A Bayesian Analysis of the Time Through the Order Penalty in Baseball

Ryan Brill, Sameer Deshpande, and Adi Wyner

CASSIS 2022.

The Time Through the Order Penalty

- In Game 6 of the 2020 World Series, pitcher Blake Snell was removed from the game at the top of the third time through the batting order because of the third *Time Through the Order Penalty* (TTOP)
- Snell was pitching exceptionally well; Snell's replacement promptly gave up two runs, and the Rays lost the series that night
- 3TTOP: there is a dropoff in pitcher performance from the 2^{nd} to the 3^{rd} time a batter faces a pitcher because the batters have learned the tendencies of the pitcher (i.e. within-game batter learning)
- TTOP was first identified by Tom Tango in *The Book: Playing the Percentages in Baseball*, and is now considered canon in the baseball community

The Time Through the Order Penalty

TTOP is widely believed:

- **Jeff Banister, Rangers**: "There are certain times that third time through does get problematic"
- **Brad Ausmus, Tigers**: "the more times a hitter sees a pitcher, the more success that hitter is going to have"
- **Kevin Cash, Rays**: "Times through the order, we value that"

MLB teams frequently use the TTOP to justify pulling pitchers at the start of the 3^{rd} TTO

The Original Analysis from The Book

<u>The original analysis:</u>

- Essentially, bin and average pitcher performance, via wOBA, over the first 27 batters of a game and examine this trajectory
- *Weighted On-Base Average* (wOBA),

outcome k	Out	BB	HBP	1B	2B	3B	HR
weight w_k	0	0.69	0.719	0.87	1.217	1.528	1.94

<u>Issues with the original analysis:</u>

1. Doesn't disentangle pitcher fatigue from batter learning
2. No statistical model
   - No uncertainty quantification on their estimated TTOP

Data & Model

- Scraped data from every plate appearance from 1990-2020 from Retrosheet
- Restrict our study to plate appearances featuring a starting pitcher in one of the first three times through the order, using the 2018 season as our primary example
- 7 plate appearance outcomes,

$$\{\text{Out, Walk, Hit by pitch, Single, Double, Triple, Home Run}\}$$

- Model the outcome of a plate appearance over the course of a game using multinomial logistic regression, e.g.

$$\log\left(\frac{\mathbb{P}(\text{Single})}{\mathbb{P}(\text{Out})}\right) = \text{ a function of pitcher fatigue}$$
$$\text{and batter learning}$$

Model

- There are 2 potential mechanistic reasons for a change in pitcher performance over the course of a game, after adjusting for confounders: *pitcher fatigue* and *batter learning*

$$\begin{aligned}\log\left(\frac{\mathbb{P}(\text{Single})}{\mathbb{P}(\text{Out})}\right) = \; & \alpha_0 + \alpha_1 \cdot t \; + \\ & \beta_2 \cdot \mathbb{1}\{t \in 2TTO\} + \\ & \beta_3 \cdot \mathbb{1}\{t \in 3TTO\} + \\ & \eta_1 \cdot (\text{Bat. Qual.}) \; + \eta_2 \cdot (\text{Pit. Qual.}) + \\ & \eta_3 \cdot (\text{Hand}) \; + \eta_4 \cdot (\text{Home})\end{aligned}$$

- $\alpha_0 + \alpha_1 \cdot t$ models continuous decline within a game
- Represents pitcher fatigue, as pitchers fatigue continuously

Model

- There are 2 potential mechanistic reasons for a change in pitcher performance over the course of a game, after adjusting for confounders: *pitcher fatigue* and *batter learning*

$$\log\left(\frac{\mathbb{P}(\text{Single})}{\mathbb{P}(\text{Out})}\right) = \alpha_0 + \alpha_1 \cdot t + \boxed{\beta_2 \cdot \mathbb{1}\{t \in 2TTO\}} + \boxed{\beta_3 \cdot \mathbb{1}\{t \in 3TTO\}} + \eta_1 \cdot (\text{BQ.}) + \eta_2 \cdot (\text{PQ.}) + \eta_3 \cdot (\text{Hand}) + \eta_4 \cdot (\text{Home})$$

- model discontinuous decline from one TTO to the next
- A large positive β_2 corresponds to a 2TTOP due to batter learning
- A large positive $\beta_3 - \beta_2$ corresponds to a 3TTOP due to batter learning

