

# Modelling Baseball Exit Velocities with the Block Maxima Method

*Jules Lanari-Collard*



MATH 598 Final Project  
McGill University  
Montréal, Québec, Canada

December 20, 2024

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Process over Outcomes . . . . .	1
1.2	The Importance of Exit Velocity . . . . .	2
1.3	Summarising Exit Velocity Data . . . . .	4
1.4	Applying Extreme Value Theory . . . . .	6
<b>2</b>	<b>Methods</b>	<b>8</b>
2.1	Data . . . . .	8
2.2	Block Maxima . . . . .	9
<b>3</b>	<b>Results and Discussion</b>	<b>12</b>
3.1	Return Level as an EV Metric . . . . .	12
3.2	Discussion . . . . .	16
3.3	Finite Upper Endpoints . . . . .	17

# Chapter 1

## Introduction

### 1.1 Process over Outcomes

Baseball is the second-most popular sport in the United States [1], and the average Major League Baseball (MLB) franchise is worth \$2.4 billion. Teams invest significant amounts on player wages every year, with the median projected payroll for 2025 being \$144 million, and some teams projected for over \$270 million [2]. Salaries are often disproportionately concentrated in a small number of high-value players; Shohei Ohtani signed the largest contract in sports history, \$700 million over 10 years, in 2024 [3], and Juan Soto recently broke that record with a \$765 million, 15 year contract with the New York Mets [4].

With the amounts of money involved, franchises employ dedicated analytics teams within their front offices to inform decisions and evaluate and project players. Having a good

understanding of a player's current and future ability is not only important when signing them to a large free agent contract, but also when trying to identify undervalued players.

Since the introduction of Statcast technology to MLB in 2015, allowing high-resolution ball and player tracking every game, there has been a shift in perspective on player evaluation. The technology allows analysts to evaluate events with respect to the *process*, not just the results. Instead of considering a ball in play as a binary event, either resulting in a hit or an out, we can consider the *expected outcome* in a probabilistic sense, given various factors such as the launch speed and launch angle. Over the course of many seasons, the wealth of data generated with this perspective has enabled a better separation between a player's true talent and luck and random variance; parsing the signal from the noise.

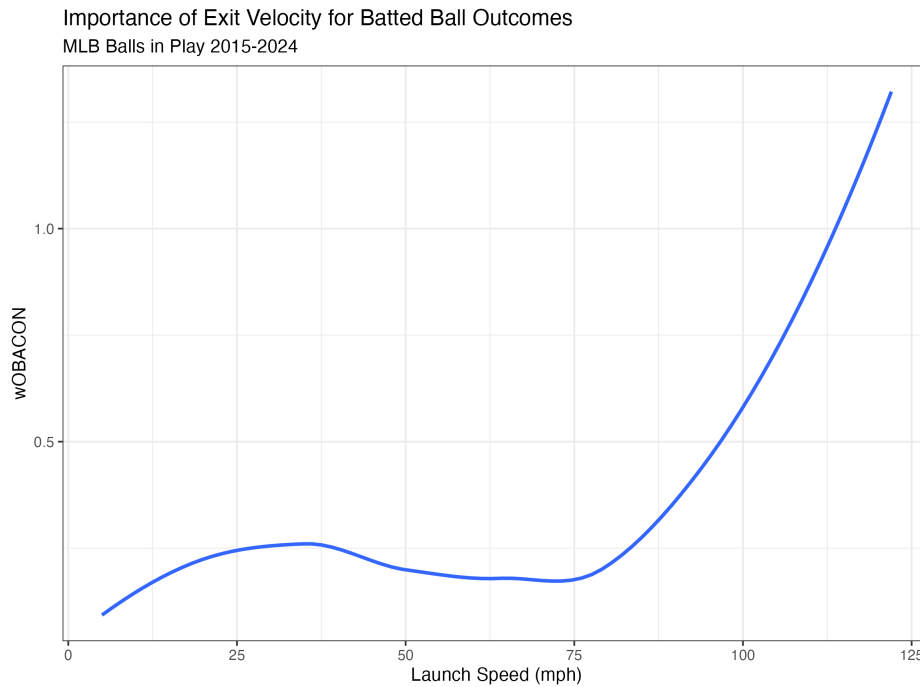
## 1.2 The Importance of Exit Velocity

One such Statcast metric is launch speed (more commonly known as **exit velocity**), simply measuring the speed (in mph) at which the baseball travels immediately after being hit by the bat. On a player level this has helped to quantify player 'power'. The use of this metric is rather intuitive; harder hit balls are more likely to turn into hits and even home runs. Figure 1.1 demonstrates this effect, using **wOBACON**<sup>1</sup> as a measure for outcome value, demonstrating an interesting trend; balls in play under 75mph all perform similarly, but

