REDUCING STRUGGLES OF FRUSTRATION IN MAJOR LEAGUE BASEBALL THROUGH EPIDEMICS

by

DANIEL JAMES HRUSOVSKY

A manuscript submitted to the Faculty of John Carroll University

in partial fulfillment of the requirements for

graduation with a Bachelor of Science of Mathematics

2025

Hrusovsky, Daniel James (Bachelor of Science, Mathematics)

REDUCING STRUGGLES OF FRUSTRATION IN MAJOR LEAGUE BASEBALL THROUGH EPIDEMICS

ABSTRACT

This paper represents mathematical models regarding the Major League Baseball matchup between batters attempting to overcome the frustration and struggles when facing pitchers. The aim of this study is to determine the maximum the number of dominating batters in Major League Baseball in order to optimize run production and league revenue through the interaction between two traditional SIS models. The results obtained from determining the stability analysis and manipulating starting populations of both batters and pitchers found that batters are forever struggling against pitchers and cannot dominate purely in terms of their relationship.

TABLE OF CONTENTS

INTRODUCTION

In the sports atmosphere, professional athletes have a high tendency to struggle in their respective sport during their career. A 2019 study reported that 35% of superior athletes struggle with mental health issues [1]. This is due to the immense demand from fans and coaches to perform well. As individuals struggle on the field, they become increasingly frustrated, causing them to play worse. This inferior play can lead to being traded, released, or even retirement.

This struggle pertains most to Major League Baseball, specifically the batting side rather than the pitching. Many batters struggle due to the extremely difficult nature of hitting a moving ball at an other-worldly speed. During the 2024 MLB season, the mean batting average for the league was .243 [3]. Thus, it is considered above average to get a base hit once every four at-bats. Compared to percentage-based statistics in sports, such as field goal percentage or completion percentage, this number is substantially lower.

This creates a problem for Major League Baseball, as a lower batting average creates less action during the game, creating a less exciting environment for fans. This lack of excitement draws fans away from the sport, allowing them to spend less money on tickets and television subscriptions. Our goal is to determine minimize the spread of frustration through batters while optimizing the number of "dominating" batters.

In this project, we will look at the interaction between batters and hitters through an SIS mathematical model to find conclusions on batter dominance through simulations determined by an eigenvalue stability analysis. This approach will allow us to understand how each population interacts with each and how we can determine a solution for the Major League batters to overtake the pitchers.

CHAPTER 1

DEVELOPMENT OF MODEL

This model will use two SIS models, one for batters and one for pitchers. The susceptible class will represent the athletes who are dominating at their respective position. This will be represented in the batting population as B_D, with P_D representing the pitching population. The infected class will represent the athletes who are not dominating. In this scenario, we will assume that those who are not dominating are struggling. Thus, the struggling batting population is denoted as B_S and the struggling pitching population is represented as P_S. Since an athlete always has the potential to struggle in their respective sport, we will exclude the idea that an athlete can "recover". Thus, we only have the susceptible and infected classes.

Since pitchers and hitters interact and share the common goal of winning, we can assume that the spread of frustration and the spread of dominance affect each other. Thus, as a pitcher or hitter begins to win the at-bat, or dominate, the counter begins to grow frustrated. Similarly, as the pitcher or hitter begins to lose the at-bat, they become increasingly frustrated, while the latter begins to dominate.

Lastly, we will assume for simplicity that an MLB roster contains 13 pitchers and 13 hitters. Since a roster can only hold 26 players at a time, most teams hold an equal number of batters and pitchers. This means that the league total will there will be 360 total athletes in each population. Below is the flow diagram of the model.

