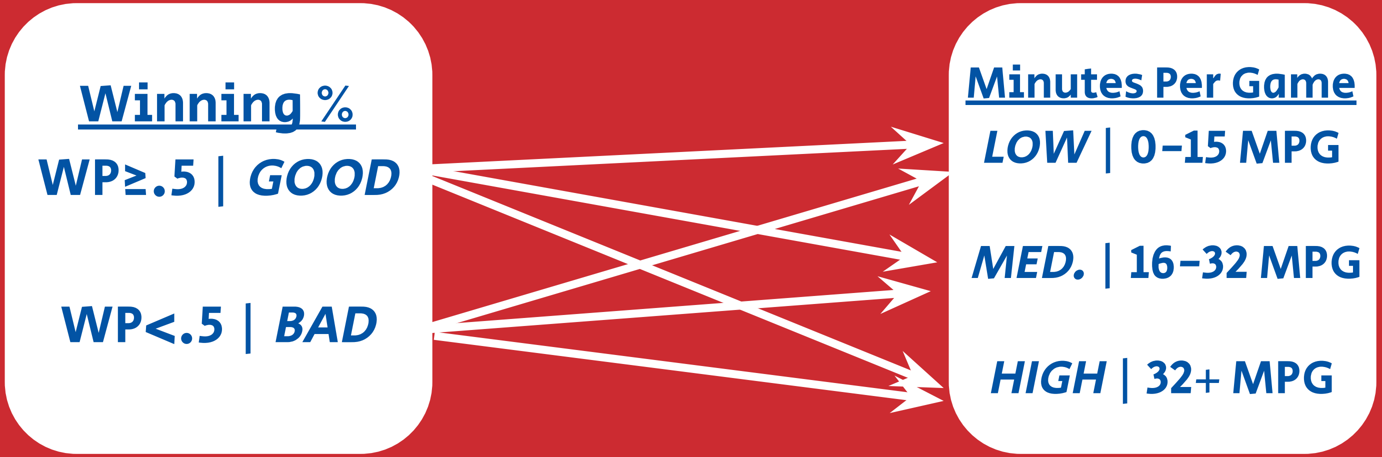
Generated Data: Assumptions for Randomization

- Draft pick numbers for teams are assigned based on winning percentage
- Trading and other transactions are minimal where it does not significantly affect our estimate
- No major change in team performance over last couple of years
- Teams make the best choice for themselves whenever they draft players



Creating Our Generated Data

1. We first grouped players into 6 categories based on team WP and player MPG.




```r
goodhigh <- fakeData1$Winning.Percentage >= 0.5 & fakeData1$MPGCat == "High"
goodmedium <- fakeData1$Winning.Percentage >= 0.5 & fakeData1$MPGCat == "Medium"
goodlow <- fakeData1$Winning.Percentage >= 0.5 & fakeData1$MPGCat == "Low"
badhigh <- fakeData1$Winning.Percentage < 0.5 & fakeData1$MPGCat == "High"
badmedium <- fakeData1$Winning.Percentage < 0.5 & fakeData1$MPGCat == "Medium"
badlow <- fakeData1$Winning.Percentage < 0.5 & fakeData1$MPGCat == "Low"
```

Creating Our Generated Data

2. For players within each of these categories, we assume salary follows a uniform distribution. For each, we generated 60 possible salaries based on draft pick. We assume [*DP*] is a vector of length 60 containing all integers 1–60.

Base Salary (from MPG Category)		
LOW	*MEDIUM*	*HIGH*
$1 Million	$2 Million	$3 Million

$$[GeneratedSalariesMPG] = base \times (2 - ([DP]/60))$$

This assumes that players drafted earlier and with a higher MPG will have a higher salary.

Creating Our Generated Data

3. We created another uniform distribution of possible salaries based on the team's performance as winning percentage (WP). We will once again assume [DP] is a vector of length 60 containing all integers 1–60.

multiplier (from Win % Category)	
GOOD	*BAD*
2	1

$$[\mathit{GeneratedSalariesWP}] = \mathit{multiplier} \times (1.1 - ([\mathit{DP}]/60))$$

This function assumes that teams with higher winning percentages value their players more and are willing to pay a higher salary.

Creating Our Generated Data

4. Lastly, we combined both functions from parts 2 and 3 to get a salary range for each draft pick 1–60 for each of our 6 categories.

Salary Range

Max = [*GeneratedSalariesMPG*] + [*GeneratedSalariesWP*]

Min = [*GeneratedSalariesMPG*] – [*GeneratedSalariesWP*]

We randomly selected a salary from a uniform distribution of salary between minimum and maximum salary.

Creating Our Generated Data

We now have a matrix of 500x60, with the rows representing each player and the columns representing the player's salary at each draft pick.

So how do we determine draft pick for each player?

$$P([DP]) = \frac{1 + MPG/[DP]}{\Sigma(1 + MPG/[DP])}$$

P([DP]) creates 60 probabilities (1 for each draft pick). We assume that the higher a player's MPG, the more likely they will be picked earlier in the draft.

Using our new probability distribution, we randomly select a draft pick for each player. We now have a generated draft pick and salary for each player.