---

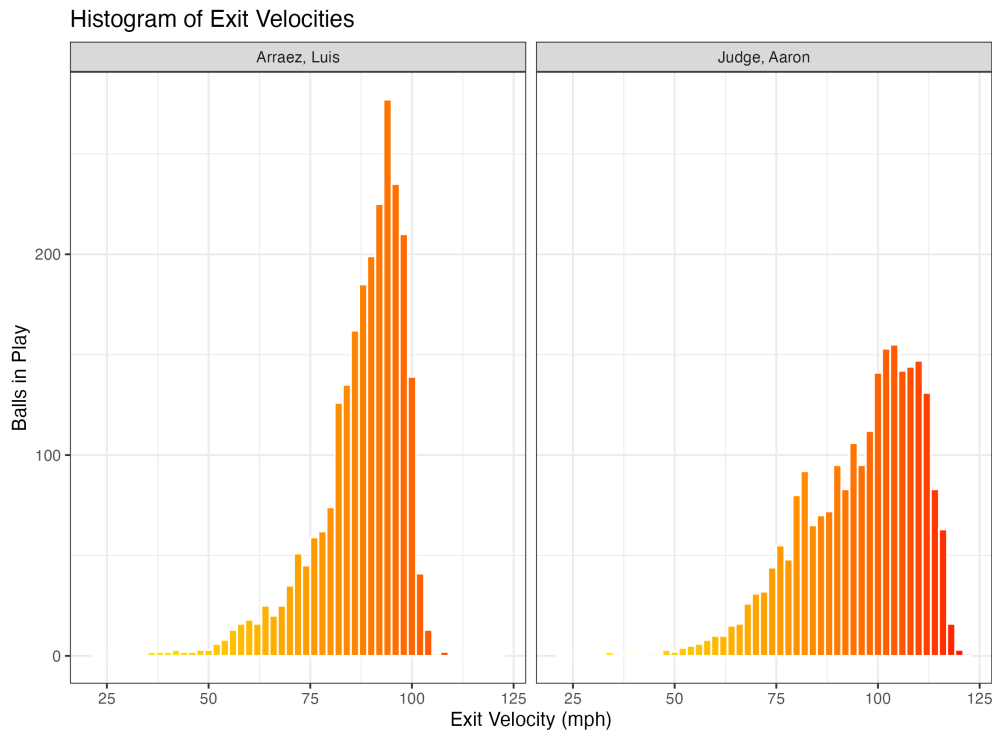
<sup>1</sup>wOBACON stands for *Weighted On-Base Average on CONTACT*. We will not go into the details of its calculation, but the main takeaway is that it measures the value of the outcomes of balls in play. For example, hits and home runs are positive, whilst outs are 0.

above 75mph there is a clear positive correlation between exit velocity (EV) and quality of outcome.



**Figure 1.1:** wOBACON and Exit Velocity

From the perspective of player evaluation, it follows that we can expect hitters who are able to consistently hit the ball harder to perform better. The question then turns to how can we best interpret each player's exit velocity data on a larger scale. To give an extreme example, suppose player A hits every ball at 100mph, whilst player B hits 80% of their batted balls at 95mph and the remaining 20% at 120mph. It is unclear which player provides more value; they have equal *average* exit velocities, but the value of their outcomes is not necessarily the same. Indeed, players across the league exhibit very different exit



**Figure 1.2:** Comparison of Contrasting EV Distributions

velocity distributions, and designing metrics to describe these distributions is a topic of great debate. Figure 1.2 illustrates the difference in distribution shape between two highly contrasting players.

### 1.3 Summarising Exit Velocity Data

Attempting to effectively summarise player-level exit velocity data, many metrics have already been formulated, with differing degrees of complexity. Before discussing the existing metrics, it is important to discuss what qualifies as an ‘informative’ or ‘useful’

metric. Metrics can be generally evaluated on three main criteria:

- Correlation with outcomes: how well does a metric represent value?
- Correlation with future outcomes: how well does a metric predict future outcomes and value?
- Variance: how much does a metric vary from year-to-year? In other words, how susceptible is the metric to random noise? In the search to separate true talent from randomness, low-variance (so-called ‘sticky’) metrics are much more reliable due to their robustness against random variation of the underlying data.

