Swinging, Fast and Slow

Scott Powers[1] and Ron Yurko[2]

[1]Department of Sport Management, Rice University

[2]Department of Statistics & Data Science, Carnegie Mellon University

Saberseminar 2024

RICE UNIVERSITY
Sport Analytics

Question

How much do batters change their swings according to the count?

And what are the consequences of these approaches?

Conclusions (Preview)

1. Estimating batter intention is a helpful framework for reasoning about the sources of variation in swing metrics
2. Batters can reduce their strikeout rate by modulating their swing length according to the count ...
3. ... but it's not worth the power tradeoff for the average batter (from a linear weights perspective)

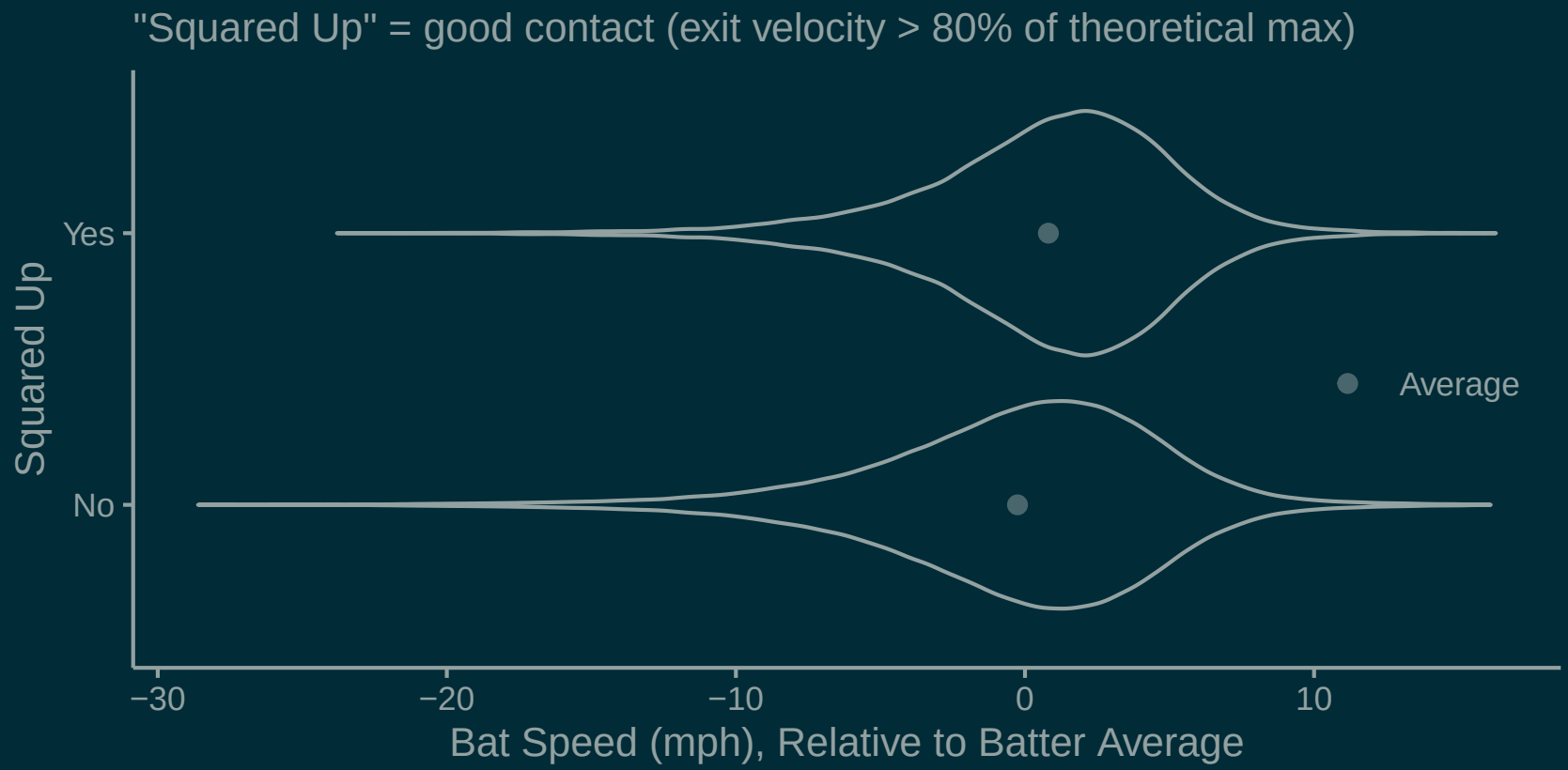
Data

May 2024: MLB releases exciting new swing tracking metrics

- For every swing, we get:
  - Swing length (distance traveled by bat head from "start")
  - Bat speed (speed of bat "sweet spot")
- Measured at point of contact (or point of nearest contact)
- Additional derived metrics:
  - Squared-up rate (exit velocity $>$ 80% of theoretical max)
  - Fast-swing rate
  - Blasts
  - Swords
- More in the future pipeline? (miss distance, contact depth, ...)

THANK YOU, MLBAM!

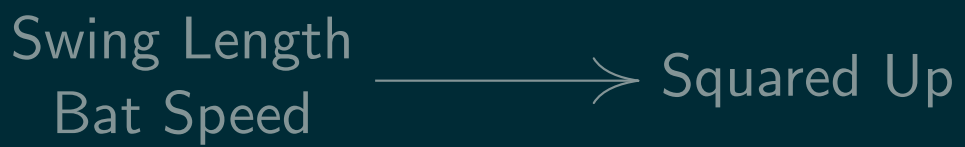
When batters swing slower, do they make better contact?



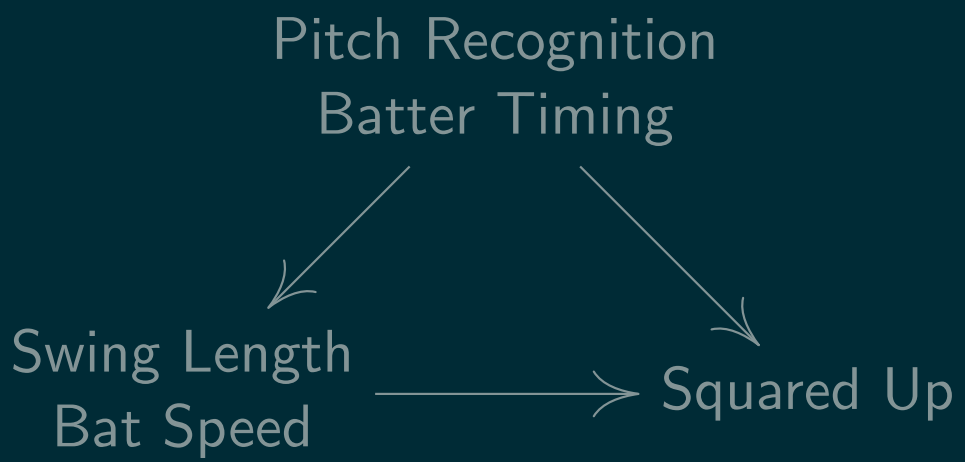
- Squared Up swings are faster on average, but ...