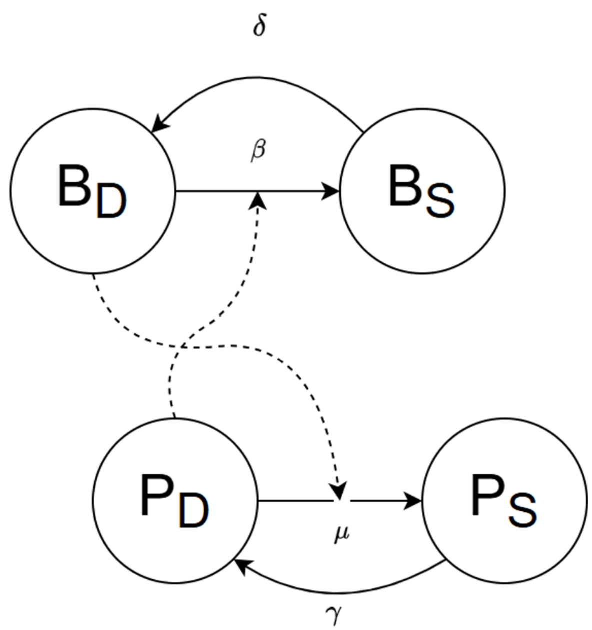


Figure 1.1: The flow diagram of the model between the population of hitters and the population of pitchers.

In this model, β represents the rate of spread for frustration in dominant batters upon interaction. The rate of partial recovery for the struggling batters is represented as δ. Similarly, μ represents the rate of spread of frustration in dominant pitchers upon interaction and γ represents the rate of partial recovery for struggling pitchers. We can assume that these parameters are greater than 0. Below is the system of differential equations.

$$\begin{cases} B_D' = -\beta B_D P_D + \delta B_S \\ B_S' = \beta B_D P_D - \delta B_S \\ P_D' = -\mu B_D P_D + \gamma P_S \\ P_S' = \mu B_D P_D - \gamma P_S \end{cases}$$

Figure 1.2: The system of differential equations of the model.

Since we know that the fixed population is made of their susceptible and infected states, we can figure that the population is the sum of the aforementioned classes. Thus,

$$B = B_D + B_S$$

$$P = P_D + P_S,$$

where B represents the batting population and P represents the pitching population. These equations will be important when simplifying our equations and determining the equilibria.

CHAPTER 2

RESULTS

To analyze the model, we will be using an eigenvalue approach. To use this approach, we will find the equilibria of the system's variables. After this, we will find the Jacobian matrix of the system's discrete equations at the system's equilibria. Lastly, we will use the Jacobian matrix to find and appropriately classify its eigenvalues. These eigenvalues will help determine the stability of our model at our equilibria.

First, we will find the equilibria of B_D. This equilibria occurs when the derivative of each population is equal to 0. Thus,

$$\mu B_D P_D - \gamma P_S = 0$$
$$\mu B_D P_D = \gamma P_S$$
$$B_D = \frac{\gamma}{\mu}(\frac{P_S}{P_D}).$$

Since $P = P_D + P_S$, then $P_S = P - P_D$. So,

$$\frac{P_S}{P_D} = \frac{P - P_D}{P_D} = \frac{P}{P_D} - 1.$$

Through substitution,

$$B_D = \frac{\gamma}{\mu}(\frac{P_S}{P_D})$$
$$B_D = \frac{\gamma}{\mu}(\frac{P}{P_D} - 1).$$

Since $P = P_D + P_S$, $P_S = P - P_D$. Thus,

$$B_S = P - P_D$$
$$B_S = P - \frac{\gamma}{\mu}(\frac{P}{P_D} - 1).$$

For the pitching population, the equilibria of the variables works in a similar fashion. Thus,

$$\beta B_D P_D - \delta B_S = 0$$
$$\beta B_D P_D = \delta B_S$$
$$P_D = \frac{\delta}{\beta}(\frac{B_S}{B_D})$$
$$P_D = \frac{\delta}{\beta}(\frac{B}{B_D} - 1),$$

since $B = B_S + B_D$. Thus, since $P = P_D + P_S$,

$$P_S = P - P_D$$
$$P_S = P - \frac{\delta}{\beta}(\frac{B}{B_D} - 1).$$

Thus, the equilibria of the system of differential equations is $(B_D, B_S, P_D, P_S) = (\frac{\gamma}{\mu}(\frac{P}{P_D} - 1), P - \frac{\gamma}{\mu}(\frac{P}{P_D} - 1), \frac{\delta}{\beta}(\frac{B}{B_D} - 1), P - \frac{\delta}{\beta}(\frac{B}{B_D} - 1))$. Notice how we do get a constant from our solution. Rather, our equilibrium is based on the values we give our parameters and initial populations.