### 1.3.1 Existing Metrics

#### Average Exit Velocity

Self-explanatory in its definition, average EV is generally viewed as a flawed statistic, as it is equally affected by both the lower and upper quantiles of a player’s EV distribution. As seen in Figure 1.1, balls hit slower than 75mph should be viewed as roughly equal in value, and do not provide as much information as the upper quantiles.

#### Best Speed

In response to the flaws of average EV, Tom Tango stated that we “learn nothing about a batter on their slow hit batted balls” [5], and instead proposed **best speed**, defined as the

average of a batter's top 50% hardest-hit balls. Best speed correlates better with outcomes and exhibits less year-to-year variation than average EV [6].

### Exit Velocity Percentiles

Alternatively, one can consider the empirical quantiles of a player's exit velocities. Commonly used quantiles are the 80th, 90th and 95th percentiles, known as **EV80**, **EV90** and **EV95** respectively [7]. They perform similarly to best speed in correlation with outcomes and predictive power, and are even lower variance than best speed.

### Maximum Exit Velocity

Also self-explanatory, maximum EV is simply the speed of a player's hardest-hit ball over a given period of time, typically a season. Unsurprisingly, due to its inefficient use of information, it performs worse than the aforementioned metrics in correlation with outcomes, predictive power, and is only lower in variance than average EV.

## 1.4 Applying Extreme Value Theory

In summary, exit velocity data has a number of traits which render it difficult to analyse with traditional methods:

- EV distributions vary drastically from player to player, not only in parameters but also potentially in type of distribution.

- We learn very little about a player's true talent from their slowly-hit batted balls (i.e. bottom 50%).
- Information about player talent lies in upper quantiles of their EV distribution.

Given that individual batted ball events (and thus also their block maxima) can safely be assumed to be i.i.d.<sup>2</sup>, extreme value theory is particularly suited to this problem. Although player-level EV distributions may differ greatly, their block-maxima will follow a Generalised Extreme Value (GEV) distribution, allowing us to fit a model for each player without restrictive assumptions on their underlying EV distribution.

---

<sup>2</sup>This was verified using an auto-correlation plot.

# Chapter 2

## Methods

### 2.1 Data

Data was collected using queries to the MLB Statcast Search API<sup>1</sup>, querying all balls in play from 2015 (implementation of Statcast tracking) to 2024. The full dataset consists of 1,157,634 batted ball events involving 2,373 unique hitters. 2,020 player seasons consisted of at least 250 batted ball events (BBEs), a threshold set to ensure sufficient sample size for good model fits in the ensuing analysis.

---

<sup>1</sup>Accessible at [https://baseballsavant.mlb.com/statcast\\_search](https://baseballsavant.mlb.com/statcast_search)

## 2.2 Block Maxima

Seeking to model the upper quantiles of each player’s exit velocity distribution, we use the block-maxima method to fit a GEV model to the data. An alternative approach would be the Peaks-Over-Threshold (POT) method, however threshold selection is a complex problem, and not generalisable when looking to fit a model for each player. With the block-maxima method, we can set a constant block size for all models without overly adverse effects on the model fit.