Caution #1: Correlation does not imply causation

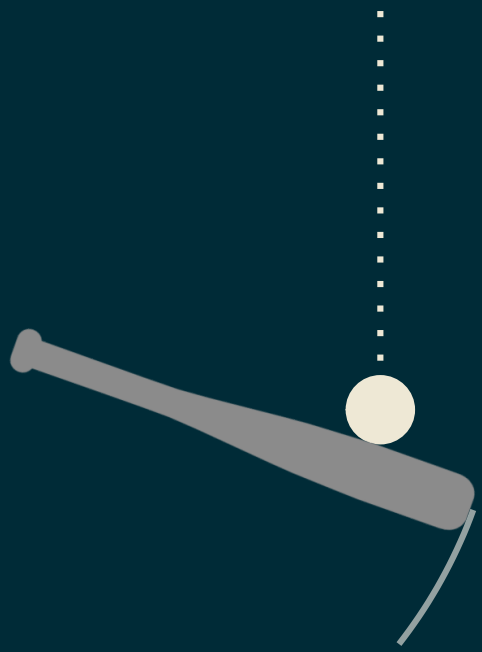
One possible causal model:



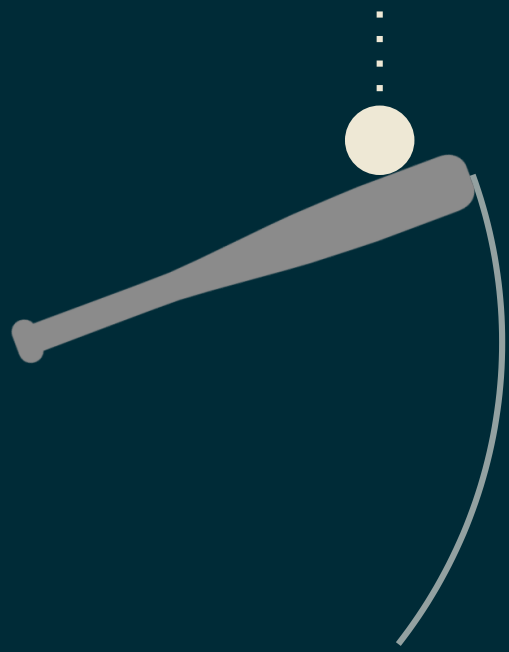
Another (more plausible) causal model with confounders:



Caution #2: How do we get our measurement?