Model

- There are 2 potential mechanistic reasons for a change in pitcher performance over the course of a game, after adjusting for confounders: *pitcher fatigue* and *batter learning*

$$
\begin{aligned}
\log\left(\frac{\mathbb{P}(\text{Single})}{\mathbb{P}(\text{Out})}\right) = \alpha_0 + \alpha_1 \cdot t \, + & \\
& \beta_2 \cdot \mathbb{1}\{t \in 2TTO\} \, + \\
& \beta_3 \cdot \mathbb{1}\{t \in 3TTO\} \, + \\
& \eta_1 \cdot (\text{Batter Quality}) \, + \\
& \eta_2 \cdot (\text{Pitcher Quality}) \, + \\
& \eta_3 \cdot (\text{Pitcher-Batter Handedness Match}) \, + \\
& \eta_4 \cdot (\text{Batter at Home})
\end{aligned}
$$

- Adjust for confounders

Model (Priors)

$$\log\left(\frac{\mathbb{P}(\text{Single})}{\mathbb{P}(\text{Out})}\right) = \alpha_0 + \alpha_1 \cdot t + \beta_2 \cdot \mathbb{1}\{t \in 2TTO\} + \beta_3 \cdot \mathbb{1}\{t \in 3TTO\} + \eta_1 \cdot (\text{Bat. Qual.}) + \eta_2 \cdot (\text{Pit. Qual.}) + \eta_3 \cdot (\text{Hand}) + \eta_4 \cdot (\text{Home})$$

- Positive slope prior so that f represents fatigue

$$\alpha_1 \sim \text{half } t_7$$

- Results are similar for other choices of positive slope prior
- Standard normal priors for non-slope parameters $\alpha_0, \beta_2, \beta_3, \eta$
- Parameters are estimated jointly in Stan

Simulation Study

1. Set “true” values for the parameters of our model.
   e.g., parameters that reflect
   - no TTOP
   - a TTOP consistent with Tango’s alleged effect size
   - a huge TTOP
2. Draw one season of plate appearance outcomes from our model ($108,545$ plate appearances)
3. Fit our model to these simulated plate appearance outcomes
4. Show that we recovered the parameters, and that we recovered the change in expected wOBA over the course of a game

Simulation Study

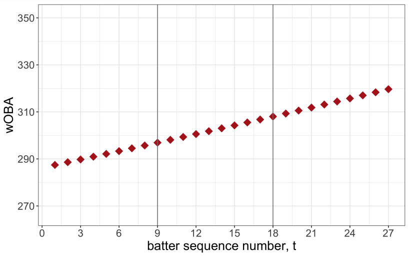
- Show that we recovered the change in expected wOBA over the course of a game
- *Weighted On-Base Average* (wOBA),

outcome k	Out	BB	HBP	1B	2B	3B	HR
weight w_k	0	0.69	0.719	0.87	1.217	1.528	194

- *Expected wOBA*,

$$\begin{aligned} xWOBA = & w_{\text{single}} \cdot \mathbb{P}(\text{single}) + \\ & w_{\text{double}} \cdot \mathbb{P}(\text{double}) + \\ & w_{\text{triple}} \cdot \mathbb{P}(\text{triple}) \ + \cdots \end{aligned}$$

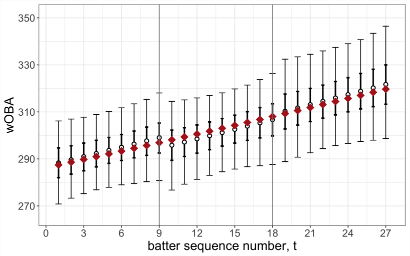
Simulation Study: No TTOP



- Red: the "true" underlying $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$

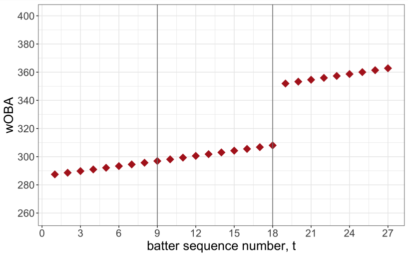
Simulation Study: No TTOP

- Our model recovers $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$



- Red: the "true" underlying $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$
- White dots: Posterior mean
- Black bars: 95% posterior credible interval