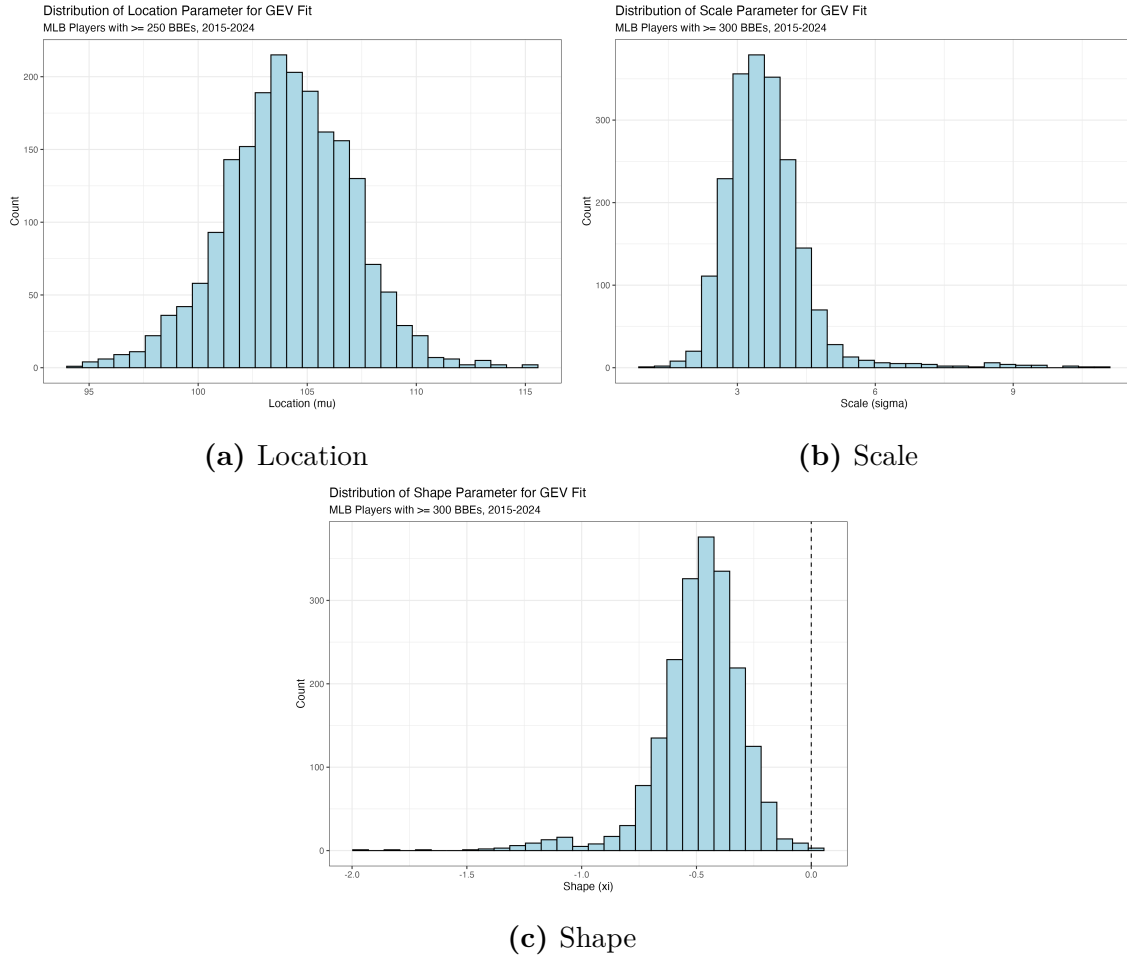
The GEV parameters can be estimated using maximum likelihood methods with the `ismev` package [8] in *R*. Using only player seasons with at least 250 BBEs, a model was fitted for each player season (2,020 total) using a block size of 10 balls in play. The distributions of the estimated parameters are shown in Figure 2.1. The parameter of interest for evaluating the quality of the models is the shape (denoted  $\xi$ ); MLE is not well-behaved for  $\xi \in (-1, -\frac{1}{2}]$  and potentially unobtainable when  $\xi \leq -1$ . The software implementation did not encounter issues, although over 50% of the estimates having  $\hat{\xi} \leq -\frac{1}{2}$  is a cause for concern.

When evaluating the fit of an individual GEV model, one can use visual diagnostic tools such as q-q plots and probability plots, however this becomes unfeasible when evaluating such a large number of models. Alternatively, we can use the well-known fact that the transform of a random sample drawn from a CDF  $F$  with  $F$  itself is standard uniformly distributed<sup>2</sup>.

With this in mind, for each player season, we transform the block maxima with the GEV

