

## Lecture 6: Models Do What They're Told

Instructor: Ryan Brill

Scribe: Jonathan Pipping

## 6.1 Case Study: Time Through the Order Penalty

In Game 6 of the 2020 World Series, the Tampa Bay Rays' manager, Kevin Cash, pulled his starting pitcher, Blake Snell, midway through the sixth inning. When he was pulled, Snell had been pitching extremely well; he had allowed just two hits and struck out nine batters on 73 pitches. Moreover, the Rays had a one run lead. Snell's replacement, Nick Anderson, promptly gave up two runs, which ultimately proved decisive: the Rays went on to lose the game and the World Series. After the game, Cash justified his decision to pull Snell, remarking that he "didn't want Mookie [Betts] or [Corey] Seager seeing Blake a third time" (Rivera, 2020).

In his justification, Cash cites the *Time Through the Order Penalty* (TTOP), which is formally identified in Tango et al. (2007, pp. 187-190) and recently popularized by Lichtman (2013). It has long been observed that, on average, batters tend to perform better the more times they face a pitcher; for instance, they tend to get on base more often on their third time facing a pitcher than their second. Tango et al. (2007) quantified the corresponding drop-off in pitcher performance as increases in *weighted on-base average* (wOBA). They observed that the average wOBA of a plate appearance in the first time through the order (1TTO) is about 9 wOBA points less than in the second TTO (2TTO). Further, the average wOBA of a plate appearance in the second TTO is about 8 wOBA points less than that in the third TOO (3TTO) (Tango et al. 2007, Table 81).

wOBA overcomes many limitations of traditional metrics like batting average, on-base percentage, and slugging percentage. Briefly, batting average and on-base percentage treat all hits equally, with singles being worth as much as triples. Slugging percentage attempts to reward different types of hits differently, but does so in too simplistic of a fashion: in computing slugging percentage, a triple is worth three times what a single is worth. Such weighting is arbitrary, and is not tied to the relative impact of a triple over a single with regard to, say, run scoring or win probability. wOBA combines the different aspects of offensive production into one metric, weighing each offensive action in proportion to its actual run value (Slowinski, 2010). The wOBA of a plate appearance is simply the weight associated with the offensive action of the outcome. Specifically, the 2019 wOBA weight of each offensive action in decreasing order is 1.940 for a home run (HR), 1.529 for a triple (3B), 1.217 for a double (2B), 0.870 for a single (1B), 0.719 for hit-by-pitch (HBP), 0.690 for unintentional walks (uBB), and 0 for an out (OUT) (Fangraphs, 2021). wOBA is rescaled so that the league average wOBA equals the league average on-base percentage. Throughout this paper, we use 2019 wOBA weights for each season. Additionally, we usually refer to wOBA points, which is wOBA multiplied by 1000, to be consistent with the baseball community's use of wOBA.

The TTOP is considered canon by much of the baseball community. Announcers routinely mention the 3TTO during broadcasts and several managers regularly use the 3TTO to justify their decisions to pull starting pitchers at the start of the third TTO. For instance, A.J. Hinch, who managed the Houston Astros from 2015 to 2019, noted "the third time through is very difficult for a certain caliber of pitchers to get through." Brad Ausmus, who managed the Detroit Tigers from 2014 to 2017, explained "the more times a hitter sees a pitcher, the more success that hitter is going to have" (Laurila, 2015).

Let's dig into the data ourselves to see if we can corroborate this claim!

## 6.2 Modeling the Time Through the Order Penalty

### 6.2.1 Data

We have a dataset every plate appearance from 2018-2019 featuring a starting pitcher in the first three times through the order. In total, we have 214,386 plate appearances. We have the following variables:

- `i`: The  $i^{th}$  plate appearance.
- `y_i`: wOBA of the  $i^{th}$  plate appearance (using 2019 wOBA weights).
- `t_i`: Batter sequence number  $(1, 2, \dots, 27)$ .

Note that the first time through the order corresponds to  $t_i \in \{1, 2, \dots, 9\}$ , the second time through the order corresponds to  $t_i \in \{10, 11, \dots, 18\}$ , and the third time through the order corresponds to  $t_i \in \{19, 20, \dots, 27\}$ .

### 6.2.2 Exploring the Data

Let's start by looking at the average wOBA in each time through the order. We do this in R by binning the data by time through the order and then computing the average wOBA in each bin. We observe in Table 6.1 that a starting pitcher performs worse on average in 2TTO than in 1TTO, and worse on average in 3TTO than in 2TTO. This is consistent with the time through the order narrative.