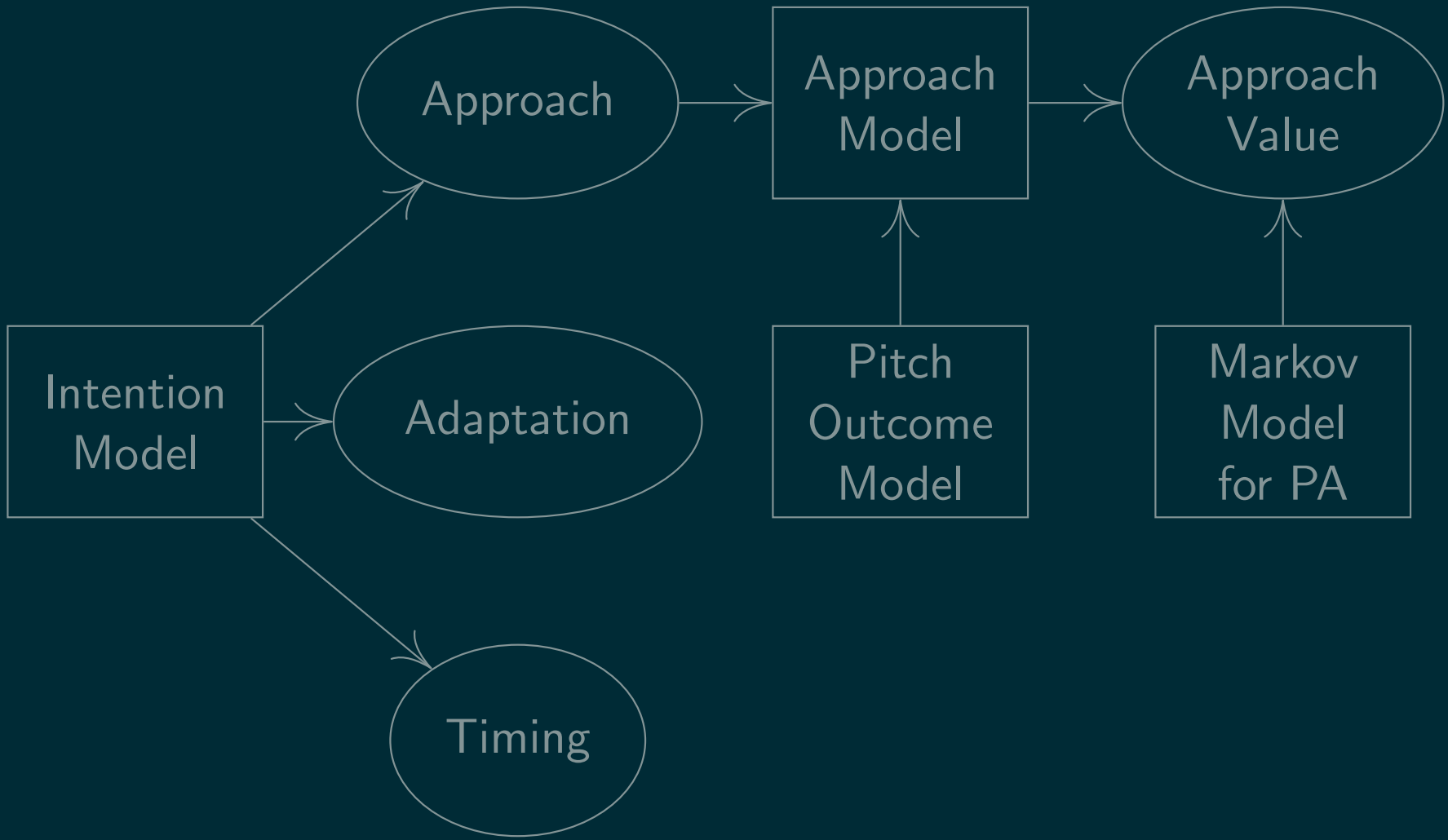
LATE Swing



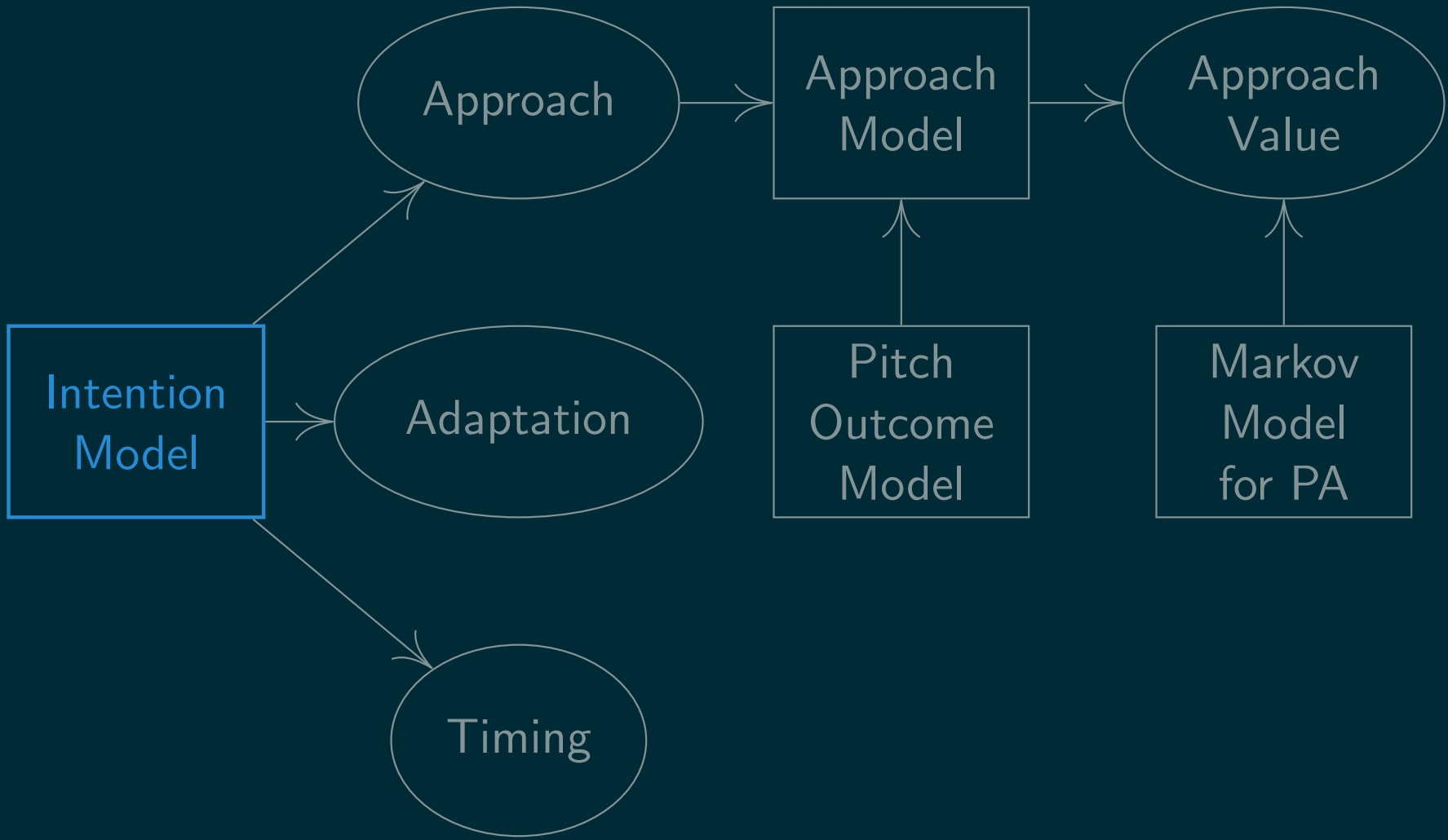
EARLY Swing

- For identical swings, timing determines point of measurement
- Swing length and bat speed are outcomes, not purely process

Outline


Approach
Approach Model
Approach Value
Intention Model
Adaptation
Pitch Outcome Model
Markov Model for PA
Timing

Outline


Approach
Approach Model
Approach Value
Intention Model
Adaptation
Pitch Outcome Model
Markov Model for PA
Timing