Next, we will find the Jacobian matrix. The Jacobian matrix is found by taking the partial derivative of each variable for each differential equation. For our model, the order of the differential equations follow the same pattern as the ordered quadruple of the equilibria, (B_D, B_S, P_D, P_S). This is the same order for the partial derivatives. Shown below is the Jacobian at any point of the model.

Next, we will substitute the variables in the Jacobian matrix with the equilibria, as shown below.

$$
\mathrm{J} = \begin{bmatrix} -\beta P_D & \delta & -\beta B_D & 0 \\ \beta P_D & -\delta & \beta B_D & 0 \\ -\mu P_D & 0 & -\mu B_D & \gamma \\ \mu P_D & 0 & \mu B_D & \gamma \end{bmatrix}
$$

Figure 2.1: The Jacobian matrix, denoted as J, of the model.

$$
J_E = \begin{bmatrix} -\delta\left(\frac{B}{B_D} - 1\right) & \delta & -\frac{\beta\gamma}{\mu}\left(\frac{P}{P_D} - 1\right) & 0 \\ \delta\left(\frac{B}{B_D} - 1\right) & -\delta & \frac{\beta\gamma}{\mu}\left(\frac{P}{P_D} - 1\right) & 0 \\ -\frac{\mu\delta}{\beta}\left(\frac{B}{B_D} - 1\right) & 0 & -\gamma\left(\frac{P}{P_D} - 1\right) & \gamma \\ \frac{\mu\delta}{\beta}\left(\frac{B}{B_D} - 1\right) & 0 & \gamma\left(\frac{P}{P_D} - 1\right) & \gamma \end{bmatrix}
$$

Figure 2.2: The Jacobian matrix at our equilibria, denoted as J_E.

This Jacobian matrix can be used to find the eigenvalues, which will help us determine the stability of our equilibrium.

Definition 2.1 *An equilbrium (x^*, y^*) is:*

- stable *if a solution which starts nearby stays nearby.*
- unstable *if it is not stable, or at least one solution diverges from the equation.*
- asymptotically stable *if it is stable and all solutions near equations converge to it.*

Finding stability will help us determine if the variables in our model decide to converge towards a point or diverge. In terms of our model, this will help us determine if our populations of hitters and pitchers will level out or deviate.

To find the eigenvalues, we will need to solve the characteristic equation of the model, which is

$$\det(J_E - \lambda I) = 0.$$

For simplification purposes we will let $B^* = \frac{B}{B_D} - 1$ and $P^* = \frac{P}{P_D} - 1$. Thus,

$$\begin{aligned}\det(J_E - \lambda I) =& (-\delta B^* - \lambda)(-\delta - \lambda)[(-\gamma P^* - \lambda)(-\gamma - \lambda) - \gamma^2 P^*] \\ & - \delta^2 B^*[(-\gamma P^* - \lambda)] - (\gamma^2 P^*) + (\frac{\beta\gamma\delta}{\mu} P^*)[(\frac{-\mu\gamma}{\beta} B^*)(-\gamma - \lambda) - \gamma(\frac{\mu\delta}{\beta} B^*)] \\ & - \frac{\beta\gamma}{\mu} P^*[(-\delta - \lambda)(\frac{-\mu\delta}{\beta} B^*)(-\gamma - \lambda) - (\frac{\mu\gamma\delta}{\beta} B^*)].\end{aligned}$$