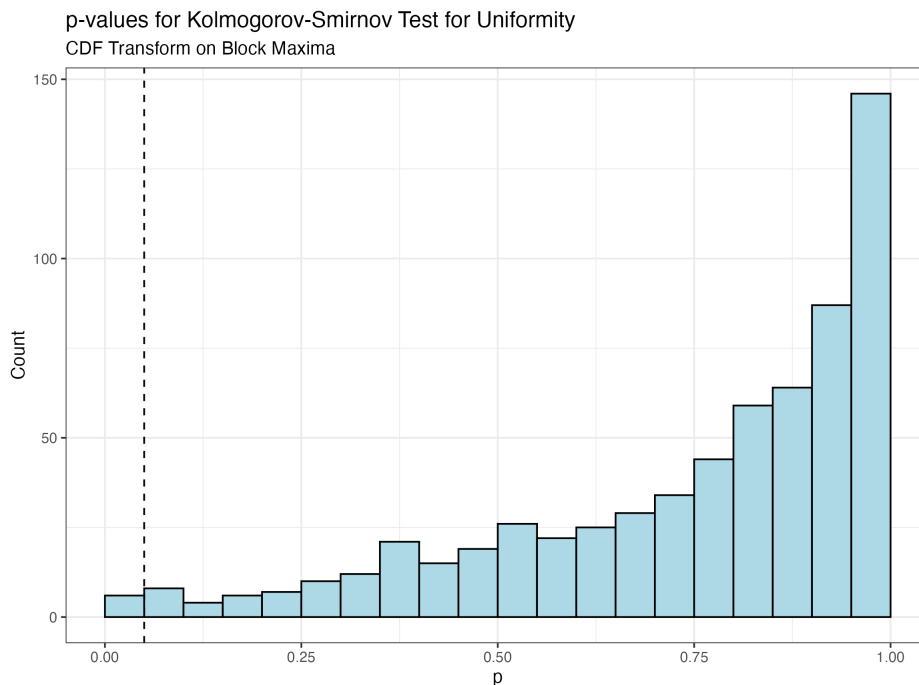
---

<sup>2</sup>The inverse of this process is known as the inverse transform method, used for generating random samples from different distributions.



**Figure 2.1:** Distribution of Estimated Parameters

CDF with the estimated parameters. For a well-fitted model, the resulting points should be approximately standard uniform, which we can test with the Kolmogorov-Smirnov Test for Uniformity. The p-values this test on each model are displayed in Figure 2.2, where we see that most of the transforms are in fact normally distributed at the 5% significance level.



**Figure 2.2:** Evaluating Model Fit

### 2.2.1 Return Levels

A quantity of interest with the block maxima method is the **return level**. Usually interpreted in financial contexts, a  $k$ -block return level represents the quantity one can expect to be exceeded once every  $k$  blocks on average. It can be estimated directly using a GEV model, and in this case provides a useful, easy-to-understand statistic to interpret each player EV model. In particular, we can estimate a 5-block return level for each player, representing the EV we can expect them to exceed once every 50 balls in play (BIP), on average.