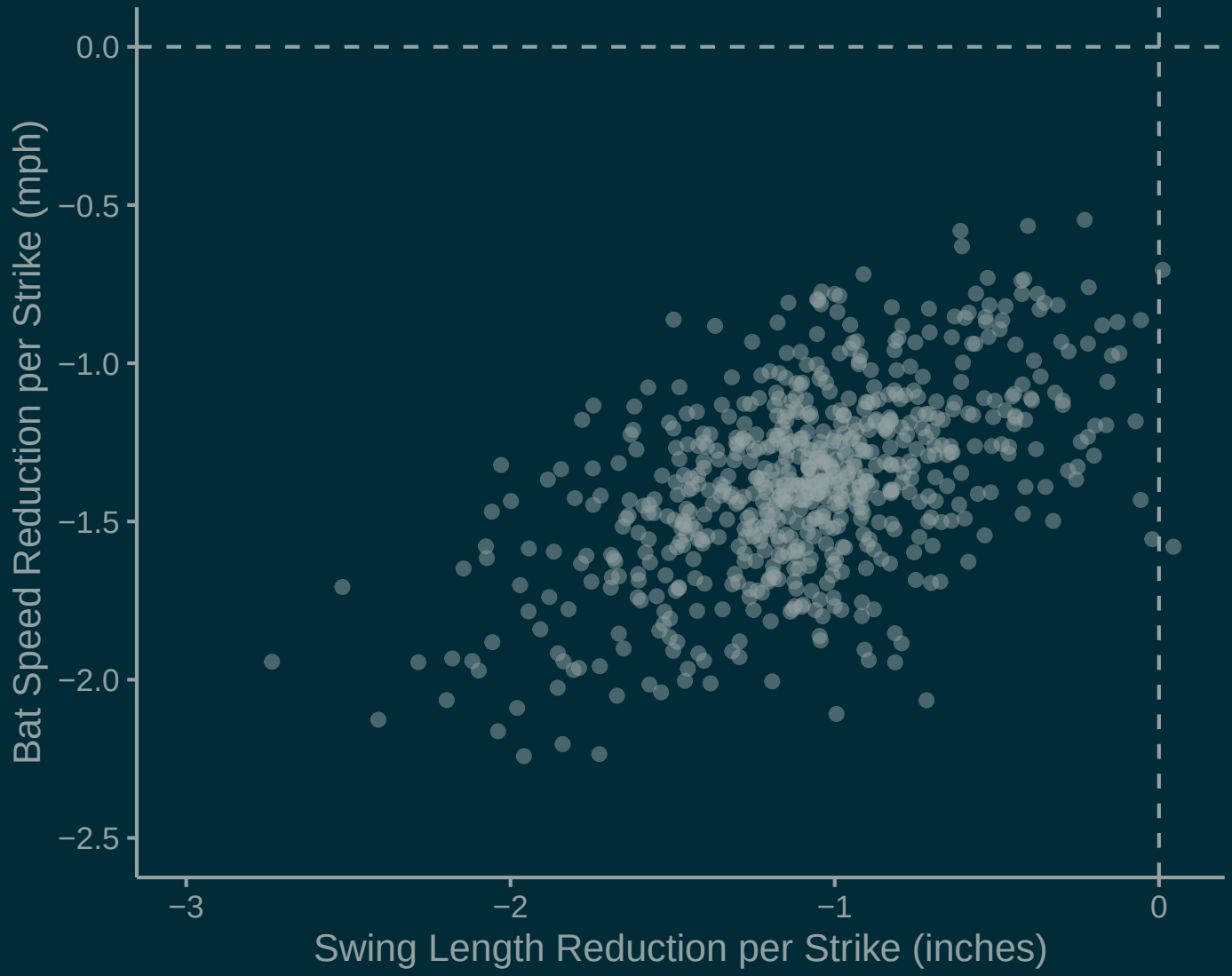
Intention Model

Goal: Estimate swing metrics at intended point of contact

1. Filter on swings above 50 mph (avoid bunts, check-swings)
2. Filter on swings that are squared-up (avoid bad timing)
3. Fit mixed-effects linear models for swing length and bat speed
    - Fixed and random effects for intercept, count, pitch location

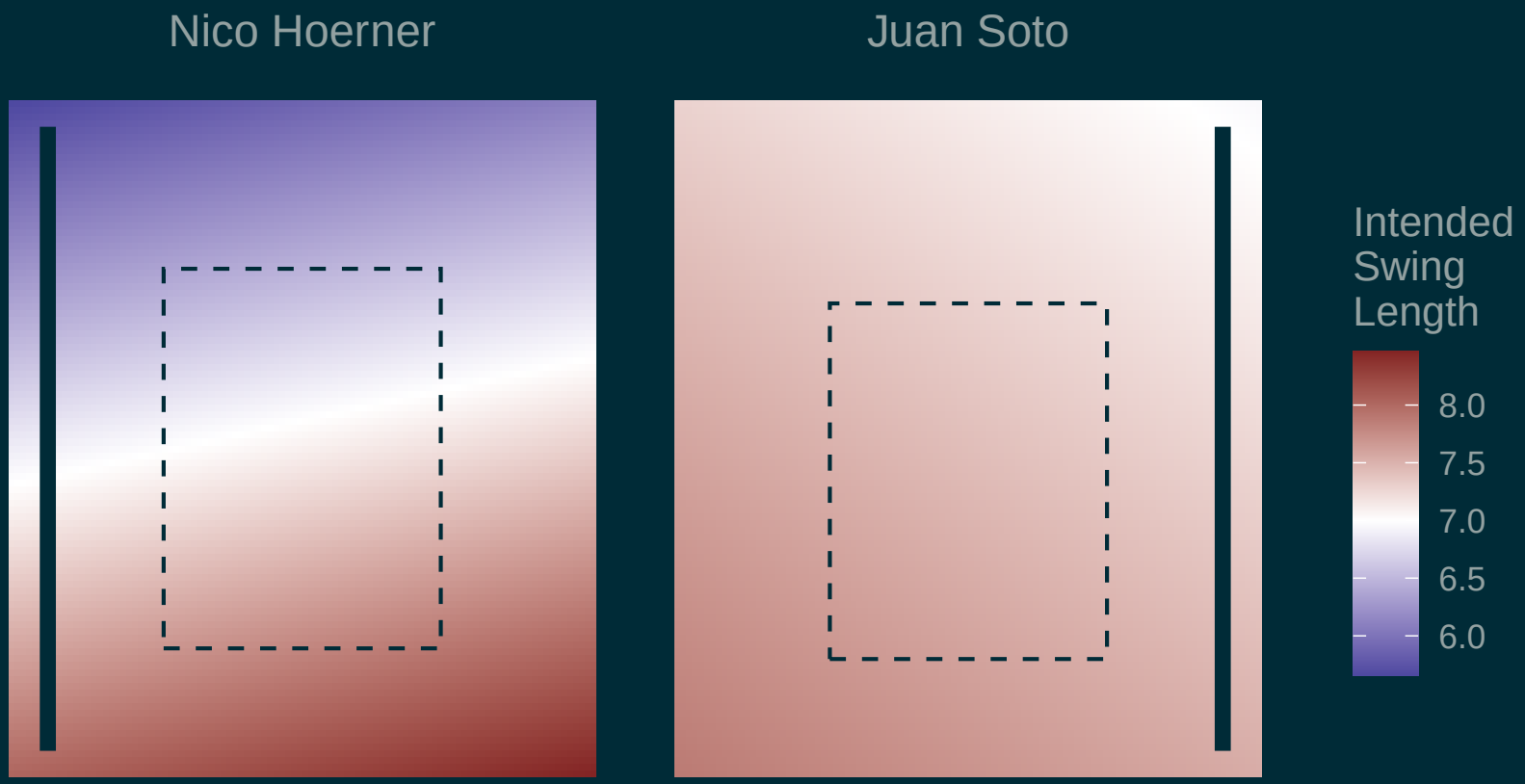
$$
\begin{aligned}
(\text{swing length})_i = \alpha + \gamma^A_{b_i} & \\
& + \beta^B \cdot (\text{balls})_i \\
& + (\beta^S + \gamma^S_{b_i}) \cdot (\text{strikes})_i \quad \Big\} \text{approach} \\
& + (\beta^X + \gamma^X_{b_i}) \cdot (\text{pitch loc x})_i \\
& + (\beta^Z + \gamma^Z_{b_i}) \cdot (\text{pitch loc z})_i \quad \Big\} \text{adaptation} \\
& + \epsilon_i
\end{aligned}
$$

Approach: How do intended swings vary by count?