Through distribution and simplification,

$$\begin{aligned}\det(J_E - \lambda I) =& \gamma^2 P^* + \gamma P^* \lambda + \gamma^3 P^{*2} + \gamma\lambda + \lambda^2 + \gamma^2 P^* \lambda - \delta^2 \gamma^2 B^* P^* \\ & - \gamma\delta^2 B^* \lambda - \gamma\delta^2 B^* P^* \lambda - \gamma^2 B^* \lambda^2 + \gamma^3 \delta B^* P^* - \gamma^2 \delta B^* P^* \lambda \\ & - \gamma\delta^2 B^* P^* \lambda - \lambda\delta B^* P^* \lambda^2 \\ =& (1 - \delta^2 B^* - \gamma\delta B^* P^*)\lambda^2 \\ & + \gamma(1 + P^* + \gamma P^* - \delta B^* - \delta^2 B^* P^* - \gamma\delta B^* P^* - \delta^2 B^* P^*)\lambda \\ & + \gamma^2 P^*(1 + \gamma P^* - \delta^2 B^* + \gamma\delta B^*).\end{aligned}$$

This results in a quadratic formula with the intent to solve for λ. Through simplification,

$$\lambda = -\delta(B^* + 1) - \gamma(P^* + 1) \pm \sqrt{-\delta(B^* + 1) - \gamma(P^* + 1)^2 - 4\delta\gamma(P^* + B^* + 1)}$$

Let $a = \delta(B^* + 1) - \gamma(P^* + 1)$ and $b = 4\delta\gamma(P^* + B^* + 1)$. Thus, we have a simplified equation as shown below.

$$-a \pm \sqrt{a^2 - b}$$

Since $a^2 > a^2 - b$, and $b > 0$ since all parameters are positive, then $\sqrt{a^2} > \sqrt{a^2 - b}$. Since $\sqrt{a^2} = a$, then through substitution,

$$\sqrt{a^2} > \sqrt{a^2 - b}$$
$$a > \sqrt{a^2 - b}.$$

Let us look at the positive section of the quadratic equation. Thus,

$$-a + \sqrt{a^2 - b} < -a + a$$
$$-a + \sqrt{a^2 - b} < 0$$

So λ is a real negative for the first part of the quadratic equation. When looking at the negative part,

$$-a - \sqrt{a^2 - b} < -a - a < 0.$$

Thus, λ is a real negative for the second part of the quadratic equation. So our eigenvalues are real negatives, regardless of the equilibria.

Since our eigenvalues our real negatives, we can classify our equilibria as a sink, which is always stable. Thus, our equilibria are always stable, which means that our solutions will stay nearby if we start near the equilibria.

CHAPTER 3

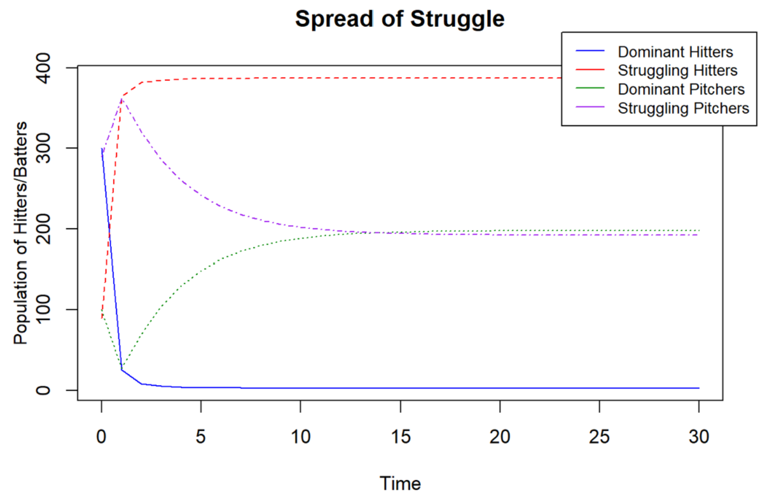
SIMULATIONS

Now that we have a general grasp on how our model will perform through stability analysis, we can view our results through simulations.