## Chapter 3

# Results and Discussion

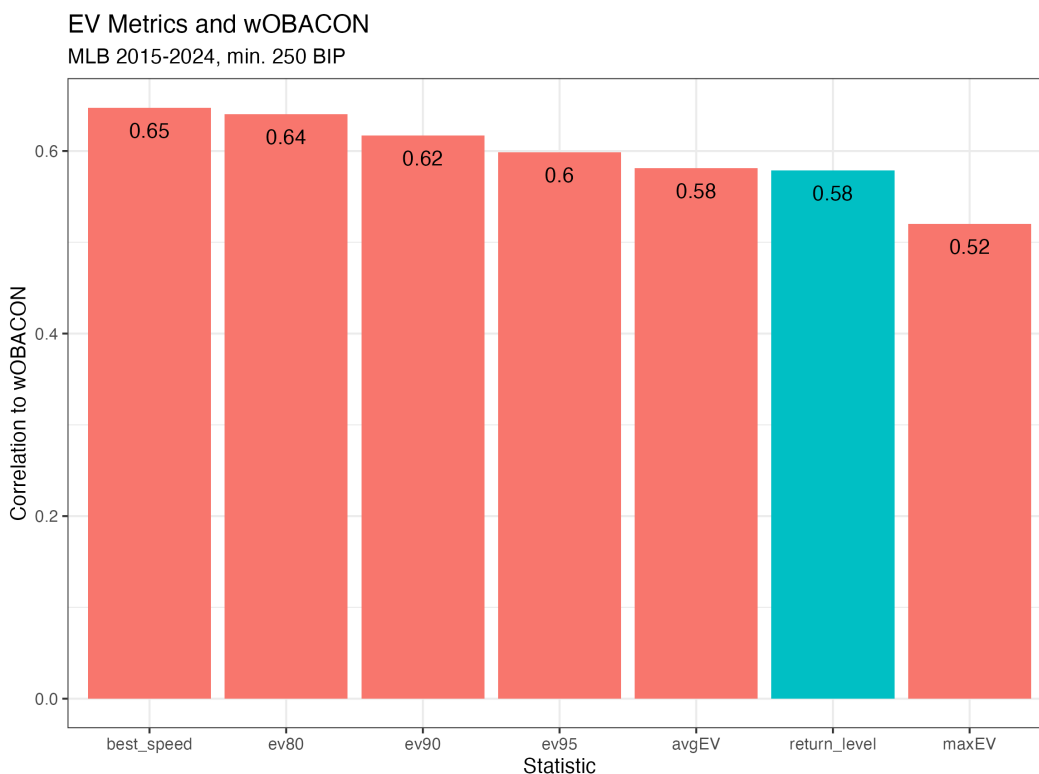
### 3.1 Return Level as an EV Metric

We can now evaluate the quality of 50-BIP return levels as a metric for summarising each player's EV distribution, in particular with respect to the criteria described in section 1.3.

Using wOBACON as the outcome of choice, we first consider how well return levels reflect value of outcomes. 50-BIP return level exhibits a 0.58 correlation coefficient<sup>1</sup> with same-season wOBACON. As shown in Figure 3.1, this is quite poor performance when compared to other metrics. This is perhaps unsurprising; one big drawback to the block-maxima method is that it ignores a lot of potentially useful information by only using the maximum observation from each block. Given that we are simply evaluating how well return levels describe the data *from which they were estimated*, it is not entirely unusual that they don't

---

<sup>1</sup>Using Pearson's Correlation Coefficient.

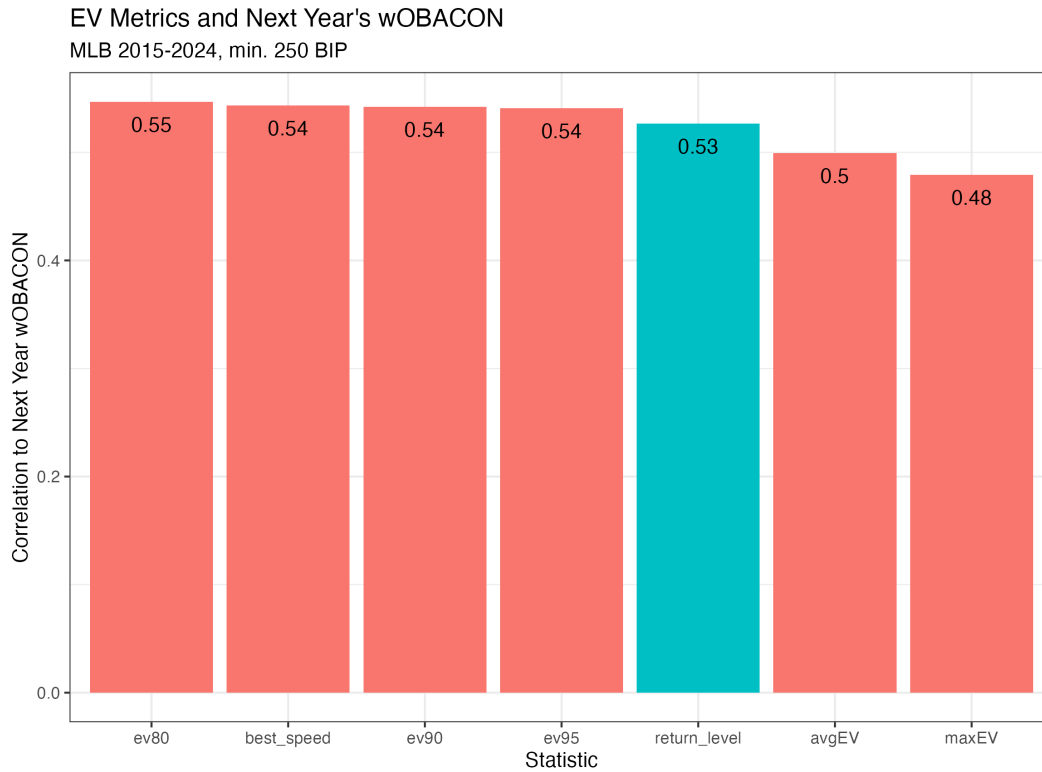


**Figure 3.1:** Correlation with Outcomes

perform particularly well.