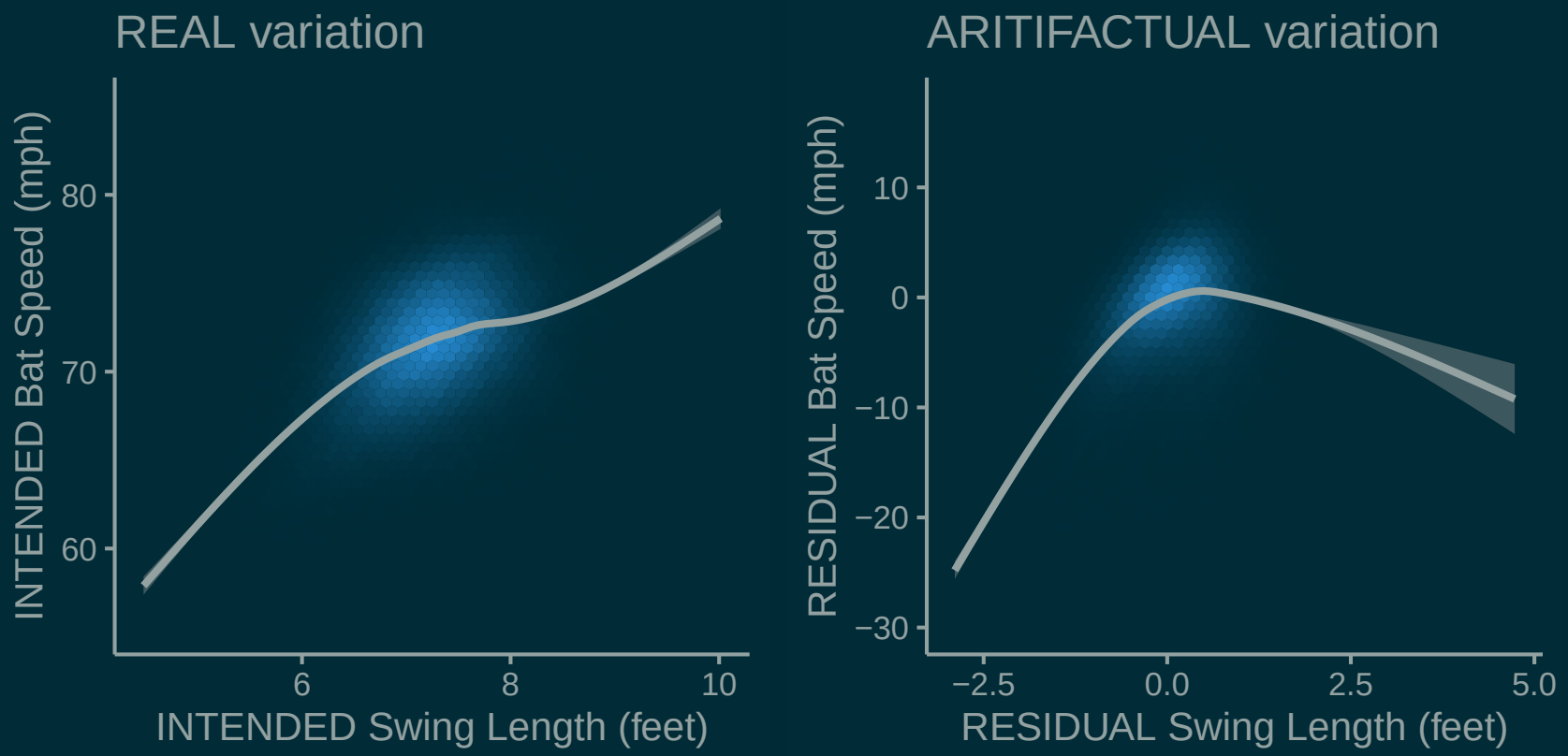
Adaptation: How do intended swings vary by location?

Swing length predicted by intention model, assuming 0–0 count



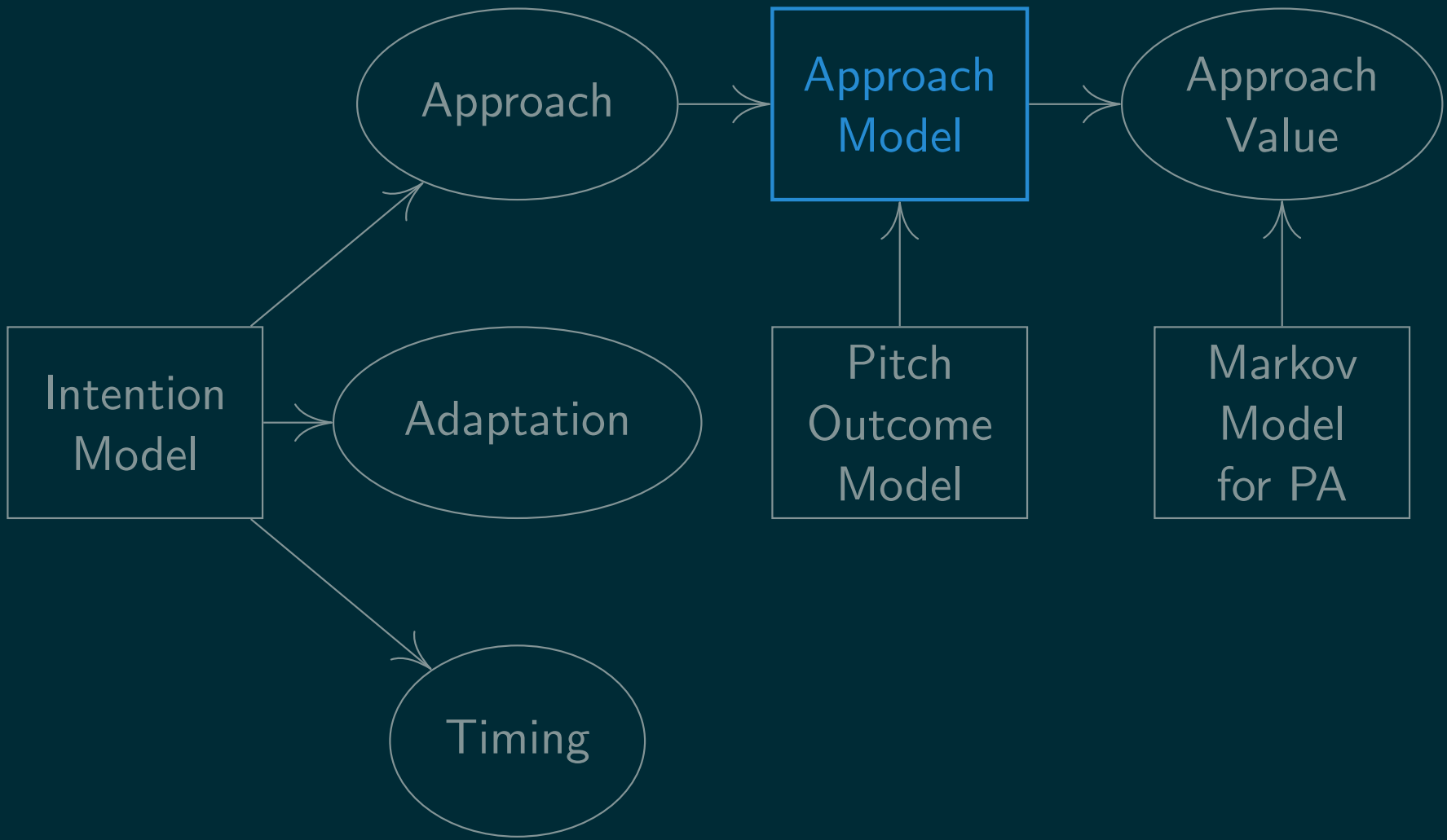
- There's clearly a difference, but it's hard to say why