For this model, we will be using fixed parameters and varying initial conditions for the populations. This will allow us to see how the hitting and pitching populations work with (or against) each other.

Let $\beta = 0.226$, or the league strikeout rate in 2024, $\delta = 0.312$ for the league on-base percentage in 2024, $\mu = 0.117$ for the average runs given up in an at-bat in 2024, and $\gamma = 0.3257$ to represent how often a pitcher prevents a hitter from getting on-base three consecutive times [2].

We will look at two different simulations; one which is hitter dominant and one which is not. Since we are looking directly at the hitting population, we will keep the pitching population constant at $P_D = 100$ and $P_S = 290$. The first simulation, shown below, has a starting dominant hitting population of $B_D = 300$ with a struggling hitting population of $B_S = 90$.

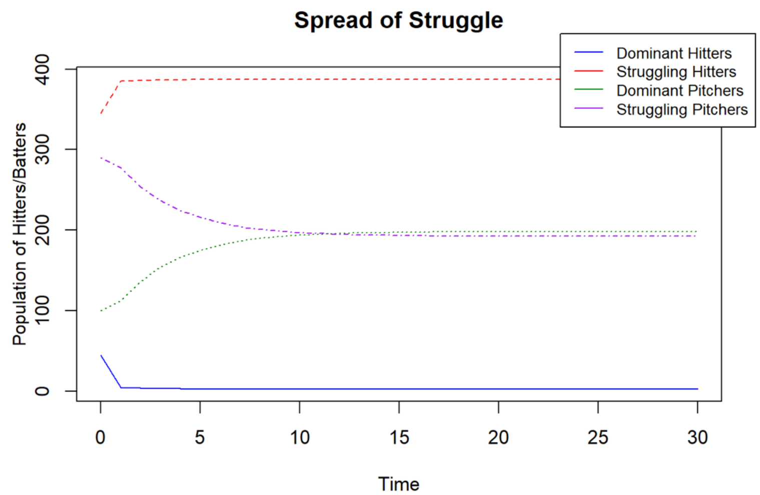


$\beta = 0.226, \delta = 0.312, \mu = 0.117, \gamma = 0.3257$

Figure 3.1: The simulation of our model with $B_D = 300$ in blue, $B_S = 90$ in red, $P_D = 100$ in green, and $P_S = 290$ in purple.

You can find the script used to calculate this in Appendix A. From this simulation, we can see that the dominant hitting population dies down to 0, while the struggling hitters overtake the entire population. With this, we can see that the pitching population mediates, with a slightly larger population in dominance.

For the second simulation where hitters are struggling, we will keep the same parameters and initial pitching populations with an initial dominant hitter population of $B_D = 50$ and struggling hitter population of $B_S = 340$. Below shows the simulation.



$\beta = 0.226, \delta = 0.312, \mu = 0.117, \gamma = 0.3257$

Figure 3.2: The simulation of our model with $B_D = 50$ in blue, $B_S = 340$ in red, $P_D = 100$ in green, and $P_S = 290$ in purple.

You can find the script used to calculate this in Appendix B. In this simulation, we receive the same result, where the struggling hitters continue to reign supreme compared to the dominant hitters, with the pitchers trending towards equal populations. One difference in this simulation is the trend of the pitching population. For the "hitter dominant" simulation, the population of struggling pitchers increased after the initial time before quickly decreasing. In this simulation, the struggling pitchers began decreasing as time increased.

CHAPTER 4

DISCUSSION

Based on our simulations of the model, it does not seem apparent for the batters to prevent struggle against pitchers. Pitchers will always have the upper hand against batters because their dominant population never dies out. This could be because of two different reasons.