In terms of predictive value, the 50-BIP return level has a 0.53 correlation with *next* season wOBACON, performing similarly to the best metrics, as shown in Figure 3.2. We notice more generally that the EV percentiles and best speed perform significantly better than average EV and maximum EV, with return levels in between.

Return level separates itself from the other metrics when we consider year-to-year variations. Only 1.5% of player seasons showed a year-to-year change in return level of more than 1 Standard Deviation. When compared to the variability of the other metrics,



**Figure 3.2:** Predictive Power

return level is second only to EV95 (see Figure 3.3). Whilst best speed is a better descriptor of outcomes and slightly better predictor, it varies more from year to year, rendering it less reliable when evaluating ‘true’ talent. An illustration of the stability of 50-BIP return level is provided in figure 3.4.

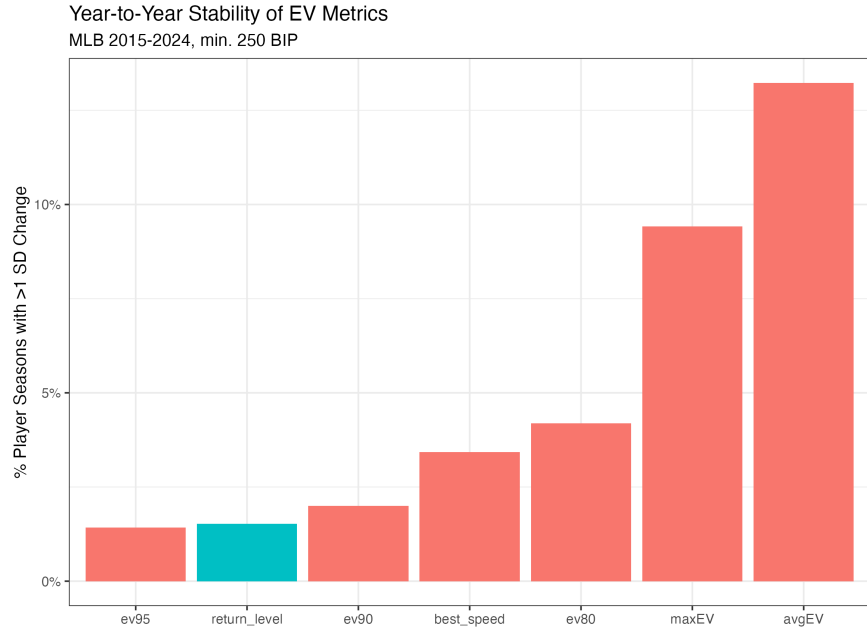


Figure 3.3: Variance

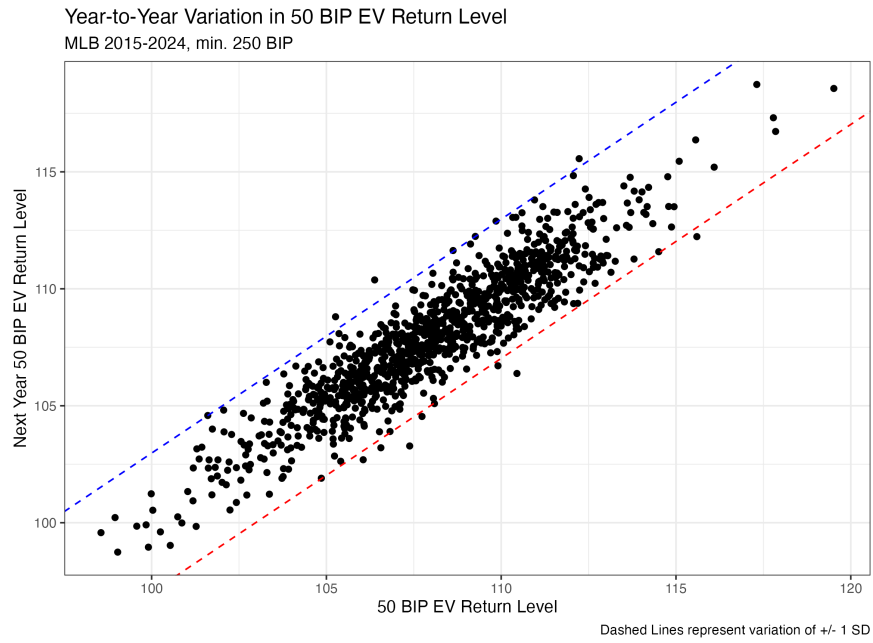


Figure 3.4: Stability of 50-BIP Return Level

## 3.2 Discussion

50-BIP EV return level performs moderately well as a descriptive metric. It performs similarly to best speed and EV percentiles in predictive power, and is below average for correlation with outcomes. However, its stability is encouraging and sets it apart from the other metrics.