Timing: What swing variation remains unexplained?



- Left: Predicted intention for all swings (not just squared up)
- Right: Residuals (relative to intention) for all swings
  - Pattern matches bat speed through path of swing!

Outline


Approach
Approach Model
Approach Value
Intention Model
Adaptation
Pitch Outcome Model
Markov Model for PA
Timing

Approach Model

1. Start with pitch outcome model which estimates $\mathbb{P}(\text{swing})$, $\mathbb{P}(\text{contact} \mid \text{swing})$, $\mathbb{P}(\text{fair} \mid \text{contact})$, $\mathbb{E}(\text{xwOBA} \mid \text{fair})$, etc.
    - Gradient boosting w/ 2022-2024 MLB data
2. Refit contact, fair and xwOBA models on batter approach, with previous predictions as offset

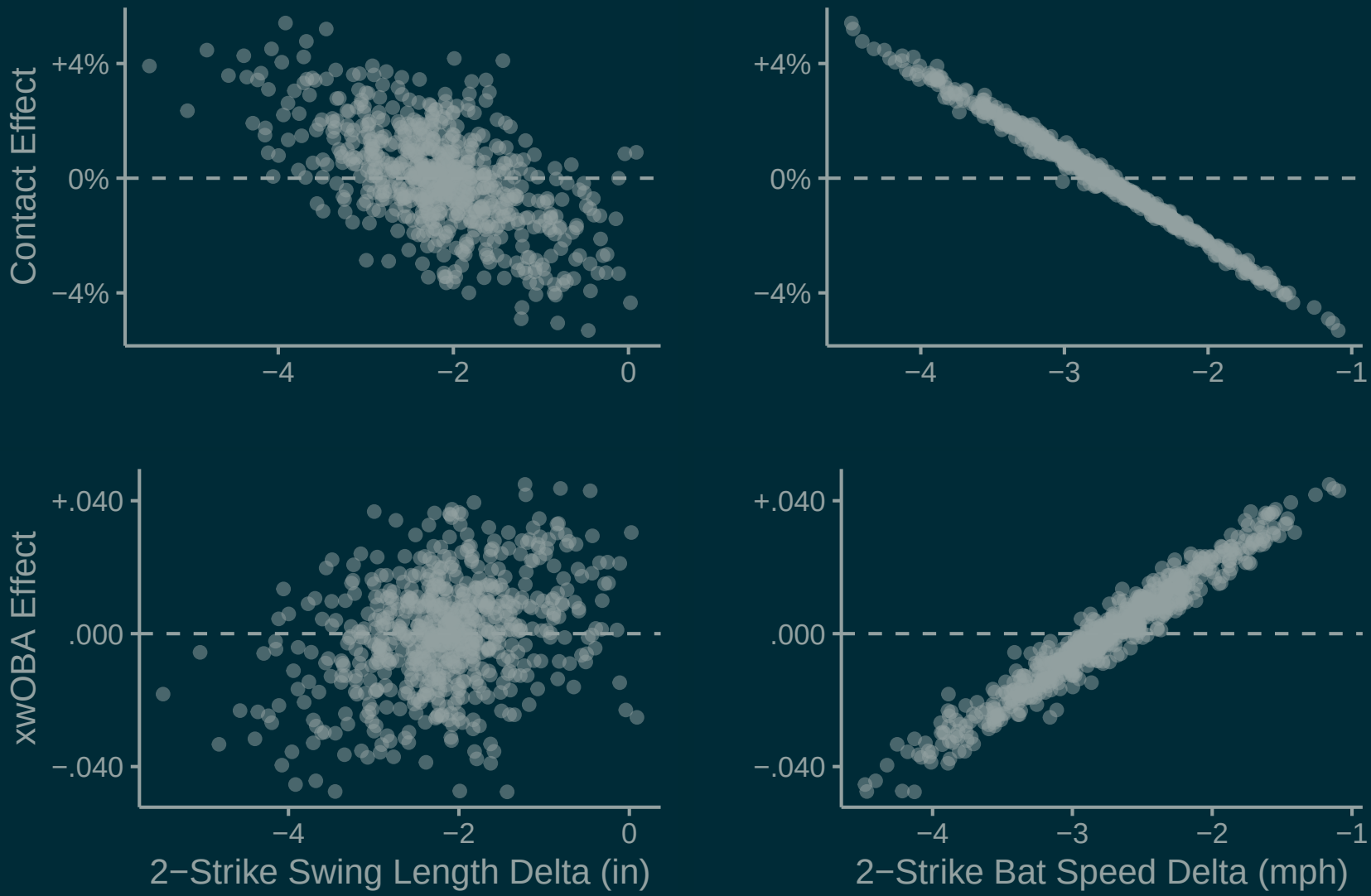
Example: p_i is contact prob, $\hat{p}_i$ is pitch outcome model estimate

$$\log\left(\frac{p_i}{1-p_i}\right) = \log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) + \beta^{SL} \cdot \hat{\gamma}^{SL}_{b_i} \cdot (\text{strikes})_i + \beta^{BS} \cdot \hat{\gamma}^{BS}_{b_i} \cdot (\text{strikes})_i$$