The first reason is that the parameters set were too "pitcher friendly". Due to the aforementioned difficult nature of hitting, the set parameters may need to be adjusted so both populations are on equal playing fields. By changing the parameters, we could see a change in our simulations to prevent near-immediate extinction of dominant batters.

The second reason is that hitters are not actually struggling despite being in the struggling population. If the entire Major League Baseball population is struggling, then the threshold of struggle may be different than initially anticipated. If this model was used for a different sport, such as basketball or football, where rates of success are generally higher, then we may see different results.

For future, work, I believe that slight adjustment to the models can be made. Instead of interaction between the dominant populations, there can be additional interactions for the struggling populations to make the athletes dominant. There could also be additional parameters, such as retirement or getting called up, that could influence the model.

APPENDIX: Hitter Dominant Simulation

Here is the script behind the hitter dominant simulation.

```r
SIR <- function(t, y, parameters) {
  bd <- y[1]
  bs <- y[2]
  pd <- y[3]
  ps <- y[4]

  a  <- parameters[1]
  b <- parameters[2]
  c <- parameters[3]
  d <- parameters[4]

  dy    <- numeric(4)
  dy[1] <- -a*bd*pd + c*bs
  dy[2] <- a*bd*pd - c*bs
  dy[3] <- -b*bd*pd + d*ps
  dy[4] <- b*bd*pd - d*ps
  list(dy)
}

# Use built-in numerical methods to approximate solution
# Begin by specifying initial conditions and parameter values
time <- seq(0, 30, by = 1)
init <- c(300, 90, 100, 290)
params <- c(0.226, 0.117, 0.312, 0.3257)
soln <- ode(init, time, SIR, params)
names <- c("Dominant Hitters", "Struggling Hitters", "Dominant Pitchers", "Struggling Pitchers")
matplot.0D(soln, xlab = "Time", ylab = "Population of Hitters/Batters",
           main = "Spread of Struggle", col = c("blue", "red", "green4","purple"), legend = FALSE)

legend("topleft",
       legend = names,
      col = c("blue", "red", "green4", "purple"),
       lty = 1,
       cex = 0.8,
       xpd = TRUE,
       inset = c(0.75, -0.1))
```

APPENDIX: Struggling Hitter Simulation

Here is the script behind the struggling hitter simulation.

```
SIR <- function(t, y, parameters) {
  bd <- y[1]
  bs <- y[2]
  pd <- y[3]
  ps <- y[4]

  a  <- parameters[1]
  b <- parameters[2]
  c <- parameters[3]
  d <- parameters[4]

  dy    <- numeric(4)
  dy[1] <- -a*bd*pd + c*bs
  dy[2] <- a*bd*pd - c*bs
  dy[3] <- -b*bd*pd + d*ps
  dy[4] <- b*bd*pd - d*ps
  list(dy)
}

# Use built-in numerical methods to approximate solution
# Begin by specifying initial conditions and parameter values
time <- seq(0, 30, by = 1)
init <- c(50, 340, 100, 290)
params <- c(0.226, 0.117, 0.312, 0.3257)
soln <- ode(init, time, SIR, params)
names <- c("Dominant Hitters", "Struggling Hitters", "Dominant Pitchers", "Struggling Pitchers")
matplot.0D(soln, xlab = "Time", ylab = "Population of Hitters/Batters",
           main = "Spread of Struggle", col = c("blue", "red", "green4","purple"), legend = FALSE)

legend("topleft",
       legend = names,
      col = c("blue", "red", "green4", "purple"),
       lty = 1,
       cex = 0.8,
       xpd = TRUE,
       inset = c(0.75, -0.1))
```

REFERENCES

[1] M. G. Brigham. Athlete mental health: What you need to know. *McLean Hospital*, page 1, 2024.

[2] P. B. Reference. 2024 major league baseball advanced pitching. page 1, 2024.

[3] P. B. Reference. 2024 major league baseball team statistics. page 1, 2024.