On the other hand, return level has some flaws. It is more computationally intensive than the other metrics, and the sample size required to achieve its demonstrated performance is relatively large. The choice of 10-BIP blocks and 5-block return levels was somewhat arbitrary, and improvements could be made by tuning those ‘parameters’. Although the metric performs relatively well across the whole sample, some individual models are extremely bad fits due to the estimated  $\hat{\xi} < -1$  and non-uniformity of the CDF transform. One solution would be to simply filter out badly-specified models, however a metric which cannot be specified for certain players is definitely problematic. Bayesian inference for GEV models is an active area of research and could aid in the issue of  $\hat{\xi} < -1$ , by allowing for the imposition of constraints (in this case one could impose  $\xi > -1$ ).

Another area for improvement would be the inclusion of covariates; age in particular. Methods exist for incorporating covariates in GEV models [9], however it is unclear how one could implement a general trend over many models. That is, although we fit a separate model for each player-season, an ageing trend is a population-wide effect, not player-specific. Such a problem is beyond the scope of this project, but nevertheless is an interesting area

to explore.

### 3.3 Finite Upper Endpoints

A final point for discussion, unrelated to return levels, is to recall that a negative shape parameter bounds the GEV distribution from above. Given that almost all the models estimate a negative shape parameter (see figure 2.1), we can infer that each player has a *finite upper endpoint* to their EV distribution. This is an interesting if unsurprising result; we would expect there to be some physical limit to each player's capabilities. Methods for the estimation of the upper endpoint are available [10], however the estimation is highly dependent on the shape parameter estimate. For  $\xi$  close to 0, the endpoint estimate becomes unreasonably high.

Estimating the upper endpoint for each player would be an alternative metric of interest; representing an estimate of their power limits. However, when estimating the upper endpoint on the fitted player models, the estimates were indeed very unstable with the estimated shape parameter, and as a result don't provide a useful interpretation.

## Bibliography

- [1] J. M. Jones. “Football retains dominant position as favorite u.s. sport.” (2023), [Online]. Available: <https://news.gallup.com/poll/610046/football-retains-dominant-position-favorite-sport.aspx> (visited on 11/23/2024).
- [2] Fangraphs. “Payroll breakdown, rosterresource.” (2024), [Online]. Available: <https://www.fangraphs.com/roster-resource/breakdowns/payroll> (visited on 11/23/2024).
- [3] S. Wexler. “\$700m stunner: Ohtani to dodgers on biggest deal in sports history.” (2023), [Online]. Available: <https://www.mlb.com/news/shohei-ohtani-contract-with-dodgers> (visited on 12/17/2024).
- [4] A. DiComo. “Mets, soto finalize record-breaking 15-yr, \$765m deal.” (2024), [Online]. Available: <https://www.mlb.com/news/juan-soto-agrees-to-contract-with-mets> (visited on 12/17/2024).

- 
- [5] T. Tango. “Statcast metric: Best speed.” (2022), [Online]. Available: <http://tangotiger.com/index.php/site/comments/statcast-metric-best-speed> (visited on 11/23/2024).
- [6] D. Andrews. “The doomed search for a perfect way to interpret exit velocity data.” (2023), [Online]. Available: <https://blogs.fangraphs.com/the-doomed-search-for-a-perfect-way-to-interpret-exit-velocity-data/> (visited on 11/23/2024).
- [7] B. Clemens. “You can’t fake exit velocity.” (2023), [Online]. Available: <https://blogs.fangraphs.com/you-cant-fake-exit-velocity/> (visited on 11/23/2024).
- [8] J. E. Heffernan and A. G. Stephenson., *Ismev: An introduction to statistical modeling of extreme values*, R package version 1.42, 2018. [Online]. Available: <https://CRAN.R-project.org/package=ismev>.
- [9] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.
- [10] I. F. Alves, P. Rosario, and C. Neves, “A general estimator for the right endpoint with an application to supercentenarian women’s records,” *Extremes*, 2016.