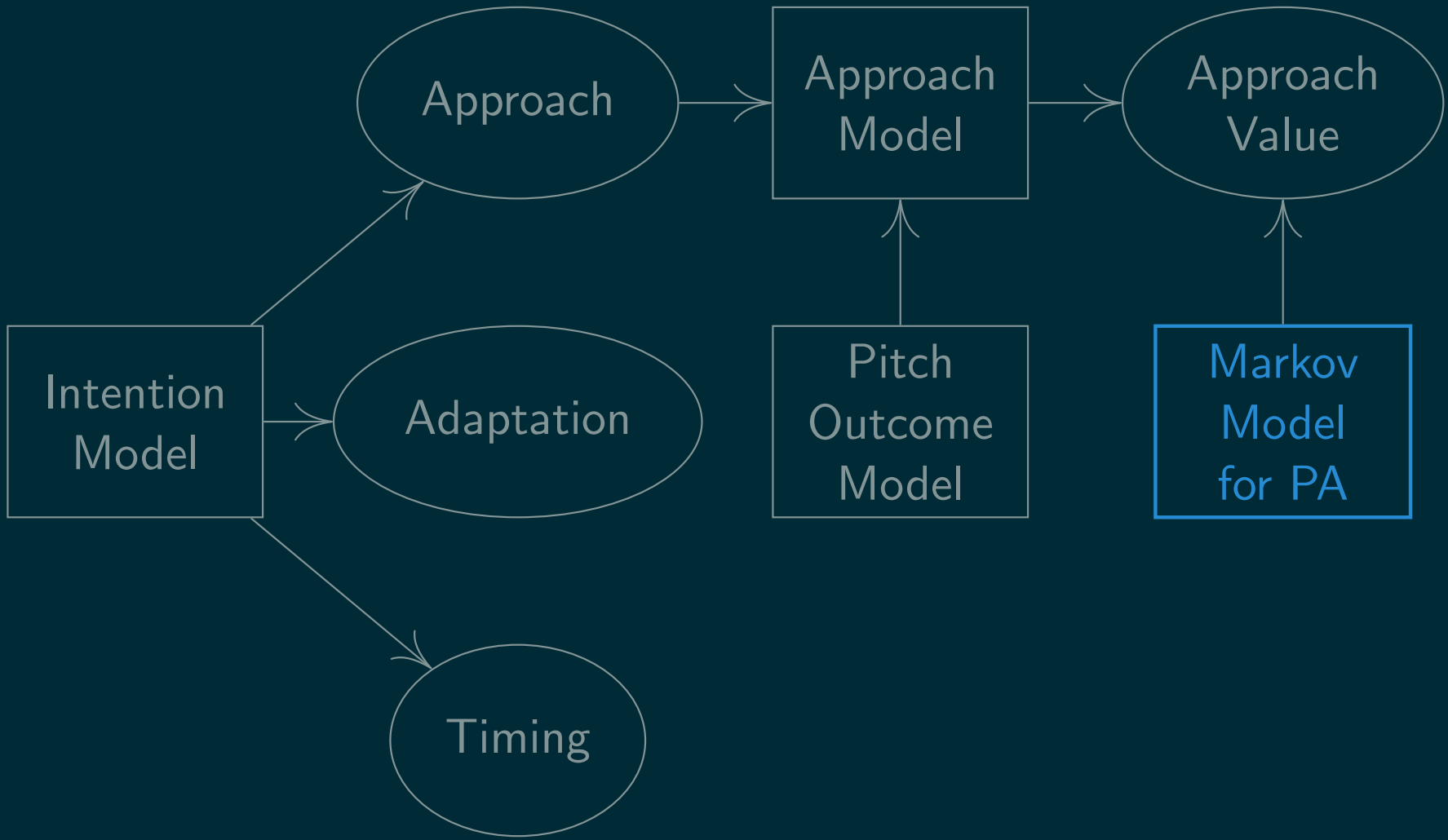
- This is instrumental variable regression from causal inference, using count as the instrument

The Effect of Approach

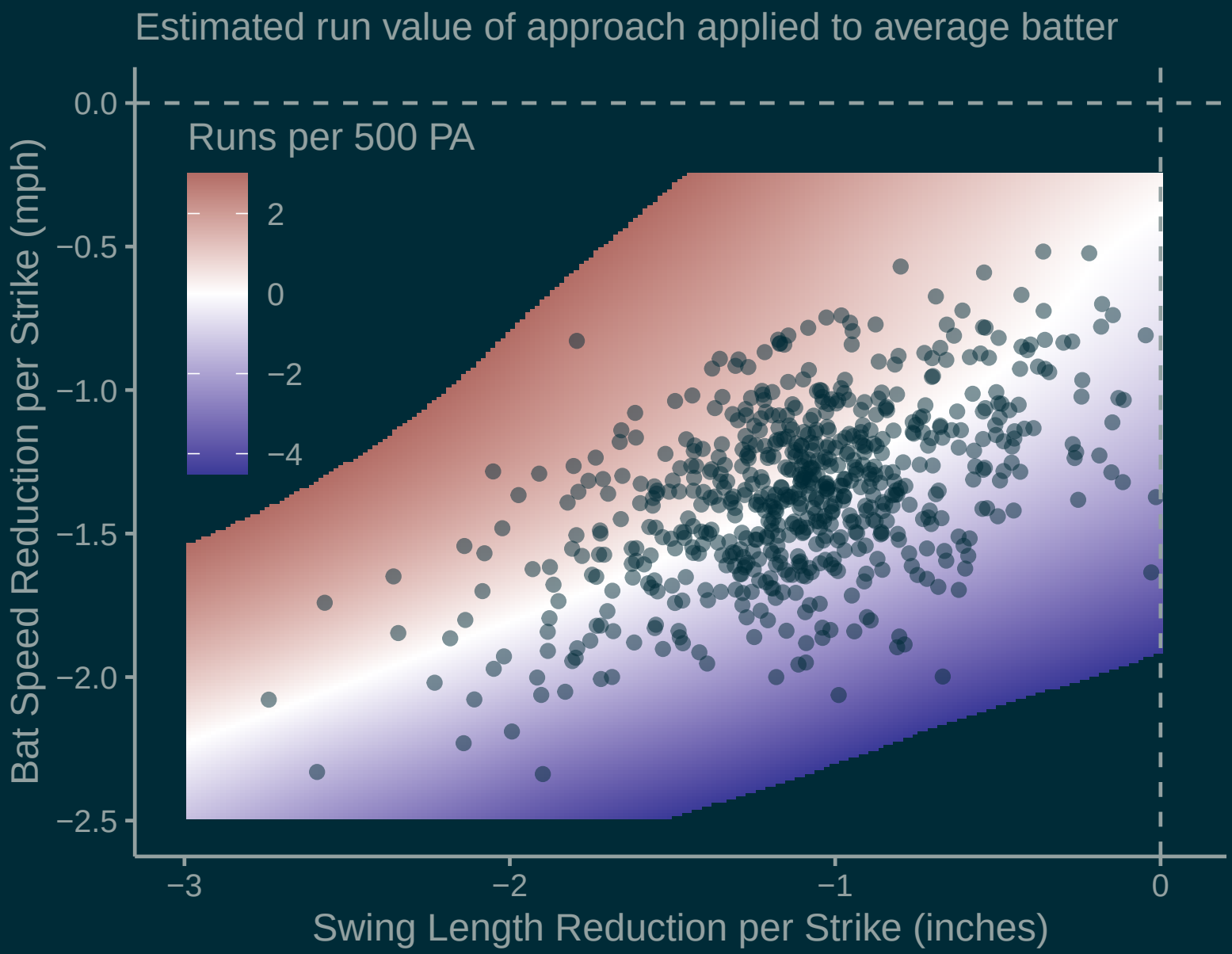
"Effect" is approach-aware vs. approach-agnostic prediction w/ 2 strikes