Time Through Order	wOBA
First	.304
Second	.318
Third	.333

Table 6.1: Average wOBA by time through the order. We can see that batters perform progressively better each time they face a pitcher.

### 6.2.3 Fitting the Model

Binning and averaging the data is equivalent to fitting the following regression model:

$$\mathbb{E}[y_i | t_i] = \beta_1 \mathbb{1}\{t_i \in 1TTO\} + \beta_2 \mathbb{1}\{t_i \in 2TTO\} + \beta_3 \mathbb{1}\{t_i \in 3TTO\} \quad (6.1)$$

where  $\mathbb{1}\{z\}$  is the indicator function that is 1 if  $z$  is true and 0 otherwise.

*Math HW: Prove this via  $\hat{y} = (X^T X)^{-1} X^T y$ .*

We can verify this by fitting the model in R and checking the coefficients.

```
# fit the model
model = lm(formula = woba_19 ~ factor(TTO) + 0, data = mlb_data)
# get model summary
summary(model)
```

We can also re-parametrize the model slightly so  $\beta_2$  represents pitcher decline from 1TTO to 2TTO, and  $\beta_3$  represents pitcher decline from 2TTO to 3TTO. This is equivalent to fitting the following model:

$$\mathbb{E}[y_i | t_i] = \beta_1 + \beta_2 \mathbb{1}\{t_i \geq 2TTO\} + \beta_3 \mathbb{1}\{t_i \geq 3TTO\} \quad (6.2)$$

What about the [trajectory](#) across the batter sequence number  $t = 1, 2, \dots, 27$ ? We bin and average for each  $t$  and plot the results against our fitted model in Figure 6.1. Note this is equivalent to fitting the following model:

$$\mathbb{E}[y_i | t_i] = \sum_{t=1}^{27} \beta_t \mathbb{1}\{t_i = t\} \quad (6.3)$$

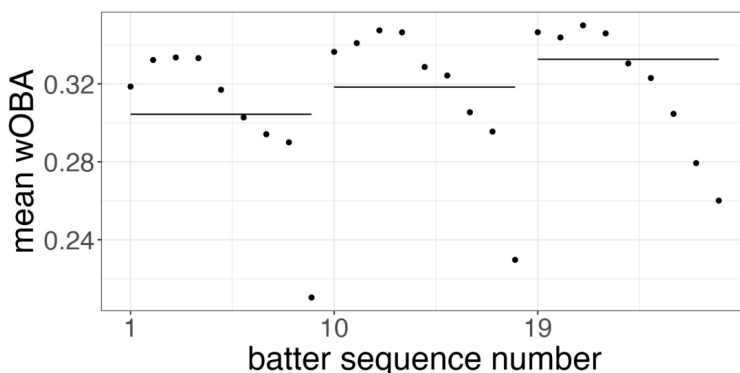


Figure 6.1: wOBA by batter sequence number plotted against our fitted model.

What's going on here? Why does the trajectory not fit along the fitted model? This is a very important reminder that regression fits observed data, it is **NOT** causation. We must first [adjust for confounders!](#)

## 6.3 Adjusting for Confounders

We need to disentangle the effect of [pitcher decline within a game](#) from the following effects:

- $BQ_i$ : Quality of the  $i^{th}$  batter.
- $PQ_i$ : quality of the  $i^{th}$  pitcher.
- $hand_i$ : Handedness match of the  $i^{th}$  pitcher-batter pair.
- $home_i$ : Home field advantage.

For now, define a batter's quality by his end-of-season wOBA averaged across all his plate appearances, and likewise for pitchers.

Later in the course we will learn a better way to estimate batter and pitcher quality that doesn't use future data: **Empirical Bayes**.

### 6.3.1 Re-fitting the Model

We now specify a model that captures pitcher decline from one time through the order to the next *after adjusting for confounders*. The model is given by:

$$\begin{aligned} \mathbb{E}[y_i | t_i] = & \beta_1 + \beta_2 \mathbb{1}\{t_i \geq 2TTO\} + \beta_3 \mathbb{1}\{t_i \geq 3TTO\} \\ & + \beta_{BQ} \cdot BQ_i + \beta_{PQ} \cdot PQ_i + \beta_{hand} \cdot hand_i + \beta_{home} \cdot home_i \end{aligned} \quad (6.4)$$

We proceed by fitting the model in R as follows:

```
# refit model with confounders
new_model = lm(formula = woba_19 ~ 1 + as.numeric(TTO >= 2) + as.numeric(TTO >= 3)
               + full_batter_woba_19 + full_pitcher_woba_19
               + hand_match + batter_home, data = mlb_data)

# get model summary
summary(new_model)
```

As we can see in Figure 6.2, we still estimate that pitchers decline from one time through the order to the next. Tom Tango's original analysis is based off of this model and is a big part of the reason that starting pitchers are often removed in the 6th or 7th inning at the start of the third time through the order.