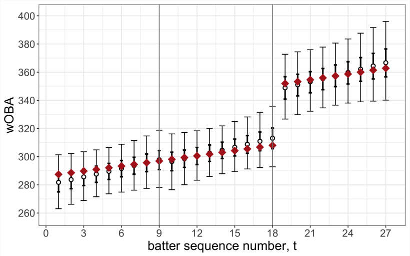
Simulation Study: Huge 3TTOP



- Red: the "true" underlying $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$

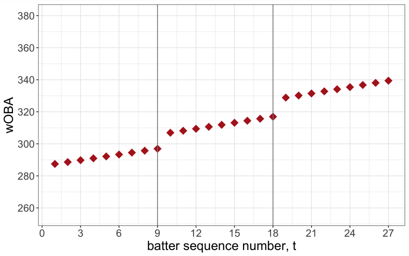
Simulation Study: Huge 3TTOP

- Our model recovers $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$



- Red: the "true" underlying $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$
- White dots: Posterior mean
- Black bars: 95% posterior credible interval

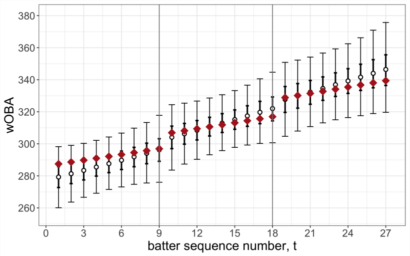
Simulation Study: TTOP Consistent with Tango's Effect Size



- Red: the "true" underlying $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$

Simulation Study: TTOP Consistent with Tango's Effect Size

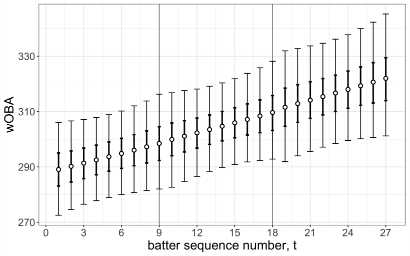
- Our model recovers $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$



- Red: the "true" underlying $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$
- White dots: Posterior mean
- Black bars: 95% posterior credible interval

Results: Little evidence for a strong Batter Learning Effect

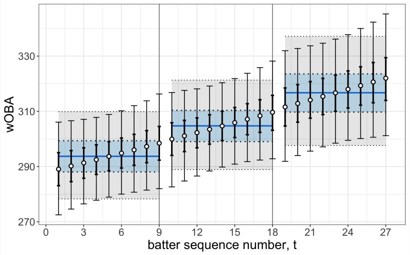
- Fit the model on all plate appearances from 2018, and plot the posterior distribution of $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$
- Expected wOBA increases steadily over the course of a game, without discontinuity due to batter learning



- White dots: Posterior mean
- Black bars: 95% posterior credible interval
- Similar plots in other years

Results: Tango's Conclusions Fit Within Our Framework

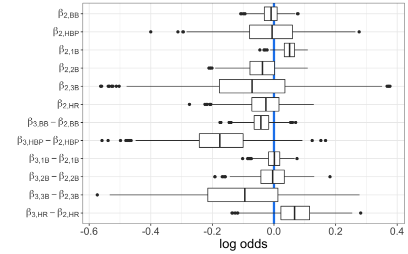
- Tango: from one TTO to the next, pitchers decline by about 10 wOBA points, on average
  - We see this is an artifact of continuous pitcher decline



- Posterior dist. of $t \mapsto xWOBA(t, x|\alpha, \beta, \eta)$, fit on 2018 data
- Blue line: *xWOBA* post. means, averaged within each TTO
- Blue region: *xWOBA* 50% intervals, avgd. within each TTO
- Gray region: *xWOBA* 95% intervals, avgd. within each TTO

Results: Little evidence for a strong Batter Learning Effect

- For each plate appearance outcome ($k = \text{out}$, 1B, 2B, etc...),
  - 2TTOP test: is β_{2k} significantly positive?
  - 3TTOP test: is $\beta_{3k} - \beta_{2k}$ significantly positive?



- Each posterior boxplot covers positive and negative values
- Similar results in other years

Conclusion

- Little evidence for a strong batter learning effect
- Pitcher fatigue appears to be the primary driver behind pitcher decline over the course of a game
- Therefore, end the rule of thumb that pitchers should be pulled at the start of the third TTO
- Pulling a pitcher should be a decision based on pitcher fatigue and pitcher quality

Thank you!

- Twitter: @RyanBrill_
- Email: ryguy123@sas.upenn.edu