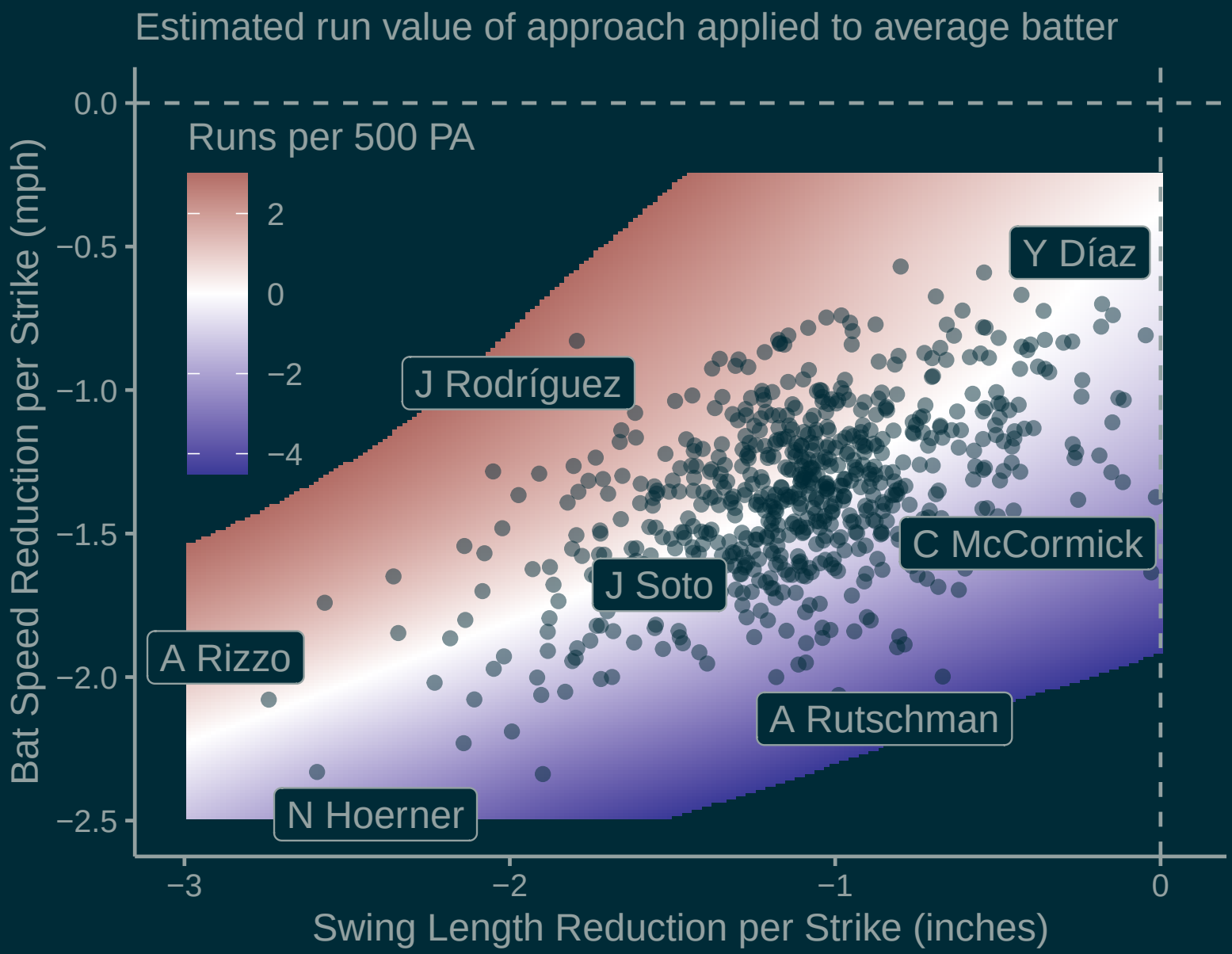
Outline


Approach
Approach Model
Approach Value
Intention Model
Adaptation
Pitch Outcome Model
Markov Model for PA
Timing

Is it good to modulate your swing by count?



Is it good to modulate your swing by count?



Conclusions

1. Estimating batter intention is a helpful framework for reasoning about the sources of variation in swing metrics
2. Batters can reduce their strikeout rate by modulating their swing length according to the count ...
3. ... but it's not worth the power tradeoff for the average batter (from a linear weights perspective)

Thank You!

Acknowledgements: Nick Wan, Noah Woodward

All code for this project (and these slides) available at:
`github.com/saberpowers/swinging-fast-and-slow`

You can learn more about each of our research at:
`saberpowers.github.io`
`stat.cmu.edu/~ryurko`