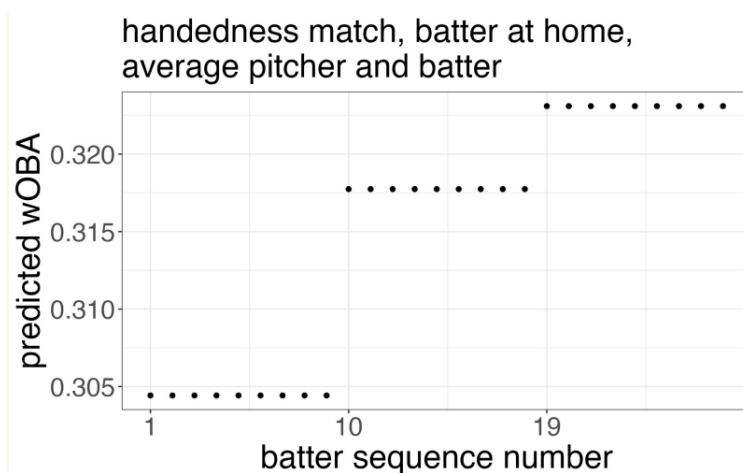


Figure 6.2: Predicted wOBA from our fitted model that adjusts for confounders.

### 6.3.2 Investigating Trajectory of Pitcher Decline

After adjusting for confounders, what does the [trajectory](#) of pitcher decline look across  $t$  (over the course of the game)? We set up a few more models to investigate this.

**Indicator Model:**

$$\mathbb{E}[y_i | t_i] = \sum_{t=1}^{27} \beta_t \mathbb{1}\{t_i = t\} + \beta_{BQ} \cdot BQ_i + \beta_{PQ} \cdot PQ_i + \beta_{hand} \cdot hand_i + \beta_{home} \cdot home_i \quad (6.5)$$

**Linear Model:**

$$\mathbb{E}[y_i | t_i] = \beta_0 + \beta_1 \cdot t_i + \beta_{BQ} \cdot BQ_i + \beta_{PQ} \cdot PQ_i + \beta_{hand} \cdot hand_i + \beta_{home} \cdot home_i \quad (6.6)$$

Additionally, we fit a smoothing spline over  $\beta_1, \dots, \beta_{27}$ . After fitting these models in R, we plot them together in Figure 6.3.

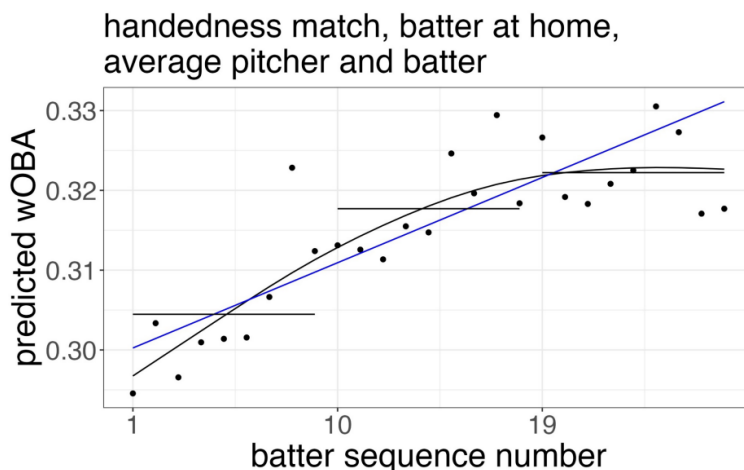


Figure 6.3: Fitted models of wOBA by batter sequence number adjusted for confounders. The black dots are  $\beta_t$  from the indicator model, the 3 black lines are the mean  $\beta_t$  in each TTO, the blue line is the  $\beta_0 + \beta_1 t$  from the linear model, and the black curve is the smoothing spline fit.

Note that pitchers **do** decline on average from one time through the order to the next as Tango showed (3 black lines), but these models actually reveal that pitchers on average **decline continuously** throughout the game. What does this mean?

*The rule of thumb to pull pitchers prior to the start of 3TTO might not make sense.*

### 6.3.3 Takeaways

Please note that regression is about **fitting patterns from observational data** and does **NOT** imply causation. We're not saying anything about the causes of pitcher decline (fatigue or batter learning), we're only saying that after adjusting for confounders, pitchers appear to predominantly decline continuously, on average.