Generated Data Distribution

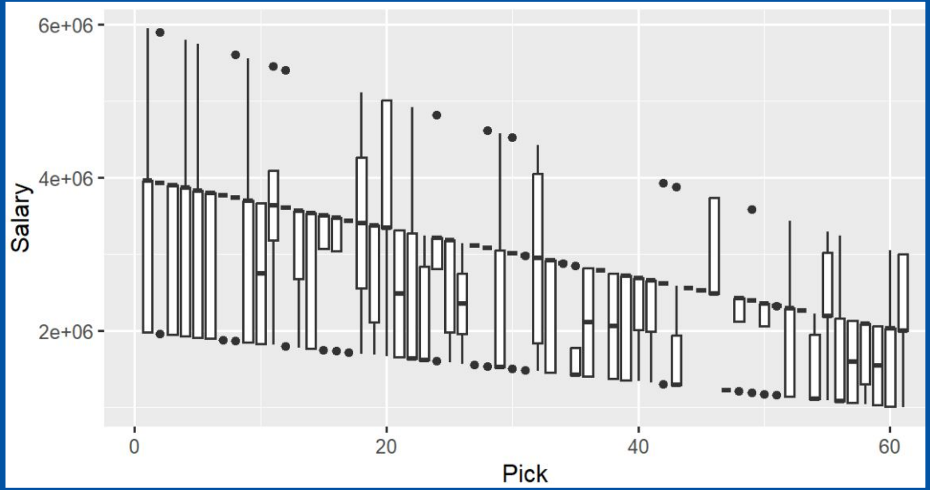


Figure 3: Generated Data Distribution of Salaries at Each Pick

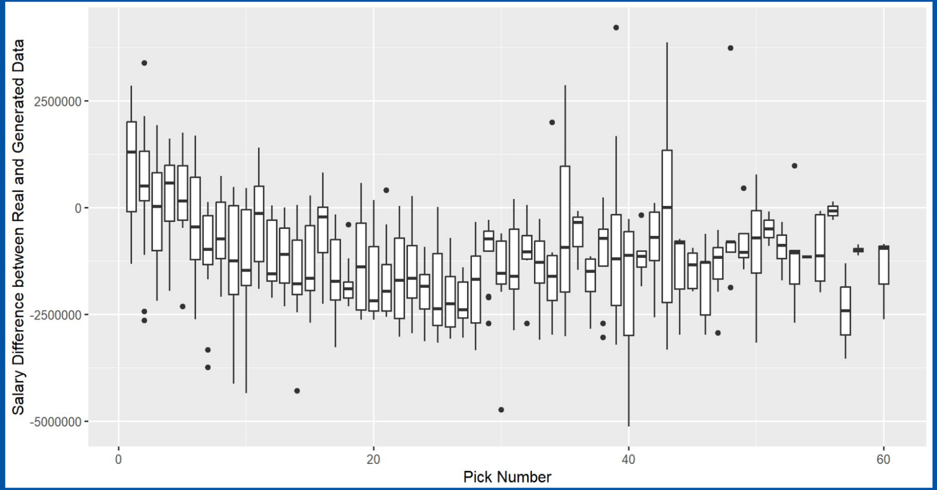


Figure 4: Comparison of Real and Generated Salaries at Each Pick

Estimates

95% Confident:
$36,329.97 ± 293.4874

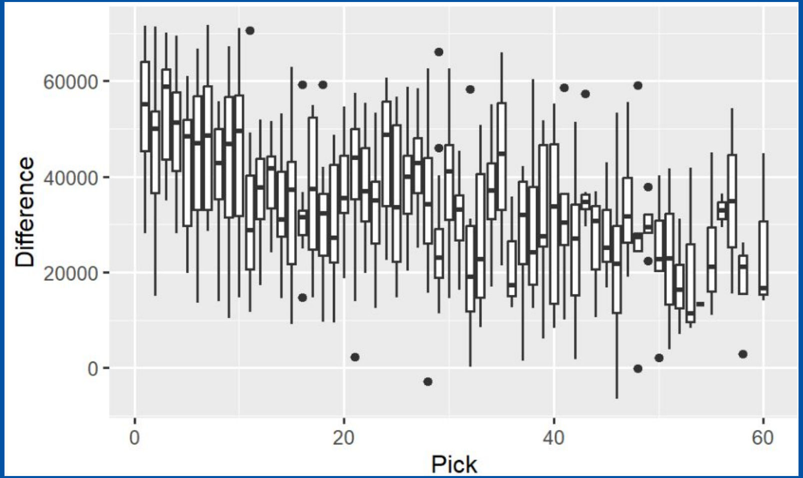


Figure 5: Distribution of Salary Differences in Real Data

95% Confident:
$28,488.98 ± 102.209

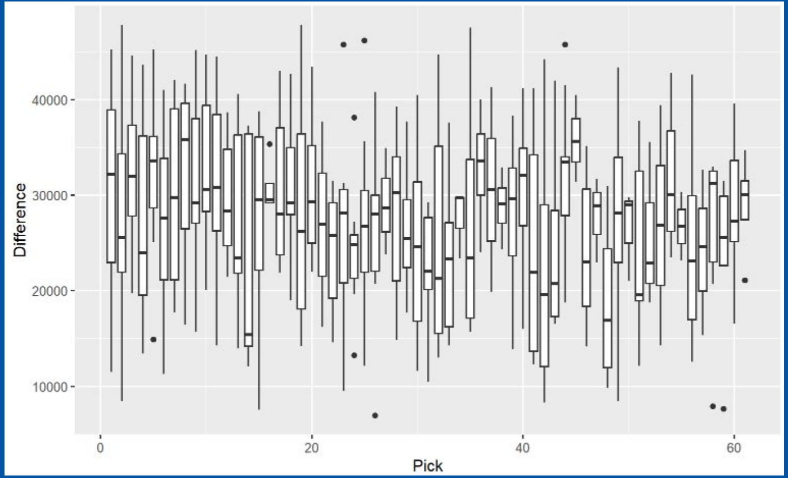
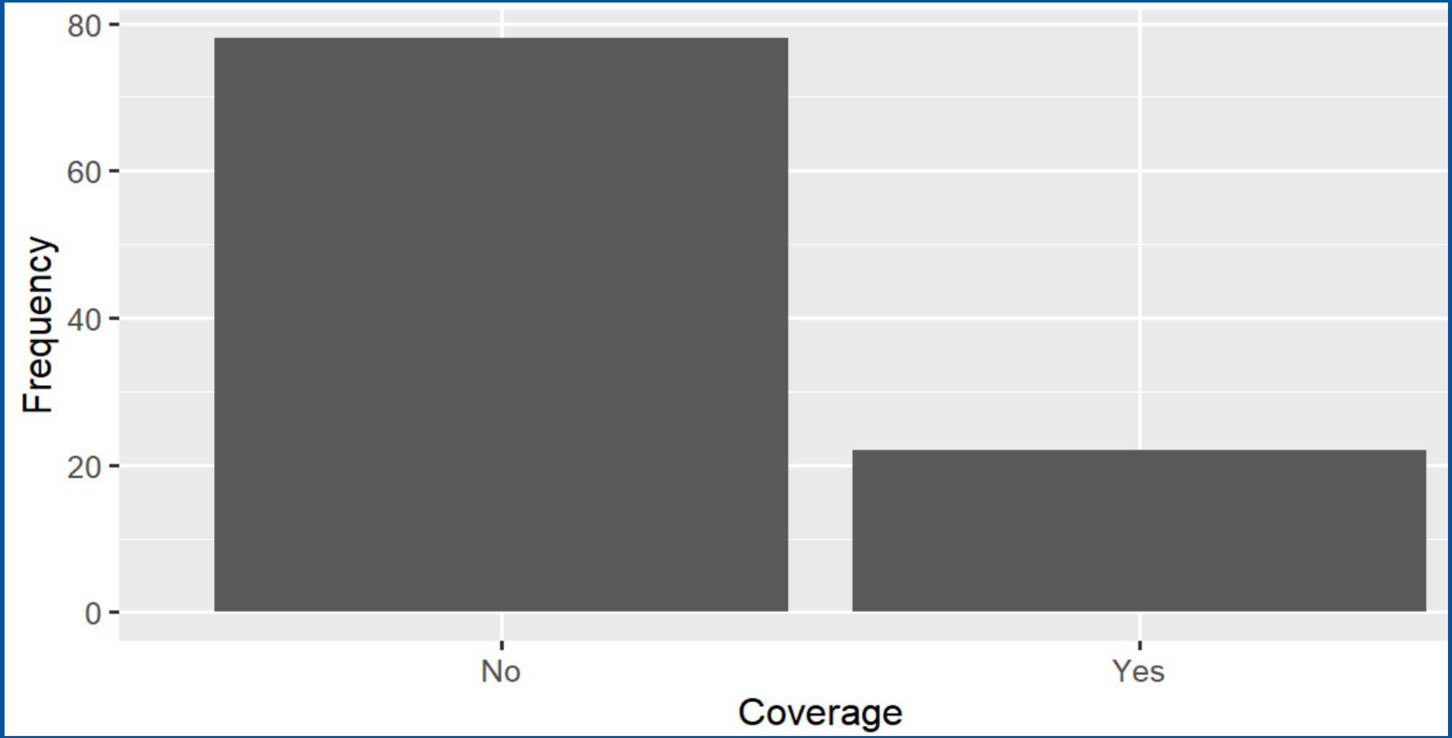


Figure 6: Distribution of Salary Differences in Generated Data

Coverage



Conclusions And Future Directions

- Difference in salary is about $30,000 between draft picks (based on real and generated data)
- Difference in salary between picks that are two or more apart is significant
- Assuming that salary is a viable approximation of player performance, our estimate suggests that those who are drafted earlier are generally better.
- Find a way to take more factors into account when talking about player performance / skill
- Getting picked later decreases salary and thus, would suggest a smaller second contract.
- In the future, we would like to look at pre-draft/college player stats as well.

Thank you!

Questions? Comments?