A potential causes of continuous pitcher decline is pitcher fatigue, and a potential cause of discontinuous pitcher decline is batter learning. Is it possible that both are at play? We address this question in the next section.

## 6.4 Continuous vs Discontinuous Pitcher Decline

### 6.4.1 Specifying the Combined Model

After adjusting for confounders and controlling for continuous pitcher decline within a game, do pitchers decline discontinuously from one time through the order to the next? We set up the following model to answer this question:

$$\begin{aligned} \mathbb{E}[y_i | t_i] = & \beta_0 + \beta_1 \cdot t_i + \beta_2 \cdot \mathbb{1}\{t_i \geq 2TTO\} + \beta_3 \cdot \mathbb{1}\{t_i \geq 3TTO\} \\ & + \beta_{BQ} \cdot BQ_i + \beta_{PQ} \cdot PQ_i + \beta_{hand} \cdot hand_i + \beta_{home} \cdot home_i \end{aligned} \quad (6.7)$$

Note that the **blue terms** represent continuous pitcher decline, the **red terms** represent discontinuous pitcher decline, and the black terms represent confounders.

We fit the model in R as follows:

```
# refit model with confounders
combined_model = lm(formula = woba_19 ~ 1 + batter_sequence_number
                    + as.numeric(TTO >= 2) + as.numeric(TTO >= 3)
                    + full_batter_woba_19 + full_pitcher_woba_19
                    + hand_match + batter_home, data = mlb_data)

# get model summary
summary(combined_model)
```

When we look at the model summary, neither  $\beta_2$  nor  $\beta_3$  are statistically-significant. As a result, we do not find statistically-significant evidence of discontinuous pitcher decline between times through the order. Figure 6.4 illustrates this, where it's clear that continuous pitcher decline dominates any discontinuous decline.

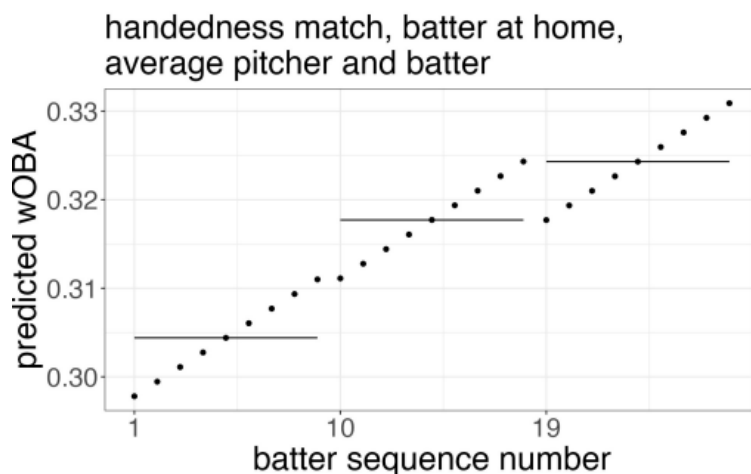


Figure 6.4: Plot of predicted wOBA by batter sequence number, adjusted for confounders. Also included are the 3 black lines representing the mean wOBA in each TTO.

### 6.4.2 Takeaways

Models do what they're told! If you only tell the model to look for discontinuous pitcher decline, then that is what you'll find. When we look a bit deeper, we find that the expected wOBA forecast by our model increases steadily over the course of a game and does not display sharp discontinuities between times through the order. Based on these results, we recommend managers cease pulling starting pitchers at the beginning of the third time through the order.

## References

- [FG] Fangraphs, *wOBA and FIP Constants*, Fangraphs, 2021.
- [JR] Rivera, J., *Rays' Kevin Cash explains decision to pull Blake Snell in World Series: 'I regret it because it didn't work out'*, Sporting News, 2020.
- [DL] Laurila, D., *Managers on the Third Time Through the Order*, FanGraphs, 2015.

- [ML] Lichtman, M., *Baseball ProGUESTus: Everything You Always Wanted to Know About the Times Through the Order Penalty*, Baseball Prospectus, 2013.
- [PS] Slowinski, P., *wOBA*, FanGraphs, 2010.
- [RB] Brill, R., Deshpande, S., & Wyner, A.J., *A Bayesian analysis of the time through the order penalty in baseball*, Wharton Sports Analytics & Business Initiative, 2023.
- [TT] Tango, T., Lichtman, M., and Dolphin, A., *The Book: Playing the Percentages in Baseball*, Potomac Books, 2007.